The XSB System
Version 3.6.x

Terrance Swift    David S. Warren

Konstantinos Sagonas
Juliana Freire
Prasad Rao
Baoqiu Cui
Ernie Johnson
Luis de Castro
Rui F. Marques
Diptikalyan Saha
Steve Dawson
Michael Kifer

April 23, 2015
Credits

Day-to-day care and feeding of XSB including bug fixes, ports, and configuration management is currently done by David Warren and Terrance Swift with the help of Michael Kifer and others. In the past Kostis Sagonas, Prasad Rao, Steve Dawson, Juliana Freire, Ernie Johnson, Baoqiu Cui, Bart Demoen and Luis F. Castro have provided tremendous help.

In Version 3.6, the core engine development of the SLG-WAM has been mainly implemented by Terrance Swift, David Warren, Kostis Sagonas, Prasad Rao, Juliana Freire, Ernie Johnson, Luis Castro and Rui Marques. The breakdown, very roughly, was that Terrance Swift wrote the initial tabling engine, the SLG-WAM, and its built-ins; and leads the current development of the tabling subsystem. Prasad Rao reimplemented the engine’s tabling subsystem to use tries for variant-based table access and Ernie Johnson extended and refactored these routines in a number of ways, including adding call subsumption. Kostis Sagonas implemented most of tabled negation. Juliana Freire revised the table scheduling mechanism starting from Version 1.5.0 to create the batched and local scheduling that is currently used. Baoqiu Cui revised the data structures used to maintain delay lists, and added attributed variables to the engine. Luis Castro rewrote the emulator to use jump tables and wrote a heap-garbage collector for the SLG-WAM. Rui Marques was responsible for the concurrency control algorithms used for shared tables, and mainly responsible for making the XSB engine multi-threaded. The incremental table maintenance subsystem was designed and first implemented by Diptikalyan Saha, and its design and development has been continued by Terrance Swift. Answer subsumption was written by David Warren and Terrance Swift. David Warren implemented hash-consed, or “interned” tables. Call abstraction and answer abstraction (restraint) were written by Terrance Swift.

Other engine work includes the following. Memory expansion code for WAM stacks was written by Ernie Johnson, Bart Demoen and David S. Warren. Heap garbage collection was written by Luis de Castro, Kostis Sagonas and Bart Demoen. Atom space garbage collection was written by David Warren; table garbage collection was written by Terrance Swift based in part on space reclamation code written by Prasad Rao. Rui Marques rewrote much of the engine to make it compliant with 64-bit architectures. Assert and retract code was based on code written by Jiyang Xu; it significantly revised by David S. Warren, who added alternative,
multiple, and star indexing and by Terrance Swift who implemented
dynamic clause garbage collection. Trie assert/retract code, and trie interning
code was written by Prasad Rao. Neng-fa Zhou, Terrance Swift and
David Warren upgraded XSB from ASCII to the character sets UTF-8,
C1253, and LATIN-1. The current version of findall/3 was re-written
from scratch by Bart Demoen, as was XSB’s original throw and catch
mechanism. 64-bit floats were added by Charles Rojo.

In terms of core system Prolog code, Kostis Sagonas was responsible for
HiLog compilation and associated built-ins as well as coding or revising
many standard predicates. Steve Dawson implemented Unification Fac-
toring. The revision of XSB’s I/O into ISO-compatible streams was done
by Michael Kifer and Terrance Swift. The auto_table and suppl_table
directives were written by Kostis Sagonas. The DCG expansion mod-
ule was written by Kostis Sagonas for non-tabled code and by Baoqiu
Cui, David Warren and Terrance Swift for tabled code. The handling
of the multifile directive was written by Baoqiu Cui and David War-
ren. C.R. Ramakrishnan wrote the mode analyzer for XSB. Michael Kifer
implemented the storage module. The multi-threaded API was written
by Terrance Swift and Rui Marques. Walter Wilson has written several
of XSB’s library predicates for tabling. Paulo Moura has added several
predicates to make XSB more consistent with other Prologs.

Michael Kifer has been in charge of XSB’s installation procedures, rewrit-
ing parts of the XSB code to make XSB configurable with GNU’s Au-
toconf, implementing XSB’s package system, and integrated GPP with
XSB’s compiler. GPP, the source code preprocessor used by XSB, was
written by Denis Auroux, who also wrote the GPP manual reproduced in
Appendix A.

The starting point of XSB (in 1990) was PSB-Prolog 2.0 by Jiyang Xu and
David Warren. PSB-Prolog in its turn was based on SB-Prolog, primarily
designed and written by Saumya Debray, David S. Warren, and Jiyang
Xu. Thanks are also due to Weidong Chen for his work on Prolog clause
indexing for SB-Prolog, to Richard O’Keefe, who contributed the Prolog
code for the Prolog reader and the C code for the tokenizer, to Ciao Prolog
whose write_term/[2,3] we use, and to SWI Prolog for their CLP(R)
package.

... Now what did I forget this time?
## Contents

1 Introduction 1

1.1 Using This Manual .......................... 7

2 Getting Started with XSB 8

2.1 Installing XSB under UNIX .................. 8

2.1.1 Possible Installation Problems .......... 12

2.2 Installing XSB under Windows ............ 13

2.2.1 Using Cygnus Software’s CygWin32 .... 13

2.2.2 Using Microsoft Visual C++ ............. 14

2.3 Invoking XSB ................................ 17

2.4 Compiling XSB programs ........................ 17

2.5 Sample XSB Programs .......................... 18

2.6 Exiting XSB ................................. 19

3 System Description 20

3.1 Entering and Exiting XSB from the Command Line .................. 20

3.2 The System and its Directories ............... 21

3.3 How XSB Finds Files: Source File Designators ................. 22

3.4 The Module System of XSB .................... 24

3.5 Standard Predicates in XSB .................... 30

3.6 The Dynamic Loader and its Search Path ................. 30

3.6.1 Changing the Default Search Path and the Packaging System 31
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6.2</td>
<td>Dynamically loading predicates in the interpreter</td>
<td>33</td>
</tr>
<tr>
<td>3.7</td>
<td>Command Line Arguments</td>
<td>34</td>
</tr>
<tr>
<td>3.8</td>
<td>Memory Management</td>
<td>39</td>
</tr>
<tr>
<td>3.9</td>
<td>Compiling, Consulting, and Loading</td>
<td>40</td>
</tr>
<tr>
<td>3.9.1</td>
<td>Static Code</td>
<td>40</td>
</tr>
<tr>
<td>3.9.2</td>
<td>Dynamic Code</td>
<td>42</td>
</tr>
<tr>
<td>3.9.3</td>
<td>The multifile directive</td>
<td>43</td>
</tr>
<tr>
<td>3.10</td>
<td>The Compiler</td>
<td>43</td>
</tr>
<tr>
<td>3.10.1</td>
<td>Invoking the Compiler</td>
<td>44</td>
</tr>
<tr>
<td>3.10.2</td>
<td>Compiler Options</td>
<td>46</td>
</tr>
<tr>
<td>3.10.3</td>
<td>Specialization</td>
<td>54</td>
</tr>
<tr>
<td>3.10.4</td>
<td>Compiler Directives</td>
<td>56</td>
</tr>
<tr>
<td>3.10.5</td>
<td>Conditional Compilation</td>
<td>57</td>
</tr>
<tr>
<td>3.10.6</td>
<td>Inline Predicates</td>
<td>64</td>
</tr>
<tr>
<td>3.11</td>
<td>A Note on ISO Compatibility</td>
<td>65</td>
</tr>
<tr>
<td>4</td>
<td>Syntax</td>
<td>67</td>
</tr>
<tr>
<td>4.1</td>
<td>Terms</td>
<td>67</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Integers</td>
<td>67</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Floating-point Numbers</td>
<td>69</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Atoms</td>
<td>69</td>
</tr>
<tr>
<td>4.1.4</td>
<td>Variables</td>
<td>71</td>
</tr>
<tr>
<td>4.1.5</td>
<td>Compound Terms</td>
<td>71</td>
</tr>
<tr>
<td>4.1.6</td>
<td>Lists</td>
<td>73</td>
</tr>
<tr>
<td>4.2</td>
<td>From HiLog to Prolog</td>
<td>74</td>
</tr>
<tr>
<td>4.3</td>
<td>Operators</td>
<td>76</td>
</tr>
<tr>
<td>5</td>
<td>Using Tabling in XSB: A Tutorial Introduction</td>
<td>80</td>
</tr>
<tr>
<td>5.1</td>
<td>Tabling in the Context of a Prolog System</td>
<td>81</td>
</tr>
</tbody>
</table>
CONTENTS

5.2 Definite Programs .................................................. 81
  5.2.1 Call Variance vs. Call Subsumption ......................... 85
  5.2.2 Tabling with Interned Terms .................................. 88
  5.2.3 Table Scheduling Strategies .................................. 89
  5.2.4 Interaction Between Prolog Constructs and Tabling .......... 91
  5.2.5 Potential Pitfalls in Tabling ................................. 94
5.3 Normal Programs .................................................. 96
  5.3.1 Stratified Normal Programs .................................. 96
  5.3.2 Non-stratified Programs .................................... 100
  5.3.3 On Beyond Zebra: Implementing Other Semantics for Non-
       stratified Programs ........................................... 104
5.4 Answer Subsumption ............................................... 107
  5.4.1 Types of Answer Subsumption ................................ 107
  5.4.2 Examples of Answer Subsumption ............................ 109
  5.4.3 Term-Sets .................................................... 112
5.5 Tabling for Termination .......................................... 116
  5.5.1 Subgoal Abstraction ......................................... 118
  5.5.2 Bounded Rationality through Radial Restraint ............... 119
5.6 Incremental Table Maintenance ................................. 121
  5.6.1 Transparent Incremental Tabling ........................... 122
  5.6.2 Updating in a Three-Valued Logic ......................... 124
  5.6.3 Incremental Tabling using Interned Tries ................. 126
  5.6.4 Abstracting the IDG for Better Performance .............. 127
  5.6.5 Summary and Implementation Status ....................... 129
  5.6.6 Predicates for Incremental Table Maintenance ............ 129
5.7 Compatibility of Tabling Modes and Predicate Attributes .... 135

6 Standard and General Predicates ................................ 137
  6.1 Input and Output .............................................. 137
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1.1 I/O Streams in XSB</td>
<td>137</td>
</tr>
<tr>
<td>6.1.2 Character Sets in XSB</td>
<td>140</td>
</tr>
<tr>
<td>6.1.3 Predicates for ISO Streams</td>
<td>140</td>
</tr>
<tr>
<td>6.1.4 DEC-IO Style File Handling</td>
<td>147</td>
</tr>
<tr>
<td>6.1.5 Character I/O</td>
<td>150</td>
</tr>
<tr>
<td>6.1.6 Term I/O</td>
<td>156</td>
</tr>
<tr>
<td>6.1.7 Special I/O</td>
<td>165</td>
</tr>
<tr>
<td>6.2 Interactions with the Operating System</td>
<td>171</td>
</tr>
<tr>
<td>6.2.1 The <code>path_sysop/2</code> interface</td>
<td>173</td>
</tr>
<tr>
<td>6.3 Evaluating Arithmetic Expressions through <code>is/2</code></td>
<td>176</td>
</tr>
<tr>
<td>6.3.1 Evaluable Functors for Arithmetic Expressions</td>
<td>177</td>
</tr>
<tr>
<td>6.4 Convenience</td>
<td>180</td>
</tr>
<tr>
<td>6.5 Negation and Control</td>
<td>181</td>
</tr>
<tr>
<td>6.6 Unification and Comparison of Terms</td>
<td>185</td>
</tr>
<tr>
<td>6.6.1 Sorting of Terms</td>
<td>191</td>
</tr>
<tr>
<td>6.7 Meta-Logical</td>
<td>193</td>
</tr>
<tr>
<td>6.8 Cyclic Terms</td>
<td>208</td>
</tr>
<tr>
<td>6.8.1 Unification with and without Occurs Check</td>
<td>208</td>
</tr>
<tr>
<td>6.8.2 Cyclic Terms</td>
<td>208</td>
</tr>
<tr>
<td>6.9 Manipulation of Atomic Terms</td>
<td>210</td>
</tr>
<tr>
<td>6.10 All Solutions and Aggregate Predicates</td>
<td>223</td>
</tr>
<tr>
<td>6.11 Meta-Predicates</td>
<td>228</td>
</tr>
<tr>
<td>6.12 Information about the System State</td>
<td>235</td>
</tr>
<tr>
<td>6.13 Execution State</td>
<td>252</td>
</tr>
<tr>
<td>6.14 Asserting, Retracting, and Other Database Modifications</td>
<td>260</td>
</tr>
<tr>
<td>6.14.1 Reading Dynamic Code from Files</td>
<td>271</td>
</tr>
<tr>
<td>6.14.2 The <code>storage</code> Module: Associative Arrays and Backtrackable Updates</td>
<td>276</td>
</tr>
<tr>
<td>6.15 Tabling Declarations and Builtinsin</td>
<td>278</td>
</tr>
</tbody>
</table>
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.15.1 Declaring and Modifying Tabled Predicates</td>
<td>280</td>
</tr>
<tr>
<td>6.15.2 Predicates for Table Inspection</td>
<td>283</td>
</tr>
<tr>
<td>6.15.3 Abolishing Tables and Table Components</td>
<td>300</td>
</tr>
<tr>
<td><strong>7 Multi-Threaded Programming in XSB</strong></td>
<td><strong>310</strong></td>
</tr>
<tr>
<td>7.1 Getting Started with Multi-Threading</td>
<td>310</td>
</tr>
<tr>
<td>7.2 Communication among Threads</td>
<td>312</td>
</tr>
<tr>
<td>7.3 Thread Statuses: Joinable and Detached Threads</td>
<td>315</td>
</tr>
<tr>
<td>7.4 Prolog Message Queues</td>
<td>317</td>
</tr>
<tr>
<td>7.5 Thread Cancellation and Signalling</td>
<td>319</td>
</tr>
<tr>
<td>7.6 Performance and other Considerations</td>
<td>321</td>
</tr>
<tr>
<td>7.7 Examples of Multi-Threaded Programs in XSB</td>
<td>321</td>
</tr>
<tr>
<td>7.8 Configuring the Multi-threaded Engine under Windows</td>
<td>322</td>
</tr>
<tr>
<td>7.9 Predicates for Multi-Threading</td>
<td>325</td>
</tr>
<tr>
<td>7.9.1 Predicates for Thread Synchronization and Communication</td>
<td>332</td>
</tr>
<tr>
<td><strong>8 Storing Facts in Tries</strong></td>
<td><strong>340</strong></td>
</tr>
<tr>
<td>8.1 Examples of Using Tries</td>
<td>342</td>
</tr>
<tr>
<td>8.2 Space Management for Tries</td>
<td>343</td>
</tr>
<tr>
<td>8.3 Predicates for Tries</td>
<td>344</td>
</tr>
<tr>
<td>8.4 Low-level Trie Manipulation Utilities</td>
<td>351</td>
</tr>
<tr>
<td>8.4.1 A Low-Level API for Interned Tries</td>
<td>351</td>
</tr>
<tr>
<td><strong>9 Hooks</strong></td>
<td><strong>355</strong></td>
</tr>
<tr>
<td>9.1 Adding and Removing Hooks</td>
<td>355</td>
</tr>
<tr>
<td>9.2 Hooks Supported by XSB</td>
<td>356</td>
</tr>
<tr>
<td><strong>10 Debugging and Profiling</strong></td>
<td><strong>358</strong></td>
</tr>
<tr>
<td>10.1 Prolog-style Tracing and Debugging</td>
<td>358</td>
</tr>
<tr>
<td>10.2 Low-Level Tracing</td>
<td>363</td>
</tr>
</tbody>
</table>
10.3 Analyzing the Execution of Tabled Programs ........................................ 364
  10.3.1 Tracing a tabled evaluation through forest logging ......................... 365
  10.3.2 Analyzing the log; seeing the forest through the trees ................. 371
  10.3.3 Discussion .................................................................................. 377
  10.3.4 Predicates for Forest Logging ...................................................... 377

11 Definite Clause Grammars ........................................................................ 379
  11.1 General Description .......................................................................... 379
  11.2 Translation of Definite Clause Grammar rules .................................. 381
    11.2.1 Definite Clause Grammars and Tabling .................................... 383
  11.3 Definite Clause Grammar predicates ................................................. 384
  11.4 Two differences with other Prologs .................................................. 388

12 Exception Handling .................................................................................. 390
  12.1 The Mechanics of Exception Handling .............................................. 390
    12.1.1 Exception Handling in Non-Tabled Evaluations ....................... 390
    12.1.2 Exception Handling in Tabled Evaluation ............................... 395
  12.2 XSB’s Standard Format for Errors .................................................... 397
    12.2.1 Error Tags .............................................................................. 398
    12.2.2 XSB-Specific Information in Error Terms ............................... 399
  12.3 Predicates to Throw and Handle Errors ............................................ 400
    12.3.1 Predicates to Throw Errors ...................................................... 400
    12.3.2 Predicates used in Handling Errors ....................................... 402
  12.4 Convenience Predicates .................................................................. 403
  12.5 Backtraces ...................................................................................... 404

13 Foreign Language Interface ................................................................. 406
  13.1 Foreign Language Modules ............................................................... 407
  13.2 Lower-Level Foreign Language Interface ......................................... 408
    13.2.1 Context Parameters ............................................................... 412
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.2.2</td>
<td>Exchanging Basic Data Types</td>
</tr>
<tr>
<td>13.2.3</td>
<td>Exchanging Complex Data Types</td>
</tr>
<tr>
<td>13.3</td>
<td>Foreign Modules That Call XSB Predicates</td>
</tr>
<tr>
<td>13.4</td>
<td>Foreign Modules That Link Dynamically with Other Libraries</td>
</tr>
<tr>
<td>13.5</td>
<td>Higher-Level Foreign Language Interface</td>
</tr>
<tr>
<td>13.5.1</td>
<td>Declaration of high level foreign predicates</td>
</tr>
<tr>
<td>13.6</td>
<td>Compiling Foreign Modules on Windows and under Cygwin</td>
</tr>
<tr>
<td>13.7</td>
<td>Functions for Use in Foreign Code</td>
</tr>
<tr>
<td>14</td>
<td>Embedding XSB in a Process</td>
</tr>
<tr>
<td>14.1</td>
<td>Calling XSB from C</td>
</tr>
<tr>
<td>14.2</td>
<td>Examples of Calling XSB</td>
</tr>
<tr>
<td>14.2.1</td>
<td>The XSB API for the Sequential Engine Only</td>
</tr>
<tr>
<td>14.2.2</td>
<td>The General XSB API</td>
</tr>
<tr>
<td>14.2.3</td>
<td>Managing Multiple XSB Threads through the API</td>
</tr>
<tr>
<td>14.2.4</td>
<td>Calling Multiple XSB Threads using Multiple C Threads</td>
</tr>
<tr>
<td>14.3</td>
<td>A C API for XSB</td>
</tr>
<tr>
<td>14.3.1</td>
<td>Initializing and Closing XSB</td>
</tr>
<tr>
<td>14.3.2</td>
<td>Passing Commands to XSB</td>
</tr>
<tr>
<td>14.3.3</td>
<td>Querying XSB</td>
</tr>
<tr>
<td>14.3.4</td>
<td>Obtaining Information about Errors</td>
</tr>
<tr>
<td>14.3.5</td>
<td>Thread Management from Calling Programs</td>
</tr>
<tr>
<td>14.4</td>
<td>The Variable-length String Data Type</td>
</tr>
<tr>
<td>14.5</td>
<td>Passing Data into an XSB Module</td>
</tr>
<tr>
<td>14.6</td>
<td>Creating an XSB Module that Can be Called from C</td>
</tr>
<tr>
<td>15</td>
<td>Restrictions and Current Known Bugs</td>
</tr>
<tr>
<td>15.1</td>
<td>Current Restrictions</td>
</tr>
<tr>
<td>15.2</td>
<td>Known Bugs</td>
</tr>
</tbody>
</table>
CONTENTS

A  GPP - Generic Preprocessor 467
   A.1 Description .............................................. 467
   A.2 Syntax .................................................. 468
   A.3 Options .................................................. 468
   A.4 Syntax Specification .................................... 472
   A.5 Evaluation Rules ........................................ 476
   A.6 Meta-macros .............................................. 477
   A.7 Examples ................................................ 482
   A.8 Advanced Examples ..................................... 487
   A.9 Author ................................................... 489
Chapter 1

Introduction

XSB is a research-oriented, commercial-grade Logic Programming system for Unix and Windows-based platforms. In addition to providing nearly all functionality of ISO-Prolog, XSB includes the following features:

- Evaluation of queries according to the Well-Founded Semantics [81] through full SLG resolution (tabling with negation). XSB’s tabling implementation supports incremental tabling, as well as call and answer subsumption.

- A fully multi-threaded engine with thread-shared static code, and that allows dynamic code and tables to be thread-shared or thread-private. This engine fully supports the draft ISO standard for multi-threading [36].

- Constraint handling for tabled programs based on an engine-level implementation of annotated variables and various constraint packages, including clpqr for handling real constraints, and bounds a simple finite domain constraint library.

- A package for Constraint Handling Rules [30] which can be used to implement user-written constraint libraries.

- A variety of indexing techniques for asserted code including variable-depth indexing on several alternate arguments, fixed-depth indexing on combined arguments, trie-indexing.

- A set of mature packages, to extend XSB to evaluate F-logic [40] through the FLORA-2 package (distributed separately from XSB), to model check concurrent systems through the XMC system, to manage ontologies through the Cold
Dead Fish package, to support literate programming through the xsbdoc package, and to support answer set programming through the XASP package among other features.

- A number of interfaces to other software systems, such as C, Java, Perl, ODBC, SModels [54], and Oracle.
- Fast loading of large files by the load_dyn predicate, and by other means.
- A compiled HiLog implementation;
- Backtrackable updates through XSB’s storage module that support the semantics of transaction logic [6].
- Extensive pattern matching packages, and interfaces to libwww routines, all of which are especially useful for Web applications.
- A novel transformation technique called unification factoring that can improve program speed and indexing for compiled code;
- Macro substitution for Prolog files via the xpp preprocessor (included with the XSB distribution).
- Preprocessors and Interpreters so that XSB can be used to evaluate programs that are based on advanced formalisms, such as extended logic programs (according to the Well-Founded Semantics [2]); Generalized Annotated Programs [41].
- Source code availability for portability and extensibility under the GNU General Public Library License.

Though XSB can be used as a Prolog system, we avoid referring to XSB as such, because of the availability of SLG resolution and the handling of HiLog terms. These facilities, while seemingly simple, significantly extend its capabilities beyond those of a typical Prolog system. We feel that these capabilities justify viewing XSB as a new paradigm for Logic Programming. We briefly discuss some of these features; others are discussed in Volumes 1 and 2 of the XSB manual, as well as the manuals for various XSB packages such as FLORA, XMC, Cold Dead Fish, xsbdoc, and XASP.
Well-Founded Semantics  To understand the implications of SLG resolution [15], recall that Prolog is based on a depth-first search through trees that are built using program clause resolution (SLD). As such, Prolog is susceptible to getting lost in an infinite branch of a search tree, where it may loop infinitely. SLG evaluation, available in XSB, can correctly evaluate many such logic programs. To take the simplest of examples, any query to the program:

\[
\text{ancestor}(X,Y) \leftarrow \text{ancestor}(X,Z), \text{parent}(Z,Y).
\]

will terminate in XSB, since \textit{ancestor/2} is compiled as a tabled predicate; Prolog systems, however, would go into an infinite loop. The user can declare that SLG resolution is to be used for a predicate by using \texttt{table} declarations, as here. Alternatively, an \texttt{auto_table} compiler directive can be used to direct the system to invoke a simple static analysis to decide what predicates to table (see Section 3.10.5). This power to solve recursive queries has proven very useful in a number of areas, including deductive databases, language processing [42, 43], program analysis [21, 16, 7], model checking [57] and diagnosis [31]. For efficiency, we have implemented SLG at the abstract machine level so that tabled predicates will be executed with the speed of compiled Prolog. We finally note that for definite programs SLG resolution is similar to other tabling methods such as OLDT resolution [79] (see Chapter 5 for details).

Example 1.0.1 The use of tabling also makes possible the evaluation of programs with non-stratified negation through its implementation of the well-founded semantics [81]. When logic programming rules have negation, paradoxes become possible. As an example consider one of Russell’s paradoxes — the barber in a town shaves every person who does not shave himself — written as a logic program.

\[
\text{:- table shaves/2.}
\]

\[
\text{shaves(barber,Person)} \leftarrow \text{person(Person)}, \text{tnot(shaves(Person,Person))}.
\]

\[
\text{person(barber)}.
\]

\[
\text{person(mayor)}.
\]

\[
\text{Logically speaking, the meaning of this program should be that the barber shaves the mayor, but the case of the barber is trickier. If we conclude that the barber does not shave himself our meaning does not reflect the first rule in the program. If we conclude that the barber does shave himself, we have reached that conclusion using information}
\]
beyond what is provided in the program. The well-founded semantics, does not treat
shaves(barber,barber) as either true or false, but as undefined. Prolog, of course,
would enter an infinite loop. XSB’s treatment of negation is discussed further in
Chapter 5.

Multi-threading From Version 3.0 onward, XSB has been thoroughly revised to
support multi-threading using POSIX or Windows threads. Detached XSB threads
can be created to execute specific tasks, and these threads will exit when the query
succeeds (or fails, or throws an exception) and all thread memory reclaimed. While
a thread’s execution state is, of course, private, it shares many resources with other
threads, such as static code and I/O streams. Dynamic code and tables can be either
thread-shared or thread-private by default or by explicit declaration.

Constraint Support XSB supports logic-based constraint handling at a low level
through attributed variables and associated packages (e.g. setarg/3. In addition,
constraints may be handled through Constraint Handling Rules. Constraint logic
programs that use attributed variables may be tabled; those that use Constraint
Handling Rules may be efficiently tabled if the CHRd package is used. Constraint
programming in XSB is mainly covered in Volume 2.

Indexing Methods Data oriented applications may require indices other than Pro-
log’s first argument indexing. XSB offers a variety of indexing techniques for asserted
code. Clauses can be indexed on a group of arguments or on alternative arguments.
For instance, the executable directive index(p/4,[3,2+1]) specifies indexes on the
(outer functor symbol of) the third argument or on a combination of (the outer func-
tion symbol of) the second and first arguments. If data is expected to be structured
within function symbols and is in unit clauses, the directive index(p/4,trie) con-
structs an indexing trie of the p/4 clauses using a depth-first, left-to-right traversal
through each clause. Representing data in this way allows discrimination of informa-
tion nested arbitrarily deep within clauses. Advantages of both kinds of indexing can
be combined via star-indexing. Star-indexing indicates that up to the first 5 fields in
an argument will be used for indexing (the ordering of the fields is via a depth-first
traversal). For instance, index(p/4,[*(4),3,2+1]) acts as above, but looks within
4th argument of p/4 before examining the outer functor of argument 3 (and finally
examining the outer functors of arguments 2 and 1 together. Using such indexing,
XSB routinely performs efficiently intensive analyses of in-memory knowledge bases
with millions of highly structured facts. Indexing techniques for asserted code are
Interfaces  A number of interfaces are available to link XSB to other systems. In UNIX systems XSB can be directly linked into C programs; in Windows-based system XSB can be linked into C programs through a DLL interface. On either class of operating system, C functions can be made callable from XSB either directly within a process, or using a socket library. XSB can also inter-communicate with Java through the InterProlog interface \(^1\) or using YJXSB. Within Interprolog, XSB and Java can be linked either through Java’s JNI interface, or through sockets. XSB can access external data in a variety of ways: through an ODBC interface, through an Oracle interface, or through a variety of mechanisms to read data from flat files. These interfaces are all described in Volume 2 of this manual.

Fast Loading of Code  A further goal of XSB is to provide in implementation engine for both logic programming and for data-oriented applications such as in-memory deductive database queries and data mining \([63]\). One prerequisite for this functionality is the ability to load a large amount of data very quickly. We have taken care to code in C a compiler for asserted clauses. The result is that the speed of asserting and retracting code is faster in XSB than in any other Prolog system of which we are aware, even when some of the sophisticated indexing mechanisms described above are employed. At the same time, because asserted code is compiled into SLG-WAM code, the speed of executing asserted code in XSB is faster than that of many other Prologs as well. We note however, that XSB does not follow the ISO-semantics of assert \([46]\).

HiLog  XSB also supports HiLog programming \([13, 66]\). HiLog allows a form of higher-order programming, in which predicate “symbols” can be variable or structured. For example, definition and execution of generic predicates like this generic transitive closure relation are allowed:

\[
\text{closure}(R)(X,Y) :- R(X,Y).
\]

where \text{closure}(R)/2 is (syntactically) a second-order predicate which, given any relation \(R\), returns its transitive closure relation \text{closure}(R). XSB supports reading and writing of HiLog terms, converting them to or from internal format as necessary (see Section 4.2). Special meta-logical standard predicates (see Section 6.7) are also provided for inspection and handling of HiLog terms. Unlike earlier versions of XSB (prior to version 1.3.1) the current version automatically provides full compilation of HiLog predicates. As a result, most uses of HiLog execute at essentially the speed

\(^1\)InterProlog is available at [www.declarativa.com/InterProlog/default.htm](http://www.declarativa.com/InterProlog/default.htm).
of compiled Prolog. For more information about the compilation scheme for HiLog employed in XSB see [66].

HiLog can also be used with tabling, so that the program above can also be written as:

```prolog
:- hilog closure.
:- table apply/3.

closure(R)(X,Y) :- R(X,Y).
closure(R)(X,Y) :- closure(R)(X,Z), R(Z,Y).
```

as long as the underlying relations (the predicate symbols to which \( R \) will be unified) are also declared as Hilog. For example, if \( a/2 \) were a binary relation to which the closure predicate would be applied, then the declaration `:- hilog a.` would also need to be included.

**Unification Factoring** For compiled code, XSB offers unification factoring, which extends clause indexing methods found in functional programming into the logic programming framework. Briefly, unification factoring can offer not only complete indexing through non-deterministic indexing automata, but can also factor elementary unification operations. The general technique is described in [20], and the XSB directives needed to use it are covered in Section 3.10.

**XSB Packages** Based on these features, a number of sophisticated packages have been implemented using XSB. For instance, XSB supports a sophisticated object-oriented interface called Flora. Flora ([http://flora.sourceforge.net](http://flora.sourceforge.net)) is available as an XSB package and is described in its own manual, available from the same site from which XSB was downloaded. Another package, XMC [http://www.cs.sunysb.edu/~lmc](http://www.cs.sunysb.edu/~lmc) depends on XSB to perform sophisticated model-checking of concurrent systems. Within the XSB project, the Cold Dead Fish package supports maintenance of, and reasoning over ontologies; xsbdoc supports literate programming in XSB, and XASP provides an interface to Smodels to support Answer Set programming. XSB packages also support Perl-style pattern matching and POSIX-style pattern matching. In addition, experimental preprocessing libraries currently supported are Extended logic programs (under the well-founded semantics), and Annotated Logic Programs. These latter libraries are described in Volume 2 of this manual.
1.1 Using This Manual

We adopt some standard notational conventions, such as the name/arity convention for describing predicates and functors, + to denote input arguments, - to denote output arguments, ? for arguments that may be either input or output and # for arguments that are both input and output (can be changed by the procedure). See Section 3.10.5 for more details. Also, the manual uses UNIX syntax for files and directories except when it specifically addresses other operating systems such as Windows.

Finally, we note that XSB is under continuous development, and this document — intended to be the user manual — reflects the current status (Version 3.6) of our system. While we have taken great effort to create a robust and efficient system, we would like to emphasize that XSB is also a research system and is to some degree experimental. When the research features of XSB — tabling, HiLog, and Indexing Techniques — are discussed in this manual, we also cite documents where they are fully explained. All of these documents can be found without difficulty on the web.

While some of Version 3.6 is subject to change in future releases, we will try to be as upward-compatible as possible. We would also like to hear from experienced users of our system about features they would like us to include. We do try to accommodate serious users of XSB whenever we can. Finally, we must mention that the use of undocumented features is not supported, and at the user’s own risk.
Chapter 2

Getting Started with XSB

This section describes the steps needed to install XSB under UNIX and under Windows.

2.1 Installing XSB under UNIX

If you are installing on a UNIX platform, the version of XSB that you received may not include all the object code files so that an installation will be necessary. The easiest way to install XSB is to use the following procedure.

1. Decide in which directory in your file system you want to install XSB and copy or move XSB there.

2. Make sure that after you have obtained XSB, you have uncompressed it by following the instructions found in the file README.

3. Note that after you uncompress and untar the XSB tar file, a subdirectory XSB will be created in the current directory. All XSB files will be located in that subdirectory. In the rest of this manual, we use $XSB_DIR to refer to this subdirectory. Note the original directory structure of XSB must be maintained, namely, the directory $XSB_DIR should contain all the subdirectories and files that came with the distribution. In particular, the following directories are required for XSB to work: emu, syslib, cmplib, lib, packages, build, and etc.

4. Change directory to $XSB_DIR/build and then run these commands:
configure
makexsb

This is it!

In addition, it is now possible to install XSB in a shared directory (e.g., /usr/local) for everyone to use. In this situation, you should use the following sequence of commands:

configure -prefix=$SHARED_XSB
makexsb
makexsb install

where $SHARED_XSB denotes the shared directory where XSB is installed. In all cases, XSB can be run using the script

$XSB_DIR/bin/xsb

However, if XSB is installed in a central location, the script for general use is:

<central-installation-directory>/<xsb-version>/bin/xsb

**Important:** The XSB executable determines the location of the libraries it needs based on the full path name by which it was invoked. The “smart script” bin/xsb also uses its full path name to determine the location of the various scripts that it needs in order to figure out the configuration of your machine. Therefore, there are certain limitations on how XSB can be invoked.

Here are some legal ways to invoke XSB:

1. invoking the smart script bin/xsb or the XSB executable using their absolute or relative path name.

2. using an alias for bin/xsb or the executable.

3. creating a new shell script that invokes either bin/xsb or the XSB executable using their full path names.

Here are some ways that are guaranteed to not work in some or all cases:

1. creating a hard link to either bin/xsb or the executable and using it to invoke XSB. (Symbolic links should be ok.)
2. changing the relative position of either `bin/xsb` or the XSB executable with respect to the rest of the XSB directory tree.

The configuration script allows many different options to be specified. A full listing can be obtained by typing `$XSB_DIR/build/configure --help`.

**Type of Machine.** The configuration script automatically detects your machine and OS type, and builds XSB accordingly. On 64-bit platforms, the default compilation of XSB will reflect the default for the C compiler (e.g. gcc) on that platform. Moreover, you can build XSB for different architectures while using the same tree and the same installation directory provided, of course, that these machines are sharing this directory, say using NFS or Samba. All you will have to do is to login to a different machine with a different architecture or OS type, and repeat the above sequence of commands – or configure with different parameters.

The configuration files for different architectures reside in different directories, and there is no danger of an architecture conflict. In fact, you can keep using the same `./bin/xsb` script regardless of the architecture. It will detect your configuration and will use the right files for the right architecture!

If XSB is being built on a machine running Windows in which Cygwin is installed, Cygwin and Windows are treated as separate operating systems, as their APIs are completely different. If no previous configuration has been made, the configure script will attempt to use gcc and other Unix facilities, and therefore will compile the system under Cygwin. If this behavior is not desired, the option `-with-win` (equivalently, `-with-os=win`) uses a Window compiler and API. If a user wants to ensure the Cygwin compiler is used (say after a previous configuration for Windows), the option `-without-win` can be used. See Section 2.2.2 for more details.

**Choice of the C Compiler and compiler-related options** On Unix systems, XSB is developed and tested mainly using gcc. Accordingly, the configure script will attempt to use gcc, if it is available. Otherwise, it will revert to cc or acc. Some versions of gcc are broken for particular platforms or gcc may not have been installed; in which case you would have to give configure an additional directive `-with-cc` (or `-with-acc`). If you must use some special compiler, use `-with-cc=your-own-compiler`. You can also use the `-with-optimization` option to change the default C compiler optimization level. (or `-disable-optimization` to disable all compiler optimizations). `-enable-debug` is mainly a development option that allows XSB to be debugged using gdb – there are many other
compiler-based options options. Type configure -help to see them all. Also see the file $XSB_DIR/INSTALL for more details.

**Word Size** XSB’s configuration script checks whether the default compilation mode of a platform is 32- or 64-bits, and will build a version of XSB accordingly. Some platforms, however, support both 32-bit and 64-bit compilation. On such a platform, a user can explicitly specify the type of compilation using the options with-bits32 and with-bits64.

**XSB and Site-specific Information** Using the option -prefix=PREFIX installs architecture-independent files in the directory PREFIX, e.g. /usr/local, which can be useful if XSB is to be shared at a site. Using the option -site-prefix=DIR installs site-specific libraries in DIR/site. Other options indicate directories in which to search for site-specific static and dynamic libraries, and for include files.

**Multi-threading** Version 3.0 of XSB was the first version that supports multi-threading. On some platforms, the multi-threaded engine is slightly slower than the single-threaded engine, mostly due to its need for concurrency control. To obtain the benefits of multiple threads on a platform that supports either POSIX or Windows threads (i.e. nearly all platforms) users must configure XSB with the directive enable-mt (see Section 7.8 for instructions specific to Windows. The multi-threaded engine works with other configuration options, multi-threading can be compiled with batched or local scheduling, with the ODBC or Interprolog interfaces, and so on.

**Interfaces** Certain interfaces must be designated at configuration time, including those to Oracle, ODBC, Smodels, Tck/Tk, and Libwww. However, the XSB-calling-C interface interface does not need to be specified at configuration time. If you wish to use the InterProlog Java interface that is based on JNI, you must specify this at configuration time; otherwise if you wish to use the sockets-based Interprolog interface, it does not need to be specified at configuration time. See Volume 2 and the InterProlog site www.declarativa.com for details of specific interfaces

While the XSB configuration mechanism can detect most include and library paths, use of certain interfaces may require information about particular directories. In particular the -with-static-libraries option might be needed if compiling with support for statically linked packages (such as Oracle) or if your standard C libraries are in odd places. Alternately, dynamic libraries on odd places may need to be specified at configuration time using the -with-dynamic-libraries
option. and finally, the `–with-includes` option might be needed if your standard header files (or your jni.h file) are in odd places, or if XSB is compiled with ODBC support. Type `configure -help` for more details.

**Type of Scheduling Strategy.** The ordering of operations within a tabled evaluation can drastically affect its performance. XSB provides two scheduling strategies: Batched Evaluation and Local Evaluation. Local Evaluation ensures that, whenever possible, subgoals are fully evaluated before there answers are returned, and provides superior behavior for programs in which tabled negation is used. Batched Evaluation evaluates queries to reduce the time to the first answer of a query. Both evaluation methods can be useful for different programs. Since Version 2.4, Local Evaluation has been the default evaluation method for XSB. Batched Evaluation can be chosen via the `–enable-batched-scheduling` configure option. Detailed explanations of the scheduling strategies can be found in [28], and further experimentation in [11].

Other options are of interest to advanced users who wish to experiment with XSB, or to use XSB for large-scale projects. In general, however users need not concern themselves with these options.

### 2.1.1 Possible Installation Problems

**Lack of Space for Optimized Compilation of C Code** When making the optimized version of the emulator, the temporary space available to the C compiler for intermediate files is sometimes not sufficient. For example on one of our SPARC-stations that had very little `/tmp` space the "-04" option could not be used for the compilation of files `emuloop.c`, and `tries.c`, without changing the default `tmp` directory and increasing the swap space. Depending on your C compiler, the amount and nature of `/tmp` and swap space of your machine you may or may not encounter problems. If you are using the SUN C compiler, and have disk space in one of your directories, say `dir`, add the following option to the entries of any files that cannot be compiled:

```
-temp=dir
```

If you are using the GNU C compiler, consult its manual pages to find out how you can change the default `tmp` directory or how you can use pipes to avoid the use of temporary space during compiling. Usually changing the default directory can be done by declaring/modifying the `TMPDIR` environment variable as follows:

```
setenv TMPDIR dir
```
CHAPTER 2. GETTING STARTED WITH XSB

Missing XSB Object Files When an object (*.xwam) file is missing from the lib directories you can normally run the make command in that directory to restore it (instructions for doing so are given in Chapter 2). However, to restore an object file in the directories syslib and cmplib, one needs to have a separate Prolog compiler accessible (such as a separate copy of XSB), because the XSB compiler uses most of the files in these two directories and hence will not function when some of them are missing. For this reason, distributed versions normally include all the object files in syslib and cmplib.

XSB on 64-bit platforms XSB has been fully tested on 64-bit Debian Linux, 64-bit and Mac OS X. However, the sockets library may have problems in Version 3.6. If this limitation proves a problem, please contact xsb-development@lists.sourceforge.net.

Typically, if the 64-bit system generates 32-bit code by default, XSB will run just as in 32-bit mode (including 64-bit floats). 64-bit compilation can be forced for XSB by configuring with the option --with-bits64, and in a similar manner 32-bit compilation can be forced with the option --with-bits32. Users who employ either option should be aware of issues that may arise when linking XSB to external C code.

• When XSB calls C code the C file must have been compiled with the same memory option as XSB. This is done automatically if the C file is compiled via a call from XSB’s compiler, but must be handled by the user otherwise. For instance, if XSB were configured --with-bits32 on a 64-bit machine defaulting to 64-bits, then C files called by XSB require the -m32 option in gcc (if not compiled by XSB).

• The appropriate memory option must be used when embedding XSB into a C or Java process. For instance, if a XSB is to be linked into a 32-bit application on a 64-bit platform defaulting to 64-bits, XSB must be configured --with-bits32, and the linking of xsb.o/so to the calling program must specify -m32.

2.2 Installing XSB under Windows

2.2.1 Using Cygnus Software’s CygWin32

This is easy: just follow the Unix instructions. This is the preferred way to run XSB under Windows, because this ensures that all features of XSB are available.

\footnote{64-bit XSB was broken in a recent releases prior to Version 3.1 because for a time the developers did not have access to a 64-bit machine.}
2.2.2 Using Microsoft Visual C++

1. XSB will unpack into a subdirectory named \texttt{xsb}. Assuming that you have \texttt{XSB.ZIP} in the \texttt{$XSB\_DIR$} directory, you can issue the command

\begin{verbatim}
unzip386 xsb.zip
\end{verbatim}

which will install XSB in the subdirectory \texttt{xsb}.

2. If you decide to move XSB to some other place, make sure that the entire directory tree is moved — XSB executable looks for the files it needs relatively to its current position in the file system.

You can compile XSB under Microsoft Visual C++ compiler by following these steps:

1. Download the free of charge Microsoft Visual C++ Express Edition from

\begin{verbatim}
http://www.microsoft.com/express/vc/
\end{verbatim}

By default, this program is installed in \texttt{C:\Program Files\Microsoft Visual Studio 10.0}, and we shall assume this directory below (at the time of this writing, the latest version was 10.0, but the version number may change).

2. Go to Start Menu then Control Panel then System (depending on your version of Windows, the System panel might not be directly inside Control Panel, but one or two levels below. Then click “Change Settings,” select the “Advanced” tab, and then click the “Environment Variables” button. In the panel that is now selected, choose the PATH variable and click Edit. At the end of the string that represents the value of PATH, add

\begin{verbatim}
;C:\Program Files\Microsoft Visual Studio 10.0\VC\BIN
\end{verbatim}

On a 64-bit machine, add both of these:

\begin{verbatim}
;C:\Program Files\Microsoft SDKs\Windows\v7.1\bin
;C:\Program Files (x86)\Microsoft Visual Studio 10.0\VC\BIN
\end{verbatim}
Note: to compile XSB as 64 bit application you must install Microsoft Windows SDK found at http://msdn.microsoft.com/en-us/windows/bb980924.aspx. The version numbers, v7.1 and 10.0, may vary, of course.

Visual C++ has a command file called vcvars32.bat, which you should find and drag into the command window (and press Return). This will set all the necessary environment variables. On a 64 bit machine, this command file is called vcvarsx86_amd64.bat or vcvarsx86_ia64.bat — whichever is appropriate for your configuration. In Visual Studio Express 9.0, these files are in

C:\Program Files (x86)\Microsoft Visual Studio 9.0\VC\BIN\vcvars32.bat
C:\Program Files (x86)\Microsoft Visual Studio 9.0\VC\BIN\ia64\vcvarsx86_ia64.bat
C:\Program Files (x86)\Microsoft Visual Studio 9.0\VC\BIN\amd64\vcvarsx86_amd64.bat

At some point, Microsoft eliminated vcvarsx86_* in Visual Studio 10.0 and introduced SetEnv.cmd instead, requiring the users to download Microsoft Windows SDK. This command file is usually found in

C:\Program Files\Microsoft SDKs\Windows\v7.1\bin\SetEnv.cmd

As far as we know, SetEnv.cmd is the only file from the entire SDK that is necessary to build XSB as a 64-bit application. For 32 bit applications, the file

C:\Program Files (x86)\Microsoft Visual Studio 10.0\VC\BIN\vcvars32.bat

is still there and installation does not require the Windows SDK.

3. cd $XSB_DIR\build

4. On a 32 bit machine, type:

makexsb ["CFG=opt"] ["ORACLE=yes"] ["MY_LIBRARY_DIRS=libs"] ["MY_INCLUDE_DIRS=opts"]

- The items in square brackets are optional and usually are not necessary.
- The options for CFG are: release (default) or debug. The latter is used when you want to compile XSB with debugging enabled.
- The ORACLE parameter (default is “no”) compiles XSB with native support for Oracle DBMS. If ORACLE is specified, you must also specify the necessary Oracle libraries using the parameter SITE_LIBS. Native Oracle support is rarely used and ODBC is the recommended way to connect to databases.
• **MY_LIBRARY_DIRS** is used to specify the external libraries and **libs** there has the form /LIBPATH:"libdir1" /LIBPATH:"libdir2" ....

• **MY_INCLUDE_DIRS** is used to specify additional directories for included files. Here **opts** has the form /I"incdir1" /I"incdir2" ....

Instead of specifying the options on command line, it might be more convenient, however, to create the file

```
XSB\build\windows\custom_settings.mak
```

and put the options there. For instance,

```
XSB_INTERPROLOG=yes
MY_INCLUDE_DIRS=/I"C:\Program Files\Java\jdk1.6.0_26\include" \
   /I"C:\Program Files\Java\jdk1.6.0_26\include\win32"
MY_LIBRARY_DIRS=/LIBPATH:"C:\pthreads\pthreadVC1.lib" /libpath:"C:\oracle"
ORACLE=yes
```

5. The above command will compile XSB as requested and will put the XSB executable and its DLL in:

```
$XSB_DIR\config\x86-pc-windows\bin\xsb.exe
$XSB_DIR\config\x86-pc-windows\bin\xsb.dll
```

6. On a 64 bit machine, use **makexsb64** instead of **makexsb**. The compiled code will be installed in

```
$XSB_DIR\config\x64-pc-windows\bin\xsb.exe
$XSB_DIR\config\x64-pc-windows\bin\xsb.dll
```

The **custom_settings.mak** file must be in

```
XSB\build\windows64\custom_settings.mak
```

Make sure you do not misspell the name of that file or else none of the specified options will take effect!

**Note:** if you compiled XSB with one set of parameters and then want to recompile with a different set, it is recommended that you run

```
makexsb  clean
```

in between the compilations (or **makexsb64 clean** in the 64-bit case). This also applies to recompilations for 32/64 bits.
2.3 Invoking XSB

Under Unix, XSB can be invoked by the command:

\$XSB\_DIR/bin/xsb

if you have installed XSB in your private directory. If XSB is installed in a shared directory (e.g., \$SHARED\_XSB for the entire site (UNIX only), then you should use

\$SHARED\_XSB/bin/xsb

In both cases, you will find yourself in the top level interpreter. As mentioned above, this script automatically detects the system configuration you are running on and will use the right files and executables. (Of course, XSB should have been built for that architecture earlier.)

Under Windows, you should invoke XSB by typing:

\$XSB\_DIR/bin/xsb

This script tries to find the XSB executable and invoke it. If, for some reason, it fails to do so, the user should call the executable directly.

\$XSB\_DIR/config/x86-pc-windows/bin/xsb.exe

You may want to make an alias such as xsb to the above commands, for convenience, or you might want to put the directory where the XSB command is found in the \$PATH environment variable. However, you should not make hard links to this script or to the XSB executable. If you invoke XSB via such a hard link, XSB will likely be confused and will not find its libraries. That said, you can create other scripts and call the above script from there.

ISO “standard” Prolog predicates are supported by XSB, in addition to many other predicates: so those of you who consider yourselves champion entomologists, can try to test them for bugs now. Details are in Chapter 6.

2.4 Compiling XSB programs

One way to compile a program from a file, such as myfile.P in the current directory and load it into memory, is to type the query:
[my_file].

where my_file is the name of the file. Chapter 3 contains a full discussion of the compiling and consulting.

If you are eccentric (or you don’t know how to use an editor) you can also compile and load predicates input directly from the terminal by using the command:

[user].

A CTRL-d or the atom end_of_file followed by a period terminates the input stream.

2.5 Sample XSB Programs

There are several sample XSB source programs in the directory: $XSB_DIR/examples illustrating a number of standard features, as well as a number of non-standardized or XSB-specific features including plain tabling, incremental tabling, tabling with negation, attributed variables, annotated programs, constraint handling rules, XSB embedded in a C program, XSB calling C functions, sockets, and various semantic web appliation

Hence, a sample session might look like (the actual times shown below may vary and some extra information is given using comments after the % character):

my_favourite_prompt> cd $XSB_DIR/examples
my_favourite_prompt> $XSB_DIR/bin/xsb
XSB Version 3.1 (Incognito) of August 10, 2007
[i386-apple-darwin8.9.1; mode: optimal; engine: slg-wam; scheduling: local; word size: 32]
| ?- [queens].
[queens loaded]

yes
| ?- demo.

% ...... output from queens program .......

Time used: 0.4810 sec

yes
| ?- statistics.

memory (total) 1906488 bytes: 203452 in use, 1703036 free
permanent space 202552 bytes
2.6 Exiting XSB

If you want to exit XSB, issue the command `halt` or simply type `CTRL-d` at the XSB prompt. To exit XSB while it is executing queries, strike `CTRL-c` a number of times.
Chapter 3

System Description

Throughout this chapter, we use $XSB\_DIR$ to refer to the directory in which XSB was installed.

3.1 Entering and Exiting XSB from the Command Line

After the system has been installed, the emulator’s executable code appears in the file:

$XSB\_DIR/bin/xsb

If, after being built, XSB is later installed at a central location, $SHARED\_XSB$, the emulator’s executable code appears in

$SHARED\_XSB/bin/xsb

Either of these commands invokes XSB’s top-level interpreter, which is the most common way of using XSB.

XSB can also directly execute object code files from the command line interface. Suppose you have a top-level routine go in a file foo.P that you would like to run from the UNIX or Windows command line. As long as foo.P contains a directive, e.g. :- go., and foo.P has been compiled to an object file (foo.xwam), then

$XSB\_DIR/bin/xsb foo
will execute go (and any other directives), loading the appropriate files as needed. In fact the command $XSB_DIR/bin/xsb is equivalent to the command:

$XSB_DIR/bin/xsb -B $XSB_DIR/syslib/loader.xwam

There is one other way to execute XSB from a command line. Using the -e command-line option any goal can be be executed, up to 1024 characters. For instance

$XSB_DIR/bin/xsb -e "writeln(‘hello world’),halt."

writes “hello world” and exits XSB. Within the 1024 character limit, any query or command can be executed, including consulting files, so this method is actually quite general .

Various options can suppress XSB’s startup and end messages, as discussed below.

There are several ways to exit XSB. A user may issue the command halt. or end_of_file., or simply type CTRL-d at the XSB prompt. To interrupt XSB while it is executing a query, strike CTRL-c.

3.2 The System and its Directories

When installed, the XSB system resides in a single directory that contains several subdirectories. For completeness, we review the information in all subdirectories. Normally, only the documentation and files in the Prolog subdirectories, particularly examples, lib, and packages will be of interest to users.

1. bin contains scripts that call XSB executables for various configurations.

2. build contains XSB configuration scripts. You may already be familiar with the build directory, which is used to build XSB.

3. config contains executables and other files specific to particular configurations.

4. docs contains the user manuals and other documentation, including the technical documentation manual for developers.

In XSB, all extensions except ‘.pl’ and ‘.prolog’ — (default ‘.P’, ‘.H’, ‘.xwam’, ‘.D’ (output by mode inferencing), and ‘.A’ (assembly dump) — are defined in C and Prolog code using macros in $XSB_DIR/emu/extensions_xsb.h and can be changed by a user if desired. Of course, such a step should not be taken lightly, as it can cause severe compatibility problems.
CHAPTER 3. SYSTEM DESCRIPTION

5. **emu** contains the C source code for the XSB emulator, for I/O and for various interfaces.

6. **etc** contains miscellaneous files used by XSB.

7. **examples** contains some examples for Prolog, tabling, HiLog and various interfaces.

8. **cmplib** contains Prolog source and object code for the compiler.

9. **gpp** contains a copy of the Gnu pre-processor used to preprocess Prolog files.

10. **lib** contains Prolog source and object code for extended libraries.

11. **packages** The directory **packages** contains the various applications, such as FLORA, the XMC model checker and many others. These applications are written in XSB and can be quite useful, but are not part of the XSB system per se.

12. **Prolog_includes** contains include files for the Prolog libraries, which are preprocessed using GPP.

13. **syslib** contains Prolog source and object code for core XSB libraries.

All Prolog source programs are written in XSB, and all object (byte code) files contain SLG-WAM instructions that can be executed by the emulator. These byte-coded instructions are machine-independent, so usually no installation procedure is needed for the byte code files.

If you are distributing an application based on XSB and need to cut down space, the **packages**, **examples** and **docs** directories are not usually needed (unless of course you are using one of the packages in your application). **lib** may not be needed, (most core system files are in syslib) nor are Prolog source files necessary. Unless your application needs to rebuild XSB, the **emu** and **build** directories do not need to be distributed.

### 3.3 How XSB Finds Files: Source File Designators

Three files are associated with Prolog source code in XSB.

---

2Other types of files may be associated with foreign code — see Volume 2.
• A single source file, whose name is the base file name plus an optional extension suffix .P, .pl, or .prolog.

• An object (byte-code) file, whose name consists of the base file name plus the suffix .xwam.

• An optional header file, whose name is the base file name plus the suffix “.H”. When used, the header file normally contains file-level declarations and directives while the source file usually contains the actual definitions of the predicates defined in that module. However, such information can be equivalently put into the .P (.pl, or .prolog) file.

Most of the XSB system predicates for compiling, consulting, and loading code, such as consult/[1,2], compile/[1,2], load_dyn/1 and others are somewhat flexible in how they designate the file of interest. Each of these predicates take as input a source file designator which can be a base file name, a source file name; or the relative or absolute paths to a base or source file name. Unfortunately, the exact semantics of a file designator differs among system predicates in Version 3.6, as well as among platforms.

In general, however, when given a source file designator, system predicates perform name resolution. There are two steps to name resolution: determining the proper directory prefix and determining the proper file extension. When FileName is absolute (i.e. it contains a path from the file to the root of the file system) determining the proper directory prefix is straightforward. If FileName is relative, i.e. it contains a ‘/’ in Unix or ‘/’ in Windows, FileName is expanded to a directory prefix in an OS-dependent way, resolving symbols like ‘.’, ‘..’ and ‘˜’ when applicable. However, the user may also enter a name without any directory prefix. In this case, XSB tries to determine the directory prefix using a set of directories it knows about: those directories in the dynamic loader path (see Section 3.6). As it searches through directory prefixes, different forms of the file name may be checked. If the source file designator has no extension the loader first checks for a file in the directory with the .P extension, (or .c for foreign modules) before searching for a file without the extension, and finally for a file with a .pl or .prolog extension. Note that since directories in the dynamic loader path are searched in a predetermined order (see Section 3.6), if the same file name appears in more than one of these directories, the first one encountered will be used.
3.4 The Module System of XSB

XSB has been designed as a module-oriented Prolog system. Modules provide a small step towards logic programming “in the large” that facilitates the construction of large programs or projects from components that are developed, compiled and tested separately. Also, module systems support the principle of information hiding and can provide a basis for data abstraction. The module system of XSB is file based – one module per file – and flat – modules cannot be nested. In addition, XSB’s module system is to some extent atom-based, where any symbol in a module can be imported, exported or be a local symbol, as opposed to the predicate-based ones where this can be done only for predicate symbols\(^3\). As we will discuss, this leads to certain differences of XSB’s module system from those of some other Prologs, and to certain incompatibilities with the ISO standard for modules (which is not supported by most Prologs). At the same time, XSB’s module system has enough commonalities with those of other Prologs to be able to support Prolog commons libraries.

Module Syntax  By default, files are not treated as modules. In order for a file to be treated as a module, it must contain one or more module/2 or export/1 declarations, which specify that a set of symbols appearing in that module is visible and therefore can be used by any other module. In XSB, the module name must be equal to the base file name in which the module is defined. Any file (either module or not) may also contain use_module/2 or import/1 declarations, which allow symbols defined in and exported by other modules to be used in the current module. In addition, a module can also contain local declarations, which specify that a set of symbols is visible by this module only, and therefore cannot be accessed by any other module. Module declarations can appear anywhere in the source or header files and have the following forms:

\[
\begin{align*}
  &\text{:- export } \text{sym}_1, \ldots, \text{sym}_l. \\
  &\text{:- import } \text{sym}_1, \ldots, \text{sym}_n \text{ from } \text{module}. \\
  &\text{:- import } \text{sym} \text{ from } \text{module} \text{ as } \text{sym}'. \\
  &\text{:- local } \text{sym}_1, \ldots, \text{sym}_m.
\end{align*}
\]

where \(\text{sym}_i\) has the form functor/arity, and \(\text{module}\) is a Prolog atom representing a module name.

In XSB, the declaration

\(^3\)Operator symbols can be exported as any other symbols, but their precedence must be redeclared in the importing module.
:- module(filename,[sym1, ..., syml]).
can be seen as syntactic sugar for

:- export sym1, ..., syml.

as long as the filename is the same as the name of the file in which it was contained. Similarly,

:- use_module(module,[sym1, ..., syml]).
is treated as semantically equivalent to

:- import sym1, ..., symn from module.

Accordingly, use_module/2 and module/1 can be used interchangibly with import/2 and export/1. However the declaration

:- use_module(module).

which is often used in other Prolog systems, is not equivalent to an XSB import statement, as each XSB import statement must explicitly declare a list of predicates that are used from each module. Such a declaration will raise a compilation error.

The declaration

:- import sym from module as sym’.

allows a predicate to be imported from a module, but renamed as sym’ within the importing module. Such a feature is useful when porting a library written for another Prolog (e.g. a constraint library) to XSB.

For modules, the base file name is stored in its byte code file, so that renaming a byte-code file for a mule may cause problems, as the renaming will not affect the information within the byte-code file. However, byte code files generated for non-modules can be safely renamed.

Module Semantics  In XSB’s atom-based module system, the name of each predicate and function symbol p/n is identified as if it were prefixed with its module name (i.e. base file name). Hence the occurrence of p/n in two different modules, m1 and m2 are distinct symbols that can be denoted as m1:p/n and m2:p/n.

Normally, only exported symbols can be imported; if a non-exported symbol p/2 is imported from a module m1 by module m2 an environment conflict warning will be issued as soon as m1 and m2 are loaded in the same session – i.e. the conflict is detected at run-time. When a non-module file is loaded, its predicates and symbols are loaded into the module usermod, which is the working module of the XSB command-line interpreter and C-calling XSB interface. Dynamically asserted code is also loaded.
A predicate symbol \( p/n \) is defined in a module \( m \) if \( m \) contains a clause with head \( p/n \) or a dynamic declaration for \( p/n \). Any predicate symbol \( p/n \) defined in a module \( m \), whether exported or not, can be called by prepending the module prefix using the \( :/2 \) functor, e.g. \( m:p(A,\ldots) \). For brevity, we call this an explicit module call to \( p/n \). The following example illustrates these principles.

<table>
<thead>
<tr>
<th>Exported and Non-Exported Predicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m1 )</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>X = a</td>
</tr>
<tr>
<td>Y = b</td>
</tr>
<tr>
<td>:- export p/2.</td>
</tr>
<tr>
<td>exported(a,b).</td>
</tr>
<tr>
<td>/* Existence Error */</td>
</tr>
<tr>
<td>local(c,d).</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>X = c</td>
</tr>
<tr>
<td>Y = d</td>
</tr>
</tbody>
</table>

Every predicate symbol defined in a module is assumed by default to be local to a module unless it is declared otherwise by an export or import declaration. Symbols that are local to a given module are not visible to other modules except through explicit module calls. The following example shows how different declarations for dynamic predicates within a module may be global (usermod) or local. Calls to statically defined predicates behave similarly.
## Visibility of Dynamic Predicates

<table>
<thead>
<tr>
<th>m1</th>
<th>m2</th>
</tr>
</thead>
<tbody>
<tr>
<td>:- dynamic d1/2.</td>
<td>:- dynamic d1/2.</td>
</tr>
<tr>
<td>:- import d2/2 from usermod.</td>
<td>:- import d2/2 from usermod.</td>
</tr>
<tr>
<td>p10:- a1.</td>
<td>X = a</td>
</tr>
<tr>
<td>p1(X,Y):- assert(d1(X,Y)).</td>
<td>Y = b</td>
</tr>
<tr>
<td>p2(X,Y):- d1(X,Y).</td>
<td>yes</td>
</tr>
<tr>
<td>p2(X,Y):- assert(d2(X,Y)).</td>
<td>X = a</td>
</tr>
<tr>
<td>p3(X,Y):- d2(X,Y).</td>
<td>Y = b</td>
</tr>
<tr>
<td>X = 1</td>
<td></td>
</tr>
<tr>
<td>Y = 2</td>
<td></td>
</tr>
<tr>
<td>X = 1</td>
<td></td>
</tr>
<tr>
<td>Y = 2</td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>?- [m1].</th>
</tr>
</thead>
<tbody>
<tr>
<td>[m1 loaded]</td>
</tr>
</tbody>
</table>

| ?- p1(a,b). |
| yes |

| ?- p2(X,Y). |
| X = a |
| Y = b |

| /* Existence Error */ |
| ?- m1:d1(X,Y). |

| ?- d1(X,Y). |
| yes |

| ?- p3(1,2). |
| yes |

| ?- p4(X,Y). |
| X = 1 |
| Y = 2 |

| ?- d2(X,Y). |
| X = 1 |
| Y = 2 |

| yes |
Functors that occur as literals in the bodies of clauses, are treated as predicate symbols.

- Standard predicates are taken to be a part of usermod, and are implicitly imported into user-defined modules. Standard predicates include ISO predicates along with many other XSB predicates for tabling, indexing and other functions. The current listing of standard predicates can be found in the index of this manual under Standard predicates.

- Other predicates are taken to be local to the module in which they occur.

Functors that do not occur as literals in the body of clauses in a module are taken to be structure symbols. These symbols are assumed to be global and do not require an explicit module call to be used, unless declared otherwise through a local/1 declaration. In addition, terms that are dynamically created by standard predicates such as read/1, functor/3, ’=..’/2, etc) are taken to be structure symbols and are contained in usermod.

All atoms are assumed to be global and do not require an explicit module call to be used. This can occasionally lead to unexpected results if a token is used both as an atom and a 0-ary function symbol. In the following table, the query \( ?- p10 \) will call \( a1/0 \), while \( ?- p11 \). will throw an existence error.

<table>
<thead>
<tr>
<th>Atoms and 0-ary Predicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
</tr>
<tr>
<td>:- export p10/0, p11/0.</td>
</tr>
<tr>
<td>:- import a1/0 from m2.</td>
</tr>
<tr>
<td>p10:- a1.</td>
</tr>
<tr>
<td>p11:- atom_chars(A1,[a,’1’]), call(A1).</td>
</tr>
</tbody>
</table>

For clarity, we state a few consequences of these rules.

- In Version 3.6, a module cannot export predicate symbols that are imported from other modules. This happens because an import declaration is considered a request for permission to use a symbol from a module where its definition and an export declaration appear.
• The implicit module for a particular symbol appearing in a module must be uniquely determined. As a consequence, a symbol of a specific \textit{functor/arity cannot} be declared as both exported and local, or (as just discussed) both exported and imported from another module, or declared to be imported from more than one module, etc. These types of environment conflicts are detected at compile-time and abort the compilation.

• If a module \texttt{m1} imports a predicate \texttt{p/n} from a module \texttt{m2}, but \texttt{m2} does not export \texttt{p/n}, nothing is detected at the time of compilation. As discussed above, if \texttt{p/n} is defined in \texttt{m2} a runtime warning about an environment conflict will be issued. However, if \texttt{p/n} is not defined in \texttt{m2}, a runtime existence error will be thrown \textsuperscript{4}.

• Only one definition of a symbol \texttt{p/n} can appear in a module, without being explicitly associated with a module using the \texttt{:/2} functor. Accordingly only one default definition of \texttt{p/n} can be loaded into the interpreter’s module (\texttt{usermod}). An attempt to load a module that redefines \texttt{p/n} results in a warning to the user and the newly loaded symbol \textit{redefines} the definition of the previously loaded one.

\textbf{Usage inference and the module system} \hspace{1em} The import and export statements of a module \textit{M} are used by the compiler for inferring usage of predicates. At compilation time, if a predicate \textit{P/N} occurs as callable in the body of a clause defined in \textit{M}, but \textit{P} is neither defined in \textit{M} nor imported into \textit{M} from some other module, a warning is issued that \textit{P/N} is undefined. Here “occurs as callable” means that \textit{P/N} is found as a literal in the body of a clause, or within a system meta-predicate, such as \texttt{assert/1, findall/3}, etc. Currently, occurrences of a term inside user-defined meta-predicates are not considered as callable by XSB’s usage inference algorithm. Alternatively, if \textit{P/N} is defined in \textit{M}, it is \textit{used} if \textit{P/N} is exported by \textit{M}, or if \textit{P/N} occurs as callable in a clause for a predicate that is used in \textit{M}. The compiler issues warnings about all unused predicates in a module. On the other hand, since all modules are compiled separately, the usage inference algorithm has no way of checking whether a predicate imported from a given module is actually exported by that module.

Usage inference can be highly useful during code development for ensuring that all predicates are defined within a set of files, for eliminating dead code, etc. In addition, import and export declarations are used by the \texttt{xsbdoc} documentation system to generate manuals for code \textsuperscript{5}. For these reasons, it is sometimes the case that usage

\textsuperscript{4}This behavior can be altered through the Prolog flag \texttt{unknown}.

\textsuperscript{5}Further information on \texttt{xsbdoc} can be found in \texttt{$XSB\_DIR/packages/xsbdoc$}.
inference is desired even in situations where a given file is not ready to be made into a module, or it is not appropriate for the file to be a module for some other reason. In such a case the directives `document_export/1` and `document_import/1` can be used, and have the same syntax as `export/1` and `import/1`, respectively. These directives affect only usage inference and `xsbdoc`. A file is treated as a module if and only if it includes an `export/1` statement, and only `import/1` statements affect dynamic loading and name resolution for predicates.

### 3.5 Standard Predicates in XSB

Whenever XSB is invoked, a large set of standard predicates are defined and can be called from the interpreter or other interface. These predicates include the various ISO predicates, along with predicates for tabling, I/O, for interaction with the operating system, for HiLog, and for much other functionality. Standard predicates are listed in this manual under the index heading `Standard predicates` and at an implementation level are declared in the file `$XSB_DIR/syslib/std_xsb.P`. If a user wishes to redefine a standard predicate, she has several choices. First, the appropriate fact in `$XSB_DIR/syslib/std_xsb.P` should be commented out. Once this is done, a user may define the predicate as any other user predicate. Alternately, the compiler option `allow_redefinition` can be used to allow the compiler to redefine a standard predicate (Section 3.10.2). If a user wants to make a new definition or new predicate standard, the safest course is to put the predicate into a module in the `lib` directory, and add or modify an associated fact in `$XSB_DIR/syslib/std_xsb.P`.

### 3.6 The Dynamic Loader and its Search Path

XSB differs from some other Prolog system in its ability to dynamically load modules. In XSB, the loading of user modules and Prolog libraries (such as the XSB compiler) is delayed until predicates in them are actually needed, saving program space for large Prolog applications. Dynamic loading is done by default, unlike other systems where it is not the default for non-system libraries.

When a predicate imported from another module (see Section 3.4) is called during execution, the dynamic loader is invoked automatically if the module is not yet loaded into the system. The default action of the dynamic loader is to search for the byte code file of the module first in the system library directories (in the order `lib, syslib`,

---

6Such predicates are sometimes called “built-ins” in other Prologs.
and then cmplib), and finally in the current working directory. If the module is found in one of these directories, then it will be loaded (on a first-found basis). Otherwise, an error message will be displayed on the current error stream reporting that the module was not found. Because system modules are dynamically loaded, the time it takes to compile a file is slightly longer the first time the compiler is invoked in a session than for subsequent compilations.

### 3.6.1 Changing the Default Search Path and the Packaging System

The default search path of the dynamic loader is based on the dynamic predicate library_directory/1 so it can easily be changed. For instance, to make sure a user's home directory is loaded, the goal `assert(library_directory('~/'))` needs to be executed from the command line or from within a program. If you always want XSB to search particular directories, the easiest way is to have a file named `.xsb/xsbc.P` in the user's home directory. User-supplied library directories are searched by the dynamic loader before searching the default library directories. The `.xsb/xsbc.P` file, which is automatically consulted by the XSB interpreter, might look like the following:

```
:- assert(library_directory('~/')).
:- assert(library_directory('/usr/lib/sbprolog')).
```

After loading the module of the above example the user's home directory is searched first, then "/usr/lib/sbprolog/", and finally XSB's system library directories (lib, syslib, cmplib) as well as the current working directory. XSB also uses library_directory/1 for internal purposes. For instance, before the user's `.xsb/xsbc.P` is consulted, XSB puts the packages directory and the directory `.xsb/config/$CONFIGURATION` on the library search path. The directory `.xsb/config/$CONFIGURATION` is used to store user libraries that are machine or OS dependent. ($CONFIGURATION for a machine is something that looks like `sparc-sun-solaris2.6` or `pc-linux-gnu`, and is selected by XSB automatically at run time). If a user wished, say, to search the current working directory before her home directory, she could simply add

```
:- asserta(library_directory('./')).
```
to her .xsb/xsbrc.P file (or anywhere else). The file .xsb/xsbrc.P is not limited to setting the library search path. In fact, arbitrary Prolog code can go there so that XSB can be initialized in any manner desired.

We emphasize that in the presence of a .xsb/xsbrc.P file it is the user’s responsibility to avoid module name clashes with modules in XSB’s system library directories. Such name clashes can cause unexpected behavior as system code may try to load a user’s predicates. The list of module names in XSB’s system library directories can be found by looking through the directories $XSB_DIR/{syslib,cmplib,lib}.

Apart from the user libraries, XSB now has a simple packaging system. A package is an application consisting of one or more files that are organized in a subdirectory of one of the XSB system or user libraries. The system directory $XSB_DIR/packages has a number of examples of such packages, many of which are documented in Volume 2 of this manual, or contain their own manuals. Packages are convenient as a means of organizing large XSB applications, and for simplifying user interaction with such applications. User-level packaging is implemented through the predicate

\[
\text{bootstrap_userpackage}(+LibraryDir, +PackageDir, +PackageName).
\]

which must be imported from the packaging module.

To illustrate, suppose you wanted to create a package, foobar, inside your own library, my_lib. Here is a sequence of steps you can follow:

1. Make sure that my_lib is on the library search path by putting an appropriate assert statement in your xsbrc.P.

2. Make a subdirectory ~/my_lib/foobar and organize all the package files there. Designate one file, say, foo.P, as the entry point, i.e., the application file that must be loaded first.

3. Create the interface program ~/my_lib/foobar.P with the following content:

\[
:\text{- bootstrap_userpackage('~/my_lib', 'foobar', foobar), [foo].}
\]

The interface program and the package directory do not need to have the same name, but it is convenient to follow the above naming schema.

4. Now, if you need to invoke the foobar application, you can simply type [foobar]. at the XSB prompt. This is because both and ~/my_lib/foobar have already been automatically added to the library search path.
5. If your application files export many predicates, you can simplify the use of your package by having ~/$my_lib/foobar.P import all these predicates, renaming them, and then exporting them. This provides a uniform interface to the foobar module, since all the package predicates are can now be imported from just one module, foobar.

In addition to adding the appropriate directory to the library search path, the predicate bootstrap_userpackage/3 also adds information to the predicate package_configuration/3, so that other applications could query the information about loaded packages.

Packages can also be unloaded using the predicate unload_package/1. For instance,

```prolog
:- unload_package(foobar).
```

removes the directory ~/$my_lib/foobar from the library search path and deletes the associated information from package_configuration/3.

Finally, if you have developed and tested a package that you think is generally useful and you would like to distribute it with XSB, please contact xsb-development@sourceforge.net.

### 3.6.2 Dynamically loading predicates in the interpreter

Modules are usually loaded into an environment when they are consulted (see Section 3.9). Specific predicates from a module can also be imported into the run-time environment through the standard predicate import PredList from Module. Here, PredList can either be a Prolog list or a comma list. (The import/1 can also be used as a directive in a source module (see Section 3.4).

We provide a sample session for compiling, dynamically loading, and querying a user-defined module named quick_sort. For this example we assume that quick_sort.P is a file in the current working directory, and contains the definitions of the predicates concat/3 and qsort/2, both of which are exported.

```prolog
| ?- compile(quick_sort).
[Compiling ./quick_sort]
[quick_sort compiled, cpu time used: 1.439 seconds]
yes
| ?- import concat/3, qsort/2 from quick_sort.
yes
```
| ?- concat([1,3], [2], L), qsort(L, S).

L = [1,3,2]
S = [1,2,3]

yes.

The standard predicate import/1 does not load the module containing the imported predicates, but simply informs the system where it can find the definition of the predicate when (and if) the predicate is called.

### 3.7 Command Line Arguments

There are several command line options for the emulator. The general synopsis obtained by the command `$XSB_DIR/bin/xsb -help` is:

```
xsb [flags] [-l]
xsb [flags] module
xsb [flags] -B boot_module [-D cmd_loop_driver] [-t]
xsb [flags] -B module.suffix -d
xsb [-h | -v | --help | --version]
```

**module:**

- Module to execute after XSB starts up.
- Module should have no suffixes, and either be an absolute pathname or the file module.xwam must be on the library search path.

**boot_module:**

- This is a developer’s option.
- The -B flags tells XSB which bootstrapping module to use instead of the standard loader. The loader must be specified using its full pathname, and boot_module.xwam must exist.

**module_to_disassemble:**

- This is a developer’s option.
- The -d flag tells XSB to act as a disassembler.
- The -B flag specifies the module to disassemble.

**cmd_loop_driver:**

- The top-level command loop driver to be used instead of the standard one. Usually needed when XSB is run as a server.

- B : specify the boot module to use in lieu of the standard loader
- D : Sets top-level command loop driver to replace the default
CHAPTER 3.  SYSTEM DESCRIPTION

- `t`: trace execution at the SLG-WAM instruction level
  (for this to work, build XSB with the --debug option)
- `d`: disassemble the loader and exit
- `v`, `--version`: print the version and configuration information about XSB
- `h`, `--help`: print this help message

Flags:

- `e` `goal`: evaluate goal when XSB starts up
- `p`: enable Prolog profiling through use of `profile_call/1`
- `l`: the interpreter prints unbound variables using letters
- `--nobanner`: don’t show the XSB banner on startup
- `--quietload`: don’t show the ‘module loaded’ messages
- `--noprompt`: don’t show prompt (for non-interactive use)
- `S`: set default tabling method to call-subsumption
- `--max_subgoal_depth N`: set maximum tabled subgoal depth to N (default is maximum integer)
- `--max_subgoal_action A`: set action on maximum subgoal depth: e(rror)/a(bstract)/w(arn)
- `--max_tries N`: allow up to N tries for interning terms
- `--max_threads N`: maintain information for up to N threads (MT engine only)
- `--max_mqueues N`: allow up to N message queues (MT engine only)
- `--shared_predicates`: make predicates thread-shared by default
- `g gc_type`: choose heap garbage collection ("indirection","none" or "copying")
- `c N` [unit]: initially allocate N units (default KB) for the trail/choice-point stack
- `m N` [unit]: initially allocate N units (default KB) for the local/global stack
- `o N` [unit]: initially allocate N units (default KB) for the SLG completion stack
- `-r`: turn off automatic stack expansion
- `-T`: print a trace of each called predicate

unit: k/K memory in kilobytes; m/M in megabytes; g/G in gigabytes

Command-line Options  These options tend to be most useful for developers.

- `t` Traces through code at SLG-WAM instruction level. This option is intended for
  developers and is not fully supported. It is also not available when the system
  is being used at the non-debug mode (see Section 10).

- `D` Tells XSB to use a top-level command loop driver specified here instead of the
  standard XSB interpreter. This is most useful when XSB is used as a server.

- `d` Produces a disassembled dump of `byte_code_file` to `stdout` and exits.

Flags  The order in which flags appear makes no difference.
General Flags

-e goal  Pass goal to XSB at startup. This goal is evaluated right before the first prompt is issued. For instance, `xsb -e "write(Hello!), nl."` will print a heart-warming message when XSB starts up.

-p Enables the engine to collect information that can be used for profiling. See Volume 2 of this manual for details.

-l Forces the interpreter to print unbound variables as letters, as opposed to the default setting which prints variables as memory locations prefixed with an underscore. For example, starting XSB’s interpreter with this option will print the following:

```

Y = A
X = A
Z = 3
W = foo(A,3)
```
as opposed to something like the following:

```

Y = _h118
X = _h118
Z = 3
W = foo(_h118,3);
```

-nobanner  Start XSB without showing the startup banner. Useful in batch scripts and for interprocess communication (when XSB is launched as a subprocess). For instance,

`xsb -e "writeln(’hello world’),halt."
[xsb_configuration loaded]
[sysinitrc loaded]

XSB Version 3.1 (Incognito) of August 10, 2007
[i386-apple-darwin8.9.1; mode: optimal; engine: slg-wam; scheduling: local; word size: 32]

Evaluating command line goal:
| ?- writeln(’hello world’),halt.
| ?- hello world

End XSB (cputime 0.02 secs, elapsetime 0.02 secs)

Prints out quite a bit of verbiage. Using the \texttt{--nobanner} option reduces this verbiage somewhat.

\begin{verbatim}
xsb --nobanner -e "writeln('hello world'),halt."
[xsb_configuration loaded]
[sysinitrc loaded]
\end{verbatim}

Evaluating command line goal:
| ?- writeln('hello world'),halt.

| ?- hello world

\texttt{--quietload} Do not tell when a new module gets loaded. Again, is useful in non-interactive activities and for interprocess communication. Continuing our example:

\begin{verbatim}
xsb --quietload --nobanner -e "writeln('hello world'),halt."
| ?-
| ?- hello world
\end{verbatim}

\texttt{--noprompt} Do not show the XSB prompt.

\texttt{--nofeedback} Do not print the feedback messages such as \texttt{"yes"} and \texttt{"no"} after queries. This and the \texttt{--noprompt} options are useful only in batch mode and in interprocess communication when you do not want the prompt to clutter the picture. Putting all this together, we finally get:

\begin{verbatim}
xsb --noprompt --quietload --nobanner --nofeedback -e "writeln(hello),halt."
\end{verbatim}

\texttt{hello world}

So that XSB can be used to write reasonable scripts.

\texttt{--max_threads N} Allows XSB to maintain information for up to \texttt{N} threads. This means that XSB can currently run \texttt{N} threads that are active, or that are inactive, non-detached, and not yet joined. Has no effect if the engine has been configured without multi-threading.

\texttt{-S} Indicates that tabled predicates are to be evaluated using subsumption-based tabling as a default for tabled predicates whose tabling method is not
specified by using `table Predspec as subsumptive` or `table Predspec as variant` (see Section 6.15.1). If this option is not specified, variant-based tabling will be used as the default tabling method by XSB.

`-shared_predicates` In the multi-threaded engine, makes all predicates thread-shared by default; has no effect in the single-threaded engine.

`-T` Generates a trace at entry to each called predicate (both system and user-defined). This option is available mainly for people who want to modify and/or extend XSB, and it is not the normal way to trace XSB programs. For the latter, the standard predicates `/trace/0 or debug/0` should be used (see Chapter 10). Note: This option is not available when the system is being used at the non-tracing mode (see Section 10).

`--max_subgoal_depth N` : set maximum tabled subgoal depth to $N$ (default is maximum integer). This flag sets the depth of a subgoal upon which an action may be taken (such as throwing an error, abstracting, or issuing a warning).

`--max_subgoal_action A` : set action on maximum subgoal depth: e(rror)/a(bstract)/w(arn)

**Memory Management Flags**

`-g gc_type` Chooses the heap garbage collection strategy that is employed; choice of the strategy is between the default **indirection; copying**, which is not fully supported; or **none**. See [10] for a description of the indirection garbage collector, and [23] for the copying garbage collector.

`-c size [units]` Allocates *initial size* units of space to the trail/choice-point stack area. The trail stack grows upward from the bottom of the region, and the choice point stack grows downward from the top of the region. If units is not provided or is `k` or `K`, the size is allocated in kilobytes; if `m` or `M` in megabytes; and if `g` or `G` in gigabytes. Because this region is expanded automatically, this option is rarely needed. If this option is not specified the default initial size is 768 KBytes.

`-m size [units]` Allocates *initial size* units of space to the local/global stack area. The global stack grows upward from the bottom of the region, and the local stack grows downward from the top of the region. If units is not provided or is `k` or `K`, the size is allocated in kilobytes; if `m` or `M` in megabytes; and if `g` or `G` in gigabytes. Because this region is expanded automatically, this option is rarely needed. If this option is not specified the default initial size is 768 KBytes.

`-o size [units]` Allocates *initial size* units of space to the completion stack area. If units is not provided or is `k` or `K`, the size is allocated in kilobytes;
if \( m \) or \( M \) in megabytes; and if \( g \) or \( G \) in gigabytes. Because this region is expanded automatically, this option is rarely needed. If this option is not specified the default initial size is 768 KBytes.

- \( -r \) Turns off automatic stack expansion. This can occasionally be useful for isolating memory management problems.

### 3.8 Memory Management

All execution stacks are automatically expanded in Version 3.6, including the local stack/heap region, the trail/choice point region, and the completion stack region. Execution stacks increase their size until it is not possible to do so with available system memory. At that point XSB tries to find the maximal amount of space that will still fit in system memory. For the main thread, each of these regions begin with an initial value set by the user at the command-line or with a default value (see Section 3.7). When a thread is created within an XSB process, the size of the thread’s execution stacks may be set by \texttt{thread_create/3}, otherwise the default values indicated in Section 3.7 are used. Once XSB is running, these default values may be modified using the appropriate Prolog flags (see Section 6.12). In addition, whenever a thread exits, memory specific to that thread is reclaimed.

Heap garbage collection is automatically included in XSB [10, 23]. (To change the algorithm used for heap garbage collection or to turn it off altogether, see the predicate \texttt{garbage_collection/1} or Section 3.7 for command-line options). In Version 3.6 the default behavior is indirect garbage collection. Starting with Version 3.0, heap garbage collection may automatically invokes garbage collection of XSB’s “string” table, which stores Prolog’s atomic constants. Expansion and garbage collection of execution stacks can occur when multiple threads are active; however atom garbage collection will not be invoked if there is more than one active XSB thread.

The program area (the area into which XSB byte-code is loaded) is also dynamically expanded as needed. For dynamic code (created using \texttt{assert/1}, or standard predicates such as \texttt{load_dyn/1} and \texttt{load_dyn/1}) index size is also automatically reconfigured. Space reclaimed for dynamic code depends on several factors. If there is only one active thread, space is reclaimed for retracted clauses and abolished predicates as long as (1) there are no choice points that may backtrack into the retracted or abolished code, and (2) if the dynamic predicate is tabled, all of its tables are completed. Otherwise, the code is marked for later garbage collection. If more than one thread is active, private predicates behave as just described, however space reclamation for shared predicates will be delayed until there is a single active thread. See
Section 6.14 for details.

Space for tables is dynamically allocated as needed and reclaimed through use of `abolish_all_tables/0, abolish_table_pred/1, abolish_table_call/1` and other predicates. As with dynamic code, space for tables may be reclaimed immediately or marked for later garbage collection depending on whether choice points may backtrack into the abolished tables, on the number of active threads, etc. Tabling also includes various stacks used to copy information into or out of tables, most of which are dynamically allocated and expanded. These stacks may be thread-private or shared among threads: space for thread-private stacks is reclaimed when a thread exits. See Section 6.15.3 for details.

Perhaps more than a standard Prolog system, XSB is used to evaluate queries in knowledge representation languages that have a higher level of declarativity than Prolog and as a result may consume a great deal of space. If XSB needs memory that is unobtainable from the operating system, it will usually abort with a resource error, and become ready for a new query from its command line or API. In such a case, a user or program can use `statistics/[0,1,2]` to investigate whether and how XSB is consuming memory. Other options to bounding memory include the use of `bounded_call/4` or the use of the `max_memory` flag. Use of the `max_memory` flag is recommended in cases where XSB is embedded in a C program through the C/XSB interface, or is embedded in or communicating with a java program through Interprolog. In such a case, XSB will abort with a resource error whenever a memory allocation would exceed the user-defined threshold.

### 3.9 Compiling, Consulting, and Loading

Like other Prologs, XSB provides for both statically compiled code and dynamically asserted code. Static compiled code may be more optimized than asserted code, particularly for clauses that have large bodies, but certain types of indexing, such as trie and star indexing are (currently) available only for dynamically asserted predicates (see `index/2`).

#### 3.9.1 Static Code

In XSB, there is no difference between compiled and consulted static code: “compiling” in XSB means creation of a file containing SLG-WAM byte-code; “consulting”

---

7In rare cases, XSB will exit if the inability to allocate more memory will leave it in an inconsistent state (e.g. if XSB cannot allocate needed memory during heap garbage collection).
means loading such a byte-code file, after compiling it (if the source file was altered later than the object file).

consult(+Files,+OptionList)
consult(+Files)
[+Files]

The standard predicate consult/[1,2] is the most convenient method for entering static source code rules into XSB’s database. Files is either a source file designator (see Section 3.3) or a list of source file designators, and Options is a list of options to be passed to XSB’s compiler if the file needs to be compiled (see Section 3.10). consult(Files) is defined as consult(Files,[]), as is [Files].

Consulting a file File (module) conceptually consists of the following five steps which are described in detail in the following paragraphs.

Name Resolution: determine the file that File designates, including directory and drive location and extension, as discussed in Section 3.3.

Compilation: if the source file or header has changed later than the object file (or if there is no byte-code file) compile the file using compile/2 with the options specified, creating a byte-code file. This strategy is used whether the source file is Prolog, C, or C++.

Loading: load the byte-code file into memory.

Importing: if the file is a module, import any exported predicates of that module to usermod.

Query Execution: execute any queries that the file may contain, i.e. any terms with principal functor ’?-’/1, or with the principal functor ’:-’/1 and that are not directives like the ones described in Section 3.10. The queries are executed in the order in which they appear in the source file.

Error conditions for consult(+File,+Options) are as follows:

- File is not instantiated
  - instantiation_error
- File is not an atom
  - type_error(atom,File)

\(^8\)In XSB, reconsult/[1,2] is defined to have the same actions as consult/[1,2].
• File does not exist in the current set of library directories
  – existence_error(file,File)
• File has an object code extension (e.g. .xwam)
  – permission_error(compile,file,File)
• File has been loaded previously in the session and there is more than one active thread.
  – misc_error

Error conditions of compiler options are determined by compile/2 which consult/[1,2] calls.

In addition, ensure_loaded/[1,2] acts much like consult/[1,2]

ensure_loaded(+FileName)  
ISO
This predicate checks to see whether the object file for FileName is newer than the source code and header files for FileName, and compiles FileName if not. If FileName is loaded into memory, ensure_loaded/1 does not reload it, unlike consult/1 which will always reload. In addition, ensure_loaded/2 can be used to load a file with dynamic code. It is fully documented in Section 6.14.1.

3.9.2 Dynamic Code

In XSB, most source code file can also be “consulted” dynamically via the predicates load_dyn/[1,2], load_dync/[1,2] and ensure_loaded/2. These predicates act as consult/2 in that if a given file File has already been dynamically loaded, old versions of predicates defined in File will be retracted and their new definitions made to correspond to those in File (except for predicates in which a multifile/1 declaration is present in File). Dynamic loading can be performed using XSB’s reader of canonical terms (which does not include operators, but does allow list and comma-list notation) via load_dync/2; dynamic loading using XSB’s general reader for Hilog terms is performed via load_dyn/2.

The predicates mentioned above are described more fully in Chapter 6. Here, we simply compare the tradeoffs of static and dynamic loading.

• Advantages for Dynamic Loading
  – For large files, containing $10^4 – 10^7$ clauses, dynamic loading is much faster than XSB’s compiler, especially when the canonical reader is used.
Dynamically loaded files have advantages of dynamic code including star-, trie, compound, and alternate indexes, as well as being modifiable via assert and retract.

- Advantages for Static Compilation
  
  - Although dynamically loaded predicates are compiled into SLG-WAM code, compiled static clauses are more optimized than dynamically predicates, particularly when the clauses have large bodies or when arithmetic is used. For facts and pure binary predicates (those containing a single literal in their body) however, static and dynamic byte code is essentially the same.

  - Dynamic loading does not allow module/export declarations, mode declarations, or unification factoring. It does however, allow files to import predicates, allows tabling and dynamic declarations (except for auto_table and suppl_table, and operator declarations (when a canonical read is not used).

### 3.9.3 The multifile directive

The default action upon loading a file or module is to delete all previous byte-code for predicates defined in the file. If this is not the desired behavior, the user may add to the file a declaration

```prolog
:- multifile Predicate_List .
```

where Predicate_List is a list of predicates in functor/arity form. The effect of this declaration is to delete only those clauses of predicate/arity that were defined in the file itself. *If a predicate P is to be treated as multifile, the multifile/1 directive for P must appear in all files that contain clause definitions for P.* If P is dynamic, this means that the multifile declaration for P must appear in files defining P whether they are compiled and consulted, or dynamically loaded via load_dyn/[1,2] or load_dync/[1,2].

### 3.10 The Compiler

The XSB compiler translates XSB source files into byte-code object files. It is written entirely in Prolog. Both the sources and the byte code for the compiler can be found in the XSB system directory cmplib. Prior to compiling, XSB
filters the programs through *GPP*, a preprocessor written by Denis Auroux (auroux@math.polytechnique.fr). This preprocessor maintains high degree of compatibility with the C preprocessor, but is more suitable for processing Prolog programs. The preprocessor is invoked with the compiler option *xpp_on* as described below. The various features of GPP are described in Appendix A.

XSB also allows the programmer to use preprocessors other than GPP. However, the modules that come with XSB distribution require GPP. This is explained below (see *xpp_on/1* compiler option).

The following sections describe the various aspects of the compiler in more detail.

### 3.10.1 Invoking the Compiler

In addition to invoking the compiler through *consult/[1,2]*, the compiler can be invoked directly at the interpreter level (or in a program) through the Prolog predicates *compile/[1,2]*.

```
compile(+Files,+OptionList)
```

```
compile(+Files)
```

*compile/2* compiles all files specified, using the compiler options specified in *OptionList* (see Section 3.10.2 below for the precise details.) *Files* is either an absolute or relative filename, or a ground list of absolute or relative file names; and *OptionList* is a ground list of compiler options. Since options can be set globally via the predicate *set_global_compiler_options/1*, each option in *OptionsList* can optionally be prefixed by *+* or *-*, indicating that the option is to be turned on, or off, respectively. (No prefix turns the option on.)

```
| ?- compile(Files).
```

```
| ?- compile(Files, []).
```

is just a notational shorthand for the query:

```
| ?- compile(Files, []).
```

For a given, *File* to be compiled, the source file name corresponding to *File* is obtained by concatenating a directory prefix and the extension *.P*, *.pl*, *.prolog*, or other filenames as discussed in Section 3.3. The directory prefix must be in the dynamic loader path (see Section 3.6). Note that these directories are searched in a predetermined order (see Section 3.6), so if a module with the same name appears in more than one of the directories searched, the compiler will compile the first one it encounters. In such a case, the user can override the search order by providing an absolute path name. If *File*
contains no extension, an attempt is made to compile the file File.P, File.pl, File.prolog, or other extensions before trying compiling the file with name File.

We recommend use of the extension .P for Prolog source file to avoid ambiguity. Optionally, users can also provide a header file for a module (denoted by the module name suffixed by .H). In such a case, the XSB compiler will first read the header file (if it exists), and then the source file. Currently the compiler makes no special treatment of header files. They are simply included in the beginning of the corresponding source files, and code can, in principle, be placed in either.

The result of the compilation (an SLG-WAM object code file) is stored in (filename).xwam, but compile/[1,2] does not load the object file it creates. (The standard predicate consult/[1,2] loads the object file into the system, after recompiling the source file if needed.) The object file created is always written into the directory where the source file resides: the user must therefore have write permission in that directory to avoid an error.

If desired, when compiling a module (file), clauses and directives can be transformed as they are read. This is indeed the case for definite clause grammar rules (see Chapter 11), but it can also be done for clauses of any form by providing a definition for predicate term_expansion/2 (see Section 11.3).

Predicates compile/[1,2] can also be used to compile foreign language modules. In this case, the names of the source files should have the extension .c and a .P file must not exist. A header file (with extension .H) must be present for a foreign language module (see the chapter Foreign Language Interface in Volume 2).

**Error Cases** In the cases below, File refers to an element of Files if Files is a list and otherwise refers to Files itself.

- Files is a variable, or a list containing a variable element.
  - instantiation_error.
- File is a neither an atom nor a list of atoms.
  - type_error(atom_or_list_of_atoms,File)
- File does not exist in the current set of library directories
  - existence_error(file,File)
- File has an object code extension (e.g. .xwam)
  - permission_error(compile,file,File)
• File has been loaded previously in the session and there is more than one active thread.
  – misc_error
• OptionList is a partial list or contains an option that is a variable
  – instantiation_error
• OptionList is neither a list nor a partial list
  – type_error(list,OptionsList)
• OptionList contains an option, Option not described in Section 3.10.2
  – domain_error(xsb_compiler_option,Option)

3.10.2 Compiler Options

Compiler options can be set in three ways: from a global list of options (set_global_compiler_options/1), from the compilation command (compile/2 and consult/2), and from a directive in the file to be compiled (see compiler directive compiler_options/1).

set_global_compiler_options(+OptionsList)

OptionsList is a list of compiler options (described below). Each can optionally be prefixed by + or -, indicating that the option is to be turned on, or off, respectively. (No prefix turns the option on.) This evaluable predicate sets the global compiler options in the way indicated. These options will be used in any subsequent compilation, unless they are reset by another call to this predicate, overridden by options provided in the compile invocation, or overridden by options in the file to be compiled.

The following options are currently recognized by the compiler:

singleton_warnings_off Does not print out any warnings for singleton variables during compilation. This option can be useful for compiling XSB programs that have been generated by some other program.

optimize When specified, the compiler tries to optimize the object code. In Version 3.6, this option optimizes predicate calls, among other features, so execution may be considerably faster for recursive loops. However, due to the nature of the optimizations, the user may not be able to trace all calls to predicates in the program. As expected, the compilation phase will also be slightly longer. For these reasons, the use of the optimize option may not be suitable for the development phase, but is recommended once the code has been debugged.
**allow_redefinition** By default the compiler refuses to compile a file that contains clauses that would redefine a standard predicate (unless the **sysmod** option is in effect.) By specifying this option, the user can direct the compiler to quietly allow redefinition of standard predicates.

**xpp_on** Filter the program through a preprocessor before sending it to the XSB compiler. By default (and for the XSB code itself), XSB uses GPP, a preprocessor developed by Denis Auroux (auroux@math.polytechnique.fr) that has high degree of compatibility with the C preprocessor, but is more suitable for Prolog syntax. In this case, the source code can include the usual C preprocessor directives, such as `#define`, `#ifdef`, and `#include`. This option can be specified both as a parameter to `compile/2` and as part of the `compiler_options/1` directive inside the source file. See Appendix A for more details on GPP.

When an `#include "file"` statement is encountered, XSB directs GPP to search for the files to include in the directories `$XSB_DIR/emu` and `$XSB_DIR/prolog_includes`. However, additional directories can be added to this search path by asserting into the predicate `gpp_include_dir/1`, which must be imported from module `parse`. For example if you want additional directories to be searched, then the following statements must be executed:

```prolog
:- import gpp_include_dir/1 from parse.
:- assert(gpp_include_dir('some-other-dir')).
```

Note that when compiling XSB programs, GPP searches the current directory and the directory of the parent file that contains the include-directive last. If you want GPP to search directories in a different order, `gpp_options/1` can be used (see below).

Note: if you assert something into `gpp_include_dir/1` then you must also execute `retractall(gpp_include_dir(_))` later on or else subsequent Prolog compilations might not work correctly.

---

9For compatibility, XSB also supports the ISO predicate `include/1` which also allows extra files to be included during compilation.
XSB predefines the constant `XSB_PROLOG`, which can be used for conditional compilation. For instance, you can write portable program to run under XSB and other prologs that support C-style preprocessing and use conditional compilation to account for the differences:

```c
#ifdef XSB_PROLOG
   XSB-specific stuff
#else
   other Prolog’s stuff
#endif
common stuff
```

**gpp_options** This dynamic predicate must be imported from module `parse`. If some atom is asserted into `gpp_options` then this atom is assumed to be the list of command line options to be used by the preprocessor (only the first asserted atom is ever considered). If this predicate is empty, then the default list of options is used (which is `-P -m -nostdinc -nocurinc`, meaning: use Prolog mode and do not search the standard C directories and the directory of the parent file that contains the include-instruction).

As mentioned earlier, when XSB invokes Gpp, it uses the option `-nocurinc` so that Gpp will not search the directory of the parent file. If a particular application requires that the parent file directory must be searched, then this can be accomplished by executing `assert(gpp_options(''-P -m -nostdinc''))`.

Note: if you assert options into `gpp_options/1` then do not forget to also execute `retractall(gpp_options(_))` after that or else subsequent Prolog compilations might not work correctly.

**xpp_dump** This causes XSB to dump the output from the GPP preprocessor into a file. If the file being compiled is named `file.P` then the dump file is named `file.P_gpp`. This option can be included in the list of options in the `compiler_options/1` directive, but usually it is used for debugging, as part of the `compile/2` predicate. If `xpp_dump` is specified directly in the file using `compiler_options/1` directive, then it should not follow the `gpp_on` option in the list (or else it will be ignored).

**Note:** multiple occurrences of `xpp_on` and `xpp_dump` options are allowed, but only the first one takes effect—all the rest are ignored!
xpp_on/N and xpp_dump/N

XSB also allows one to filter program files through a pipeline of external preprocessors in addition to or instead of GPP. This can be specified with the N-ary versions of xpp_on and xpp_dump:

\[
\begin{align*}
\text{xpp\_on}(\text{spec1}, \ldots, \text{specN}) \\
\text{xpp\_dump}(\text{spec1}, \ldots, \text{specN})
\end{align*}
\]

Each \text{spec1}, \ldots, \text{specN} is a preprocessor specification of the form \text{preprocessor\_name} or \text{preprocessor\_name}(\text{options}). Each preprocessor is applied in a pipeline passing its output to the next preprocessor. The first preprocessor is applied to the file being compiled. The preprocessor name is an atom or a function symbol and \text{options} must be an atom. If \text{preprocessor\_name} is \text{gpp}, then the GPP preprocessor will be invoked. Note that \text{gpp} can appear anywhere in the aforesaid sequence of specs (or not appear at all), so it is possible to preprocess XSB files before and/or after (or instead of) GPP. Note that \text{xpp\_on(gpp)} and \text{xpp\_dump(gpp)} are equivalent to the earlier 0-ary compiler options \text{xpp\_on} and \text{xpp\_dump}, respectively.

To use a preprocessor other than GPP two things must be done:

- A 4-ary Prolog predicate must be provided, which takes three input arguments and produces in its 4th argument a syntactically correct shell (Unix or Windows) command for invoking the preprocessor. The first preprocessor in the pipeline must be taking its input from a file, but the subsequent preprocessors must expect their input from the standard input. All preprocessors must send their results to the standard output. The arguments to the 4-ary predicate in question are:
  - File: this is the XSB input file to be processed. Usually this argument is left unused (unbound), but might be useful for producing error messages or debugging.
  - Preprocessor name: this is the name under which the preprocessor is registered (see below). It is the same as \text{processor\_name} referred to above. This name is up to the programmer; it is to be used to refer to the preprocessor (it does not need to be related in any way to the shell-command-producing predicate or to the OS’s pathname for the preprocessor).
  - Options: these are the command-line options that the preprocessor might need. If the preprocessor spec mentioned above is \text{foo(bar)} then the preprocessor name (argument 2) would be bound to \text{foo} and options (argument 3) to \text{bar}. 

Shell command: this is the only output argument. It is supposed to be the shell command to be used to invoke the preprocessor. The shell command must not include the file name to be processed—that name is added automatically as the last option to the shell command.

Special considerations for using XSB as a preprocessor. XSB can be used as a preprocessor for XSB programs by constructing a shell command that invokes XSB. However, several conventions need to be observed. First, the file to be preprocessed is automatically attached as the last argument of the aforesaid shell command, but XSB does not accept file names in that place as a command-line option (except with special flags used by XSB developers only and for other purposes). Therefore, the file name to be read and preprocessed by XSB must be passed to XSB by some other means (e.g., using the \texttt{-e "command"} option). In addition, the last command line option for that XSB-based command must be \texttt{-ignore}, which will cause XSB to ignore the remaining options, including the aforesaid file name. Also, if a preprocessor appears in the pipeline as the second preprocessor or later (i.e., after the first argument in \texttt{xpp_dump}), that preprocessor’s shell command line must expect to receive the output of the preceding preprocessor \textit{on the standard input}. Therefore, in order to serve as the second or later preprocessor in the pipeline, XSB must be invoked with the \texttt{-e "see(userin)."} option followed by a call to the predicate that would actually do the preprocessing.

Here are a few examples. To invoke XSB as the first preprocessor in the pipeline, one could construct the following shell command (shown below as an atom of the kind that one needs to construct in the “Shell command” argument being discussed):

\begin{verbatim}
'.../xsb options -e "preprocessPred(''MyFile''),halt." --ignore'
\end{verbatim}

Note that here the file to be preprocessed, \texttt{MyFile}, needs to be passed to the preprocessing predicate as an argument. To use XSB as the second and later preprocessor in the pipeline, the appropriate command could be

\begin{verbatim}
'.../xsb options -e "see(userin),preprocessPred,halt."'
\end{verbatim}

Here the file to be preprocessed will come on the standard input of XSB. No need for the \texttt{-ignore} option here because no file names would be attached at the end of this command (since the file is piped through the standard input).

In both cases, the file passed to \texttt{preprocessPred/1} or \texttt{preprocessPred/0}
could be processed using `read/1` and `write_canonical/1`. The typical options that one would want to pass in both cases (to replace `options`) are

    --noprompt --quietload --nobanner --nofeedback

Note that other commands might need to be executed under the `-e` option in order to bootstrap the preprocessor (e.g., additional XSB files might need to be loaded).

- The preprocessor must be registered using the following query:

```
:- import register_xsb_preprocessor/2 from parse.
?- register_xsb_preprocessor(preproc_name,preproc_predicate(_,_,_,_)).
```

Here the argument `preproc_name` is the user-given name for the preprocessor, while `preproc_predicate` is the 4-ary shell-command-producing predicate described earlier.

The registration query must be executed before the start of the preprocessing of the input XSB file. Clearly, this implies that the shell-command-producing predicate must be in a different file than the one being preprocessed.

Note: one cannot register the same preprocessor twice. The second time the same name is used, it is ignored. However, it is possible to register the same shell-command-producing predicate twice, if the user registers the these shell-command-producing predicates under different preprocessor names.

```
The difference between `xpp_on/N` and `xpp_dump/N` is that the latter also saves the output of each preprocessing stage in a separate file. For instance, if the XSB file to be preprocessed is `abc.P` and the `xpp_dump/N` option has the form `xpp_dump(foo,gpp,bar)` then three files will be produced: `abc.P_foo`, `abc.P_gpp`, `abc.P_bar`, each containing the result of the respective stage in preprocessing.

Here is an example. Suppose that `foobar.P` includes the definition of the following predicate

```
make_append_cmd(_File,_Name,Options,ResultingCmd) :-
    fmt_write_string(ResultingCmd, '/bin/cat "%s"', arg(Options)).
```

and also has the following registration query:

```
?- parse:register_xsb_preprocessor(appendfile,make_append_cmd(_,_,_,_)).
```
Suppose that the file `abc.P` includes the following compiler directive:

```prolog
:- compiler_options([xpp_on(appendfile('data.P'),gpp)]).
```

If the file `foobar.P` is loaded before compiling `abc.P` then the file `data.P` will be first appended to `abc.P` and then the result will be processed by GPP. The final result will be parsed and compiled by XSB.

Note that although the parameters `File` and `Name` are not used by `make_append_cmd/4` in our example, when this predicate is called they will be bound to `foobar.P` and `appendfile`, respectively, and could be used by the shell-command-producing predicates for various purposes.

**quit_on_error** This causes XSB to exit if compilation of a program end with an error. This option is useful when running XSB from a makefile, when it is necessary to stop the build process after an error has been detected. For instance, XSB uses this option during its own build process.

**auto_table** When specified as a compiler option, the effect is as described in Section 3.10.5. Briefly, a static analysis is made to determine which predicates may loop under Prolog’s SLD evaluation. These predicates are compiled as tabled predicates, and SLG evaluation is used instead.

**suppl_table** The intention of this option is to direct the system to table for efficiency rather than termination. When specified, the compiler uses tabling to ensure that no predicate will depend on more than three tables or EDB facts (as specified by the declaration `edb` of Section 3.10.5). The action of `suppl_table` is independent of that of `auto_table`, in that a predicate tabled by one will not necessarily be tabled by the other. During compilation, `suppl_table` occurs after `auto_table`, and uses table declarations generated by it, if any.

**spec_repr** When specified, the compiler performs specialization of partially instantiated calls by replacing their selected clauses with the representative of these clauses, i.e. it performs folding whenever possible. In general specialization with replacement is correct only under certain conditions. XSB’s compiler checks for sufficient conditions that guarantee correctness, and if these conditions are not met, specialization with replacement is not performed for the violating calls.

**spec_off** When specified, the compiler does not perform specialization of partially instantiated calls.
**unfold_off** When specified, singleton sets optimizations are not performed during specialization. This option is necessary in Version 3.6 for the specialization of **table** declarations that select only a single chain rule of the predicate.

**spec_dump** Generates a **module.spec** file, containing the result of specializing partially instantiated calls to predicates defined in the **module** under compilation. The result is in Prolog source code form.

**ti_dump** Generates a **module.ti** file containing the result of applying unification factoring to predicates defined in the **module** under compilation. The result is in Prolog source code form. See page 63 for more information on unification factoring.

**ti_long_names** Used in conjunction with **ti_dump**, generates names for predicates created by unification factoring that reflect the clause head factoring done by the transformation.

**modeinfer** This option is used to trigger mode analysis. For each module compiled, the mode analyzer creates a **module.D** file that contains the mode information. **WARNING:** Occasionally, the analysis itself may take a long time. As far as we have seen, the analysis times are longer than the rest of the compilation time only when the module contains recursive predicates of arity $\geq 10$. If the analysis takes an unusually long time (say, more than 4 times as long as the rest of the compilation) you may want to abort and restart compilation without **modeinfer**.

**mi_warn** During mode analysis, the .D files corresponding to the imported modules are read in. The option **mi_warn** is used to generate warning messages if these .D files are outdated — *i.e.*, older than the last modification time of the source files.

**mi_foreign** This option is used *only* when mode analysis is performed on XSB system modules. This option is needed when analyzing **standard** and **machine** in **syslib**.

**sysmod** Mainly used by developers when compiling system modules and used for bootstrapping. If specified, standard predicates (see `/$XSB_DIR/syslib/std_xsb.P`) are automatically available for use only if they are primitive predicates (see the file `$XSB_DIR/syslib/machine.P` for a current listing of primitive predicates.) When compiling in this mode, non-primitive standard predicates must be explicitly imported from the appropriate system module. Also standard predicates are permitted to be defined.
verbo  Compiles the files (modules) specified in “verbose” mode, printing out information about the progress of the compilation of each predicate.

profile  This option is usually used when modifying the XSB compiler. When specified, the compiler prints out information about the time spent in certain phases of the compilation process.

asm_dump, compile_off  Generates a textual representation of the SLG-WAM assembly code and writes it into the file module.A where module is the name of the module (file) being compiled.

   Warning: This option was created for compiler debugging and is not intended for general use. There might be cases where compiling a module with these options may cause generation of an incorrect .A and .xwam file. In such cases, the user can see the SLG-WAM instructions that are generated for a module by compiling the module as usual and then using the -d module.xwam command-line option of the XSB emulator (see Section 3.7).

index_off  When specified, the compiler does not generate indices for the predicates compiled.

3.10.3  Specialization

From Version 1.4.0 on, the XSB compiler automatically performs specialization of partially instantiated calls. Specialization can be thought as a source-level program transformation of a program to a residual program in which partially instantiated calls to predicates in the original program are replaced with calls to specialized versions of these predicates. The expectation from this process is that the calls in the residual program can be executed more efficiently that their non-specialized counterparts. This expectation is justified mainly because of the following two basic properties of the specialization algorithm:

 Compile-time Clause Selection  The specialized calls of the residual program directly select (at compile time) a subset containing only the clauses that the corresponding calls of the original program would otherwise have to examine during their execution (at run time). By doing so, laying down unnecessary choice points is at least partly avoided, and so is the need to select clauses through some sort of indexing.

 Factoring of Common Subterms  Non-variable subterms of partially instantiated calls that are common with subterms in the heads of the selected clauses are factored out from these terms during the specialization process. As a result, some
head unification (get_* or unify_*) and some argument register (put_*) WAM instructions of the original program become unnecessary. These instructions are eliminated from both the specialized calls as well as from the specialized versions of the predicates.

Though these properties are sufficient to get the idea behind specialization, the actual specialization performed by the XSB compiler can be better understood by the following example. The example shows the specialization of a predicate that checks if a list of HiLog terms is ordered:

\[
\text{ordered([]}). \\
\text{ordered([X])}. \\
\text{ordered([X,Y|Z]) :- } \\
\quad X @=< Y, \_\text{ordered}(Y, Z). \\
\]

The transformation (driven by the partially instantiated call \text{ordered([Y|Z])}) effectively allows predicate \text{ordered/2} to be completely deterministic (when used with a proper list as its argument), and to not use any unnecessary heap-space for its execution. We note that appropriate \text{:- index} directives are automatically generated by the XSB compiler for all specialized versions of predicates.

The default specialization of partially instantiated calls is without any folding of the clauses that the calls select. Using the \text{spec_repr} compiler option (see Section 3.10.2) specialization with replacement of the selected clauses with the representative of these clauses is performed. Using this compiler option, predicate \text{ordered/2} above would be specialized as follows:

\[
\text{ordered([]}). \\
\text{ordered([X|Y]) :- } \_\text{ordered}(X, Y). \\
\quad \text{:- index } \_\text{ordered}/2-2. \\
\quad \_\text{ordered}(X, []). \\
\quad \_\text{ordered}(X, [Y|Z]) :- X @=< Y, \_\text{ordered}(Y, Z). \\
\]

We note that in the presence of cuts or side-effects, the code replacement operation is not always sound, i.e. there are cases when the original and the residual program
are not computationally equivalent (with respect to the answer substitution semantics). The compiler checks for sufficient (but not necessary) conditions that guarantee computational equivalence, and if these conditions are not met, specialization is not performed for the violating calls.

The XSB compiler prints out messages whenever it specialises calls to some predicate. For example, while compiling a file containing predicate \texttt{ordered/1} above, the compiler would print out the following message:

\begin{verbatim}
% Specialising partially instantiated calls to ordered/1
\end{verbatim}

The user may examine the result of the specialization transformation by using the \texttt{spec_dump} compiler option (see Section 3.10.2).

Finally, we have to mention that for technical reasons beyond the scope of this document, specialization cannot be transparent to the user; predicates created by the transformation do appear during tracing.

### 3.10.4 Compiler Directives

Consider a directive

\begin{verbatim}
:- foo(a).
\end{verbatim}

That occurs in a file that is to be compiled. There are two logical interpretations of such a directive.

1. \texttt{foo(a)} is to be executed upon loading the file; or

2. \texttt{foo(a)} provides information used by the compiler in compiling the file.

By default, the interpretation of a directive is as in case (1) except in the case of the compiler directives listed in this section, which as their name implies, are taken to provide information to the compiler. Some of the directives, such as the \texttt{mode/1} directive, have no meaning as an executable directive, while others, such as \texttt{import/2} do. In fact as an executable directive \texttt{import/2} imports predicates into \texttt{usermod}. For such a directive, a statement beginning with \texttt{?-}, such as

\begin{verbatim}
?- import foo/1 from myfile.
\end{verbatim}
indicates that the directive should be executed upon loading the file, and should have no meaning to the compiler. On the other hand, the statement

\[ \text{- import foo/1 from myfile.} \]

Indicates that \text{foo/1} terms in the file to be compiled are to be understood as \text{myfile:foo/1}. In other words, the statement is used by the compiler and will not be executed upon loading. For non-compiler directives the use of \text{-} and \text{:} has no effect — in both cases the directive is executed upon loading the file.

The following compiler directives are recognized in Version 3.6 of XSB

Including Files in a Compilation

\text{include(+FileName)} \quad \text{ISO}

The ISO directive

\[ \text{- include(FileName)} \]

Causes the compiler to act as if the code from \text{FileName} were contained at the position where the directive was encountered. XSB’s preprocessor can perform the same function via the command \texttt{#include FileName} and can support more sophisticated substitutions, but \text{include/1} should be used if code portability is desired.

3.10.5 Conditional Compilation

Section 3.10.2 described a way of performing conditional compilation using XSB’s interaction with GPP. Conditional compilation can also be done through XSB’s compiler, using the directives \text{- if(+Condition)}, \text{- elseif(+Condition)}, \text{- else}, and \text{- endif}. For instance the fragment

\[
\begin{align*}
\text{- if(current_prolog_flag(dialect,xsb)).} \\
\text{- include('file2.P').} \\
\text{- elseif(current_prolog_flag(dialect,swi)).} \\
\text{- include('file3.P').} \\
\text{- endif.}
\end{align*}
\]
allows different Prolog code to be included for XSB and for another Prolog. This framework is very general: for instance, as long as `if...elseif...endif` blocks are not nested, any Prolog code can be used in the consequents of the (else)if. The condition of `if/1` or `elseif/1` can be any Prolog goal, although care should be used in selecting `Condition`. For instance, the goal

```prolog
:- if(file_exists('file1.P')).
```

might be true during compilation, but if the object file produced by the compilation is moved, the condition might no longer be true.

```prolog
if(?Condition)  
elseif(?Condition)  
else  
endif
```

Directives to invoke conditional compilation as described above. If `Condition` is a “changeable” goal such as `file_exists/1`, a warning will be issued but no error will be raised.

### Mode Declarations

The XSB compiler accepts `mode` declarations of the form:

```prolog
:- mode ModeAnnot₁,...,ModeAnnotₙ.
```

where each `ModeAnnot` is a `mode` annotation (a `term indicator` whose arguments are elements of the set `{+,−,#,?}`). From Version 1.4.1 on, `mode` directives are used by the compiler for tabling directives, a use which differs from the standard use of modes in Prolog systems\(^{10}\). See Section 3.10.5 for detailed examples.

Mode annotations have the following meaning:

+ This argument is an input to the predicate. In every invocation of the predicate, the argument position must contain a non-variable term. This term may not necessarily be ground, but the predicate is guaranteed not to alter this argument.

```prolog
:- mode see(+), assert(+).
```

\(^{10}\)The most common uses of `mode` declarations in Prolog systems are to reduce the size of compiled code, or to speed up a predicate’s execution.
- This argument is an output of the predicate. In every invocation of the predicate the argument position will always be a variable (as opposed to the # annotation below). This variable is unified with the value returned by the predicate. We note that Prolog does not enforce the requirement that output arguments should be variables; however, output unification is not very common in practice.

    :- mode cputime(-).

# This argument is either:

    - An output argument of the predicate for which a non-variable value may be supplied for this argument position. If such a value is supplied, the result in this position is unified with the supplied supplied value. The predicate fails if this unification fails. If a variable term is supplied, the predicate succeeds, and the output variable is unified with the return value.

    :- mode '='(#,#).

    - An input/output argument position of a predicate that has only side-effects (usually by further instantiating that argument). The # symbol is used to denote the ± symbol that cannot be entered from the keyboard.

? This argument does not fall into any of the above categories. Typical cases would be the following:

    - An argument that can be used both as input and as output (but usually not with both uses at the same time).

        :- mode functor(?,-,?).

    - An input argument where the term supplied can be a variable (so that the argument cannot be annotated as +), or is instantiated to a term which itself contains uninstantiated variables, but the predicate is guaranteed not to bind any of these variables.

        :- mode var(?), write(?).

We try to follow these mode annotation conventions throughout this manual.

Finally, we warn the user that mode declarations can be error-prone, and since errors in mode declarations do not show up while running the predicates interactively, unexpected behavior may be witnessed in compiled code, optimized to take modes into account (currently not performed by XSB). However, despite this danger, mode annotations can be a good source of documentation, since they express the programmer’s intention of data flow in the program.
Tabling Directives

Memoization is often necessary to ensure that programs terminate, and can be useful as an optimization strategy as well. The underlying engine of XSB is based on SLG, a memoization strategy, which, in our version, maintains a table of calls and their answers for each predicate declared as tabled. Predicates that are not declared as tabled execute as in Prolog, eliminating the expense of tabling when it is unnecessary.

The simplest way to use tabling is to include the directive

\[ :- \text{auto_table}. \]

anywhere in the source file. auto_table declares predicates tabled so that the program will terminate.

To understand precisely how auto_table does this, it is necessary to mention a few properties of SLG. For programs which have no function symbols, or where function symbols always have a limited depth, SLG resolution ensures that any query will terminate after it has found all correct answers. In the rest of this section, we restrict consideration to such programs.

Obviously, not all predicates will need to be tabled for a program to terminate. The auto_table compiler directive tables only those predicates of a module which appear to static analysis to contain an infinite loop, or which are called directly through \texttt{tnot/1}. It is perhaps more illuminating to demonstrate these conditions through an example rather than explaining them. For instance, in the program.

\[ :- \text{auto_table}. \]

\[ p(a) :- s(f(a)). \]

\[ s(X) :- p(f(a)). \]

\[ r(X) :- q(X,W), r(Y). \]

\[ m(X) :- \text{tnot}(f(X)). \]

\[ :- \text{mode ap1}(-,-,+). \]

\[ \text{ap1}([H|T],L,[H|L1]) :- \text{ap1}(T,L,L1). \]

\[ :- \text{mode ap}(+,+,-). \]

\[ \text{ap}([],F,F). \]

\[ \text{ap}([H|T],L,[H|L1]) :- \text{ap}(T,L,L1). \]
mem(H,[H|T]).
mem(H,[_|T]) :- mem(H,T).

The compiler prints out the messages

\% Compiling predicate s/1 as a tabled predicate
\% Compiling predicate r/1 as a tabled predicate
\% Compiling predicate m/1 as a tabled predicate
\% Compiling predicate mem/2 as a tabled predicate

Terminating conditions were detected for \texttt{ap1/3} and \texttt{ap/3}, but not for any of the other predicates.

\texttt{auto_table} gives an approximation of tabled programs which we hope will be useful for most programs. The minimal set of tabled predicates needed to ensure termination for a given program is undecidable. It should be noted that the presence of meta-predicates such as \texttt{call/1} makes any static analysis useless, so that the \texttt{auto_table} directive should not be used in such cases.

Predicates can be explicitly declared as tabled as well, through the \texttt{table/1}. When \texttt{table/1} is used, the directive takes the form

\begin{verbatim}
:~ table(F/A).
\end{verbatim}

where \texttt{F} is the functor of the predicate to be tabled, and \texttt{A} its arity.

Another use of tabling is to filter out redundant solutions for efficiency rather than termination. In this case, suppose that the directive \texttt{edb/1} were used to indicate that certain predicates were likely to have a large number of clauses. Then the action of the declaration \texttt{:~ suppl_table} in the program:

\begin{verbatim}
:~ edb(r1/2).
:~ edb(r2/2).
:~ edb(r3/2).
:~ suppl_table.
\end{verbatim}

\texttt{join(X,Z)} :- \texttt{r1(X,X1),r2(X1,X2),r3(X2,Z)}.

would be to table \texttt{join/2}. The \texttt{suppl_table} directive is the XSB analogue to the deductive database optimization, \textit{supplementary magic templates} \cite{5}. \texttt{suppl_table/0} is shorthand for \texttt{suppl_table(2)} which tables all predicates containing clauses with two
or more edb facts or tabled predicates. By specifying `suppl_table(3)` for instance, only predicates containing clauses with three or more edb facts or tabled predicates would be tabled. This flexibility can prove useful for certain data-intensive applications.

**Indexing Directives**

The XSB compiler by default generates an index on the principal functor of the first argument of a predicate. Indexing on the appropriate argument of a predicate may significantly speed up its execution time. In many cases the first argument of a predicate may not be the most appropriate argument for indexing and changing the order of arguments may seem unnatural. In these cases, the user may generate an index on any other argument by means of an indexing directive. This is a directive of the form:

```
:- index Functor/Arity-IndexArg.
```

indicating that an index should be created for predicate `Functor/Arity` on its `IndexArg` argument. One may also use the form:

```
:- index(Functor/Arity, IndexArg, HashTableSize).
```

which allows further specification of the size of the hash table to use for indexing this predicate if it is a *dynamic* (i.e., asserted) predicate. For predicates that are dynamically loaded, this directive can be used to specify indexing on more than one argument, or indexing on a combination of arguments (see its description on page 266). For a compiled predicate the size of the hash table is computed automatically, so `HashTableSize` is ignored.

All of the values `Functor`, `Arity`, `IndexArg` (and possibly `HashTableSize`) should be ground in the directive. More specifically, `Functor` should be an atom, `Arity` an integer in the range 0..255, and `IndexArg` an integer between 0 and `Arity`. If `IndexArg` is equal to 0, then no index is created for that predicate. An index directive may be placed anywhere in the file containing the predicate it refers to.

As an example, if we wished to create an index on the third argument of predicate `foo/5`, the compiler directive would be:

```
:- index foo/5-3.
```
CHAPTER 3. SYSTEM DESCRIPTION

Unification Factoring

When the clause heads of a predicate have portions of arguments common to several clauses, indexing on the principal functor of one argument may not be sufficient. Indexing may be improved in such cases by the use of unification factoring. Unification Factoring is a program transformation that “factors out” common parts of clause heads, allowing differing parts to be used for indexing, as illustrated by the following example:

\[
\begin{align*}
    p(f(a),X) & : - q(X). & \quad p(f(X),Y) & : - _p(X,Y). \\
    p(f(b),X) & : - r(X). & \quad _p(a,X) & : - q(X). \\
    & \quad _p(b,X) & : - r(X).
\end{align*}
\]

The transformation thus effectively allows \( p/2 \) to be indexed on atoms \( a/0 \) and \( b/0 \). Unification Factoring is transparent to the user; predicates created by the transformation are internal to the system and do not appear during tracing.

The following compiler directives control the use of unification factoring:\footnote{Unification factoring was once called transformational indexing, hence the abbreviation \texttt{ti} in the compiler directives}:

\[- \texttt{ti} (F/A). \quad \text{specifies that predicate} \quad F/A \quad \text{should be compiled with unification factoring enabled.} \]

\[- \texttt{ti\_off} (F/A). \quad \text{specifies that predicate} \quad F/A \quad \text{should be compiled with unification factoring disabled.} \]

\[- \texttt{ti\_all}. \quad \text{specifies that all predicates defined in the file should be compiled with unification factoring enabled.} \]

\[- \texttt{ti\_off\_all}. \quad \text{specifies that all predicates defined in the file should be compiled with unification factoring disabled.} \]

By default, higher-order predicates (more precisely, predicates named \textit{apply} with arity greater than 1) are compiled with unification factoring enabled. It can be disabled using the \texttt{ti\_off} directive. For all other predicates, unification factoring must be enabled explicitly via the \texttt{ti} or \texttt{ti\_all} directive. If both \(- \texttt{ti}(F/A). \) (\(- \texttt{ti\_all}. \)) and \(- \texttt{ti\_off}(F/A). \) (\(- \texttt{ti\_off\_all}. \)) are specified, \(- \texttt{ti\_off}(F/A). \) (\(- \texttt{ti\_off\_all}. \)) takes precedence. Note that unification factoring may have no effect when a predicate is well indexed to begin with. For example, unification factoring has no effect on the following program:
even though the two clauses have \( c/0 \) in common. The user may examine the results of the transformation by using the \texttt{ti_dump} compiler option (see Section 3.10.2).

### Other Directives

XSB has other directives not found in other Prolog systems.

\(:- \texttt{hilog \ atom}_1, \ldots, \texttt{atom}_n.\)

Declares symbols \( \texttt{atom}_1 \) through \( \texttt{atom}_n \) as HiLog symbols. The \texttt{hilog} declaration should appear \textit{before} any use of the symbols. See Chapter 4 for a purpose of this declaration.

\(:- \texttt{ldoption(Options)}.\)

This directive is only recognized in the header file (.H file) of a foreign module. See the chapter \textit{Foreign Language Interface} in Volume 2 for its explanation.

\(:- \texttt{compiler_options(OptionsList)}.\)

Indicates that the compiler options in the list \texttt{OptionsList} should be used to compile this file. This must appear at the beginning of the file. These options will override any others, including those given in the compilation command. The options may be optionally prefixed with + or - to indicate that they should be set on or off. (No prefix indicates the option should be set on.)

### 3.10.6 Inline Predicates

\textit{Inline predicates} represent “primitive” operations in the (extended) WAM. Calls to inline predicates are compiled into a sequence of WAM instructions in-line, i.e., without actually making a call to the predicate. Thus, for example, relational predicates (like \( >/2 \), \( >=/2 \), etc.) compile to, essentially, a subtraction followed by a conditional branch. As a result, calls to inline predicates will not be trapped by the debugger, and their evaluation will not be visible during a trace of program execution. Inline predicates are expanded specially by the compiler and thus cannot be redefined by the user without changing the compiler. The user does not need to import these predicates from anywhere. There are available no matter what options are specified during compiling.
Table 3.1 lists the inline predicates of XSB Version 3.6. Those predicates that start with \texttt{\_\$} are internal predicates that are also expanded in-line during compilation.

We warn the user to be cautious when defining predicates whose functor starts with \texttt{\_\$} since the names of these predicates may interfere with some of XSB’s internal predicates. The situation may be particularly severe for predicates like \texttt{\_\$builtin'/1} that are treated specially by the XSB compiler.

### 3.11 A Note on ISO Compatibility

In Version 3.6, an effort has been made to ensure compatibility with the core Prolog ISO standard [34]. In this section, we summarize the differences with the ISO standard. XSB implements almost all ISO built-ins and evaluable functions, although there are semantic differences between XSB’s implementation and that of the ISO standard in certain cases.

The main difference of XSB from the ISO semantics is that XSB does not support the logical update semantics for assert and retract, but instead supports an immediate semantics. XSB does, however, support an ISO-like semantics for incremental tables.

Version 3.6 of XSB mostly supports the full ISO syntax for Prolog, and its I/O system is based on UTF-8 encoding, which includes ASCII as a subset of its characters. Beyond XSB’s support for Hilog, most differences between ISO syntax and XSB syntax are fairly minor. However, as XSB supports only UTF-8, ISO predicates relating to different character sets, such as \texttt{char_conversion/2}, \texttt{current_char_conversion/2} and others are not supported.

A somewhat more minor difference involves XSB’s implementation of ISO streams. XSB can create streams from several First class objects, including pipes, atoms, and consoles in addition to files. However by default, XSB opens streams in binary mode, rather than text mode in opposition to the ISO standard, which opens streams in text mode. This makes no difference in UNIX or LINUX, for which text and binary
streams are identical, but does make a difference in Windows, where text files are processed more than binary files.

As a final point, XSB currently throws an \texttt{error/3} term in its error ball, rather than an \texttt{error/2} term.

Most other differences with the core standard are mentioned under portability notes for the various predicates.

XSB supports most new features mentioned in the revisions to the core standard [35], including \texttt{call_cleanup/2} and various library predicates such as \texttt{subsumes/2}, \texttt{numbervars/3} and so on. XSB also has strong support for the working multi-threading Prolog standard [36], and XSB has been one of the first Prologs to support this standard. However, because XSB has an atom-based module system it does not support the ISO standard for Prolog modules.
Chapter 4

Syntax

The syntax of XSB is based on ISO Prolog [34], although it lacks a few of the ISO standard’s somewhat arcane features. Beginning with Version 3.6, XSB supports Unicode through UTF-8 atoms as described in Section 4.1.3. XSB’s reader also contains extensions to support HiLog [13], which adds certain features of second-order syntax to Prolog.

4.1 Terms

The data objects of the HiLog language are called terms. A HiLog term can be constructed from any logical symbol or a term followed by any finite number of arguments. In any case, a term is either a constant, a variable, or a compound term.

A constant is either a number (integer or floating-point) or an atom

1 Constants are definite elementary objects, and correspond to proper nouns in natural language.

4.1.1 Integers

ISO Integers

The printed form of an integer normally consists of a sequence of digits optionally preceded by a minus sign (’-’), interpreted, of course, as base 10 integers. It is also possible to enter integers in other bases:

1This Prolog usage contradicts the usage of the word “atom” in logic as short for “atomic formula”.
• $0bnnn$ represents an integer in base 2, e.g.,

| ?- X = 0b110. 
X = 6

• $0onn$ represents an integer in base 8, e.g.,

| ?- X = 0o110. 
X = 72

• $0xnnn$ represents an integer in base 16, e.g.,

| ?- X = 0x110. 
X = 272

Character code constants are integers of the form $0'nnn$, where nnn is the decimal form of any UTF-8 codepoint. E.g.,

| ?- 0'A = X 
X = 65

Escape characters (cf. Section 4.1.3) can be written similarly (if this is ever needed):

| ?- 0'\n = X 
X = 10

Other Integer Representations

It is also possible to enter integers in bases 2 through 36; this can be done by preceding the digit sequence by the base (in decimal) followed by an apostrophe (’). If a base greater than 10 is used, the characters A–Z or a–z are used to stand for digits greater than 9.

Using these rules, examples of valid integer representations in XSB are:
representing respectively the following integers in decimal base:

\[
\begin{array}{cccccccc}
1 & -3456 & 95359 & 9'888 & 16'1FA4 & -12'A0 & 20'
\end{array}
\]

Note that the following:

\[
+525 \quad 12'2CF4 \quad 37'12 \quad 20'-23
\]

are not valid integers of XSB.

Character code constants, mentioned above, can be seen as integers in “base zero”.

### 4.1.2 Floating-point Numbers

XSB supports ISO floating-point numbers, which consist of a sequence of digits with an embedded decimal point, optionally preceded by a minus sign (‘-’), and optionally followed by an exponent consisting of uppercase or lowercase ‘E’ and an optionally signed base 10 integer.

Using these rules, examples of floating point numbers are:

\[
\begin{array}{cccccc}
1.0 & -34.56 & 817.3E12 & -0.0314e26 & 2.0E-1
\end{array}
\]

Note that in any case there must be at least one digit before, and one digit after, the decimal point.

### 4.1.3 Atoms

An atom consists of a sequence of characters that follow the following rules.

- **Non-guoted Atoms** begin with the ASCII character a-z and are followed by a sequence of ISO alphanumerical characters: a-z, A-Z, 0-9, and underscore _.

- **Quoted Atoms** begin and end with the ASCII character ’ and may contain any sequence of
  - Printable UTF-8 characters
Meta-escaped quotes. E.g.,

\[ ?- X = 'a''b'. \]

\[ X = a'b \]

(Unfortunately, the current version of XSB does not support escaped quotes (\'').)

ISO escape characters and sequences

* \b the newline character (ASCII 7).
* \b the newline character (ASCII 8).
* \f the form feed character (ASCII 12).
* \n the newline character (ASCII 10).
* \r the carriage return character (ASCII 13).
* \t a tab character (ASCII 9).
* \v a vertical tab character (ASCII 11).
* Octal escapes of the form \nnn\, where nnn is the octal number corresponding to an ASCII code. E.g.,

\[ ?- write('\60\'). \]

0

* Hexadecimal escapes of the form \xnn\, where nn is the hexidecimal number corresponding to an ASCII code\(^2\). E.g.,

\[ ?- write('\30\'). \]

0

UTF-8 escape sequences have the form \unnnn where nnnn is the hexidecimal number corresponding to a UTF-8 codepoint.

• Operator-based Atoms are defined as any sequence from the following set of characters (except of the sequence ‘/*’, which begins a comment):

\[ + - * / \^ < > = \^ : . ? @ # & \]

Examples of such atoms are:

\[ ^=., ::= === \]

\(^2\)The current version of XSB differs from the ISO specification in that hexadecimal escapes do not have a trailing slash.
• Special Atoms are

! ; [] {} 

Note that the bracket pairs are special. While '[]' and '{}' are atoms, '[:], :, {}, and :}' are not 3.

### 4.1.4 Variables

Variables may be written as any sequence of (ASCII) ISO alphanumeric characters beginning with either a capital letter or '_' . For example:

\[
\text{X HiLog Var1 _3 _List}
\]

If a variable is referred to only once in a clause, it does not need to be named and may be written as an anonymous variable, represented by a single underscore character '_. Any number of anonymous variables may appear in a clause; all of these variables are read as distinct variables.

### 4.1.5 Compound Terms

Like in Prolog, the structured data objects of HiLog are compound terms (or structures). The external representation of a HiLog compound term comprises a functor (called the principal functor or the name of the compound term) and a sequence of one or more terms called arguments. Unlike Prolog where the functor of a term must be an atom, in HiLog the functor of a compound term can be any valid HiLog term. This includes numbers, atoms, variables or even compound terms. Thus, since in HiLog a compound term is just a term followed by any finite number of arguments, all the following are valid external representations of HiLog compound terms:

<table>
<thead>
<tr>
<th>foo(bar)</th>
<th>prolog(a, X)</th>
<th>hilog(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>123(john, 500)</td>
<td>X(kostis, sofia)</td>
<td>X(Y, Z, Y(W))</td>
</tr>
<tr>
<td>f(a, (b(c))(d))</td>
<td>map(double)([], [])</td>
<td>h(map(P)(A, B))(C)</td>
</tr>
</tbody>
</table>

3The form [X] is a special notation for lists (see Section 4.1.6), while the form {X} is just "syntactic sugar" for the term '{X}'.
Like a functor in Prolog, a functor in HiLog can be characterized by its name and its arity which is the number of arguments this functor is applied to. For example, the compound term whose principal functor is \( \text{map}(P) \) of arity 2, and which has arguments \( L_1, \) and \( L_2, \) is written as:

\[
\text{map}(P)(L_1, L_2)
\]

As in Prolog, when we need to refer explicitly to a functor we will normally denote it by the form \( \text{Name}/\text{Arity} \). Thus, in the previous example, the functor \( \text{map}(P) \) of arity 2 is denoted by:

\[
\text{map}(P)/2
\]

Note that a functor of arity 0 is represented as an atom.

In Prolog, a compound term of the form \( p(t_1, t_2, \ldots, t_k) \) is usually pictured as a tree in which every node contains the name \( p \) of the functor of the term and has exactly \( k \) children each one of which is the root of the tree of terms \( t_1, t_2, \ldots, t_k \).

For example, the compound term

\[
\text{s(np(kostis), vp(v(loves), np(sofia)))}
\]

would be pictured as the following tree:

```
        s
       / \
      np  vp
     / \  /
    v np
   |   |
  kostis loves sofia
```

The principal functor of this term is \( s/2 \). Its two arguments are also compound terms. In illustration, the principal functor of the second argument is \( vp/2 \).

Likewise, any external representation of a HiLog compound term \( t(t_1, t_2, \ldots, t_k) \) can be pictured as a tree in which every node contains the tree representation of the name \( t \) of the functor of the term and has exactly \( k \) children each one of which is the root of the tree of terms \( t_1, t_2, \ldots, t_k \).

Sometimes it is convenient to write certain functors as operators. Binary functors (that is, functors that are applied to two arguments) may be declared as infix operators, and unary functors (that is, functors that are applied to one argument)
may be declared as either prefix or postfix operators. Thus, it is possible to write the following:

\[ \begin{align*}
X + Y & \quad (P; Q) \\
X < Y & \quad + X \\
P & ;
\end{align*} \]

More about operators in HiLog can be found in section 4.3.

4.1.6 Lists

As in Prolog, lists form an important class of data structures in HiLog. They are essentially the same as the lists of Lisp: a list is either the atom `'[]'`, representing the empty list, or else a compound term with functor `'.'` and two arguments which are the head and tail of the list respectively, where the tail of a list is also a list. Thus a list of the first three natural numbers is the structure:

```
     /
    / \     /
   1 .   2 .     /
  / \   / \     /
 3 []
```

which could be written using the standard syntax, as:

\[ (1,(2,(3,[]))) \]

but which is normally written in a special list notation, as:

\[ [1,2,3] \]

Two examples of this list notation, as used when the tail of a list is a variable, are:

\[ [\text{Head}|\text{Tail}] \quad [\text{foo,bar}|\text{Tail}] \]

which represent the structures:

```
     /
    / \     /
   Head   Tail foo .     /
  / \   / \     /
 bar Tail
```
respectively.

Note that the usual list notation \([H|T]\) does not add any new power to the language; it is simply a notational convenience and improves readability. The above examples could have been written equally well as:

\[
(\text{Head}, \text{Tail}) \quad .(\text{foo}, (\text{bar}, \text{Tail}))
\]

For convenience, a further notational variant is allowed for lists of integers that correspond to UTF-8 character codes. Lists written in this notation are called \textit{strings}. For example,

"I am a HiLog string"

represents exactly the same list as:

\[[73, 32, 97, 109, 32, 97, 32, 72, 105, 76, 111, 103, 32, 115, 116, 114, 105, 110, 103]\]

### 4.2 From HiLog to Prolog

From the discussion about the syntax of HiLog terms, it is clear that the HiLog syntax allows the incorporation of some higher-order constructs in a declarative way within logic programs. As we will show in this section, HiLog does so while retaining a clean first-order declarative semantics. The semantics of HiLog is first-order, because every HiLog term (and formula) is automatically \textit{encoded (converted)} in predicate calculus in the way explained below.

Before we briefly explain the encoding of HiLog terms, let us note that the HiLog syntax is a simple (but notationally very convenient) encoding for Prolog terms, of some special form. In the same way that in Prolog:

\[
1 + 2
\]

is just an (external) shorthand for the term:

\[
+(1, 2)
\]

in the presence of an infix operator declaration for + (see section 4.3), so:

\[
X(a, b)
\]
is just an (external) shorthand for the Prolog compound term:

\[
\text{apply}(X, a, b)
\]

Also, in the presence of a \texttt{hilog} declaration (see section 3.10.5) for \( h \), the HiLog term whose external representation is:

\[
h(a, h, b)
\]

is a notational shorthand for the term:

\[
\text{apply}(h, a, h, b)
\]

Notice that even though the two occurrences of \( h \) refer to the same symbol, only the one where \( h \) appears in a functor position is encoded with the special functor \texttt{apply/\( n \), \( n \geq 1 \)}.

The encoding of HiLog terms is performed based upon the existing declarations of \texttt{hilog} symbols. These declarations (see section 3.10.5), determine whether an atom that appears in a functor position of an external representation of a HiLog term, denotes a functor or the first argument of a set of special functors \texttt{apply}. The actual encoding is as follows:

- The encoding of any variable or parameter symbol (atom or number) that does not appear in a functor position is the variable or the symbol itself.

- The encoding of any compound term \( t \) where the functor \( f \) is an atom that is not one of the \texttt{hilog} symbols (as a result of a previous \texttt{hilog} declaration), is the compound term that has \( f \) as functor and has as arguments the encoding of the arguments of term \( t \). Note that the arity of the compound term that results from the encoding of \( t \) is the same as that of \( t \).

- The encoding of any compound term \( t \) where the functor \( f \) is either not an atom, or is an atom that is a \texttt{hilog} symbol, is a compound term that has \texttt{apply} as functor, has first argument the encoding of \( f \) and the rest of its arguments are obtained by encoding of the arguments of term \( t \). Note that in this case the arity of the compound term that results from the encoding of \( t \) is one more than the arity of \( t \).

Note that the encoding of HiLog terms described above, implies that even though the HiLog terms:
externally appear to have the same form, in the presence of a hi log declaration for \( h \) but not for \( p \), they are completely different. This is because these terms are shorthands for the terms whose internal representation is:

\[
\begin{align*}
    p(a, b) \\
    \text{apply}(h, a, b)
\end{align*}
\]

respectively. Furthermore, only \( h(a, b) \) is unifiable with the HiLog term whose external representation is \( X(a, b) \).

We end this short discussion on the encoding of HiLog terms with a small example that illustrates the way the encoding described above is being done. Assuming that the following declarations of parameter symbols have taken place,

\[
\begin{align*}
    \text{:-} & \quad \text{hilog} \quad h. \\
    \text{:-} & \quad \text{hilog} \quad (\text{hilog}).
\end{align*}
\]

before the compound terms of page 71 were read by XSB, the encoding of these terms in predicate calculus using the described transformation is as follows:

\[
\begin{align*}
    \text{foo}(\text{bar}) \quad & \quad \text{prolog}(\text{a}, \text{X}) \\
    \text{apply}(\text{hilog}, \text{X}) \quad & \quad \text{apply}(\text{123}, \text{john}, \text{500}) \\
    \text{apply}(\text{X}, \text{kostis}, \text{sofia}) \quad & \quad \text{apply}(\text{X}, \text{Y}, \text{Z}, \text{apply}(\text{Y}, \text{W})) \\
    \text{f}(\text{a}, \text{apply}(\text{b}(\text{c}), \text{d})) \quad & \quad \text{apply}(\text{map(double)}, [], []) \\
    \text{apply}(\text{apply}(\text{h}, \text{apply}(\text{map(P)}, \text{A}, \text{B})), \text{C})
\end{align*}
\]

4.3 Operators

From a theoretical point of view, operators in Prolog are simply a notational convenience and add absolutely nothing to the power of the language. For example, in most Prologs ‘+’ is an infix operator, so

\[
2 + 1
\]

is an alternative way of writing the term \( + (2, 1) \). That is, \( 2 + 1 \) represents the data structure:
and not the number 3. (The addition would only be performed if the structure were passed as an argument to an appropriate procedure, such as is/2).

However, from a practical or a programmer’s point of view, the existence of operators is highly desirable, and clearly handy.

Prolog syntax allows operators of three kinds: infix, prefix, and postfix. An infix operator appears between its two arguments, while a prefix operator precedes its single argument and a postfix operator follows its single argument.

Each operator has a precedence, which is an integer from 1 to 1200. The precedence is used to disambiguate expressions in which the structure of the term denoted is not made explicit through the use of parentheses. The general rule is that the operator with the highest precedence is the principal functor. Thus if ‘+’ has a higher precedence than ‘/’, then the following
\[
\begin{align*}
    a+b/c & \quad a+(b/c)
\end{align*}
\]
are equivalent, and both denote the same term \((a,/(b,c))\). Note that in this case, the infix form of the term \(/+(a,b),c\) must be written with explicit use of parentheses, as in:
\[
(a+b)/c
\]

If there are two operators in the expression having the same highest precedence, the ambiguity must be resolved from the types (and the implied associativity) of the operators. The possible types for an infix operator are
\[
yfx \quad xfx \quad xfy
\]
Operators of type ‘xfx’ are not associative. Thus, it is required that both of the arguments of the operator must be subexpressions of lower precedence than the operator itself; that is, the principal functor of each subexpression must be of lower precedence, unless the subexpression is written in parentheses (which automatically gives it zero precedence).

Operators of type ‘xfy’ are right-associative: only the first (left-hand) subexpression must be of lower precedence; the right-hand subexpression can be of the
same precedence as the main operator. *Left-associative* operators (type 'yfx') are the other way around.

An atom named Name can be declared as an operator of type Type and precedence Precedence by the command:

\[ \text{op}(\text{Precedence}, \text{Type}, \text{Name}) \]

The same command can be used to redefine one of the predefined XSB operators (obtainable via \texttt{current_op/3}). However, it is not allowed to alter the definition of the comma (',') operator. An operator declaration can be cancelled by redeclaring the Name with the same Type, but Precedence 0.

As a notational convenience, the argument Name can also be a list of names of operators of the same type and precedence.

It is possible to have more than one operator of the same name, so long as they are of different kinds: infix, prefix, or postfix. An operator of any kind may be redefined by a new declaration of the same kind. For example, the built-in operators '+' and '-' are as if they had been declared by the command:

\[ \text{:- op}(500, \text{yfx}, [+,-]). \]

so that:

\[ 1-2+3 \]

is valid syntax, and denotes the compound term:

\[ (1-2)+3 \]

or pictorially:

```
  +
 / \     
-  3
 / \     
1  2
```

In XSB, the list functor '.'/2 is one of the standard operators, that can be thought as declared by the command:
:- op(661, xfy, .).

So, in XSB,

\[ 1.2. [] \]

represents the structure

\[ . \]
\[ / \]
\[ 1 . \]
\[ / \]
\[ 2 [] \]

Contrasting this picture with the picture above for \( 1-2+3 \) shows the difference between ‘yfx’ operators where the tree grows to the left, and ‘xfy’ operators where it grows to the right. The tree cannot grow at all for ‘xfx’ type operators. It is simply illegal to combine ‘xfx’ operators having equal precedences in this way.

If these precedence and associativity rules seem rather complex, remember that you can always use parentheses when in any doubt.

In Version 3.6 of XSB the possible types for prefix operators are:

\[ fx \quad fy \quad hx \quad hy \]

and the possible types for postfix operators are:

\[ xf \quad yf \]

We end our discussion about operators by mentioning that prefix operators of type \( hx \) and \( hy \) are *proper HiLog operators*. The discussion of proper HiLog operators and their properties is deferred for the manual of a future version.
Chapter 5

Using Tabling in XSB: A Tutorial

Introduction

XSB has two ways of evaluating predicates. The default is to use Prolog-style evaluation, but by using various declarations a programmer can also use tabled resolution which can provide a different, more declarative programming style than Prolog. In this section we discuss various aspects of tabling and their implementation in XSB. Our aim in this section is to provide a user with enough information to be able to program productively with tables in XSB. It is best to read this tutorial with a copy of XSB handy, since much of the information is presented through a series of exercises.

For the theoretically inclined, XSB uses SLG resolution which can compute queries to non-floundering normal programs under the well-founded semantics [81], and is guaranteed to terminate when these programs have the bounded term-depth property. This tutorial covers only enough of the theory of tabling to explain how to program in XSB. For those interested, the web site contains papers covering in detail various aspects of tabling (often through the links for individuals involved in XSB). An overview of SLG resolution, and practical evaluation strategies for it, are provided in [15, 71, 65, 29]. The engine of XSB, the SLG-WAM, is an extension of the WAM [85, 1], and is described in [62, 58, 28, 64, 14, 22, 37, 18, 19, 11, 51, 74, 52, 77] as it is implemented in Version 3.6 and its performance analyzed. Examples of large-scale applications that use tabling are overviewed in [42, 43, 16, 57, 7, 17, 31, 78].
5.1 Tabling in the Context of a Prolog System

Before describing how to program using tabling it is perhaps worthwhile to review some of the goals of XSB’s implementation of tabling. Among them are:

1. To execute tabled predicates at the speed of compiled Prolog.

2. To ensure that the speed of compiled Prolog is not slowed significantly by adding the option of tabling.

3. To ensure that the functionality of Prolog is not compromised by support for tabling.

4. To provide Prolog functionality in tabled predicates and operators whenever it is semantically sensible to do so.

5. To provide standard predicates to manipulate tables taken as data structures in themselves.

Goals 1 and 2 are addressed by XSB’s engine, which in Version 3.6 is based on a virtual machine called the SLG-WAM. The overhead for SLD resolution using this machine is small, and usually less than 5%. Thus when XSB is used simply as a Prolog system (i.e., no tabling is used), it is reasonably competitive with other Prolog implementations based on a WAM emulator written in C or assembly. For example, when compiled as a threaded interpreter (see Chapter 3) XSB Version 3.6 is about two times slower than Quintus 3.1.1 or emulated SICStus Prolog 3.1. Goals 3, 4 and 5 have been nearly met, but there are a few instances in which interaction of tabling with a Prolog construct has not been accomplished, or is perhaps impossible. Accordingly we discuss these instances throughout this chapter. XSB is still under development however, so that future versions may support more transparent mixing of Prolog and tabled code.

5.2 Definite Programs

Definite programs, also called Horn Clause Programs, are Prolog programs without negation or aggregation. In XSB, this means without the \+1, fail_if/1, not/1, tnot/1, setof/3, bagof/3, tt findall/3 or other aggregation operators. Consider the Prolog program

\[
\text{path}(X,Y) :- \text{path}(X,Z), \text{edge}(Z,Y).
\]
\[
\text{path}(X,Y) :- \text{edge}(X,Y).
\]
together with the query \( ?- \text{path}(1,Y) \). This program has a simple, declarative meaning: there is a path from \( X \) to \( Y \) if there is a path from \( X \) to some node \( Z \) and there is an edge from \( Z \) to \( Y \), or if there is an edge from \( X \) to \( Y \). Prolog, however, enters into an infinite loop when computing an answer to this query. The inability of Prolog to answer such queries, which arise frequently, comprises one of its major limitations as an implementation of logic.

A number of approaches have been developed to address this problem by reusing partial answers to the query \( \text{path}(1,Y) \) \cite{25, 79, 4, 82, 83}. The ideas behind these algorithms can be described in the following manner. Calls to tabled predicates, such as \( \text{path}(1,Y) \) in the above example, are stored in a searchable structure together with their proven instances. This collection of tabled subgoals paired with their answers, generally referred to as a table, is consulted whenever a new call, \( C \), to a tabled predicate is issued. If \( C \) is sufficiently similar to a tabled subgoal \( S \), then the set of answers, \( A \), associated with \( S \) may be used to satisfy \( C \). In such instances, \( C \) is resolved against the answers in \( A \), and hence we refer to the call \( C \) as a consumer of \( A \) (or \( S \)). If there is no such \( S \), then \( C \) is entered into the table and is resolved against program clauses as in Prolog — i.e., using SLD resolution. As each answer is derived during this process, it is inserted into the table entry associated with \( C \) if it contains information not already in \( A \). In this second case, we refer to \( C \) as a generator, or producer, as resolution of \( C \) in this manner produces the answers stored in its table entry. If the answer is in fact added to this set, then it is additionally scheduled to be returned to all consumers of \( C \). If instead it is rejected as redundant, then the evaluation simply fails and backtracks to generate more answers.

Notice that since consuming subgoals resolve against unique answers rather than repeatedly against program clauses, tabling will terminate whenever

1. a finite number of subgoals are encountered during query evaluation, and
2. each of these subgoals has a finite number of answers.

Indeed, it can be proven that for any program with the bounded term depth property — roughly, where all terms generated in a program have a maximum depth — SLG computation will terminate. These programs include the important class of Datalog programs.

Predicates can be declared tabled in a variety of ways. A common form is the compiler directive

\[
:- \text{table } P_1, \ldots, P_n.
\]

where each \( P_i \) is a predicate indicator or callable term. More generally

\[
:- \text{table } P_1, \ldots, P_n \text{ as Options.}
\]
allows a user to specify different types of tabling through `options` along with other properties of the designated predicates. For static predicates, these directives must be added to the file containing the clauses of the predicate(s) to be tabled, and the directives cause the predicates to be compiled with tabling 1. For dynamic predicates, the executable directives

?— table P₁,…,Pₙ.

and

?— table P₁,…,Pₙ as Options.

cause a Pᵢ to be tabled (with the appropriate options) if no clauses have been asserted for Pᵢ.

**Exercises** Unless otherwise noted, the file `$XSB_DIR/examples/table_examples.P` contains all the code for the running examples in this section. Invoke XSB with its default settings (i.e., don’t supply additional options) when working through the following exercises.

**Exercise 5.2.1** Consult `$XSB_DIR/examples/table_examples.P` into XSB and try the goal

?— path(1,X).

and continue typing ;<RETURN> until you have exhausted all answers. Now, try rewriting the `path/2` predicate as it would be written in Prolog — and without a tabling declaration. Will it now terminate for the provided `edge/2` relation? (Remember, in XSB you can always hit <CTRL>-C if you go into an infinite loop). □

The return of answers in tabling aids in filtering out redundant computations — indeed it is this property which makes tabling terminate for many classes of programs. The same generation program furnishes a case of the usefulness of tabling for optimizing a Prolog program.

**Exercise 5.2.2** If you are still curious, load in the file `cyl.P` in the `$XSB_DIR/examples` directory using the command.

?— load_dyn(cyl.P).

---

1 In Version 3.6, tabling does not work together with multi-file predicates.
and then type the query

?- same_generation(X,X),fail.

Now rewrite the same_generation/2 program so that it does not use tabling and retry the same query. What happens? (Be patient — or use <ctrl>-C).

Exercise 5.2.3 The file table_examples.P contains a set of facts

ordered_goal(one).
ordered_goal(two).
ordered_goal(three).
ordered_goal(four).

Clearly, the query ?- ordered_goal(X) will return the answers in the expected order.

table_examples.P also contains a predicate

:- table table_ordered_goal/1.
table_ordered_goal(X):- ordered_goal(X).

which simply calls ordered_goal/1 and tables its answers (tabling is unnecessary in this case, and is only used for illustration). Call the query ?- table_ordered_goal(X) and backtrack through the answers. In what order are the answers returned?

The examples stress two differences between tabling and SLD resolution beyond termination properties. First, that each solution to a tabled subgoal is returned only once — a property that is helpful not only for path/2 but also for same_generation/2 which terminates in Prolog. Second, because answers are sometimes obtained using program clauses and sometimes using the table, answers may be returned in an unaccustomed order.

Tabling Dynamic Predicates Dynamic predicates may be tabled just as static predicates, as the following exercise shows.

Exercise 5.2.4 For instance, restart XSB and at the prompt type the directive

?- table(dyn_path/2).

and
?- load_dyn(dyn_examples).

Try the queries to path/2 of the previous examples. Note that it is important to dynamically load dyn_examples.P — otherwise the code in the file will be compiled without knowledge of the tabling declaration.

In general, as long as the directive table/1 is executed before asserting (or dynamically loading) the predicates referred to in the directive, any dynamic predicate can be tabled.

**Letting XSB Decide What to Table** Other tabling declarations are also provided. Often it is tedious to decide which predicates must be tabled. To address this, XSB can automatically table predicates in files. The declaration auto_table chooses predicates to table to assist in termination, while suppl_table chooses predicates to table to optimize data-oriented queries. Both are explained in Section 3.10.2.\(^2\)

### 5.2.1 Call Variance vs. Call Subsumption

The above description gives a general characterization of tabled evaluation for definite programs but glosses over certain details. In particular, we have not specified the criteria for

- **Call Similarity** – whereby a newly issued subgoal $S$ is determined to be “sufficiently similar” to a tabled subgoal $S_{tab}$ so that $S$ can use the answers from the table of $S_{tab}$ rather than re-deriving its own answers. In the first case where $S$ uses answers of a tabled subgoal it is termed a consumer; in the second case when $S$ produces its own answers it is called a generator or producer.

- **Answer Similarity** – whereby a derived answer to a tabled subgoal is determined to contain information similar to that already in the set of answers for that subgoal.

Different measures of similarity are possible. XSB’s engine supports two measures for call similarity: variance and subsumption. XSB’s engine supports a variance-based measure for answer similarity, but allows users to program other measures in certain cases. We discuss call similarity here, but defer the discussion of answer similarity until Section 5.4.

\(^2\)The reader may have noted that table/1 is referred to as a directive, while auto_table/0 and suppl_table/0 were referred to as declarations. The difference is that at the command line, user can execute a directive but not a compiler declaration.
Determining Call Similarity via Variance  By default, XSB determines that a subgoal \( S \) is similar to a tabled subgoal \( S_{\text{tab}} \) if \( S \) is a variant of \( S_{\text{tab}} \), that is if \( S \) and \( S_{\text{tab}} \) are identical up to variable renaming \(^3\). As an example \( p(X,Y,X) \) is a variant of \( p(A,B,A) \), but not of \( p(X,Y,Y) \), or \( p(X,Y,Z) \). Under variance-based call similarity, or call variance, when a tabled subgoal \( S \) is encountered, a search for a table entry containing a variant subgoal \( S_{\text{tab}} \) is performed. Notice that if \( S_{\text{tab}} \) exists, then all of its answers are also answers to \( S \), and therefore will be resolved against it. Call variance was used in the original formulation of SLG resolution [15] for the evaluation of normal logic programs according to the well-founded semantics and interacts well with many of Prolog’s extra-logical constructs.

Determining Call Similarity via Subsumption  Call similarity can also be based on call subsumption. A term \( T_1 \) subsumes a term \( T_2 \) if \( T_2 \) is more specific than \( T_1 \) \(^4\). Furthermore, we say that \( T_1 \) properly subsumes \( T_2 \) if \( T_2 \) subsumes \( T_1 \), but is not a variant of \( T_1 \). Under call subsumption, when a tabled subgoal \( S \) is encountered, a search is performed for a table entry containing a subsuming subgoal \( S_{\text{tab}} \). Notice that, if such an entry exists, then its answer set \( \mathcal{A} \) logically contains all the solutions to satisfy \( C \). The subset of answers \( \mathcal{A}' \subseteq \mathcal{A} \) which unify with \( C \) are said to be relevant to \( C \).

Notice that call subsumption permits greater reuse of computed results, thus avoiding even more program resolution, and thereby can lead to time and space performances superior to call variance. In addition, beginning with Version 3.2, call-subsumption based tabling fully supports well-founded negation under the default local scheduling strategy. However, there are downsides to this paradigm. First of all, subsumptively tabled predicates do not interact well with certain Prolog constructs with which variant-tabled predicates can (see Example 5.2.4 below). Second, call subsumption does not yet support calls with tabled attributed variables or answer subsumption \(^5\).

Example 5.2.1  The terms \( T_1: p(f(Y),X,1) \) and \( T_2: p(f(Z),U,1) \) are variants as one can be made to look like the other by a renaming of the variables. Therefore, each subsumes the other.

---

\(^3\)Formally, \( S \) and \( S_{\text{tab}} \) are variants if they have an mgu \( \theta \) such that the domain and range of \( \theta \) consists only of variables.

\(^4\)Formally, \( T_1 \) subsumes \( T_2 \) if there is a substitution \( \theta \) whose domain consists only of variables from \( T_1 \) such that \( T_1 \theta = T_2 \).

\(^5\)Beginning with Version 3.2, XSB supports attributed variables in answers under call subsumption, although not in calls.
The term $t_3$: $p(f(Y), X, 1)$ subsumes the term $t_4$: $p(f(Z), Z, 1)$. However, they are not variants. Hence $t_3$ properly subsumes $t_4$.

The above examples show how a variant-based tabled evaluation can reduce certain redundant subcomputations over SLD. However, even more redundancy can be eliminated, as the following example shows.

**Exercise 5.2.5** Begin by abolishing all tables in XSB, and then type the following query

```prolog
?- abolish_all_tables.
?- path(X,Y), fail.
```

Notice that only a single table entry is created during the evaluation of this query. You can check that this is the case by invoking the following query

```prolog
?- get_calls_for_table(path/2,Call).
```

Now evaluate the query

```prolog
?- path(1,5), fail.
```

and again check the subgoals in the table. Notice that two more have been added. Further notice that these new subgoals are subsumed by that of the original entry. Correspondingly, the answers derived for these newer subgoals are already present in the original entry. You can check the answers contained in a table entry by invoking `get_returns_for_call/2` on a tabled subgoal. For example:

```prolog
?- get_returns_for_call(p(1,_),Answer).
```

Compare these answers to those of $p(X,Y)$ and $p(1,5)$. Notice that the same answer can, and in this case does, appear in multiple table entries.

Now, let's again abolish all the tables and change the evaluation strategy of $\text{path}/2$ to use subsumption.

```prolog
?- abolish_all_tables.
?- table path/2 as subsumptive.
```

And re-perform the first few queries:
?- path(X, Y), fail.
?- get_calls_for_table(path/2, Call).
?- path(1, 5).
?- get_calls_for_table(path/2, Call).

Notice that this time the table has not changed! Only a single entry is present, that for the original query \( p(X, Y) \).

When using call subsumption, XSB is able to recognize a greater range of “redundant” queries and thereby make greater use of previously computed answers. The result is that less program resolution is performed and less redundancy is present in the table. However, subsumption is not a panacea. The elimination of redundant answers depends upon the presence of a subsuming subgoal in the table when the call to \( p(1, 5) \) is made. If the order of these queries were reversed, one would find that the same entries would be present in this table as the one constructed under variant-based evaluation.

**Declarations for Call Variance and Call Subsumption**  By default tabled predicate use call variance. However, call subsumption can be made the default by giving XSB the \(-S\) option at invocation (refer to Section 3.7). More versatile constructs are provided by XSB so that the tabling method can be selected on a per predicate basis. Use of the directive

```
table p/n as subsumptive
```

or

```
table p/n as variant
```

described in Section 6.15.1, ensures that a tabled predicate is evaluated using the desired strategy regardless of the default tabling strategy.

### 5.2.2 Tabling with Interned Terms

XSB supports, on request, a special representation of *ground* terms, known as interned terms (see `intern_term/2`). This representation is also sometimes known as a “hash-consing” representation. All interned terms are stored in a global area and each such term is stored only once, with all instances of a given interned (sub-)term pointing to that one stored representation. This can allow for a much more succinct representation of sets of ground terms that share subterms. Importantly interned ground terms, in principle, do not need to be copied into and out of tables.
To take advantage of this possibility, a table must be declared as `intern`. As an example of a possible use of this mechanism, consider a simple DCG that recognizes all strings of a’s starting with a single b:

```prolog
:- table bas/2 as intern.

bas --> [b].
bas --> bas, [a].
```

This predicate must be tabled in order to terminate, since the grammar is left-recursive. If we use the usual list representation of an input string and use variant tabling, every call to `bas/2` and every return will copy the remaining list into the table, and recognition will be quadratic. (For example on my laptop, recognizing a list of one b followed by 10,000 a’s takes about 1.84 seconds, and 20,000 a’s about 7.285 seconds.) If we table `bas/2 as intern`, the initial ground input list will be interned (copied to intern space) on the first call, and after that every subsequent call of `bas/2` will be given an interned term, which need not be copied into (or out of) the table. In this case the complexity will be linear. (For example on my laptop, recognizing a list of one b and 1,000,000 a’s takes less than a second.)

When a table is declared `as intern`, at the time of a call, all arguments are automatically interned (with `intern_term/2`) before the call is looked up in the table, and on return, every answer is interned before being added to the table. Copying an interned subterm into or out of a table requires just a pointer copy, which takes, of course, constant time.

Because an interned term is treated just like a atom (with no indexing done on its structure), tabling as intern always uses variant tabling, and thus cannot be combined with subsumptive tabling. Also it cannot be combined with answer subsumption tabling.

For more information on tabling as intern, see [86].

### 5.2.3 Table Scheduling Strategies

Recall that SLD resolution works by selecting a goal from a list of goals to be proved, and selecting a program clause $C$ to resolve against that goal. During resolution of a top level goal $G$, if the list of unresolved goals becomes empty, $G$ succeeds, while if there is no program clause to resolve against the selected goal from the list resolution against $G$ fails. In Prolog clauses are selected in the order they are asserted, while literals are selected in a left-to-right selection strategy. Other strategies are possible
for SLD, and in fact completeness of SLD for definite programs depends on a non-fixed literal selection strategy. This is why Prolog, which has a fixed literal selection strategy is not complete for definite programs, even when they have bounded term-depth.

Because tabling uses program clause resolution, the two parameters of clause selection and literal selection also apply to tabling. Tabling makes use of a dynamic literal selection strategy for certain non-stratified programs (via the delaying mechanism described in Section 5.3.2), but uses the same left-to-right literal selection strategy as Prolog for definite programs. However, in tabling there is also a choice of when to return derived answers to subgoals that consume these answers. While full discussion of scheduling strategies for tabling is not covered here (see [28]) we discuss two scheduling strategies that have been implemented for XSB Version 3.6 6.

- **Local Scheduling** Local Scheduling depends on the notion of a subgoal dependency graph. For the state of a tabled evaluation, a non-completed tabled subgoal \( S_1 \) directly depends on a non-completed subgoal \( S_2 \) when \( S_2 \) is in the SLG tree for \( S_1 \) – that is when \( S_2 \) is called by \( S_1 \) without any intervening tabled predicate. The edges of the subgoal dependency graph are then these direct dependency relations, so that the subgoal dependency graph is directed. As mentioned, the subgoal dependency graph reflects a given state of a tabled evaluation and so may changed as the evaluation proceeds, as new tabled subgoals are encountered, or encountered in different contexts, as tables complete, and so on. As with any directed graph, the subgoal dependency graph can be divided up into strongly connected components, consisting of tabled subgoals that depend on one another. Local scheduling then fully evaluates each maximal SCC (a SCC that does not depend on another SCC) before returning answers to any subgoal outside of the SCC 7.

- **Batched Scheduling** Unlike Local Scheduling, Batched Scheduling allows answers to be returned outside of a maximal SCC as they are derived, and thus resembles Prolog’s tuple at a time scheduling.

Both Local and Batched Scheduling have their advantages, and we list points of comparison.

---

6Many other scheduling strategies are possible. For instance, [27] describes a tabling strategy implemented for the SLG-WAM that emulates magic sets under semi-naive evaluation. This scheduling strategy, however, is not available in Version 3.6 of XSB.

7XSB’s implementation maintains a slight over-approximation of SCCs – see [28].
• **Time for left recursion** Batched Scheduling is somewhat faster than Local Scheduling for left recursion as Local Scheduling imposes overhead to prevent answers from being returned outside of a maximal SCC.

• **Time to first answer** Because Batched Scheduling returns answers out of an SCC eagerly, it is faster to derive the first answer to a tabled predicate.

• **Stack space** Local evaluation generally requires less space than batched evaluation as it fully explores a maximal SCC, completes the SCC’s subgoals, reclaims space, and then moves on to a new SCC.

• **Integration with cuts** As discussed in Exercise 5.2.6 and throughout Section 5.2.4, Local Scheduling integrates better with cuts, although this is partly because tabled subgoals may be fully evaluated before the cut takes effect.

• **Efficiency for call subsumption** Because Local Evaluation completes tables earlier than Batched Evaluation it may be faster for some uses of call subsumption, as subsumed calls can make use of completed subsuming tables.

• **Negation and tabled aggregation** As will be shown below, Local Scheduling is superior for tabled aggregation as only optimal answers are returned out of a maximal SCC. Local Scheduling also can be more efficient for non-stratified negation as it may allow delayed answers that are later simplified away to avoid being propagated.

On the whole, advantages of Local Scheduling outweigh the advantages of Batched Scheduling, and for this reason Local Scheduling is the default scheduling strategy for Version 3.6 of XSB. XSB can be configured to use batched scheduling via the configuration option `--enable-batched-scheduling` and remaking XSB. This will not affect the default version of XSB, which will also remain available.

### 5.2.4 Interaction Between Prolog Constructs and Tabling

Tabling integrates well with most non-pure aspects of Prolog. Predicates with side-effects like `read/1` and `write/1` can be used freely in tabled predicates as long as it is remembered that only the first call to a goal will execute program clauses while the rest will look up answers from a table. However, other extra-logical constructs like the cut (!) pose greater difficulties. Tabling with call subsumption is also theoretically precluded from correct interaction with certain meta-logical predicates.
Cuts and Tabling  The semantics for cuts in Prolog is largely operational, and is usually defined based on an ordered traversal of an SLD search tree. Tabling, of course, has a different operational semantics than Prolog – it uses SLG trees rather than SLD trees, for instance – so it is not surprising that the interaction of tabling with cuts is operational. In Prolog, the semantics for a cut can be expressed in the following manner: a cut executed in the body of a predicate \( P \) frames from the top (youngest end) of the choice point stack down to and including the call for \( P \). In XSB a cut is allowed to succeed as long as it does not cut over a choice point for a non-completed tabled subgoal, otherwise, the computation aborts. This means, among other matters, that the validity of a cut depends on the scheduling strategy used for tabling, that is on the strategy used to determine when an answer is to be returned to a consuming subgoal. Scheduling strategy was discussed Section 5.2.3: for now, we assume that XSB’s default local scheduling is used in the examples for cuts.

Exercise 5.2.6  Consider the program

\[
\begin{align*}
\text{:- table cut_p/1, cut_q/1, cut_r/0, cut_s/0.} \\
cut_p(X) & :\ cut_q(X), \ cut_r/0, \ cut_s/0. \\
cut_r & :\ cut_s. \\
cut_s & :\ cut_q(_,). \\
cut_q(1), & \ cut_q(2). 
\end{align*}
\]

What solutions are derived for the goal \( \text{?- cut_p(X)} \)? Suppose that \( \text{cut_p/1} \) were rewritten as

\[
\text{cut_p(X) :\ cut_q(X), once(cut_r).}
\]

How should this cut over a table affect the answers generated for \( \text{cut_p/1} \)? What happens if you rewrite \( \text{cut_p/1} \) in this way and compile it in XSB?  

In Exercise 5.2.6, \( \text{cut_p(1)} \) and \( \text{cut_p(2)} \) should both be true. Thus, the cut in the literal \( \text{once(cut_r)} \) must not inadvertently cut away solutions that are demanded by \( \text{cut_p/1} \). In the default local scheduling of XSB Version 3.6 tabled subgoals are fully evaluated whenever possible before returning any of their answers. Thus the first call \( \text{cut_q(X)} \) in the body of the clause for \( \text{cut_p/1} \) is fully evaluated before proceeding to the goal \( \text{once(cut_r)} \). Because of this any choice points for \( \text{cut_q(X)} \) are to a completed table. For other scheduling strategies, such as batched scheduling, non-completed choice points for \( \text{cut_p/1} \) may be present on the choice point stack so that the cut would be disallowed. In addition, it is also possible to construct examples where a cut is allowed if call variance is used, but not if call subsumption is used.
Example 5.2.2 *A further example of using cuts in a tabled predicate is a tabled meta-interpreter.*

```prolog
:- table demo/1.

demo(true).
demo((A,B)) :- !, demo(A), demo(B).
demo(C) :- call(C).
```

More elaborate tabled meta-interpreters can be extremely useful, for instance to implement various extensions of definite or normal programs.

In XSB’s compilation, the cut above is compiled so that it is valid to use with either local or batched (a non-default) evaluation. An example of a cut that is valid neither in batched nor in local evaluation is as follows.

Example 5.2.3 *Consider the program*

```prolog
:- table cut_a/1, cut_b/1.

cut_a(X) :- cut_b(X).
cut_a(a1).

cut_b(X) :- cut_a(X).
cut_b(b1).
```

For this program the goal `?- cut_a(X)` produces two answers, as expected: `a1` and `b1`. However, replacing the first class of the above program with

```prolog
cut_a(X) :- once(cut_b(X)).
```

will abort both in batched or in local evaluation.

To summarize, the behavior of cuts with tables depends on dynamic operational properties, and we have seen examples of programs in which a cut is valid in both local and batched scheduling, in local but not batched scheduling, and in neither batched nor local scheduling. In general, any program and goal that allows cuts in batched scheduling will allow them in local scheduling as well, and there are programs for which cuts are allowed in local scheduling but not in batched.
Finally, we note that in Version 3.6 of XSB a “cut” over tables implicitly occurs when the user makes a call to a tabled predicate from the interpreter level, but does not generate all solutions. This commonly occurs in batched scheduling, but can also occur in local scheduling if an exception occurs. In such a case, the user will see the warning "Removing incomplete tables..." appear. Any complete tables will not be removed. They can be abolished by using one of XSB’s predicates for abolishing tables.

Call Subsumption and Meta-Logical Predicates Meta-logical predicates like var/1 can be used to filter the choices made during an evaluation. However, this is dangerous when used in conjunction with call subsumption, since call subsumption assumes that if a specific relation holds — e.g., p(a) — then a more general query — e.g., p(X) — will also hold.

Example 5.2.4 Consider the following simple program

\[
p(X) :- \text{var}(X), X = a.
\]

to which the queries

\[
?- p(X).
\]

\[
?- p(a).
\]

are posed. Let us compare the outcome of these queries when \( p/1 \) is (1) a Prolog predicate, (2) a variant-tabled predicate, and (3) a subsumptive-tabled predicate.

Both Prolog and variant-based tabling yield the same solutions: \( X = a \) and \textbf{no}, respectively. Under call subsumption, the query \( ?- p(X) \). likewise results in the solution \( X = a \). However, the query \( ?- p(a) \). is subsumed by the tabled subgoal \( p(X) \) — which was entered into the table when that query was issued — resulting in the incorrect answer \textbf{yes}. □

As this example shows, \textit{incorrect answers} can result from using meta-logical with subsumptive predicates in this way.

5.2.5 Potential Pitfalls in Tabling

Over-Tabling While the judicious use of tabling can make some programs faster, its indiscriminate use can make other programs slower. Naively tabling \( \text{append}/3 \)
append([],L,L).
append([H|T],L,[H|T1]) :- append(T,L,T1).

is one such example. Doing so can, in the worst case, copy \( N \) sublists of the first and third arguments into the table, transforming a linear algorithm into a quadratic one.

**Exercise 5.2.7** If you need convincing that tabling can sometimes slow a query down, type the query:

\[
?- \text{genlist}(1000,L), \text{prolog_append}(L,[a],\text{Out}).
\]

and then type the query

\[
?- \text{genlist}(1000,L), \text{table_append}(L,[a],\text{Out}).
\]

append/3 is a particularly bad predicate to table. Type the query

\[
?- \text{table_append}(L,[a],\text{Out}).
\]

leaving off the call to genlist/2, and backtrack through a few answers. Will table_append/3 ever succeed for this predicate? Why not?

Suppose DCG predicates (Section 11) are defined to be tabled. How is this similar to tabling append? □

We note that XSB has special mechanisms for handling tabled DCGs. See Section 11 for details.

**Tabled Predicates and Tracing** Another issue to be aware of when using tabling in XSB is tracing. XSB’s tracer is a standard 4-port tracer that interacts with the engine at each call, exit, redo, and failure of a predicate (see Chapter 10). When tabled predicates are traced, these events may occur in unexpected ways, as the following example shows.

**Exercise 5.2.8** Consider a tabled evaluation when the query \(?- \text{a}(0,X)\) is given to the following program

\[
:- \text{table \ mut\_ret\_a}/2, \text{mut\_ret\_b}/2.
\text{mut\_ret\_a}(X,Y) :- \text{mut\_ret\_d}(X,Y).
\]
mut_ret_a(X,Y) :- mut_ret_b(X,Z),mut_ret_c(Z,Y).

mut_ret_b(X,Y) :- mut_ret_c(X,Y).
mut_ret_b(X,Y) :- mut_ret_a(X,Z),mut_ret_d(Z,Y).

mut_ret_c(2,2). mut_ret_c(3,3).

mut_ret_d(0,1). mut_ret_d(1,2). mut_ret_d(2,3).

mut_ret_a(0,1) can be derived immediately from the first clause of mut_ret_a/2. All other answers to the query depend on answers to the subgoal mut_ret_b(0,X) which arises in the evaluation of the second clause of mut_ret_a/2. Each answer to mut_ret_b(0,X) in turn depends on an answer to mut_ret_a(0,X), so that the evaluation switches back and forth between deriving answers for mut_ret_a(0,X) and mut_ret_b(0,X).

Try tracing this evaluation, using creep and skip. Do you find the behavior intuitive or not?

5.3 Normal Programs

Normal programs extend definite programs to include default negation, which posits a fact as false if all attempts to prove it fail. As shown in Example 1.0.1, which presented one of Russell’s paradoxes as a logic program, the addition of default negation allows logic programs to express contradictions. As a result, some assertions, such as shaves(barber,barber) may be undefined, although other facts, such as shaves(barber,mayor) may be true. Formally, the meaning of normal programs may be given using the well-founded semantics and it is this semantics that XSB adopts for negation (we note that in Version 3.6 the well-founded semantics is implemented only for variant-based tabling).

5.3.1 Stratified Normal Programs

Before considering the full well-founded semantics, we discuss how XSB can be used to evaluate programs with stratified negation. Intuitively, a program uses stratified negation whenever there is no recursion through negation. Indeed, most programmers, most of the time, use stratified negation.

Exercise 5.3.1 The program
win(X):= move(X,Y), tnot(win(Y)).

is stratified when the move/2 relation is a binary tree. To see this, load the files
tree1k.P and table_examples.P from the directory $XSB_DIR/examples and type
the query

?- win(1).

win(1) calls win(2) through negation, win(2) calls win(4) through negation, and so
on, but no subgoal ever calls itself recursively through negation.

The previous example of win/1 over a binary tree is a simple instance of a stratified
program, but it does not even require tabling. A more complex example is presented
below.

Exercise 5.3.2 Consider the query ?- lrd_s to the following program

lrd_p:- lrd_q, tnot(lrd_r), tnot(lrd_s).
lrd_q:- lrd_r, tnot(lrd_p).
lrd_r:- lrd_p, tnot(lrd_q).
lrd_s:- tnot(lrd_p), tnot(lrd_q), tnot(lrd_r).

Should lrd_s be true or false? Try it in XSB. Using the intuitive definition of
“stratified” as not using recursion through negation, is this program stratified? Would
the program still be stratified if the order of the literals in the body of clauses for
lrd_p, lrd_q, or lrd_r were changed?

The rules for p, q and r are involved in a positive loop, and no answers are ever
produced. Each of these atoms can be failed, thereby proving s. Exercise 5.3.2
thus illustrates an instance of how tabling differs from Prolog in executing stratified
programs since Prolog would not fail finitely for this program 8.

Completely Evaluated Subgoals Knowing when a subgoal is completely eval-
uated can be useful when programming with tabling. Simply put, a subgoal S is

8LRD-stratified stratification may be reminiscent of the Subgoal Dependency Graphs of Sec-
tion 5.2.3 but differ in several respects, most notably in that stratification considers only cycles
through negative dependencies.
completely evaluated if an evaluation can produce no more answers for $S$. The computational strategy of XSB makes great use of complete evaluation so that understanding this concept and its implications can be of great help to a programmer.

Consider a simple approach to incorporating negation into tabling. Each time a negative goal is called, a separate table is opened for the negative call. This evaluation of the call is carried on to termination. If the evaluation terminates, its answers if any, are used to determine the success of failure of the calling goal. This general mechanism underlies early formulations for tabling stratified programs [39, 69]. Of course this method may not be efficient. Every time a new negative goal is called, a new table must be started, and run to termination. We would like to use information already derived from the computation to answer a new query, if at all possible — just as with definite programs.

XSB addresses this problem by keeping track of the state of each subgoal in the table. A call can have a state of complete, incomplete or not_yet_called. Calls that do have table entries may be either complete or incomplete. A subgoal in a table is marked complete only after it is determined to be completely evaluated; otherwise the subgoal is incomplete. If a tabled subgoal is not present in the table, it is termed not_yet_called. XSB contains predicates that allow a user to examine the state of a given table (Section 6.15).

There are in fact two ways that a tabled subgoal $S$ can be determined to be completely evaluated. If $S$ is part of an SCC $S$, (a mutually recursive component), then $S$ can be completed once it is ensure that all resolution steps have been done to all subgoals in $S$. Otherwise, if there is a derivation of an answer that is identical to $S$, $S$ can be completed before the rest of the subgoals in $S$ since further evaluation of $S$ itself will not produce useful information. In this case, we sometimes say that $S$ is early completed.

Using these concepts, we can overview how tabled negation is evaluated for stratified programs. If a literal $\text{tnot}(S)$ is called, where $S$ is a tabled subgoal, the evaluation checks the state of $S$. If $S$ is complete the engine simply determines whether the table contains an answer for $S$. Otherwise the engine suspends the computation path leading to $\text{tnot}(S)$ until $S$ is completed (and calls $S$ if necessary). Whenever a suspended subgoal $\text{tnot}(S)$ is completed with no answers, the engine resumes the evaluation at the point where it had been suspended. We note that because of this behavior, tracing programs that heavily use negation may produce behavior unexpected by the user.
Subject to some semantic restrictions, an XSB programmer can intermix the use of tabled negation (tnot/1) with Prolog’s negation (\'+/1, or equivalently fail_if/1 or not/1). These restrictions are discussed in detail below — for now we focus on differences in behavior or these two predicates in stratified programs. Recall that \'+(S) calls S and if S has a solution, Prolog executes a cut over the subtree created by \'+(S), and fails. tnot/1 on the other hand, does not execute a cut, so that all subgoals in the computation path begun by the negative call will be completely evaluated. The major reason for not executing the cut is to ensure that XSB evaluates ground queries to Datalog programs with negation with polynomial data complexity. As seen [15], this property cannot be preserved if negation “cuts” over tables.

There are other small differences between tnot/1 and \'+/1 illustrated in the following exercise.

**Exercise 5.3.3** In general, making a call to non-ground negative subgoal in Prolog may be unsound (cf. [48]), but the following program illustrates a case in which non-ground negation is sound.

ngr_p:- \+ ngr_p(_).
ngr_p(a).

One tabled analog is

:- table ngr_tp/1.
ngr_tp(a).

ngr_tp:- tnot(ngr_tp(_)).

Version 3.6 of XSB will flounder on the call to ngr_tp, but not on the call to ngr_p/0. On the other hand if not_exists/1 is used

ngr_skp:- not_exists(ngr_tp(_)).

the non-ground semantics is allowed.

not_exists/1 works by asserting a new tabled subgoal, abstractly

:- table '$_$ngr_tp'
'$_$skolem_ngr_tp' :- ngr_tp(_).
to avoid the problem with variables. In addition, since not_exists/1 creates a new
tabled predicate, it can be used to call non-tabled predicates as well, ensuring tabling.

The description of tnot/1 in Section 6.5 describes other small differences between
\'+/1 and tnot/1 as implemented in XSB. Before leaving the subject of stratification, we note that the concepts of stratification also underly XSB’s evaluation of
tabled findall: tfindall/3. Here, the idea is that a program is stratified if it con-
tains no loop through tabled findall (See the description of predicate tfindall/3 on
page 226).

5.3.2 Non-stratified Programs

As discussed above, in stratified programs, facts are either true or false, while in
non-stratified programs facts may also be undefined. XSB represents undefined facts
as conditional answers.

Conditional Answers

Exercise 5.3.4 Consider the behavior of the win/1 predicate from Exercise 5.3.1.

\[ \text{win}(X) :- \text{move}(X,Y), \text{tnot}(\text{win}(Y)). \]

when the move/2 relation is a cycle. Load the file $XSB\_DIR/examplescycle1k.P
into XSB and again type the query \?- win(1). Does the query succeed? Try tnot(win(1)).

Now query the table with the standard XSB predicate get_residual/2, e.g. ?-
get_residual(win(1),X). Can you guess what is happening with this non-stratified
program?

The predicate get_residual/2 (Section 6.15) unifies its first argument with a
tabled subgoal and its second argument with the (possibly empty) delay list of that
subgoal. The truth of the subgoal is taken to be conditional on the truth of the
elements in the delay list. Thus win(1) is conditional on tnot(win(2)), win(2) in
tnot(win(3)) and so on until win(1023) which is conditional on win(1).

From the perspective of the well-founded semantics, win(1) is undefined. In-
formally, true answers in the well-founded semantics are those that have a (tabled)
derivation. False answers are those for which all possible derivations fail — either
finitely as in Prolog or by failing positive loops. win(1) fits in neither of these cases
— there is no proof of win(1), yet it does not fail in the sense given above and is thus
undefined.
However this explanation does not account for why undefined answers should be represented as conditional answers, or why a query with a conditional answer and its negation should both succeed. These features arise from the proof strategy of XSB, which we now examine in more detail.

**Exercise 5.3.5** *Consider the program*

\[
\text{:- table simpl_p/1,simpl_r/0,simpl_s/0.} \\
simpl_p(X):- \text{tnot(simpl_s)}. \\
simpl_s:- \text{tnot(simpl_r)}. \\
simpl_s:- \text{simpl_p(X)}. \\
simpl_r:- \text{tnot(simpl_s)},\text{simpl_r}. \\
\]

*Try the query ?- simpl_p(X). If you have a copy of XSB defined using Batched Scheduling load the examples program and query ?- simpl_p(X) – be sure to backtrack through all possible answers. Now try the query again. What could possibly account for the difference in behavior between Local and Batched Scheduling?*

At this point, it is worthwhile to examine closely the evaluation of the program in Exercise 5.3.5. The query `simpl_p(X)` calls `simpl_s` and `simpl_r` and executes the portion of the program shown below in bold:

\[
\text{simpl_p(X):- tnot(simpl_s).} \\
\text{simpl_s:- tnot(simpl_r).} \\
\text{simpl_s:- simpl_p(X).} \\
\text{simpl_r:- tnot(simpl_s),simpl_r}. \\
\]

Based on evaluating only the bold literals, the three atoms are all undefined since they are neither proved true, nor fail. However if the evaluation could only look at the literal in italics, `simpl_r`, it would discover that `simpl_r` is involved in a positive loop and, since there is only one clause for `simpl_r`, the evaluation could conclude that the atom was false. This is exactly what XSB does, it delays the evaluation of `tnot(simpl_s)` in the clause for `simpl_r` and looks ahead to the next literal in the body of that clause. This action of looking ahead of a negative literal is called delaying. A delayed literal is moved into the delay list of a current path of computation.
Whenever an answer is derived, the delay list of the current path of computation is copied into the table. If the delay list is empty, the answer is unconditional; otherwise it is conditional. Of course, for definite programs any answers will be unconditional — we therefore omitted delay lists when discussing such programs.

In the above program, delaying occurs for the negative literals in clauses for `simpl_p(X)`, `simpl_s`, and `simpl_r`. In the first two cases, conditional answers can be derived, while in the third, `simpl_r` will fail as mentioned above. Delayed literals eventually become evaluated through *simplification*. Consider an answer of the form

```
simpl_p(X):- tnot(simpl_s) |
```

where the `|` is used to represent the end of the delay list. If, after the answer is copied into the table, `simpl_s` turns out to be false, (after being initially delayed), the answer can become unconditional. If `simpl_s` turns out to be true, the answer should be removed, it is false.

In fact, it is this last case that occurs in Exercise 5.3.5. The answer

```
simpl_p(X):- tnot(simpl_s) |
```

is derived, and returned to the user (XSB does not currently print out the delay list). The answer is then removed through simplification so that when the query is re-executed, the answer does not appear.

We will examine in detail how to alter the XSB interface so that evaluation of the well-founded semantics need not be confusing. It is worthwhile to note that the behavior just described is uncommon.

Version 3.6 of XSB handles dynamically stratified programs through delaying negative literals when it becomes necessary to look to their right in a clause, and then simplifying away the delayed literals when and if their truth value becomes known. However, to ensure efficiency, literals are never delayed unless the engine determines them to not to be stratified under the LRD-stratified evaluation method.

**When Conditional Answers are Needed**  A good Prolog programmer uses the order of literals in the body of a clause to make her program more efficient. However, as seen in the previous section, delaying can break the order that literals are evaluated within the body of a clause. It then becomes natural to ask if any guarantees can be made that XSB is not delaying literals unnecessarily.

Such a guarantee can in fact be made, using the concept of *dynamic stratification* [50]. Without going into the formalism of dynamic stratification, we note that a
program is dynamically stratified if and only if it has a two-valued model. It is also known that computation of queries to dynamically stratified programs is not possible under any fixed strategy for selecting literals within the body of a clause. In other words, some mechanism for breaking the fixed-order literal selection strategy must be used, such as delaying.

However, by redefining dynamic stratification to use an arbitrary fixed-order literal selection strategy (such as the left-to-right strategy of Prolog), a new kind of stratification is characterized, called Left-to-Right Dynamic Stratification, or LRD-stratification. LRD-stratified is not as powerful as dynamic stratification, but is more powerful than other fixed-order stratification methods, and it can be shown that for ground programs, XSB delays only when programs are not LRD-stratified. In the language of [65] XSB is delay minimal.

Programming in the Well-founded Semantics  XSB delays literals for non-LRD-stratified programs and later simplifies them away. In Local Scheduling, all simplification will be done before the first answer is returned to the user. In Batched Scheduling it is usually better to make a top-level call for a predicate, p as follows:

?- p,fail ; p.

when the second p in this query is called, all simplification on p will have been performed. However, this query will succeed if p is true or undefined.

Exercise 5.3.6 Write a predicate wfs_call(+Tpred,+Val) such that if Tpred is a ground call to a tabled predicate, wfs_call(+Tpred,+Val) calls Tpred and unifies Val with the truth value of Tpred under the well-founded semantics. Hint: use get_residual/2.

How would you modify wfs_call(?Tpred,?Val) so that it properly handled cases in which Tpred is non-ground.

Trouble in Paradise: Answer Completion  The engine for XSB performs both program clause and answer resolution, along with delay and simplification. What it does not do is to perform an operation called answer completion which is needed in certain (pathological?) programs.

Exercise 5.3.7 Consider the following program:
Using either the predicate from Exercise 5.3.6 or some other method, determine the truth value of \( \text{ac}_p(X) \). What should the value be? (hint: what is the value of \( \text{ac}_s(1) \)?)

For certain programs, XSB will delay a literal (such as \( \text{ac}_p(X) \)) that it will not be able to later simplify away. In such a case, an operation, called answer completion, is needed to remove the clause

\[
\text{ac}_p(X) :\neg \text{ac}_p(X) \mid
\]

Without answer completion, XSB may consider some answers to be undefined rather than false. It is thus is sound, but not complete for terminating programs to the well-founded semantics. Answer completion is not available for Version 3.6 of XSB, as it is expensive and the need for answer completion arises rarely in practice. However answer completion will be included at some level in future versions of XSB.

### 5.3.3 On Beyond Zebra: Implementing Other Semantics for Non-stratified Programs

The Well-founded semantics is not the only semantics for non-stratified programs. XSB can be used to (help) implement other semantics that lie in one of two classes. 1) Semantics that extend the well-founded semantics to include new program constructs; or 2) semantics that contain the well-founded partial model as a submodel.

An example of a semantics of class 1) is (WFSX) [3], which adds explicit (or provable) negation to the default negation used by the Well-founded semantics. The addition of explicit negation in WFSX, can be useful for modeling problems in domains such as diagnosis and hierarchical reasoning, or domains that require updates [44], as logic programs. WFSX is embeddable into the well-founded semantics; and
this embedding gives rise to an XSB meta-interpreter, or, more efficiently, to the preprocessor described in Section Extended Logic Programs in Volume 2. See [72] for an overview of the process of implementing extensions of the well-founded semantics.

An example of a semantics of class 2) is the stable model semantics. Every stable model of a program contains the well-founded partial model as a submodel. As a result, the XSB can be used to evaluate stable model semantics through the residual program, to which we now turn.

The Residual Program Given a program \( P \) and query \( Q \), the residual program for \( Q \) and \( P \) consists of all (conditional and unconditional) answers created in the complete evaluation of \( Q \).

Exercise 5.3.8 Consider the following program.

\[
:- \text{table ppgte}_p/0, \text{ppgte}_q/0, \text{ppgte}_r/0, \text{ppgte}_s/0, \text{ppgte}_t/0, \text{ppgte}_u/0, \text{ppgte}_v/0.
\text{ppgte}_p:- \text{ppgte}_q. \text{ppgte}_p:- \text{ppgte}_r.
\text{ppgte}_q:- \text{ppgte}_s. \text{ppgte}_r:- \text{ppgte}_u.
\text{ppgte}_q:- \text{ppgte}_t. \text{ppgte}_r:- \text{ppgte}_v.
\text{ppgte}_s:- \text{ppgte}_w. \text{ppgte}_u:- \text{undefined}.
\text{ppgte}_t:- \text{ppgte}_x. \text{ppgte}_v:- \text{undefined}.
\text{ppgte}_w:- \text{ppgte}(1). \text{ppgte}_x:- \text{ppgte}(0).
\text{ppgte}_w:- \text{undefined}. \text{ppgte}_x:- \text{undefined}.
\text{ppgte}(0).
\]

\[
:- \text{table undefined/0}.
\text{undefined}:- \text{tnot(undefined)}.
\]

Write a routine that uses get_residual/2 to print out the residual program for the query \( ?- \text{ppgte}_p, \text{fail} \). Try altering the tabling declarations, in particular by making \( \text{ppgte}_q/0, \text{ppgte}_r/0, \text{ppgte}_s/0 \) and \( \text{ppgte}_t/0 \) non-tabled. What effect does altering the tabling declarations have on the residual program?

When XSB returns a conditional answer to a literal \( L \), it does not propagate the delay list of the conditional answer, but rather delays \( L \) itself, even if \( L \) does not
occur in a negative loop. This has the advantage of ensuring that delayed literals are not propagated exponentially through conditional answers.

**Stable Models** Stable models are one of the most popular semantics for non-stratified programs. The intuition behind the stable model semantics for a ground program $P$ can be seen as follows. Each negative literal $\text{not} L$ in $P$ is treated as a special kind of atom called an *assumption*. To compute the stable model, a guess is made about whether each assumption is true or false, creating an assumption set, $A$. Once an assumption set is given, negative literals do not need to be evaluated as in the well-founded semantics; rather an evaluation treats a negative literal as an atom that succeeds or fails depending on whether it is true or false in $A$.

**Example 5.3.1** Consider the simple, non-stratified program

\begin{verbatim}
writes_manual(terry)←¬writes_manual(kostis), has_time(terry).
writes_manual(kostis)←¬writes_manual(terry), has_time(kostis).
has_time(terry).
has_time(kostis).
\end{verbatim}

there are two stable models of this program: in one \texttt{writes\_manual(terry)} is true, and in another \texttt{writes\_manual(kostis)} is true. In the Well-Founded model, neither of these literals is true. The residual program for the above program is

\begin{verbatim}
writes_manual(terry)←¬writes_manual(kostis).
writes_manual(kostis)←¬writes_manual(terry).
has_time(terry).
has_time(kostis).
\end{verbatim}

Computing stable models is an intractable problem, meaning that any algorithm to evaluate stable models may have to fall back on generating possible assumption sets, in pathological cases. For a ground program, if it is ensured that residual clauses are produced for *all* atoms, using the residual program may bring a performance gain since the search space of algorithms to compute stable models will be correspondingly reduced. In fact, by using XSB in conjunction with a Stable Model generator, Smodels [54], an efficient system has been devised for model checking of concurrent systems that is 10-20 times faster than competing systems [47]. In addition, using the XASP package (see the separate manual, [12] in XSB’s packages directory) a consistency checker for description logics has also been created [73].
5.4 Answer Subsumption

By default XSB adds an answer $A$ to a table $T$ only if $A$ is not a variant of some other answer already in $T$, a technique termed \textit{answer variance}. While answer variance is sufficient to allow tabling to compute the well-founded semantics and to terminate for programs with bounded term-depth, other choices of when and how to add an answer can be made. Using \textit{partial order answer subsumption}, $A$ would be added to $T$ only if $A$ is maximal with respect to other answers in $T$ according to a given partial order $>_O$. Furthermore if $A$ is added, any answers in $T$ that $A$ subsumes (i.e., is greater than in $>_O$) are deleted. When using \textit{lattice answer subsumption}, $A$ itself may not be added to $T$, rather the join is taken of $A$ and another answer $A'$ in $T$, with $A'$ being deleted. Despite its conceptual simplicity, answer subsumption can be a powerful tool. Partial order answer subsumption allows a table to retain only answers that are maximal according to a metric or to a preference relation; lattice answer subsumption can form the basis of multi-valued logics, quantitative logics, and of abstract interpretations for programs and process logics.

5.4.1 Types of Answer Subsumption

Partial Order Answer Subsumption.

We illustrate the use of partial order answer subsumption through a shortest-path predicate (Figure 5.1) that counts the number of edges between two vertices.

\begin{align*}
    \text{sp}(X,Y,1) & :\text{- edge}(X,Y). \\
    \text{sp}(X,Z,N) & :\text{- sp}(X,Y,N1),\text{edge}(Y,Z),N \text{ is } N1 + 1.
\end{align*}

\textbf{Figure 5.1: A Shortest Path Predicate}

As mentioned above, partial-order answer subsumption retains in a table $T$ only those answers that are maximal according to a given partial order $>_O$. In the case of the shortest-path predicate of Figure 5.1, $sp(A_1,A_2,A_3) >_O sp(B_1,B_2,B_3)$ if, $A_1 = B_1$, $A_2 = B_2$, and $A_3 < B_3$. Note that that minimal distances are maximal in $<_O$, and that $<_O$ is undefined if $A_3$ or $B_3$ is non-numeric. In XSB, partial order answer subsumption is specified for $sp/3$ using the declaration

\begin{verbatim}
:- table sp(_,_,po((<)/2)).
\end{verbatim}

In a given state of computation, only those answers that are maximal according to $>_O$ are available for resolution. Thus, for a finite graph with cycles, $sp/3$ will terminate.
using answer subsumption, but not with answer variance. Other partial orders beyond distance metrics may be useful. For instance, $>_{O}$ may specify a preference ordering between derived atoms so that answer subsumption provides an alternative to default-based methods for computing preferences.

The treatment of variables in calls to partial order answer subsumptive tabled predicates deserves mention. Variables in arguments not in the subsumption position are treated as "group-by" variables: i.e., for each value such a variable can take, a different aggregate is computed. So for example a call to $sp(a,X,SD)$ will succeed for each node reachable from $a$, binding $X$ to that node and $SD$ to the shortest distance from $a$ to that node. One can place a $\sim$ in a non-subsumption position of table declaration, e.g.,

$$:-\ \text{table} \ sp(_,\sim,po((<)/2)).$$

to indicate that values of that position should be aggregated over. For example, with this table declaration, the call $sp(a,X,SD)$ will find the distance to the closest node reachable from $a$, (which, if $a$ has any successors, will be 1, since a successor to $a$ will be a nearest reachable successor at distance 1 from $a$.)

Non-variables in the subsumption position in a call will be treated as selecting what answers are included in the aggregation.

**Lattice Answer Subsumption.**

An upper semi-lattice is a partial order for which any two elements have a unique least upper bound. Because the ordering for the third argument of $sp/3$ is total, it also forms an upper semi-lattice, and so can be computed using lattice answer subsumption. \footnote{The terminology lattice answer subsumption is employed even though only the join of the lattice is used.} In XSB lattice answer subsumption for $sp/3$ is declared as

$$:-\ \text{table} \ sp(_,_,\text{lattice}(\text{min}/3)).$$

with $\text{min}/3$ defined as $\text{min}(X,Y,Z) :- \ Z \ \text{is} \ \text{min}(X,Y)$. Operationally, this means that whenever an answer $sp(A_1,A_2,A_3)$ is derived, if there is another answer $sp(B_1,B_2,B_3)$ where $A_1 = B_1$ and $A_2 = B_2$ the join $J_3$ of $A_3$ and $B_3$ is taken, and only $sp(A_1,A_2,J_3)$ is available for resolution. As with a partial order, the join operation ensures termination for shortest path over a finite graph with cycles.

As the following proposition shows, lattice answer subsumption can be modeled either starting with a lattice, or starting with a function with appropriate properties.
Proposition 5.4.1 Let \( \text{op} \) be an associative, commutative, and idempotent binary function. Then there is a partial order \( P \), such that \( P \) is an upper semi-lattice with join \( \text{op} \).

Conversely, if a function does not have the above properties, it is not suitable for lattice answer subsumption. Accordingly the aggregate functions count and sum cannot be computed using lattice answer subsumption\(^1\). Lattice answer subsumption has a variety of applications. \([77]\) shows how it is used for social-network analysis and Section 5.4.2 shows its use for an application of multi-valued logics, \([72]\) describes how a similar formalism can implement a quantitative logic, and \([59, 60]\) describes how XSB’s PITA package is based on answer subsumption (see Volume 2 of this manual).

Partial Order Answer Subsumption with Abstraction.

Computation over an abstract domain may require certain maximal answers to be abstracted. In many cases, abstraction can be modeled by a join operation, but in others the abstraction represents an implicit induction step in the following sense. Given a set \( A \) of answers, it may be detected that the program computed does not have a finite model. An abstraction operation then is applied so that \( A \) and its extensions can be symbolically represented by a single answer \( A \). Using answer subsumption, this abstraction can be taken only if needed during program execution. Abstractly, partial order answer subsumption with abstraction uses the declaration

\[
\text{:- table p(_,_,po(rel/2,abs/3)).}
\]

where \( \text{rel/2} \) is a partial order, and \( \text{abs/3} \) is the abstraction operation. Section 5.4.2 provides a detailed example of how such an approach is used to analyze a process logic.

5.4.2 Examples of Answer Subsumption

Answer Subsumption and Abstract Interpretation

Net-style formalisms, such as Petri Nets, Workflow Nets, etc. have been used extensively for process modeling. Reachability is a central problem in analyzing properties

\(^{10}\)Since count and sum are not idempotent their semantics is based on multi-sets, rather than sets. Incorporating these as tabling features requires modifying their semantics to be set-based, in a manner similar to aggregation ASP systems.
of such nets, to which properties such as liveness, deadlock-freedom, and the existence of home states can be reduced. However, many interesting net formalisms cannot guarantee a finite number of configurations in a given net, so abstraction methods must be applied for their analysis.

For instance, the lack of finiteness is a problem in analyzing Place/Transition (PT) Nets. PT nets have no guard conditions or after-effects, and do not distinguish between token types. However, PT nets do allow a place to hold more than one token, leading to a potentially infinite number of configurations. This can be seen in the simple network of Figure 5.2 (from [24]) in which transitions are denoted by squares and places by circles. Each transition removes one token from the places that are the sources of its input edges and adds one token to each place at the target of each of its output edges. Starting from the configuration in Figure 5.2, repeated application of transition $t_1$ leads to place $s_2$ containing an unbounded number of tokens; repeated application of the sequence $t_1,t_2,t_3,t_4$ leads to place $s_4$ containing an unbounded number of tokens.

Despite such examples, reachability in PT nets is decidable and can be determined using an abstraction method called $\omega$-sequences, (see e.g. [24]). The main idea in determining $\omega$ sequences is to define a partial order $\geq_\omega$ on configurations as follows. If configurations $C_1$ and $C_2$ are both reachable, $C_1$ and $C_2$ have tokens in the same set $PL$ of places, $C_1$ has at least as many tokens in each place as $C_2$, and there exists a non-empty $PL_{sub} \subseteq PL$, such that for each $pl \in PL_{sub}$ $C_1$ has strictly more tokens than $C_2$, then $C_1 >_\omega C_2$. When evaluating reachability, if $C_2$ is reached first, and then $C_1$ was subsequently reached, $C_1$ is abstracted by marking each place in $PL_{sub}$ with the special token $\omega$ which is taken to be greater than any integer. If $C_1$ was reached first and then $C_2$, $C_2$ is treated as having already been seen.

Tabling combined with partial order answer subsumption requires slightly over 100 lines of code to model reachability in PT nets using $\omega$-sequences. Due to space restrictions, the program cannot be fully described here, but the top-level reachability predicate is shown in Figure 5.3. Despite its succinctness, it can evaluate reachability
CHAPTER 5. USING TABLING IN XSB: A TUTORIAL INTRODUCTION  111

:- table reachable(_,po(omega_gte/2,omega_abs/3)).
reachable(InConf,NewConf):-
    reachable(InConf,NewConf),
    hasTransition(Conf,NewConf).
reachable(InConf,NewConf):- hasTransition(InConf,NewConf).

Figure 5.3: Top-level predicate for PT net reachability

in networks with millions of states in a few minutes. This use of tabling to determine reachability in PT nets can be seen as a special case of tabling for abstract interpretation (cf. [38] and other works). However the framework for answer subsumption described here allows tabling to be used to efficiently perform abstract interpretation within a general Prolog system

Scalability for multi-valued and quantitative logics

The technique of program justification (cf. e.g. [55]) has been used for debugging tabled programs that cannot be debugged by traditional means. Here, we consider justification in the context of the Silk system, currently under development at Vulcan, Inc. Silk is a commercial knowledge representation and rule system built on top of Flora-2, which is implemented using XSB. One of the salient features of Silk is its default reasoning, which is based on a parameterized argumentation theory evaluated under the well-founded semantics [84]. One issue in using Silk is that knowledge engineers must have a way of understanding the reasoning of the system, a task complicated by the use of the well-founded semantics and the intricacies of the argumentation theory. We describe an experimental approach to justification of Silk-style argumentation theories using multi-valued logics.

As noted in [84], argumentation theories in Silk are usually extensions of the default theories of Courteous Logic Programs (CLP) and are based on two user-defined predicates: opposes/2 and overrides/2. Two atoms oppose each other if no model of a program can contain both atoms: an atom and its explicit negation oppose each other, but opposition can capture many other types of contradictions. Given two opposing atoms, one atom may override the other, and so be given preference. For atoms $A_1$ and $A_2$, if $A_1$ and $A_2$ are both derivable and oppose each other but neither overrides the other, $A_1$ and $A_2$ mutually rebut each other. If in addition $A_1$, say, overrides $A_2$, $A_1$ refutes $A_2$ 11. Within Silk and Flora-2, the compilation of an argumentation theory ensures that rebutted atoms have an undefined truth value, as

11In [84] argumentation theories are built on named rules, here we base them on derived atoms.
do atoms that refute themselves (i.e. if the overrides/2 predicate is cyclic). However, for justification, it is meaningful to distinguish those facts that are undefined due to a negative loop in the argumentation theory from those that are undefined due to a negative loop in the program itself. In addition, it is meaningful to distinguish an atom that is true because it overrides some other atom, from an atom whose derivation does not depend on the argumentation theory. Similar distinctions can be made for default false literals leading to the truth lattice shown in Figure 5.4.

5.4.3 Term-Sets

XSB provides support for a programming technique for representing sets of terms, called term-sets. (While it is not closely related to answer subsumption, it is partially implemented through tabling and a table declaration, and so this facility is documented here.)

We begin in an example. We can represent a set of Prolog terms by using a particular term of the form \{Var:Goal\} where Goal has (only) Var free in it. Then we will use this set-term to represent the set of terms obtained by evaluating Goal and taking the values of Var that are obtained. I.e., they would be the terms in the list L returned by the Prolog call to setof(Var,Goal,L). For example, the set-term:

\{X : member(X,[a,b,c])\}

represents the set of terms \{a,b,c\}.

Now a term-set is a Prolog term that may contain set-terms as subterms. For example,
m({X:member(X,[a,b,c])},g(d,{Y:member(Y,[e,f,g])}),h)

is a term-set, and it represents the set of terms obtained from it by replacing (re-
cursively) any embedded set-term by a term in that set-term. So the above term-set
represents the 9 terms:

m(a,g(d,e),h)  m(a,g(d,f),h)  m(a,g(d,g),h)
m(b,g(d,e),h)  m(b,g(d,f),h)  m(b,g(d,g),h)
m(c,g(d,e),h)  m(c,g(d,f),h)  m(c,g(d,g),h)

This example shows an advantage of this representation. Say a term-set has \( k \)
subset-terms each of which is of the member form in this example where each member
has a list of atoms of length \( n \). To represent this set of terms explicitly takes \( O(n^k) \) space,
whereas to represent them with the term-set takes only \( O(n \times k) \) space. So a term-set
representation can take exponentially less space than an explicit representation.

It is relatively easy to write a predicate, \textit{member_termset}/2, which takes a variable
and a term-set and nondeterministically generates all concrete terms represented by
the term-set, called \textit{extensionalizing} the term-set. Some care must be taken since a
call to goal to extensionalize a set-term may itself return a term-set. Also term-sets
can be self-recursive and thus represent infinitely many Prolog terms. For example,
consider the term-set:

\{X : p(X)\} where
\[ p(a). \]
\[ p(f({X:p(X)})). \]

This term-set represents the terms for which \( p/1 \) is true. Now \( p(a) \) is true, so \( a \) is
in the term-set. Since \( a \) is in \( \{X:p(X)\} \), then \( p(f(a)) \) is true because of the second
fact for \( p/1 \), and so \( f(a) \) is in the term-set. And so on. So this term-set contains the
infinitely many terms:

\[ a, f(a), f(f(a)), f(f(f(a))), \ldots \]

A particularly interesting use of term-sets is in conjunction with tabling. Consider
the term-set \( \{X:p(1,2,X)\} \) where \( p/3 \) is tabled. If \( p(1,2,\_\) \) has been called and so
its table is filled, then extensionalizing this term-set requires just a table lookup; in
some sense we can think of such a term-set as standing for a pointer into a table to a
set of terms. This can be elegantly used to solve an important problem in handling
parse trees in context-free parsing.

Consider the following DCG for the language \( a^* \):
which recognizes a string of \( a \)'s and constructs its parse trees.

To generate all answers, this DCG will take time exponential in the length of the input string; not surprising since there are exponentially many parses. But say we give it an input string of \( n \) \( a \)'s followed by one \( b \). In this case it will take exponential time to fail, since it will construct all the exponentially many partial parse trees for the initial \( k \) \( a \)'s. We would like the parser in this case to fail in polynomial time. We can do this by representing the parse trees as a term-set during the recognition the string. Then after the string is recognized, we extensionalize the set-term that represents the parse trees. In this way we can get the behavior we want. The set-term that represents the parse trees for any grammar will be constructed in polynomial time; the extensionalization of that term-set will take exponential time only if there are exponentially many parses.

We can cause XSB to automatically use the term-set representation for the grammar by adding to the above program the declaration:

\[
:- \text{table } a(\text{termset},_,_).
\]

which tells XSB to use the term-set representation of the first argument of nonterminal \( a/3 \).

With this declaration, XSB transforms the above program into the following:

\[
:- \text{table } a/3.
\]

\[
a(a(P1,P2),S0,S) :- \_\_a'(P1,S0,S1),\_\_a'(P2,S1,S).
a(a,S0,S1) --> 'C'(S0,a,S1).
\]

\[
:- \text{table } \_\_a'/3 \text{ as subsumptive.}
\_\_a'({X: \_\_a'(X,S0,S)},{S0,S}) :- a(_,S0,S).
\]

A new predicate \( \_\_a'/3 \) has been introduced, and all calls to the original predicate \( a/3 \) are replaced by calls to the new one. It is defined to call the original \( a/3 \) but to return the term-set instead of the concrete parse tree in the argument declared to be a term-set.

We can see that a call to \( a/3 \) in this new program will have exactly as many answers as the corresponding call to \( a/2 \) in the original recognizing DCG, since given
values for S0 and S, a call to '$_a$/2 returns only one value in its first argument. So a call to a/3 with have the polynomial complexity of the recognizer. So now when this representation is used, one gets the concrete parse tree for a string by writing, for example:

| ?- a(Pts,[a,a,a,a,a,a,a],[]), member_termset(Parse,Pts).

Here the term-set representing the parses for the sequence of a’s will be returned in the variable Pts, and then member_termset is used to extensionalize it to the produce the actual explicit parse tree. With this way of handling parse trees in arbitrary context-free grammars, the complexity of parsing to create the term-set is always polynomial, and then extensionalizing the term-set may be exponential if all parses are desired and there are exponentially many of them. (In fact, if the grammar contains a rule such as A --> A, there may be infinitely many parses.) Of course, if the parsing call to a/3 fails, then there is no extensionalization to do, and the process is polynomial.

Note that the transformation uses subsumptive tabling for the newly introduced auxiliary predicate. This is important for this example, since the parsing calls to '$_a$/3 will normally have S0 bound and S free, yet when extensionalizing the constructed term-set to obtain the parse trees, the calls will have both S0 and S bound. We do not want to recompute the parse during extensionalization, which would happen were we to use variant tabling, and so we use subsumptive tabling.

Problems in graph traversal provide another example of the effective use of term-sets. For graph reachability, we have the very familiar:

:- table reach/2.
reach(X,Y) :- edge(X,Y).
reach(X,Y) :- reach(X,Z), edge(Z,Y).

which is linear in the number of edges in the graph. But say that we now want to construct the path from X to Y when Y is reachable from X. One simple way to do it (collecting the intermediate nodes in the path in reverse order) is:

:- table path/3.
path(X,Y,[]) :- edge(X,Y).
path(X,Y,[Z|Path]) :- path(X,Z,Path), edge(Z,Y).

For an acyclic edge graph, this works fine, but for a graph with cycles, this will go into an infinite loop. Indeed, it must, since in a cyclic graph there are infinitely many different paths between some nodes. However, we can use term-set to handle this situation more flexibly. We modify the above program by adding:
:- table path(_,_,termset).

With this declaration, every call to `path/3` (for a finite edge graph) will terminate in time linear in the number of edges. And all the paths will be presented in the term-set returned in the third argument. Here we have an advantage similar to the one we had in the grammar example above: if there is no path from our source to our target node, we will find that out in linear time. Without the term-set declaration, this might take exponential time, while the program builds all the paths to all the nodes that are reachable from our source node. Also, if we want only one possible path from our source to our target, we can easily retrieve only one member of the term-set during extensionalization, and the whole process is still linear.

Now consider what happens with when the graph has cycles. In this case, the term-set may be recursive and represent the infinitely many paths between nodes. For example, the term-set representing all paths from `a` to `a` in the graph with a single edge from `a` to `a` will have the same structure as the example of an infinite term-set given at the beginning of this subsection. Once the path term-set is constructed (in time linear in the number of edges for a single source), producing paths reduces to processing the term-set structure. For example to generate all paths between nodes which do not contain repeated intermediate nodes, one could write an extensionalization predicate that passes a list of term-sets in the process of being expanded, and refuse to re-expand one currently being expanded. This is the technique often used in Prolog without tabling to compute reachability in cyclic graphs.

All of these examples can be seen as special cases of constructing proof trees or justifications of goals. Indeed, term-sets could be effectively used in the construction of a justification or explanation system.

### 5.5 Tabling for Termination

As noted throughout this manual, tabling adds important termination properties to programs and queries. In this section we state more precisely what these termination properties are, and how the properties can be strengthened through declarations and settings for subgoal abstraction and for bounded rationality through answer abstraction.

Before doing so, it is important to set the context for where issues of termination may arise. Consider first a pure normal program in which every predicate is tabled. This means a program where rules may only call other rules, possibly through negation (`tnot/1, not_exists/1 or u_not/1` in XSB); but where there are no calls to builtins
for arithmetic, all-solutions predicates, or other built-ins. If such a fully-tabled pure normal program does not have function symbols, XSB will always terminate for any query. For instance, XSB will terminate for fully tabled pure datalog programs – even if the head of a rule is “unsafe” in that it contains variables that do not occur in the body of that rule \(^{12}\).

While datalog programs are useful for certain kinds of knowledge representation, they are not powerful enough for general programming as they do not allow recursive structures such as lists. Thus, for the rest of this section we consider pure programs that may contain function symbols. Consider a pure definite program in which every predicate is tabled. Such a program would call each tabled subgoal (up to variance) exactly once if call variance were used, and at most once if call subsumption were used. In addition, tabling guarantees that each answer will be returned to each call to a tabled subgoal at most once. This means that there are two sources of non-termination. Either there can be an infinite number of subgoals, or there can be an infinite number of answers \(^{13}\).

**An Infinite Number of Subgoals** If a definite program produces an infinite number of subgoals but has a finite number of answers, the program can be made to terminate by abstracting the subgoal. For instance, consider the program fragment:

```prolog
:- table p/1.
p(X) :- p(f(X)).
```

The goal `?- p(1)` can create an infinite number of tabled subgoals: `p(f(1))`, `p(f(f(1)))`, `p(f(f(f(1))))` and so on. Note that since all of the subgoals are ground, none subsume one another, so that call subsumption will not help here, (although call subsumption would help if the goal were `?- p(X)`).

**Infinite Answers** Of course, subgoal abstraction can’t handle cases where there are an infinite number of answers, as in the program fragment:

```prolog
p(f(X)) :- p(X).
```

when given the query `?- p(X)`.

We consider each case in turn.

\(^{12}\)Evaluations that call non-ground negative literals will terminate through floundering, although this can be avoided in most cases by using `not_exists/1`.

\(^{13}\)Using the forest of trees model of tabling (cf. Section 10.3) non-termination requires that there are an infinite number of trees or that at least one tree have infinite size.
5.5.1 Subgoal Abstraction

In a nutshell, subgoal abstraction allows a goal like \( p(f(f(f(1)))) \) to be rewritten as

\[
p(f(f(X))), x = f(1).
\]

If all subgoals that have a term depth (or term size) over a given finite threshold are abstracted, any query can produce only a finite number of subgoals (since there are a finite number of predicate, function and constant symbols in any program). If a program is definite, it can be shown that any query to a program will terminate if that program uses subgoal abstraction [79]. For normal programs, the situation is not much different at a conceptual level. A goal such as \( \text{tnot}(p(f(f(f(1)))) \) would execute as \( p(f(f(X))) \) and then ensure that none of the answers to this goal have a binding for \( X \) that allows it to unify with \( f(1) \). Using this intuition, it can be shown that if a program has a well-founded model with a finite number of true or undefined answers it will terminate using tabling with subgoal abstraction [60, 61].

Despite its theoretical power, subgoal abstraction can also cause problems if used indiscriminately. If the second argument of the subgoal

\[
?- \text{member}(e,[a,b,c,d,e])
\]

is abstracted forming the goal

\[
?- \text{member}(e,[a,b,c|X])
\]

leading to an infinite number of answers. a goal that terminates without abstraction will not terminate after abstraction. Note that any program containing \text{member}/2 and at least one constant does not have a finite model. While an experienced programmer would never want to table \text{member}/2, he well may want to table a grammar or other program that performs recursion through a finite structure.

Declaring Subgoal Abstraction

Subgoal abstraction in XSB is performed for goals called positively; but not for goals called negatively including \text{tnot}/1 and \text{not_exists}/1; rather, such goals will throw an exception if they surpass the specified depth. In addition, subgoal abstraction is only implemented for call variance, and applies equally all terms, whether they are lists or non-lists. However, despite these restrictions, a tabled evaluation can be guaranteed to terminate for queries to safe programs (cf. [60]).
Subgoal abstraction can be declared by setting a value for the maximum depth of a subgoal and for the action to take when a subgoal is encountered that reaches that depth.

- **depth** The maximum depth can be set to \( n \) for a set of predicates \(<PredSpec>\) via including the specifier `subgoal_abstract(n)` as part of the tabling declaration

  \[
  :- \text{table } <PredSpec> \text{ as } \ldots, \text{subgoal_abstract}(n), \ldots
  \]

  Specifying `subgoal_abstract(0)` turns abstraction off for predicates in \(<PredSpec>\). The depth can also be set globally by setting the flag `max_table_subgoal_depth` to the desired maximal depth. If the subgoal depth has been set of a given predicate via a tabling declaration the declared depth will override the global depth.

- **action** When a subgoal is encountered of maximum depth, abstraction is enabled if the Prolog flag `max_table_subgoal_action` to `abstract`. Other possible values for the action are `error` and `fail` (cf. pg. 236).

Unless otherwise specified, XSB starts up with `max_table_subgoal_action` set to `error` and `max_table_subgoal_depth` set to the maximum integer possible under the compilation method for which XSB has been configured. Under this default behavior, XSB will throw an error if a subgoal has depth greater than `max_table_subgoal_depth`. As an alternative to setting flags, subgoal abstraction can be set by calling XSB with the command-line arguments `-max_subgoal_action a` and `-max_subgoal_depth n` where \( n \) is the size of the desired depth.

### 5.5.2 Bounded Rationality through Radial Restraint

In XSB, a programmer can use a form of bounded rationality called *radial restraint* to ensure that only a finite number of answers are generated by a query [32]. For instance if the query \( p(X) \) to the program

\[
p(f(X)) : - \ p(X).
p(0).
\]

were evaluated using this approach with a depth limit of 3, the answers, \( p(0), p(f(0)), p(f(f(0))) \) and \( p(f(f(f(X)))) \) would be generated; however, \( p(f(f(f(X)))) \) would have the truth value of `undefined`. Note that by doing things in this way, both of the goals \( p(f(f(f(0)))) \), and \( p(f(f(f(1)))) \) will unify with \( p(f(f(f(X)))) \) and so will succeed with a truth value of `undefined`. Similarly `\text{tnot}(p(f(f(f(0))))`, and
tnot(p(f(f(f(1))))) will both succeed with a value of undefined (perhaps better called unknown in this context). Since all predicates and function symbols have a maximum arity (256 in XSB) bounding the depth of an answer ensures that only a finite number of answers are returned 14.

Semantically when radial restraint is used, XSB is computing an approximation to the three-valued well-founded model of a program, called a restrained model. To see this, suppose the proof of a query \( Q \) does not depend on negation. If \( Q \) has a derivation that does not require any answers whose depth is greater than \( n \), it is proven as usual. Similarly, if \( Q \) is false in the well-founded model of a program, and none of the subgoals explored in the derivation of \( Q \) derive answers whose depth is greater than \( n \), XSB will derive that \( Q \) is false. Thus, the approximation that XSB computes is informationally sound in the sense that no incorrect answer will be derived, although the truth value of some atoms won’t be known that might have been if the depth bound had been set higher. Due to the halting problem, there is no way to know in general what depth to set for answer abstraction, or whether any bound needs to be set at all.

If a restrained model is derived, that are undefined through radial restraint can be distinguished from answers that are undefined in the well-founded model of a program, or for other reasons such as unsafe negation. If an answer \( A \) was abstracted due to a depth check, the query \( \text{get_residual}(A,\text{Delay}) \) would bind \( \text{Delay} \) to a list containing the atom \( \text{brat}_\text{undefined} \), where \( \text{brat}_\text{undefined}/0 \) is simply a predicate defined as

\[
\text{brat}_\text{undefined}:- \text{tnot}(\text{brat}_\text{undefined})
\]

which in a delay list indicates that an answer was made undefined through bounded-rationality based answer abstraction.

More generally, information about when a derivation has uses bounded rationality can be obtained either through the predicate \( \text{explain}_\text{u_val}/3 \), or \( \text{get_residual}_\text{sccs}/[3,5] \). Both of these predicates traverse the residual dependency graph to provide information about why a literal is undefined.

\[14\] If a program has a infinite number of true answers and a finite number of false answers, one possible approach might be to “dualize” the program so that only false answers are computed. However, since most programs with function symbols have an infinite number of both true and false answers, this approach won’t work in general.
Using Radial Restraint

Radial restraint is currently implemented only for tabling with call variance, and
lists are not currently included in the depth check. Similarly to the use of subgal
abstraction, answer abstraction is the implementational basis of radial restraint. It
can be declared by setting a value for the maximum depth of an answer and for the
action to take when an answer is encountered that reaches that depth.

- **depth** The maximum depth can be set to n for a set of predicates via including
  the specifier `answer_abstract(n)` as part of their tabling declaration

  ```prolog
  :- table <PredSpec> as ...,answer_abstract(n),...
  ```

  Specifying `answer_abstract(0)` turns answer abstraction off for predicates
  in `<PredSpec>`. The depth can also be set globally by setting the flag `max_table_answer_depth` to the desired maximal depth. If the answer depth
  of a given predicate has been set via a tabling declaration, the predicate-specific
  declared depth will override the global depth.

- **action** When an answer is encountered of maximum depth, abstraction is en-
  abled if the Prolog flag `max_table_answer_action` to `bounded_rationality`. Other possible values for the action are `error` and `fail` (cf. pg. 236).

Unless otherwise specified, XSB starts up with

- **max_table_answer_action** and

- **max_table_answer_list_action**

set to `error` and `max_table_answer_depth` and `max_table_answer_depth` set to
the maximum integer possible under the compilation method for which XSB has
been configured. Under this default behavior, XSB will throw an error if a subgoal
has depth greater than `max_table_subgoal_depth`.

5.6 Incremental Table Maintenance

XSB allows the user to declare that the system should maintain the correctness of
a given table with respect to dynamically changing facts and rules through so-called
incremental tables [68, 67, 76]. After a database update or series of updates Δ, an
incremental table T that depends on Δ is by default updated transparently: that is T
and all tables upon which $T$ depends are automatically updated (if needed) whenever a future subgoal calls $T$. In either case, incremental tabling brings XSB closer to the functionality of deductive databases. If tables are thought of as materialized database views (or snapshots), then the incremental table maintenance subsystem enables incremental view maintenance; also as discussed below, if choice points are thought of as database cursors then incremental tabling also provides view consistency.\footnote{In the current version of XSB, there are certain restrictions on how incremental tabling can be used; cf. Section 5.7.}

5.6.1 Transparent Incremental Tabling

To demonstrate incremental table maintenance (informally called \textit{incremental tabling}), consider first the following simple program that does not use incremental tabling:

\begin{verbatim}
:- table p/2.
p(X,Y) :- q(X,Y), Y =< 5.

:- dynamic q/2.
q(a,1). q(b,3). q(c,5). q(d,7).
\end{verbatim}

and the following queries and results:

\begin{verbatim}
| ?- p(X,Y), writeln([X,Y]), fail.  
  [c,5]  
  [b,3]  
  [a,1]  
  no  
  | ?- assert(q(d,4)).  
  yes  
  | ?- p(X,Y), writeln([X,Y]), fail.  
    [c,5]  
    [b,3]  
    [a,1]  
    no
\end{verbatim}

In this program, the table for \texttt{p/2} depends on the contents of the dynamic predicate \texttt{q/2}. We first evaluate a query, \texttt{p(X,Y)}, which creates a table. Then we use \texttt{assert/1
to add a fact to the q/2 predicate and re-evaluate the query. We see that the answers haven’t changed, because the table is already created and the second query just retrieves answers directly from that existing table. However the answers are inconsistent with the model of p/2 after the assert. I.e., if the table didn’t exist (e.g. if p/2 weren’t tabled), the answer [d,4] would also be derived. Without incremental table maintenance, the only solution to this problem is for the XSB programmer to explicitly abolish a table whenever changing (with assert or retract) a predicate on which the table depends. By declaring that the tables for p/2 should be incrementally maintained, XSB automatically keeps the tables for p/2 correct.

Consider a slight rewrite of the above program:

```prolog
:- table p/2 as incremental.
p(X,Y) :- q(X,Y),Y =< 5.
```

```prolog
:- dynamic q/2 as incremental.
q(a,1). q(b,3). q(c,5). q(d,7).
```

in which p/2 is declared to be incrementally tabled and q/2 is declared to be both dynamic and incremental, meaning that an incremental table depends on it. Consider the following goals and execution:

```
| ?- import incr_assert/1 from increval.
yes
| ?- p(X,Y),writeln([X,Y]),fail.
[c,5]
[b,3]
[a,1]
no
| ?- incr_assert(q(d,4)).
```

```
yes
| ?- p(X,Y),writeln([X,Y]),fail.
[d,4]
[c,5]
[b,3]
[a,1]
no
```
The transparent approach to incremental updating works as follows. When `incr_assert/1` is called, it sparks an invalidation phase in which tables that depend on `q(d,4)` are marked as invalid (i.e., possibly inconsistent with respect to underlying dynamic code). An Incremental Dependency Graph (IDG) is used to obtain the right tables to invalidate. However, if the invalidation phase finds an affected table that is incomplete, a permission error is thrown, since it is unclear whether sensible semantics can be given to updating a subgoal that is incomplete. After the invalidation phase is completed, when/if a subgoal calls an invalid table `T` the engine interrupts itself to recompute `T` and any tables upon which `T` depends. On the other hand, if no calls are ever made to an invalid incremental table `T'`, `T'` will never incur the cost of an update.

**View Consistency**

As described above, transparent incremental tablings's use of lazy updating ensures that a new query `Q` will always be consistent with the state of the dynamic code at the time `Q` is called. However, transparent incremental tabling enforces a stronger property of view consistency similar to those of database systems: that answers to a query `Q` should be those derivable at the time `Q` was called, and should not be affected by any updates. Because XSB’s incremental tabling does not allow updates that affect tables that are still being computed, supporting view consistency effectively means ensuring consistency for choice points into completed incremental tables. As such choice points correspond to database cursors, we term them Open Cursor Choice Points, (OCCPs).

XSB’s support for view consistency is designed so that no perceptable overhead in incurred if there are no OCCPs whose view needs to be maintained. Not surprisingly, numerous long-lived OCCPs whose views need to be maintained across updates causes an overhead for the engine, a situation that is in some sense similar to the cost of maintaining views for cursors in database system.

### 5.6.2 Updating in a Three-Valued Logic

As discussed earlier in this chapter, answers that are undefined in the well-founded semantics are represented as conditional answers. Beginning with version 3.3.7, incremental updates work correctly with conditional answers \(^{16}\). No special care needs

\(^{16}\)Before Version 3.3.7, incremental updates only worked correctly on stratified tables: those with only unconditional answers.
to be taken for updating in the well-founded semantics as the following example illustrates.

:- dynamic data/1 as incremental.

:- table opaque_undef/0 as opaque.
opaque_undef:- tnot(opaque_undef).

:- table p/1 as incremental.
p(_X):= opaque_undef.
p(X):= data(X).

Note that opaque_undef/1 upon which p/1 depends is explicitly declared as opaque\(^{17}\).
When the above program is loaded, XSB will behave as follows.

\[
| ?- p1(1).
\]

undefined

\[
| ?- incr_assert(data(1)).
\]

yes
\[
| ?- p1(1).
\]

yes
\[
| ?- incr_retract(data(1)).
\]

yes
\[
| ?- p1(1).
\]

undefined
\[
| ?- get_residual(p1(1),C).
\]

C = [opaque_undef]

\(^{17}\)An *opaque* predicate \(P\) is tabled and is used in the definition of some incrementally tabled predicate but should not be maintained incrementally. In this case the system assumes that the programmer will abolish tables for \(P\) in such a way so that re-calling it will always give semantically correct answers.
Declaring Predicates to be Incremental

In XSB, tables can have numerous properties: such as subsumptive, variant, incremental, opaque, dynamic, private, and shared, and can use answer subsumption, answer abstraction or call abstraction. XSB also has variations in forms of dynamic predicates: tabled, incremental, private, and shared. XSB extends the table and dynamic compiler and executable directives with modifiers that allow users to indicate the kind of tabled or dynamic predicate they want. For example,

:~ table p/3,s/1 as subsumptive,private.

:~ table q/3 as incremental,variant.

:~ dynamic r/2,t/1 as incremental.

In the current version of XSB, incremental tabling works with subgoal abstraction, answer abstraction, and well-founded negation. However several combinations involving incremental tabling are not supported and will throw an error (cf. page 280 and page 269, respectively). Incremental tabling has not yet been ported to the multi-threaded engine and it currently does not works for predicates that use call subsumption or answer subsumption.

5.6.3 Incremental Tabling using Interned Tries

Sometimes it is more convenient or efficient to maintain facts in interned tries rather than as dynamically asserted facts (cf. Chapter 8). Tables based on interned tries can be automatically updated when terms are interned or uninterned just as they can be automatically updated when a fact is asserted or retracted. Consider the example from Section 5.6.1 rewritten to use interned tries. As usual, an incrementally updated table is declared as such:

:~ table p/2 as incremental.

However, the declaration for dynamic data changes: rather than using the declaration

:~ dynamic q/2 as incremental

a trie is specified as incremental in its creation.

trie_create(Trie_handle,[incremental,alias(inctrie)])
As described in Chapter 8, the trie handle returned is an integer, but can be aliased just as with any other trie. The trie may then be initially loaded:

\[
\text{trie_intern}(q(a,1),\text{inctrie}),\text{trie_intern}(q(b,3),\text{inctrie}),\text{trie_intern}(q(c,5),\text{inctrie}),\text{trie_intern}(q(d,7),\text{inctrie}).
\]

At this stage a query to \( p/2 \) acts as before:

\[
p(X,Y) :- \text{trie_interned}(q(X,Y),\text{inctrie}),Y \leq 5.
\]

\[
| \text{?- } p(X,Y),\text{writeln([X,Y])},\text{fail}. \\
| \begin{array}{l} \\
| [c,5] \\
| [b,3] \\
| [a,1] \\
\end{array}
\]

The following sequence ensures that \( p/2 \) is incrementally updated as \( \text{inctrie} \) changes:

\[
| \text{?- import incr_trie_intern/2}. \\
| \text{yes} \\
| \text{?- incr_trie_intern(inctrie,q(d,4))}. \\
| \text{yes} \\
| \text{?- p(X,Y),writeln([X,Y]),fail}. \\
| \begin{array}{l} \\
| [d,4] \\
| [c,5] \\
| [b,3] \\
| [a,1] \\
\end{array}
\]

\[\text{no}\]

Given the proper directives to make a trie incremental, transparent incremental tabling works for changes made to interned tries just as it does for regular dynamic code and for trie-indexed dynamic code.

### 5.6.4 Abstracting the IDG for Better Performance

As mentioned above, incremental table maintenance makes use of an IDG. Specifically, the nodes of the IDG are the incrementally tabled subgoals; and each such table
contains information about its incident edges: those subgoals upon which a node directly depends or directly affects. While the IDG is a critical data structure to efficiently update incremental tables, in certain situations constructing the IDG can cause non-trivial overheads in query time and table space. These overheads can be addressed in many cases by abstracting the IDG. When a tabled subgoal \( S \) is called, rather than creating an edge between \( S \) and its nearest tabled ancestor \( S' \) (if any), one could abstract \( S, S' \) or both, potentially collapsing a large number of nodes and edges of the IDG. If \( S \) is an incremental table, then performing subgoal abstraction on \( S \) as introduced in Section 5.5, will abstract the IDG - rather than having \( n \) nodes \( S_1, \ldots, S_n \) and their associated links, the IDG will contain a single node \( abstract(S) \). However, subgoal abstraction will not work to abstract the leaf nodes of the IDG, which are subgoals to non-tabled dynamic incremental predicates.

In Version 3.6 of XSB, IDG nodes for dynamic incremental predicates may undergo depth abstraction: given a subgoal \( S \) and integer \( k \), subterms of \( S \) with depth \( k + 1 \) are replaced by unique new variables. For instance, abstracting \( q(f(1)) \) at level 1 gives \( q(f(X_1)) \); abstracting at level 0 gives \( q(X_1) \). Figure 5.5 illustrates an important case where abstracting dynamic incremental predicates can be critical to good performance for incremental tabling. In the case of left-linear recursion, if no abstraction is used a new node will be created for each call to \texttt{edge/2} as shown on the left side of this figure. If a large number of data elements are in fact reachable, the size of the IDG can be very large. If calls to the \texttt{edge/2} predicate make use of depth-0 abstraction, the graph may be much smaller as seen on the right side of Fig. 5.5. Whether abstracting a IDG in this manner is useful or not is application dependent; however, performance results indicate that for left-linear recursion, abstraction greatly reduces both query time and space.

\begin{verbatim}
:- table reach/2 as incremental.
:- dynamic edge/2 as incremental.
reach(X,Y):- edge(X,Y).
reach(X,Y):- reach(X,Z),edge(Z,Y).
\end{verbatim}

Figure 5.5: A left-linear program and schematic IDGs: Left without IDG abstraction; Right: with IDG abstraction

Abstracting the \texttt{edge/2} predicate has subtle differences from abstracting tabled
subgoals. As mentioned, the edge/2 predicate of Fig. 5.5 is not tabled. Furthermore, the actual edge/2 subgoal itself should not be abstracted to depth 0 since losing the first argument instantiation would prevent the use of indexing. Rather, only the IDG’s representation of the subgoal should be abstracted. Abstraction of dynamic code for the IDG can be specified via the declaration:

\[ \text{:-dynamic edge/2 as incremental, abstract(0).} \]

In Version 3.6 dynamic incremental code can be abstracted, but incremental interned tries (Section 5.6.3) cannot be. Also, currently only depth 0 abstraction is supported.

### 5.6.5 Summary and Implementation Status

The main design choices of incremental tabling are as usual what to table, and also what dynamic predicates or tries should be made incremental. In addition, performance optimizations may be made through a mixture of subgoal abstraction and dynamic predicate abstraction. This optimization can be informed by use of statistics/0 which includes summary information about the IDG, or using the IDG inspection predicates of Section 5.6.6 if more details are needed.

In the current version of XSB, incremental tabling has not yet been ported to the multi-threaded engine. In addition, incremental tabling only works for predicates that use both call and answer variance. However, incremental tabling does work with for the full well-founded semantics, for trie indexed dynamic code (in addition to regular dynamic code) and with interned tries as described in Section 5.6.3. The space reclamation predicates abolish_all_tables/0, abolish_table_call/[1,2] and abolish_table_pred/[1,2] can be safely used with incremental tables.

### 5.6.6 Predicates for Incremental Table Maintenance

#### A Note on Terminology

Suppose p/1 and q/1 are incrementally tabled, and that there is a clause

\[ p(X) :- q(X). \]

In this case we say that p(X) *depends_on* q(X) and that q(X) *affects* p(X). A recursive predicate both depends on and affects itself.
Declarations  The following directives support incremental tabling based on changes in dynamic code:

\[\text{table} + \text{PredSpecs as incremental}\]

is a executable predicate that indicates that each tabled predicate specified in \text{PredSpec} is to have its tables maintained incrementally. \text{PredSpec} is a list of skeletons, i.e. open terms, or \text{Pred/Arity} specifications. The tables must use call variance and answer variance and must be compiled and loaded into the single-threaded engine. If a predicate is declared with tabling attributes that are not supported with incremental tabling a permission error is thrown. This predicate implies that its arguments are tabled predicates. See page 280 for further discussion of tabling options.

We also note that any tabled predicate that is called by a predicate tabled as incremental must also be tabled as incremental or as opaque. On the other hand, a dynamic predicate \text{d/n} that is called by a predicate tabled as incremental may or may not need to be declared as incremental. However if \text{d/n} is not declared incremental, then changes to it will not be propagated to incrementally maintained tables.

\[\text{dynamic} + \text{PredSpecs as incremental}\]

is an executable predicate that indicates that each predicate in \text{PredSpecs} is dynamic and used to define an incrementally tabled predicate and will be updated using \text{incr_assert/1} and/or \text{incr_retractall/1} (or relatives.) Note that dynamic incremental predicates cannot themselves be tabled. This predicate implies that its arguments are dynamic predicates. See page 269 for further discussion of dynamic options.

\[\text{table} + \text{PredSpecs as opaque}\]

is an executable predicate that indicates that each predicate \text{P} in \text{PredSpecs} is tabled and is used in the definition of some incrementally tabled predicate but should not be maintained incrementally. In this case the system assumes that the programmer will abolish tables for \text{P} in such a way so that re-calling it will always give semantically correct answers. In other words, instead of maintaining information to support incremental table maintenance, the system re-calls the opaque predicate whenever its results are required to recompute an answer. One example of an appropriate use of opaque is for tabled predicates in a DCG used to parse some string. Rather than incrementally maintain all dependencies on all input strings, the user can declare these intermediate tables

\[\text{18No explicit module references are allowed.}\]
as opaque and abolish them before any call to the DCG. This predicate implies that its arguments are tabled predicates.

**Basic Incremental Maintenance Predicates** The following predicates are used to manipulate incrementally maintained tables:

- `incr_assert(+Clause)` module: increval
- `incr_assertz(+Clause)` module: increval
- `incr_asserta(+Clause)` module: increval
- `incr_retract(+Clause)` module: increval
- `incr_retractall(+Term)` module: increval

are versions of `assert/1` and other standard Prolog predicates. They modify dynamic code just as their Prolog counterparts, but they first invalidate all incrementally maintained tables that depend on `Clause`.

**Error Cases** are the same as `assert<\alpha/z>/1`, `retract/1` and `retractall/1` with the additional error conditions that relate to the semantics of incremental tabling. Note that if these error conditions arise, the update will not occur.

- The head of the clause `Clause` or the `Term` refers to a predicate that is not incremental and dynamic.
  - `type error(dynamic_incremental, Term)`
- `Clause` affects an incremental table that is incomplete (and so is in the course of being computed).
  - `permission_error`

`incr_invalidate_calls(+Goal)` module: increval

Let $\mathcal{T}$ be the least set of all incrementally maintained tables whose goals that unify with `Goal`, or whose tables are (transitively) affected by a goal in $\mathcal{T}$. This predicate invalidates all tables in $\mathcal{T}$. Any subsequent call to a goal $G$ associated with $\mathcal{T}$ will be automatically be incrementally updated if necessary. (As will any goals that $G$ depends on that are in need of updating.) In a similar manner, an invocation of `incr_table_update/[0,1,2]` will cause tables in $\mathcal{T}$ to be updated.

Note that this predicate is needed for exceptional cases only. Calls to `incr_assert/1` and similar predicates mentioned above perform invalidation automatically, as does `abolish_table_call/[1,2]`. However, `incr_invalidate_calls/1` is useful if a tabled predicate depends on some external data and not (only) on dynamic incremental predicates. For example, such a predicate might depend on a
relation stored in an external relational database (perhaps accessed through the ODBC interface). Of course, in such a case, the application programmer must know when the external relation changes and invoke incr_invalidate_calls/1 as necessary.

**Error Cases**

- **Goal** is tabled, but not incrementally tabled
  - permission_error(invalidate,non-incremental predicate,Goal)

**Incremental Maintenance using Interned Tries** The following predicates are used to modify incremental tries, and can be freely intermixed with predicates for modifying incremental dynamic code, as well as with predicates for invalidating or updating tables (Section 5.6.6).

```prolog
incr_trie_intern(+TrieIdOrAlias,+Term) module: intern
is a version of trie_intern/2 for tries declared as incremental. A call to this predicate interns Term in TrieIdOrAlias and then invalidates all incrementally maintained tables that depend on this trie.
```

```prolog
incr_trie_uninternall(+TrieIdOrAlias,+Term) module: intern
is a version of trie_unintern/2 for tries declared as incremental. A call to this predicate removes all terms unifying with Term in TrieIdOrAlias and then invalidates all incrementally maintained tables that depend on this trie.
```

**Inspecting the State of the Incremental Dependency Graph** The predicates in this section allow a user to inspect properties of IDG that can be useful in debugging, profiling or optimizing a computation. In addition they provide information about which subgoals in the IDG are invalid – i.e., which subgoals depend on a dynamic code that has changed, but have not been updated.

As explained below, IDG nodes can be accessed via the predicate is_incremental_subgoal/1, while IDG edges can be accessed via incr_directly_depends/2. The predicates get_incr_scc/[1,2] and get_incr_scc_with_deps/[3,4] can be used to efficiently materialize the dependency graph in Prolog, including SCC information. Similarly, the predicates incr_invalid_subgoals/1 and incr_is_invalid/1 can be used to determine which subgoals are invalid.

---

19The predicates for traversing the incremental dependency graph are somewhat analogous to those for traversing the residual dependency graph (Section 6.15.2).
is_incremental_subgoal(?Subgoal) module: increval
This predicate non-deterministically unifies \texttt{Subgoal} with incrementally tabled
subgoals that are currently table entries.

incr_directly_depends(?Goal\textsubscript{1}, ?Goal\textsubscript{2}) module: increval
accesses the edges of the IDG: the incremental goals (Tables) that directly
depend on or directly affect one another. At least one of \texttt{Goal\textsubscript{1}} or \texttt{Goal\textsubscript{2}} must
be bound.

- If \texttt{Goal\textsubscript{1}} is bound, then this predicate will return in \texttt{Goal\textsubscript{2}} through back-
tracking the goals for all incrementally maintained tables on which \texttt{Goal\textsubscript{1}}
directly depends.

- If \texttt{Goal\textsubscript{2}} is bound, then it returns in \texttt{Goal\textsubscript{1}} through backtracking the goals
for all incrementally maintained tables that \texttt{Goal\textsubscript{2}} directly affects – in other
words all goals that directly depend on \texttt{Goal\textsubscript{2}}.

**Error Cases**

- Neither \texttt{Goal\textsubscript{1}} nor \texttt{Goal\textsubscript{2}} is bound
  - instantiation_error

- \texttt{Goal\textsubscript{1}} and/or \texttt{Goal\textsubscript{2}} is bound, but is not incrementally tabled
  - table_error

incr_trans_depends(?Goal\textsubscript{1}, ?Goal\textsubscript{2}) module: increval
is similar to incr\_directly\_depends/2 except that it returns goals according
to the transitive closure of the “directly depends” relation. Error conditions are
the same as incr\_directly\_depends/2.

get_incr_sccs(?SCCList) module: increval
get_incr_sccs_with_deps(?SCCList, ?DepList) module: increval
get_incr_sccs(+SubgoalList, ?SCCList) module: increval
get_incr_sccs_with_deps(+SubgoalList, ?SCCList, ?DepList) module: increval

Most linear algorithms for SCC detection over a graph use destructive assign-
ment on a stack to maintain information about the connecteness of a component;
as a result such algorithms are difficult to write efficiently in Prolog.

get\_incr\_sccs/1 unifies \texttt{SCCList} with SCC information for the incremental
dependency graph that is represented as a list whose elements are of the form

\texttt{ret(Subgoal, SCC)}.
SCC is a numerical index for the SCCs of Subgoal. Two subgoals are in the same SCC iff they have the same index, however no other dependency information can be otherwise directly inferred from the index \(^{20}\).

If dependency information is also desired, \texttt{get_incr_scc_with_dependencies/2} should be called. In addition to the SCC information as above, \texttt{DepList} is unified with a list of dependency terms of the form

\[
\text{depends(SCC1,SCC2)}
\]

for each pair SCC1 and SCC1 such that some subgoal with index SCC1 directly depends on some subgoal with index SCC1. If it is necessary to know which subgoal(s) in SCC1 directly depends on which subgoal(s) in SCC2, the information can be easily reconstructed using \texttt{incr_directly_depends/2} above. Similarly, \texttt{incr_directly_depends/2} can be used to determine the actual edges within a given SCC.

Ordinarily a user will want to see the entire dependency graph and in such a case the predicates described above should be used. However, note that if the dependency graph is the result of several independent queries it may not be connected. \texttt{get_incr_scc/2} takes as input a list of incremental subgoals, \texttt{SubgoalList}. For each \texttt{Subgoal} in \texttt{SubgoalList}, this predicate finds the set of subgoals connected to \texttt{Subgoal} by any mixture of depends and affects relations, unions these sets together, and finds the SCCs of all subgoals in the unioned set.

SCC detection is implemented using Tarjan’s algorithm \cite{tarjan84} in C working directly on XSB’s data structures. The algorithm is \(O(|V| + |E|)\) where \(|V|\) is the number of vertices and \(|E|\) the number of edges in the dependency graph. As a result, \texttt{get_incr_sccs/[1,2]} provides an efficient means to materialize the high-level topography of the dependency graph \(^{21}\).

**Error Cases**

- \texttt{SCCList} contains a predicate that is not tabled
  
  - permission_error

\(^{20}\)The actual number for each SCC index depends on how the incremental dependency graph happens to be traversed; as a result it is best to rely on the index only as a “generated” name for each SCC.

\(^{21}\)Currently, the materialization of dependency information between SCCs is implemented in a naive manner, so that \texttt{get_incr_sccs_with_deps/[2,3]} is \(O(|V|^2)\).
CHAPTER 5. USING TABLING IN XSB: A TUTORIAL INTRODUCTION

incr_invalid_subgoals(-List) module: increval
   This predicate unifies List with a sorted list of the incremental subgoals that are currently invalid.

incr_is_invalid(+Subgoal) module: increval
   Succeeds if Subgoal is an incrementally tabled subgoal that is invalid, and fails otherwise.

5.7 Compatibility of Tabling Modes and Predicate Attributes

As discussed in this chapter, there are several choices for how to table a predicate. Either call subsumption or call variance may be used, incremental tabling might or might not be used, and answer subsumption might or might not be used. Furthermore, a tabled predicate, like any other predicate, may be static or dynamic and thread shared or thread private. Together, there are 48 different combinations, not all of which are supported in Version 3.6 of XSB. To analyze further, all combinations are supported for call-variance and for thread private predicates. However, call subsumption has not been fully integrated with dynamic code or thread shared predicates, and cannot currently be combined with incremental tabling or with answer subsumption. Similarly incremental tabling is not yet supported in the multi-threaded engine (it is supported for “thread private” computations only in the sequential engine). The compatibilities are listed in Table 5.1. Further combinations will be supported in future versions of XSB as resources allow.

The combinations in Table 5.1 allow full well-founded computation, constrained variables in calls and answers (including the residual program), and safe space reclamation, with the following exceptions. Answer subsumption does support non lrd-stratified programs; and call subsumption does not yet support attributed variables in calls.
<table>
<thead>
<tr>
<th>variant</th>
<th>static</th>
<th>private</th>
<th>nonincremental</th>
<th>no answer subsumption</th>
<th>subsumptive</th>
</tr>
</thead>
<tbody>
<tr>
<td>variant</td>
<td>static</td>
<td>private</td>
<td>nonincremental</td>
<td>answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>static</td>
<td>private</td>
<td>opaque</td>
<td>no answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>static</td>
<td>private</td>
<td>opaque</td>
<td>answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>static</td>
<td>private</td>
<td>incremental</td>
<td>no answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>static</td>
<td>private</td>
<td>incremental</td>
<td>answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>static</td>
<td>shared</td>
<td>nonincremental</td>
<td>no answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>static</td>
<td>shared</td>
<td>nonincremental</td>
<td>answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>static</td>
<td>shared</td>
<td>opaque</td>
<td>no answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>static</td>
<td>shared</td>
<td>opaque</td>
<td>answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>static</td>
<td>shared</td>
<td>incremental</td>
<td>no answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>static</td>
<td>shared</td>
<td>incremental</td>
<td>answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>nonincremental</td>
<td>no answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>nonincremental</td>
<td>answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>opaque</td>
<td>no answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>opaque</td>
<td>answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>incremental</td>
<td>no answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>incremental</td>
<td>answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>nonincremental</td>
<td>no answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>nonincremental</td>
<td>answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>opaque</td>
<td>no answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>opaque</td>
<td>answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>incremental</td>
<td>no answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>incremental</td>
<td>answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>nonincremental</td>
<td>answer subsumption</td>
<td>yes</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>opaque</td>
<td>no answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>opaque</td>
<td>answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>incremental</td>
<td>no answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>private</td>
<td>incremental</td>
<td>answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>nonincremental</td>
<td>no answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>nonincremental</td>
<td>answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>opaque</td>
<td>no answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>opaque</td>
<td>answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>incremental</td>
<td>no answer subsumption</td>
<td>no</td>
</tr>
<tr>
<td>variant</td>
<td>dynamic</td>
<td>shared</td>
<td>incremental</td>
<td>answer subsumption</td>
<td>no</td>
</tr>
</tbody>
</table>

Table 5.1: Support for different tabling modes in XSB Version 3.6
Chapter 6

Standard Predicates and Predicates of General Use

This chapter mainly describes standard predicates, which are always available to the Prolog interpreter, and do not need to be imported or loaded explicitly as do other Prolog predicates. By default, it is a compiler error to redefine standard predicates.

In the description below, certain standard predicates depend on HiLog semantics; the description of such predicates have the token HiLog at the right of the page. Similarly predicates that depend on SLG evaluation are marked as Tabling, and predicates whose semantics is defined by the ISO standard (or whose implementation is reasonably close to that definition) are marked as ISO. Occasionally, however, we include in this section predicates that are not standard. In such cases we denote their module in text font towards the middle of the page.

6.1 Input and Output

6.1.1 I/O Streams in XSB

XSB’s I/O is based on ISO-style streams, although it also supports older DEC-10 style file handling. The use of streams provides a unified interface to a number of different classes of sources and sinks. Currently these classes include textual and binary files, console input and output, pipes, and atoms; in the future sockets and urls may be handled under the stream interface. When streams are opened, certain actions may occur depending on the class of the source or sink and on the wishes of the user. For instance when a file \( F \) is opened for output mode, an existing file
F may be truncated (in write mode) or not (in append mode). In addition, various operations may or may not be valid depending on the class of stream. For instance, repositioning is valid for an atom or file but not a pipe or console.

XSB provides several default I/O streams, which make it easier for a user to embed XSB in other applications. These streams include the default input and output streams. They also include the standard error stream, to which XSB writes all error messages. By default the standard error stream is the same as the standard output stream, but it can be redirected either by UNIX shell-style I/O redirection or by the predicates file_reopen/4 and file_clone/3. Similarly there is the standard warning stream (to which all system warnings are written), the standard message stream, the standard debugging stream (to which debugging information is written), and the standard feedback stream (for interpreter prompts, yes/no answers, etc). All of these streams are aliased by default to standard output, and can be redirected by the predicates file_reopen/4 and file_clone/3. Such redirection can be useful for logging, or other purposes.

Streams may also be aliased: the default input and output streams are denoted by user_input and user_output and they refer to the standard input and standard output streams of the process \(^1\). Similarly, XSB’s error, warning and message streams uses the aliases user_error, user_warning and user_message respectively.

Streams are distinguished by their class – whether they are file or atom, etc.; as well as by various properties. These properties include whether a stream is positionable or not and whether a (file) stream is textual or binary.

- **Console**: The default streams mentioned above are console streams, which are textual and not repositionable.

- **File**: A file stream corresponds to an operating system file and is repositionable. On Windows, binary files and textual files differ, while on UNIX they are the same.

- **Atom**: XSB can read from an atom, just as it can from a file. Atoms are considered to be textual and repositionable. Writing to atoms via streams is not currently available in XSB, although the predicate term_to_atom/[2,3] contains much of the functionality that such streams would provide.

- **Pipe**: XSB can also open pipes either directly, or as part of its ability to spawn processes. When made into streams, pipes are textual and not repositionable.

\(^1\)For backwards compatibility, the default input stream can also be aliased by user or userin, and the default output stream by user or userout.
I/O Stream Implementation

A user may notice that XSB’s I/O streams are small integers, but they should not be confused with the file descriptors used by the OS. The OS file descriptors are objects returned by the C `open` function; XSB I/O streams indices into the internal XSB table of open files and associated information. The OS does not know about XSB I/O streams, while XSB (obviously) does know about the OS file descriptors. An OS file descriptor may be returned by certain predicates (e.g. `pipe_open/2` or user-defined I/O). In the former case, a file descriptor can be promoted to XSB stream by `open/{3,4}` and in the latter by using the predicate `fd2iostream/2`.

When it starts, XSB opens a number of standard I/O streams that it uses to print results, errors, debugging info, etc. The descriptors are described in the file `prolog_includes/standard.h`. This file provides the following symbolic definitions:

```c
#define STDIN 0
#define STDOUT 1
#define STDERR 2
#define STDWARN 3 /* output stream for xsb warnings */
#define STDMSG 4 /* output for regular xsb messages */
#define STDDBG 5 /* output for debugging info */
#define STDFDBK 6 /* output for XSB feedback
  (prompt/yes/no/Aborting/answers) */
#define AF_INET 0 /* XSB-side socket request for Internet domain */
#define AF_UNIX 1 /* XSB-side socket request for UNIX domain */
```

These definitions can be used in user programs, if the following is provided at the top of the source file:

```prolog
compiler_options([xpp_on]).
#include "standard.h"
```

If this header is used, the various streams can be used as any other output stream – e.g. `?- write(STDWARN,'watch it!')`. (Note: the XSB preprocessor is not invoked on clauses typed into an interactive XSB session, so the above applies only to programs loaded from a file using `consult` and such.)
6.1.2 Character Sets in XSB

Beginning in Version 3.5 of XSB, alternate character sets are supported.

- **UTF-8** which on input atomatically interprets the sequence of bytes as UTF-8 byte sequences and decodes them to obtain the unicode code points; and on output converts from the unicode code points to UTF-8 byte sequences.

- **LATIN-1** which performs no transformation on byte sequences (i.e. treats each byte directly as a unicode code point.)

- **CP1252** which implements Windows code page 1252 encoding, the default for most Windows systems.

Other character sets, in particular, UTF-16, may be supported in the future.

In the current version of XSB, UTF-8 is the default character set when XSB is configured on UNIX-style systems such as Linux and Mac OSX. CP1252 is the the default character set on Windows-style systems. The character set may be changed at any time via the Prolog flag `character_set`, whose value must be one of `utf_8`, `cp1252`, or `latin_1`. The character set in effect at the time of opening a stream is the character set that will be used to read (or write) the stream.

6.1.3 Predicates for ISO Streams

open(+SourceSink,+Mode,-Stream)

ISO open/1 creates a stream for the source or sink designated in SourceSink, and binds Stream to a structure representing that stream.

- If SourceSink is an atom, or the term `file(File)` where File is an atom, the stream is a file stream. In this case Mode can be
  - **read** to create an input stream. In Windows, whether the file is textual or binary is determined by the file’s properties.
  - **write** to create an output stream. Any previous file with a similar path is removed and a (textual) file is created which becomes a record of the output stream.
  - **write_binary** to create an output stream. Any previous file with a similar path is removed and a file is created which becomes a record of the output stream. The file created is binary in Windows, while in UNIX `write_binary` has the same effect as `write`. 
– **append** to create an output stream. In this case the output stream is appended to the contents of the file, if it exists, and otherwise a new file is created for (textual) output

– **append_binary** to create an output stream. In this case the output stream is appended to the contents of the file, if it exists, and otherwise a new file is created for (binary) output

- If `SourceSink` is the term `atom(Atom)` where `Atom` is an atom, the stream is an atom stream. In this case `Mode` currently can only be **read**. This stream class, which reads from interned atoms, is analogous to C’s `sscanf()` function.

- If `SourceSink` is the term `pipe(FileDescriptor)` where `FileDescriptor` is an integer, then a pipe stream is opened in the mode for `FileDescriptor`.

**ISO Compatibility Note:** This predicate extends the ISO definition of `open/3` to include strings and pipes as well as the file modes `write_binary` and `append_binary`.

**Error Cases**

- `SourceSink` or `Mode` is not instantiated
  - `instantiation_error`

- `Mode` is not a valid I/O mode
  - `domain_error(io_mode,Mode)`

- `SourceSink` is a file and cannot be opened, or opened in the desired mode
  - `permission_error(open,file,SourceSink)`

`open(+File,+Mode,-Stream,+Options)`

ISO

`open/4` behaves as does `open/3`, but allows a list of options to be given. The current options are a subset of ISO options and are:

- **alias(A)** allows the stream to be aliased to an atom A.

- **type(T)** has no effect on file streams in UNIX, which are always textual, but in Windows if T is binary a binary file is opened.

**Error Cases** Error cases are the same as `open/3` but with the addition:

- **Option_list** contains an option O that is not a (currently implemented) stream option.
  - `domain_error(stream_option,O)`
• An element of OptionsList is alias(A) and A is already associated with an existing thread, queue, mutex or stream
  – permission_error(create, alias, A)
• An element of OptionsList is alias(A) and A is not an atom
  – type_error(atom, A)

ISO Compatibility Note: The ISO option reposition(Boolean) currently has no effect on streams, because whether or not the stream is repositionable or not depends on the stream class. The ISO option eof_action(Action) currently has no effect on file streams. If these options are encountered in Options, a warning is issued to STDWARN.

close(+Stream_or_alias,+OptionsList)

ISO close/2 closes the stream or alias Stream_or_alias. OptionsList allows the user to declare whether a permission error will be raised in XSB upon a resource or system error from the closing function (e.g. fclose() or other system function). If OptionsList is non-empty and contains only terms unifying with force(true) then such an error will be ignored (possibly leading to unacknowledged loss of data). Otherwise, a permission error is thrown if fclose() or other system function returns an error condition. If the stream class of Stream_or_alias is an atom, then the only action taken is to close the stream itself – the interned atom itself is not affected.

Error Cases

• Stream_or_alias is a variable
  – instantiation_error
• Stream_or_alias is neither a variable, nor a stream term nor an alias.
  – domain_error(stream_or_alias, Stream_or_alias)
• Stream_or_alias is not associated with an open stream
  – existence_error(stream, Stream_or_alias)
• OptionList contains an option O that is not a closing option.
  – domain_error(close_option, O)
• OptionList contains conflicting options
  – domain_error(close_option, OptionList)
• Closing the stream produces an error (and OptionList is a non-empty list containing terms of the form force(true)).
permission_error(close,file,Stream_or_alias)

close(+Stream_or_alias)

Behaves as close(Stream_or_alias,[force(false)]).

set_input(+Stream_or_alias)

Makes file Stream_or_alias the current input stream.

Error Cases

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable, nor a a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not an open input stream
  - existence_error(stream,Stream_or_alias)

set_output(+Stream_or_alias)

Makes file Stream_or_alias the current output stream.

Error Cases

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable, nor a a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open output stream
  - existence_error(stream,Stream_or_alias)

stream_property(?Stream,?Property)

This predicate backtracks through the various stream properties that unify with Property for the stream Stream. Currently, the following properties are defined.

- stream_class(C) gives the stream class for a file: i.e. file, atom, console or pipe.
- file_name(F) is a property of Stream, if Stream is a file stream and F is the file name associate with Stream. The full operating system path is used.
• \texttt{type(T)} is a property of \texttt{Stream}, if \texttt{Stream} is a file stream and \texttt{T} is the file type of \texttt{Stream}: \texttt{text} or \texttt{binary}.

• \texttt{mode(M)} is a property of \texttt{Stream}, if \texttt{M} represents the I/O mode with which \texttt{Stream} was opened: i.e. \texttt{read}, \texttt{write}, \texttt{append}, \texttt{write_binary}, etc., as appropriate for the class of \texttt{Stream}.

• \texttt{alias(A)} is a property of \texttt{Stream}, if \texttt{Stream} was opened with alias \texttt{A}.

• \texttt{input} is a property of \texttt{Stream}, if \texttt{Stream} was opened in the I/O mode: \texttt{read}.

• \texttt{output} is a property of \texttt{Stream}, if \texttt{Stream} was opened in the I/O mode: \texttt{write}, \texttt{append}, \texttt{write_binary}, or \texttt{append_binary}.

• \texttt{reposition(Bool)} is true, if \texttt{Stream} is repositionable, and false otherwise.

• \texttt{end_of_stream(E)} returns \texttt{at} if the end of stream condition for \texttt{Stream} is true, and \texttt{not} otherwise.

• \texttt{position(Pos)} returns the current position of the stream as determined by \texttt{fseek} or the byte-offset of the current stream within an atom. In either case, if an end-of-stream condition occurs, the token \texttt{end_of_file} is returned.

• \texttt{eof_action(Action)} is \texttt{reposition} if the stream class is \texttt{console}, \texttt{eof_code} if the stream class is \texttt{file}, and \texttt{error} is the stream class is \texttt{pipe} or \texttt{atom}.

\texttt{flush_output(+Stream_or_alias)} \texttt{ISO}

Any buffered data in \texttt{Stream_or_alias} gets flushed. If \texttt{Stream} is not buffered (i.e. if it is of class \texttt{atom}), no action is taken.

\textbf{Error Cases}

• \texttt{Stream_or_alias} is a variable
  
  – \texttt{instantiation_error}

• \texttt{Stream_or_alias} is neither a variable, nor a a stream term nor an alias.
  
  – \texttt{domain_error(Stream_or_alias,Stream)}

• \texttt{Stream} is not associated with an open output stream
  
  – \texttt{existence_error(Stream_or_alias,Stream)}

• Flushing (i.e. \texttt{fflush()}) returns an error.
  
  – \texttt{permission_error(flush,stream,Stream)}

\texttt{flush_output} \texttt{ISO}

Any buffered data in the current output stream gets flushed.
set_stream_position(+Stream_or_alias,+Position) ISO

If the stream associated with Stream_or_alias is repositionable (i.e. is a file or atom), sets the stream position indicator for the next input or output operation. Position is a positive integer, taken to be the number of bytes the stream is to be placed from the origin.

Error Cases

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable, nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Position is not instantiated to a positive integer.
  - domain_error(stream_position,Position)
- Stream_or_alias is not associated with an open stream
  - existence_error(stream,Stream_or_alias)
- Stream_or_alias is not repositionable, or repositioning returns an error.
  - permission_error(resposition,stream,Stream_or_alias)

at_end_of_stream(+Stream_or_alias) ISO

Succeeds if Stream_or_alias has position at or past the end of stream.

Error Cases

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable, nor a stream term nor an alias.
  - domain_error(stream,Stream_or_alias)
- Stream_or_alias is not an open stream
  - existence_error(stream,Stream_or_alias)

at_end_of_stream ISO

Acts as at_end_of_stream/1 but using the current input stream.
Other Predicates using ISO Streams

\textbf{file\_reopen(+FileName,+Mode,+Stream,-RetCode)}
Takes an existing I/O stream, closes it, then opens it and attaches it to a file. This can be used to redirect I/O from any of the standard streams to a file. For instance,

\begin{verbatim}
| ?- file_reopen('/dev/null', w, 3, Error).
\end{verbatim}

redirects all warnings to the Unix black hole.
On success, RetCode is 0; on error, the return code is negative.

\textbf{file\_clone(+SrcStream,?DestStream,-RetCode)}
This is yet another way to redirect I/O. It is a Prolog interface to the C \texttt{dup} and \texttt{dup2} system calls. If DestStream is a variable, then this call creates a new XSB I/O stream that is a clone of SrcStream. This means that I/O sent to either stream goes to the same place. If DestStream is not a variable, then it must be a number corresponding to a valid I/O stream. In this case, XSB closes DestStream and makes it into a clone of SrcStream.
For instance, suppose that 10 is a I/O Stream that is currently open for writing to file \texttt{foo.bar}. Then

\begin{verbatim}
| ?- file_clone(10,3,_).
\end{verbatim}
causes all messages sent to XSB standard warnings stream to go to file \texttt{foo.bar}. While this could be also done with \texttt{file\_reopen}, there are things that only \texttt{file\_clone} can do:

\begin{verbatim}
| ?- file_clone(1,10,_).
\end{verbatim}
This means that I/O stream 10 now becomes clone of standard output. So, all subsequent I/O will now go to standard output instead of \texttt{foo.bar}.
On success, RetCode is 0; on error, the return code is negative.

\textbf{file\_truncate(+Stream, +Length, -Return)} \hspace{1cm} \texttt{module: file\_io}
The regular file referenced by the StreamStream is chopped to have the size of Length bytes. Upon successful completion Return is set to zero.

\textbf{Portability Note:} Under Windows (including Cygwin) \texttt{file\_truncate/2} is implemented using \texttt{_chsize()}, while on Unix \texttt{ftruncate()} is used. There are
minor semantic differences between these two system calls, which are reflected by the behavior of `file_truncate/2` on different platforms.

**Error Cases**

- **Stream_or_alias** is a variable
  - `instantiation_error`
- **Stream_or_alias** is neither a variable, nor a stream term nor an alias.
  - `domain_error(stream_or_alias,Stream_or_alias)`
- **Stream_or_alias** is not associated with an open stream
  - `existence_error(stream,Stream_or_alias)`
- **Length** is a variable
  - `instantiation_error`
- **Length** is neither a variable nor an integer
  - `type_error(integer,Length)`

```prolog
tmpfile_open(-Stream)
```

Opens a temporary file with a unique filename. The file is deleted when it is closed or when the program terminates.

```prolog
flush_all_output_streams
```

Flushes output streams, both user and system `STDOUT`, `STDERR`, etc. This convenience predicate is written as

```prolog
flush_all_open_streams:-
    stream_property(S,mode(X)),(X = append ; X = write),flush_output(S),fail.
flush_all_open_streams.
```

### 6.1.4 DEC-IO Style File Handling

```prolog
see(+File_or_stream)
```

Makes `File_or_stream` the current input stream.

- If there is an open input stream associated with the file that has `File_or_stream` as its file name, and that stream was opened previously, then it is made the current input stream.
- Otherwise, the specified file is opened for input and made the current input stream. If the file does not exist, `see/1` throws a permission error.
Note that see/1 is incompatible with ISO aliases – calling see(Alias) with an ISO alias will try to open a file named Alias rather than using the alias. Also note that different file names (that is, names which do not unify) represent different input streams (even if these different file names correspond to the same file).

**Error Cases**

- File_or_stream is a variable
  - instantiation_error
- File_or_stream is neither a variable nor an atomic file identifier nor a stream identifier.
  - domain_error(stream_or_path,F)
- File File_or_stream is directory or file is not readable.
  - permission_error(open,file,F)
- File File_or_stream does not exist.
  - existence_error(stream_or_path,F)

seeing(?F)
F is unified with the name of the current input stream. This is exactly the same with predicate current_input/1 described in Section 6.12, and it is only provided for upwards compatibility reasons.

seen
Closes the current input stream. Current input reverts to "userin" (the standard input stream).

tell(+F)
Makes file F the current output stream.

- If there is an open output stream associated with F and that was opened previously by tell/1, then that stream is made the current output stream.
- Otherwise, the specified file is opened for output and made the current output stream. If the file does not exist, it is created.

Also note that different file names (that is, names which do not unify) represent different output streams (even if these different file names correspond to the same file).

The implementation of the ISO predicate set_output/1, is essentially that of tell/1.
CHAPTER 6. STANDARD AND GENERAL PREDICATES

Error Cases

- File_or_stream is a variable
  - instantiation_error
- File_or_stream is neither a variable nor an atomic file identifier nor a stream identifier.
  - domain_error(stream_or_path,F)
- File File_or_stream is directory or file is not readable.
  - permission_error(open,file,F)
- File File_or_stream does not exist.
  - existence_error(stream_or_path,F)

telling(?F)
F is unified with the name of the current output stream. This predicate is exactly the same with predicate current_output/1 described in Section 6.12, and it is only provided for upwards compatibility reasons.

told
Closes the current output stream. Current output stream reverts to “userout” (the standard output stream).

file_exists(+F)
Succeeds if file F exists. F must be instantiated to an atom at the time of the call, or an error message is displayed on the standard error stream and the predicate aborts.

Error Cases

instantiation_error F is uninstantiated.

url_encode(+Filename,-EncodedFilename)
This predicate is useful when one needs to create a file whose name contains forbidden characters, such as >, <, and the like. It takes a string and encodes any forbidden character using an appropriate %-sequence of characters that is acceptable as a file name in any OS: Unix, Windows, or Mac. For instance,

X = http%3a%2f%2ffoo%27%3e%24
url_decode(+Filename,-EncodedFilename)
This predicate performs the inverse operation with respect to url_encode/2. For instance,

\[ \text{?- url_decode('http\%3a\%2f\%2ffoo\%27\%3e\%24',X).} \]
\[ X = \text{http://foo’}$ \]

### 6.1.5 Character I/O

Beginning with Version 3.6, XSB supports Unicode in the form of UTF-8 characters. Due to this change, we recommend using ISO-compliant character I/O predicates, rather than older predicates such as get/1, get0/1, put/1 and so on. As the use of these older predicates may sometimes give unexpected answers when used with non-ASCII characters, they are deprecated, although they are still available for backward compatibility.

get_char(+Stream_or_alias,?Char)
ISO
Unifies Char with the next UTF-8 character from Stream_or_alias, advancing the position of the stream. Char is unified with the atom \text{end_of_file} if an end of file condition is detected.

**Error Cases**

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open input stream
  - existence_error(stream,Stream_or_alias)
- Char is not a variable or character.
  - domain_error(character_or_variable,Char)

get_char(?Char)
ISO
Behaves as get_char/2, but reads from the current input stream.

**Error Cases**
• Char is not a variable or character.
  − domain_error(character_or_variable,Char)

get_code(+Stream_or_alias,?Code)
  ISO
  Code unifies with the UTF-8 code of the next character from Stream_or_alias. The position of the stream is advanced. Char is unified with -1 if an end of file condition is detected.

Error Cases
  • Stream_or_alias is a variable
    − instantiation_error
  • Stream_or_alias is neither a variable nor a stream term nor an alias.
    − domain_error(stream_or_alias,Stream_or_alias)
  • Stream_or_alias is not associated with an open input stream
    − existence_error(stream,Stream_or_alias)
  • Code is not a variable or character code
    − domain_error(character_code_or_variable,Code)

get_code(?Code)
  ISO
  Behaves as get_code/2, but reads from the current input stream.

Error Cases
  • Code is not a variable or character code
    − domain_error(character_code_or_variable,Code)

get_byte(+Stream_or_alias,?Byte)
  ISO
  Byte unifies with the value of the the next byte from Stream_or_alias. The position of the stream is advanced. Char is unified with -1 if an end of file condition is detected. If reading from ASCII text, get_byte/2 will have the same behavior as get_code/2, but in general get_code/2 may return multi-byte characters.

Error Cases
  • Stream_or_alias is a variable
    − instantiation_error
  • Stream_or_alias is neither a variable nor a stream term nor an alias.

\footnote{The obsolescent predicate get0/1 is defined as get_code/1.}
CHAPTER 6. STANDARD AND GENERAL PREDICATES

- domain_error(stream_or_alias,Stream_or_alias)
  - Stream_or_alias is not associated with an open input stream
    - existence_error(stream,Stream_or_alias)
  - Code is not a variable or character code
    - domain_error(character_code_or_variable,Code)

get_byte/1
Behaves as get_byte/2, but reads from the current input stream\(^3\).

Error Cases
- Code is not a variable or Code is not a proper value for a byte
  - domain_error(byte_code_or_variable,Code)

peek_char(+Stream_or_alias,?Char)
ISO
Unifies Char with the next UTF-8 character from Stream_or_alias. The position in Stream_or_alias is unchanged. Char is unified with the atom end_of_file if an end of file condition is detected.

Error Cases
- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open input stream
  - existence_error(stream,Stream_or_alias)
- Char is not a variable or character.
  - domain_error(character_or_variable,Char)

peek_char(?Char)
ISO
Behaves as peek_char/2, but the current input stream is used.

Error Cases
- Char is not a variable or character.
  - domain_error(character_or_variable,Char)

\(^3\) The obsolescent predicate get0/1 is defined using get_byte/1, but returns the next byte that does not match an ASCII whitespace character.
peek_code(+Stream_or_alias,?Code)

Unifies Code with the next UTF-8 code from Stream_or_alias. The position in Stream_or_alias is unchanged. Code is unified with -1 if an end of file condition is detected.

Error Cases

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open input stream
  - existence_error(stream,Stream_or_alias)
- Code is not a variable or character.
  - domain_error(character_code_or_variable,Code)

peek_code(?Code)

Behaves as peek_code/2, but the current input stream is used.

Error Cases

- Char is not a variable or character.
  - domain_error(character_code_or_variable,Code)

peek_byte(?Byte)

Unifies Byte with the next byte from Stream_or_alias. The position in Stream_or_alias is unchanged. Code is unified with -1 if an end of file condition is detected.

Error Cases

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open input stream
  - existence_error(stream,Stream_or_alias)
- Code is not a variable or character.
  - domain_error(byte_code_or_variable,Code)
peek_byte(?Byte)
Behaves as peek_byte/2, but the current input stream is used.

Error Cases
- Char is not a variable or character.
  - domain_error(byte_code_or_variable,Code)

put_char(+Stream_or_alias,+Char)
Writes a UTF-8 character Char to Stream_or_alias.

Error Cases
- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open input stream
  - existence_error(stream,Stream_or_alias)
- Char is a not a character
  - type_error(character,Char)

put_char(+Char)
Puts a UTF-8 character Char to the current output stream.

Error Cases
- Code is a not a character.
  - type_error(character,Char)

put_code(+Stream,+Code)
Puts the character for the UTF-8 code Code to Stream_or_alias.

Error Cases
- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open input stream
put_code(+Code)
ISO
Puts the character for the UTF-8 code Code to the current output stream.

Error Cases

• Code is a not a character code.
  – type_error(character_code, Code)

nl
ISO
A new line character is sent to the current output stream.

nl(+Stream_or_alias)
ISO
A new line character is sent to the designated output stream.

Error Cases

• Stream_or_alias is a variable
  – instantiation_error
• Stream_or_alias is neither a variable nor a stream term nor an alias.
  – domain_error(stream_or_alias, Stream_or_alias)
• Stream_or_alias is not associated with an open stream
  – existence_error(stream, Stream_or_alias)

tab(+N)
Puts N spaces to the current output stream.

Error Cases

• Code is a not a positiveInteger
  – domain_error(positiveInteger, Code)

\footnote{The obsolescent predicate put/1 is defined as put_code/1.}
6.1.6 Term I/O

Beginning with Version 3.6, XSB automatically supports Unicode in the form of UTF-8 characters for reading and writing.

**read(??Term)**  ISO

HiLog term is read from the current or designated input stream, and unified with Term according to the operator declarations in force. (See Section 4.1 for the definition and syntax of HiLog terms). The term must be delimited by a full stop (i.e. a “.” followed by a carriage-return, space or tab). Predicate read/1 does not return until a valid HiLog term is successfully read; that is, in the presence of syntax errors read/1 does not fail but continues reading terms until a term with no syntax errors is encountered. If a call to read(Term) causes the end of the current input stream to be reached, variable Term is unified with the term end_of_file. In that case, further calls to read/1 for the same input stream will cause an error failure.

In Version 3.6, read/[1,2] are non ISO-compliant in how they handle syntax errors or their behavior when encountering an end of file indicator.

**read(+Stream_or_alias, ??Term)**  ISO

read/2 has the same behavior as read/1 but the input stream is explicitly designated by Stream_or_alias.

**Error Cases**

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open stream
  - existence_error(stream,Stream_or_alias)

**read_canonical(-Term)**

Reads a term that is in canonical format from the current input stream and returns it in Term. On end-of-file, it returns the atom end_of_file. If it encounters an error, it prints an error message on STDERR and returns the atom read_canonical_error. This is significantly faster than read/1, but requires the input to be in canonical form.
CHAPTER 6. STANDARD AND GENERAL PREDICATES

read_canonical(+Stream_or_alias),-Term
Behaves as read_canonical/1, but reads from Stream_or_alias.

Error Cases

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open input stream
  - existence_error(stream,Stream_or_alias)

read_term(?Term,?OptionsList)
ISO
A term is read from the current input stream as in read/1; but OptionsList is a (possibly empty) list of read options that specifies additional behavior. The read options include

- variables(Vars): once a term has been read, Vars is a list of the variables in the term, in left-to-right order.
- variable_names(VN_List): once a term has been read VN_List is a list of non-anonymous variables in the term. The elements of the list have the form A = V where V is a non-anonymous variable of the term, and A is the string used to denote the variable in the input stream.
- singletons(VS_List): once a term has been read VN_List is a list of the non-anonymous singleton variables in the term. The elements of the list have the form A = V where V is a non-anonymous variable of the term, and A is the string used to denote the variable in the input stream.

Error Cases

- OptionsList is a variable, or is a list containing a variable element.
  - instantiation_error
- OptionsList contains a non-variable element O that is not a read option.
  - domain_error(read_option,O)

read_term(+Stream_or_alias, ?Term,?OptionsList)
ISO
read_term/3 has the same behavior as read_term/2 but the input stream is explicitly designated using the first argument.

Error Cases are the same as read_term/2, but with the additional errors that may arise in stream checking.
- **Stream_or_alias** is a variable
  - instantiation_error
- **Stream_or_alias** is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- **Stream_or_alias** is not associated with an open stream
  - existence_error(stream,Stream_or_alias)

```prolog
write_term(?Term,+Options)
```
ISO

Outputs +Term to the current output stream. Stream \( \text{write_term}/3 \) according to the list of write options, Options. The current set of write options which form a superset of the ISO-standard write options, are as follows:

- **quoted(+Bool)**. If Bool = true, then atoms and functors that can’t be read back by \( \text{read}/1 \) are quoted, if Bool = false, each atom and functor is written as its unquoted name. Default value is false.

- **ignore_ops(+Bool)**. If Bool = true each compound term is output in functional notation; curly brackets and list braces are ignored, as are all explicitly defined operators. If Bool = false, curly bracketed notation and list notation is enabled when outputting compound terms, and all other operator notation is enabled. Default value is false.

- **numbervars(+Bool)**. If Bool = true, a term of the form ‘$VAR’(N) where N is an integer, is output as a variable name consisting of a capital letter possibly followed by an integer. A term of the form ‘$VAR’(Atom) where Atom is an atom, is output as itself (without quotes). Finally, a term of the form ‘$VAR’(String) where String is a character string, is output as the atom corresponding to this character string. If bool is false this cases are not treated in any special way. Default value is false.

- **max_depth(+Depth)**. Depth is a positive integer or zero. If positive, it denotes the depth limit on printing compound terms. If Depth is zero, there is no limit. Default value is 0 (no limit).

- **priority(+Prio)**. Prio is an integer between 1 and 1200. If the term to be printed has higher priority than Prio, it will be printed parenthesized. Default value is 1200 (no term parenthesized).

From the following examples it can be seen that \( \text{write_term}/[2,3] \) can duplicate the behavior of a number of other I/O predicates such as \( \text{write}/[1,2] \), \( \text{writeq}/[1,2] \), \( \text{write_canonical}/[1,2] \), etc.
CHAPTER 6. STANDARD AND GENERAL PREDICATES

| ?- write_term(f(1+2,'A','string','$VAR'(3),'$VAR'('Temp'),(multifile foo)),[]).
| f(1 + 2,A,"string",$VAR(3),$VAR(Temp),(multifile foo))
| yes

| ?- write_term(f(1+2,'A','string','$VAR'(3),'$VAR'('Temp'),(multifile foo)),
  [quoted(true)]).
| f(1 + 2,'A','"string","$VAR'(3),"$VAR'('Temp'),(multifile foo))
| yes

| ?- write_term(f(1+2,'A','string','$VAR'(3),'$VAR'('Temp'),(multifile foo)),
  [quoted(true),ignore_ops(true),numbervars(true)]).
| f(+1,2,'A','.'(115,'.'(116,'.'(114,'.'(105,'.'(110,'.'(103,[[]))))))),D,Temp,(multifile foo))
| yes

| ?- write_term(f(1+2,'A','string','$VAR'(3),'$VAR'('Temp'),(multifile foo)),
  [quoted(true),ignore_ops(true),numbervars(true),priority(1000)]).
| f(+1,2,'A','.'(115,'.'(116,'.'(114,'.'(105,'.'(110,'.'(103,[[]]))))))),D,Temp,multifile(foo))
| yes

**Error Cases**

- **Options** is a variable
  - instantiation_error
- **Options** neither a variable nor a list
  - type_error(list,Options)
- **Options** contains a variable element, O
  - instantiation_error
- **Options** contains an element O that is neither a variable nor a write option.
  - domain_error(write_option,O)

**ISO Compatibility Note:** In Version 3.6, write_term/[2,3] do not properly handle operators.

write_term(+Stream_or_alias,?Term,+Options) ISO
Behaves as write_term/2, but writes to Stream_or_alias.

**Error Cases** are the same as write_term/2 but with these additions.

- **Stream_or_alias** is a variable
– instantiation_error

• \textbf{Stream\_or\_alias} is neither a variable nor a stream term nor an alias.
  
  – domain_error(Stream\_or\_alias, Stream\_or\_alias)

• \textbf{Stream\_or\_alias} is not associated with an open output stream
  
  – existence_error(stream, Stream\_or\_alias)

\textbf{write}(\text{?Term}) \quad \text{ISO}

Semantically, \texttt{write/1} behaves as if \texttt{write\_term/1} were invoked using \texttt{quoted(false)}, \texttt{ignore\_ops(false)}, and \texttt{numbervars(false)}. Attributed variables are written according to the value of the Prolog flag \texttt{write\_attributes} (cf. \texttt{current\_prolog\_flag/2}).

The HiLog term \texttt{Term} is written to the current output stream, according to the operator declarations in force. Any uninstantiated subterm of term \texttt{Term} is written as an anonymous variable (an underscore followed by a token).

All \emph{proper HiLog terms} (HiLog terms which are not also Prolog terms) are not written in their internal Prolog representation. \texttt{write/1} always succeeds without producing an error.

HiLog (or Prolog) terms that are output by \texttt{write/1} cannot in general be read back using \texttt{read/1}. This happens for two reasons:

• The atoms appearing in term \texttt{Term} are not quoted. In that case the user must use \texttt{writeq/1} or \texttt{write\_canonical/1} described below, which quote around atoms whenever necessary.

• The output of \texttt{write/1} is not terminated by a full-stop; therefore, if the user wants the term to be accepted as input to \texttt{read/1}, the terminating full-stop must be explicitly sent to the current output stream.

\texttt{write/1} treats terms of the form ‘$VAR'(N), which may be generated by \texttt{numbervars/1,3} specially: it writes ‘A’ if \(N\)=0, ‘B’ if \(N\)=1, \ldots, ‘Z’ if \(N\)=25, ‘A1’ if \(N\)=26, etc. ‘$VAR’(-1) is written as the anonymous variable ‘_’.

\textbf{write}(\text{+Stream\_or\_alias, ?Term}) \quad \text{ISO}

\texttt{write/2} has the same behavior as \texttt{write/1} but the output stream is explicitly designated using the first argument.

\textbf{Error Cases} are the same as \texttt{read\_term/2}, but with the additional errors that may arise in stream checking.

• \textbf{Stream\_or\_alias} is a variable
  
  – instantiation_error
• **Stream_or_alias** is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)

• **Stream_or_alias** is not associated with an open output stream
  - existence_error(stream,Stream_or_alias)

writeq(?Term)  ISO
Acts as write_term/1 when defined with the options quoted(true), numbervars(true), and ignore_ops(false). In other words, atoms and functors are quoted whenever necessary to make the result acceptable as input to read/1 writeq/1 also treats terms of the form ’\VAR’ (N) specially, writing A if N= 0, etc., and output is in accordance with current operator definitions. writeq/1 always succeeds without producing an error.

writeq(+Stream_or_alias, ?Term)  ISO
writeq/2 has the same behavior as writeq/1 but the output stream is explicitly designated using the first argument.

Error Cases

• **Stream_or_alias** is a variable
  - instantiation_error

• **Stream_or_alias** is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)

• **Stream_or_alias** is not associated with an open output stream
  - existence_error(stream,Stream_or_alias)

write_canonical(?Term)  ISO
This predicate is provided so that the HiLog term Term, if written to a file, can be read back using read_canonical/[1,2] or read/[1,2] regardless of special characters appearing in Term or prevailing operator declarations. Like write_prolog/1, write_canonical/1 writes all proper HiLog terms to the current output stream using the standard Prolog syntax (see Section 4.1 on the standard syntax of HiLog terms). write_canonical/1 also quotes atoms and functors as writeq/1 does, to make them acceptable as input of read/1. Except for list-notation ([ ]) and infix comma-list notation, operator declarations are not taken into consideration, so that apart from these exceptions compound terms are written in the form:

\[
\langle \text{predicate name}\rangle((\langle\text{arg}_1\rangle,\ldots,\langle\text{arg}_n\rangle))
\]
Unlike writeq/1, write_canonical/1 does not treat terms of the form \$VAR\(N\) specially. It writes square bracket lists using \'(\).
and \([\]
(that is, \([\text{foo}, \text{bar}]\)
is written as \'(\text{foo},'(\text{bar},[\]))\).

Finally, write canonical/2 writes attributed variables as simple variables.

**ISO Compatibility Note:** In XSB, list notation and infix comma-list notation are considered canonical both for reading and writing. We find that this improves readability, and that these operators are so standard that there is little likelihood that they will not be in effect by any Prolog reader. We therefore deviate from the ISO standard definition of canonical in these cases.

\[\text{write canonical}(+\text{Stream_or_alias}, ?\text{Term})\]

write_canonical/2 has the same behavior as write_canonical/1 but the output stream is explicitly designated using the first argument.

**Error Cases**

- **Stream_or_alias** is a variable
  - instantiation_error
- **Stream_or_alias** is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- **Stream_or_alias** is not associated with an open output stream
  - existence_error(stream,Stream_or_alias)

\[\text{writeln}(?\text{Term})\]

writeln(Term) can be defined as write(Term), nl.

\[\text{writeln}(+\text{Stream},?\text{Term})\]

writeln(Term) can be defined as write(Stream,Term), nl(Stream).

\[\text{write prolog}(?\text{Term})\]

write_prolog/1 acts as write/1 except that any proper HiLog term Term is written using Prolog syntax – i.e. as a term whose outer functor is apply.

write_prolog/1 outputs Term according to the operator declarations in force. Because of this, it differs from writeCanonical/1 described above, despite the fact that both predicates write HiLog terms as Prolog terms.

write_prolog/2 has the same behavior as write_prolog/1 but the output stream is explicitly designated using the first argument. Error Cases for write_prolog/2 are the same as for write/2.

**Examples:**
numbervars(+Term, +FirstN, ?LastN, +Options)  module: num_vars

This predicate provides a mechanism for grounding a (HiLog) term so that it may be analyzed. Each variable in the (HiLog) term Term is instantiated to a term of the form '$VAR'(N), where N is an integer starting from FirstN. FirstN is used as the value of N for the first variable in Term (starting from the left). The second distinct variable in Term is given a value of N satisfying "N is FirstN + 1" and so on. The last variable in Term has the value LastN-1.

In numbervars/4, Options can be used to indicate the action to take upon encountering an attributed variable. Currently, Options must be either be the empty list, or the list [attvar(Action)] or the term attvar(Action), where Action is

- **error** Throw a type error if an attributed variable is encountered.
- **bind** Bind attributed variables by unifying them with terms of the form '$VAR'(N).
- **skip** Skip over attributed variables, performing no action on these variables.

**Error Cases**

- Options is a variable
  - instantiation_error
- Options is not an empty list, the list [attvar(Action)] or the term attvar(Action) where Action is one of bind, error or skip:
  - domain_error

numbervars(+Term, +FirstN, ?LastN)  module: num_vars

Acts as numbervars(+Term, +FirstN, ?LastN,attvar(error)).
numbervars(+Term)  
module: num_vars

This predicate is defined as: numbervars(Term, 0, _). It is included solely for convenience.

unnumbervars(+Term, +FirstN, ?Copy)  
module: num_vars

This predicate is a partial inverse of predicate numbervars/3. It creates a copy of Term in which all subterms of the form 'VAR'(int) where int is not less than FirstN are uniformly replaced by variables. 'VAR' subterms with the same integer are replaced by the same variable. Also a version unnumbervars/2 is provided which calls unnumbervars/3 with the second parameter set to 0.

**Term Writing to Designated I/O Streams**

While XSB has standard I/O streams for errors, warnings, messages, and feedback (cf. Section 6.1.1), the predicates above write to STDOUT which is the standard output for the process. Most of the time there is no issue with this as these streams are aliased to STDOUT. However in a number of circumstances, STDOUT may be redirected: a user may have invoked tell/1, XSB may be invoked through C or interprolog, etc. In such cases, it may be useful to ensure that output goes to one of the other I/O streams.

error_write(?Message)  
module: standard

warning(?Message)  
module: standard

These predicates output Message to XSB’s STDERR stream, rather than to XSB’s STDOUT stream, as does write/1 and writeln/1. In addition, if Message is a list or comma list, the elements in the comma list are output as if they were concatenated together. Each of these predicates must be imported from the module standard.

As above, but writes to STDFDBK, the console feedback stream.
The default behavior for warnings can be altered by setting the value of the Prolog flag `warning_action` to either `silent_warning` which performs no action when `warning/1` is called, or `error_warning` which throws a miscellaneous exception when `warning/1` is called (WARNING: this includes compiler warnings). The default behavior can be restored by setting `warning_action` to `print_warning`.

```prolog
message(?Message) module: standard
message(?Message) module: standard
```

As above, but writes to `STDMSG` the standard stream for messages.

### 6.1.7 Special I/O

```prolog
fmt_read(+Fmt,-Term,-Ret)
fmt_read(+Stream,+Fmt,-Term,-Ret)
```

These predicates provide a routine for reading data from the current input file (which must have been already opened by using `see/1`) according to a C format, as used in the C function `scanf`. `Fmt` must be a string of characters (enclosed in `'`) representing the format that will be passed to the C call to `scanf`. See the C documentation for `scanf` for the meaning of this string. The usual alphabetical C escape characters (e.g., `\n`) are recognized, but not the octal or the hexadecimal ones. Another difference with C is that, unlike most C compilers, XSB insists that a single `%` in the format string signifies format conversion specification. (Some C compilers might output `%` if it is not followed by a valid type conversion spec.) So, to output `%` you must type `%%`. Format can also be an atom enclosed in single quotes. However, in that case, escape sequences are not recognized and are printed as is.

`Term` is a term (e.g., `args(X,Y,Z)`) whose arguments will be unified with the field values read in. (The functor symbol of `Term` is ignored.) Special syntactic sugar is provided for the case when the format string contains only one format specifier: If `Term` is a variable, `X`, then the predicate behaves as if `Term` were `arg(X)`.

If the number of arguments exceeds the number of format specifiers, a warning is produced and the extra arguments remain uninstantiated. If the number of format specifiers exceeds the number of arguments, then the remainder of the format string (after the last matching specifier) is ignored.

Note that floats do not unify with anything. `Ret` must be a variable and it will be assigned a return value by the predicate: a negative integer if end-of-file is encountered; otherwise the number of fields read (as returned by `scanf`.)
fmt_read cannot read strings (that correspond to the \%s format specifier) that are longer than 16K. Attempting to read longer strings will cause buffer overflow. It is therefore recommended that one should use size modifiers in format strings (e.g., \%2000s), if such long strings might occur in the input.

Error Cases

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open output stream
  - existence_error(stream,Stream_or_alias)

If the number of arguments in Term is greater than the number of conversion specifiers in Fmt no error is thrown, but a warning is issued.

fmt_write(+Fmt,+Term)

fmt_write(+Stream_or_alias,+Fmt,+Term)

These predicates provide routines for writing formatted data to a given output stream (fmt_write/3) or the current output stream (fmt_write/2).

Fmt should be a Prolog character list (string) or atom. A Prolog character list is preferred, as space can be more easily reclaimed for character lists than for atoms. Term is a Prolog term (e.g., args(X,Y,Z)) whose arguments will be output. The number of arguments in Term should equal the number of conversion specifiers in Fmt. The functor symbol of Term is ignored.

Allowable syntaxes for Fmt reflect the syntax of the C function printf() on a given platform, with the following exceptions

- The usual alphabetical C escape characters (e.g., \n) are recognized, but not the octal or the hexadecimal ones.
- \%S is supported, in addition to the usual C conversion specifiers. The corresponding argument can be any Prolog term. This provides an easy way to print the values of Prolog variables, etc.
- \%! is supported and indicates that the corresponding argument is to be ignored and will generate nothing in the output.

\footnote{In the case where Fmt contains only a single conversion specifier, Term may be a string, integer or a float, and is considered to be equivalent to specifying \texttt{arg(Term)}.}
• A single % in the format string must be followed by a conversion operator (e.g. d, s, etc.). (Some C compilers output % if the percentage character is not followed by a valid type conversion spec.) However, to output %, fmt_write must contain %%. 

Example

| ?- fmt_write("%d %f %s %S \n",args(1,3.14159,ready,hello(world))). |
| 1 3.141590 ready hello(world) |

yes

XSB also offers an alternate version of formatted output in the format library described in volume 2. While not as efficient as fmt_write/[2,3], the format library is more compatible with the formatted output found in other Prologs.

Error Cases

• Stream_or_alias is a variable
  – instantiation_error

• Stream_or_alias is neither a variable nor a stream term nor an alias.
  – domain_error(stream_or_alias,Stream_or_alias)

• Stream_or_alias is not associated with an open output stream
  – existence_error(stream,Stream_or_alias)

• Fmt is uninstantiated or not a character string or atom
  – type_error('character string or atom',Fmt)

• A format specifier in Fmt and its corresponding argument in Term are of incompatible types.
  – misc_error

• Term contains fewer arguments than Fmt has format specifiers or Term is uninstantiated
  – misc_error

If the number of arguments in Term is greater than the number of conversion specifiers in Fmt no error is thrown, but a warning is issued.

Caution for 64-bit Platforms As discussed, fmt_write/[2,3] calls printf() and inherits the flexibility of that function, but also its “features”. One of
these features is that in most 64-bit platforms, large integers that behave perfectly well otherwise are not printed out properly by `printf()` with the `%d` format – rather another format string needs to be used (such as `%ld` on Linux). `fmt_write/[1,2]` recognizes the `%ld` option and passes it onto `fprintf()`, but the proper format string for 64-bit integers may be different on other platforms.

`fmt_write_string(-String,+Fmt,+Term)`
This predicate works like the C function `sprintf`. It takes the format string and substitutes the values from the arguments of `Term` (e.g., `args(X,Y,Z)`) for the formatting instructions `%s`, `%d`, etc. Additional syntactic sugar, as in `fmt_write`, is recognized. The result is available in `String`. `Fmt` is a string or an atom that represents the format, as in `fmt_write`.

If the number of format specifiers is greater than the number of arguments to be printed, an error is issued. If the number of arguments is greater, then a warning is issued.

`fmt_write_string` requires that the printed size of each argument (e.g., X,Y,and Z above) must be less than 16K. Longer arguments are cut to that size, so some loss of information is possible. However, there is no limit on the total size of the output (apart from the maximum atom size imposed by XSB).

`file_read_line_list(-String)`
A line read from the current input stream is converted into a list of character codes. This predicate avoids interning an atom as does `file_read_line_atom/3`, and so is recommended when speed is important. This predicate fails on reaching the end of file.

`file_read_line_list(Stream_or_alias,-CharList)`
Acts as does `file_read_line_list`, but uses `Stream_or_atom`.

Error Cases
- `Stream_or_alias` is a variable
  - `instantiation_error`
- `Stream_or_alias` is neither a variable nor a stream term nor an alias.
  - `domain_error(stream_or_alias,Stream_or_alias)`
- `Stream_or_alias` is not associated with an open input stream
  - `existence_error(stream,Stream_or_alias)`

`file_read_line_atom(-Atom)`
Reads a line from the current (textual) input stream, returning it as `Atom`. This predicate fails on reaching the end of file.
file_read_line_atom(+Stream_or_alias,-Atom)

Like file_read_line_atom/1 but reads from Stream_or_alias. **Error Cases**

- **Stream_or_alias** is a variable
  - instantiation_error
- **Stream_or_alias** is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- **Stream_or_alias** is not associated with an open input stream
  - existence_error(stream,Stream_or_alias)

file_write_line(+String, +Offset)

module: file_io

These predicates write String beginning with character Offset to the current output stream. String can be an atom or a list of UTF-8 character codes. This does not put the newline character at the end of the string (unless String already had this character). Note that escape sequences, like \n, are recognized if String is a character list, but are output as is if String is an atom.

**Error Cases**

- **Stream_or_alias** is a variable
  - instantiation_error
- **Stream_or_alias** is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- **Stream_or_alias** is not associated with an open input stream
  - existence_error(stream,Stream_or_alias)
- **String** is neither a Prolog character list not an atom
  - misc_error

file_getbuf_list(+Stream_or_alias, +BytesRequested, -CharList, -BytesRead)

module: file_io

Read BytesRequested bytes from file represented by Stream_or_alias (which must already be open for reading) into variable String as a list of character codes. This is analogous to fread in C. This predicate always succeeds. It does not distinguish between a file error and end of file. You can determine if either of these conditions has happened by verifying that BytesRead < BytesRequested.
file_getbuf_list(+BytesRequested, -String, -BytesRead) module: file_io
Like file_getbuf_list/3, but reads from the currently open input stream (i.e., with see/1).

file_getbuf_atom(+Stream_or_alias, +BytesRequested, -String, -BytesRead) module: file_io

Read BytesRequested bytes from file represented by Stream_or_alias (which must already be open for reading) into variable String. This is analogous to fread in C. This predicate always succeeds. It does not distinguish between a file error and end of file. You can determine if either of these conditions has happened by verifying that BytesRead < BytesRequested.

Note: because XSB does not have an atom table garbage collector yet, this predicate should not be used to read large files. Use read_getbuf_list or another predicate in this case.

Error Cases

- Stream_or_alias is a variable
  - instantiation_error
- Stream_or_alias is neither a variable nor a stream term nor an alias.
  - domain_error(stream_or_alias,Stream_or_alias)
- Stream_or_alias is not associated with an open input stream
  - existence_error(stream,Stream_or_alias)

file_getbuf_atom(+BytesRequested, -String, -BytesRead) module: file_io
Like file_getbuf_atom/4, but reads from the currently open input stream.

file_putbuf(+Stream_or_alias, +BytesRequested, +String, +Offset, -BytesWritten) module: file_io

Write BytesRequested bytes into file represented by I/O port Stream_or_alias (which must already be open for writing) from variable String at position Offset. This is analogous to C fwrite. The value of String can be an atom or a list of UTF-8 characters.

Error Cases

- Stream_or_alias is a variable
  - instantiation_error
• *Stream_or_alias* is neither a variable nor a stream term nor an alias.
  - `domain_error(stream_or_alias,Stream_or_alias)`
• *Stream_or_alias* is not associated with an open input stream
  - `existence_error(stream,Stream_or_alias)`

```prolog
file_putbuf(+BytesRequested, +String, +Offset, -BytesWritten)  
module: file_io
Like `file_putbuf/3`, but output goes to the currently open output stream.
```

### 6.2 Interactions with the Operating System

XSB provides a number of facilities for interacting with the UNIX and Windows operating systems. This section describes basic facilities for invoking shell commands and file manipulation. Chapter 1 of Volume 2 discusses more advanced commands for process spawning and control, along with interprocess communication.

```prolog
shell(+SystemCall)
Calls the operating system with the atom *SystemCall* as argument, using the *libc* function `system()`. The predicate succeeds if *SystemCall* is executed successfully, otherwise it fails. As a notational convenience, the user can also supply *SystemCall* either as a list. In this case, elements of the list will be concatenated together to form the system call.

For example, the call:

```prolog
| ?- shell('echo $HOME').
```

will output in the current output stream of XSB the name of the user’s home directory; while the call:

```prolog
| ?- File = 'test.c', shell(['cc -c ', File]).
```

will call the C compiler to compile the file *test.c*.

Note that in UNIX systems, since `system()` (and `shell/1`) executes by forking off a shell process. Thus it cannot be used, for example, to change the working directory of the program. For that reason the standard predicate `cd/1` described below should be used.

**Error Cases**

• *SystemCall* is a variable
CHAPTER 6. STANDARD AND GENERAL PREDICATES

- instantiation_error

- SystemCall is neither an atom nor a list
  - type_error(atom_or_list, SystemCall)

- SystemCall is longer than the maximum command length allowed by shell/1
  - resource_error(memory)

shell(+SystemCall, -Result)

Calls the operating system with the atom SystemCall as argument, using the libc function system(). As a notational convenience, the user can also supply SystemCall as a list. In this case, elements of the list will be concatenated together to form the system call. shell/2 always succeeds instantiating Result to the exit code of system(). Thus Result will be 0 if SystemCall executed properly, and non-0 otherwise: the specific return values of system() may be platform-dependent.

Error Cases

- SystemCall is a variable
  - instantiation_error

- SystemCall is neither an atom nor a list
  - type_error(atom_or_list, SystemCall)

- Result is not a variable
  - type_error(variable, Result)

- SystemCall is longer than the maximum command length allowed by shell/2
  - resource_error(memory)

shell_to_list(+SystemCall, -StdOut, -ErrOut, -Result)

Behaves as shell/2 in its 1st and 4th arguments, and like shell/2 always succeeds. Both StdOut and ErrOut are lists of lists: each element of the outer list corresponds to a line of output from SystemCall, while each element of an inner list corresponds to a token in that line. shell_to_list/3 is thus a sort of Prolog analog of the shell command ‘SystemCall’.

Examples:
?- shell_to_list(sw_vers,Stdout,Ret).
Stdout = [[ProductName:,Mac,OS,X],[ProductVersion:,10.4.9],[BuildVersion:,8P2137]]
Ret = 0

?- shell_to_lists('gcc -c nofile.c',StdOut,StdErr,Ret).
Stdout = []
StdErr = [[i686-apple-darwin8-gcc-4.0.1:,nofile.c:,No,such,file,or,directory]]
Ret = 256

Error cases are as with shell/2

datetime(?Date)

Unifies Date to the current UTC date, returned as a Prolog term, suitable for term comparison. Note that datime/1 must be explicitly imported from the module standard.
Example:

> date
Mon Aug 9 16:19:44 EDT 2004
> xsb
XSB Version 2.6 (Duff) of June 24, 2003
|i686-pc-cygwin; mode: optimal; engine: slg-wam; gc: indirection; scheduling: local

| ?- import datime/1 from standard

yes
| ?- datime(F).
F = datime(2004,8,9,20,20,23)

yes

local_datime(?Date)

Acts as datime/1, but returns the local, rather than the UTC date.

6.2.1 The path_sysop/2 interface

In addition, XSB provides the following unified interface to the operations on files. All these calls succeed iff the corresponding system call succeeds. These calls work on both Windows and Unixes unless otherwise noted.
path_sysop(isplain, +Path)
    Succeeds, if Path is a plain file.

path_sysop(isdir, +Path)
    Succeeds, if Path is a directory.

path_sysop(rename, +OldPath, +NewPath)
    Renames OldPath into NewPath.

path_sysop(copy, +FromPath, +ToPath)
    Copies FromPath into ToPath.

path_sysop(rm, +Path)
    Removes the plain file Path.

path_sysop(rmdir, +Path)
    Deletes the directory Path, succeeding only if the directory is empty.

path_sysop(rmdir_rec, +Path)
    Deletes the directory Path along with any of its contents.

path_sysop(link, +SrsPath, +DestPath)
    Creates a hard link from SrsPath to DestPath. UNIX only.

path_sysop(cwd, -Path)
    Binds Path to the current working directory.

path_sysop(chdir, +Path)
    Changes the current working directory to Path.

path_sysop(mkdir, +Path)
    Creates a new directory, Path.

path_sysop(exists, +Path)
    Succeeds if the file Path exists.

path_sysop(readable, +Path)
    Succeeds if Path is a readable file.

path_sysop(writable, +Path)
    Succeeds if Path is a writable file.

path_sysop(executable, +Path)
    Succeeds if Path is an executable file.
**path_sysop(modtime, +Path, -Time)**

Returns a list that represents the last modification time of the file. Succeeds if file exists. In this case, Time is bound to a list \([\text{high}, \text{low}]\) where \text{low} is the least significant 24 bits of the modification time and \text{high} is the most significant bits (25th) and up. Time represents the last modification time of the file. The actual value is thus \text{high} \times 2^{24} + \text{low}, which represents the number of seconds elapsed since 00:00:00 on January 1, 1970, Coordinated Universal Time (UTC).

**path_sysop(newerthan, +Path1, +Path2)**

Succeeds if the last modification time of Path1 is higher than that of Path2. Also succeeds if Path1 exists but Path2 does not.

**path_sysop(size, +Path, -Size)**

Returns a list that represents the byte size of Path. Succeeds if the file exists. In this case Size is bound to the list of the form \([\text{high}, \text{low}]\) where \text{low} is the least significant 24 bits of the byte-size and \text{high} is the most significant bits (25th) and up. The actual value is thus \text{high} \times 2^{24} + \text{low}.

**path_sysop(tmpfilename, -Name)**

Returns the name of a new temporary file. This is useful when the application needs to open a completely new temporary file.

**path_sysop(extension, +Name, -Ext)**

Returns file name extension.

**path_sysop(basename, +Name, -Base)**

Returns the base name of the file name (i.e., the name sans the directory and the extension).

**path_sysop(dirname, +Name, -Dir)**

Returns the directory portion of the filename. The directory is slash or backslash terminated.

**path_sysop(isabsolute, +Name)**

Succeeds if Name is an absolute path name. File does not need to exist.

**path_sysop(expand, +Name, -ExpandedName)**

Binds ExpandedName to the expanded absolute path name of Name. The file does not need to exist. Duplicate slashes, references to the current and parent directories are factored out.
6.3 Evaluating Arithmetic Expressions through is/2

Before describing is/2 and the expressions that it can evaluate, we note that in Version 3.6 of XSB, integers in XSB are represented using a single word of 32 or 64 bits, depending on the machine architecture. Floating point values are, by default, stored as word-sized references to double precision values, regardless of the target machine. Direct (non-referenced, tagged) single precision floats can be activated for speed purposes by passing the option –enable-fast-floats to the configure script at configuration time. This option is not recommended when any sort of precision is desired, as there may be as little as 28 bits available to represent a given number value under a tagged architecture.

All of the evaluable functors described below throw an instantiation error if one of their evaluated inputs is a variable, and an evaluation(undefined) error if one of their evaluated inputs is instantiated but non-numeric. With this in mind, we describe below only their behavior on correctly typed input.

**ISO Compatibility Note:** In addition, evaluation of arithmetic expressions through is/2 does not check for overflow or underflow. As a result, XSB’s floating point operations do not conform to IEEE floating point standards, and deviates in this regard from the ISO Prolog standard (see [34] Section 9). We hope to fix these problems in a future release.

\[
is(?\text{Result}, +\text{Expression})
\]

\[
is(\text{Result}, \text{Expression})\text{ is true iff the result of evaluating } \text{Expression}\text{ as a sequence of evaluable functors unifies with } \text{Result}. \text{ As mentioned in Section 3.10.6, }\text{is/2}\text{ is an inline predicate, so calls to }\text{is/2}\text{ within compiled code will not be visible during a trace of program execution.}
\]

**Error Cases**

- **instantiation_error** \text{Expression} contains an uninstantiated value

- **domain_error**(< function >, < value > \text{Expression} contains a function applied to value, but value is not part of the domain of function).

For is/2 the action for the above error cases can be altered so that the is/2 literal is treated as having a truth value of undefined in the well-founded semantics. This is done via the Prolog flag exception_action.

We also note that the ISO Prolog evaluable functors float_integer_part/1 (which can be obtained via truncate/1), float_fractional_part/1 (which can be obtained via \(X - \text{truncate}(X)\)), and bitwise complement (which is implementation dependent in the ISO standard) are not implemented in Version 3.6.
6.3.1 Evaluable Functors for Arithmetic Expressions

\(+(+\text{Expr1},+\text{Expr2})\) \hspace{1cm} \text{Evaluable Functor (ISO)}
If \(+\text{Expr1}\) evaluates to \text{Number1}, and \text{Expr2} evaluates to \text{Number2}, returns \text{Number1} + \text{Number2}, performing any necessary type conversions.

\(-(+\text{Expr1},+\text{Expr2})\) \hspace{1cm} \text{Evaluable Functor (ISO)}
If \(+\text{Expr1}\) evaluates to \text{Number1}, and \text{Expr2} evaluates to \text{Number2}, returns \text{Number1} - \text{Number2}, performing any necessary type conversions.

\(*(+\text{Expr1},+\text{Expr2})\) \hspace{1cm} \text{Evaluable Functor (ISO)}
If \(+\text{Expr1}\) evaluates to \text{Number1}, and \text{Expr2} evaluates to \text{Number2}, returns \text{Number1} * \text{Number2} (i.e. multiplies them), performing any necessary type conversions.

\(/(+\text{Expr1},\text{Expr2})\) \hspace{1cm} \text{Evaluable Functor (ISO)}
If \(+\text{Expr1}\) evaluates to \text{Number1}, and \text{Expr2} evaluates to \text{Number2}, returns \text{Number1} / \text{Number2} (i.e. divides them), performing any necessary type conversions.

div(+\text{Expr1},\text{Expr2}) \hspace{1cm} \text{ISO}
//(+\text{Expr1},\text{Expr2}) \hspace{1cm} \text{Evaluable Functor}
If \(+\text{Expr1}\) evaluates to \text{Number1}, and \text{Expr2} evaluates to \text{Number2}, returns \text{Number1} // \text{Number2} (i.e. integer division), performing any necessary type conversions, and rounding to 0 if necessary.
Example:

| ?- X is 3/2. |
| X = 1.5000 |
| yes |
| ?- X is 3 // 2. |
| X = 1 |
| yes |
| ?- X is -3 // 2. |
| X = -1 |
| yes |

\(-(+\text{Expr1})\) \hspace{1cm} \text{Evaluable Functor (ISO)}
If \(+\text{Expr}\) evaluates to \text{Number}, returns \(-\text{Number1}\), performing any necessary type conversions.
'∧'(Expr1,Expr2)  Evaluable Functor (ISO)
    If Expr1 evaluates to Number1, and Expr2 evaluates to Number2, returns the bitwise conjunction of Number1 and Number2.

'∨'(Expr1,Expr2)  Evaluable Functor (ISO)
    If Expr1 evaluates to Number1, and Expr2 evaluates to Number2, returns the bitwise disjunction Number1 and Number2.

'≫'(Expr1,Expr2)  Evaluable Functor (ISO)
    If Expr1 evaluates to Number1, and Expr2 evaluates to Number2, returns the logical shift right of Number1, Number2 places.

'≪'(Expr1,Expr2)  Evaluable Functor (ISO)
    If Expr1 evaluates to Number1, and Expr2 evaluates to Number2, returns the logical shift left of Number1, Number2 places.

xor(Expr1,Expr2)  ISO

'>>'(Expr1,Expr2)  Evaluable Functor
    If Expr1 evaluates to Number1, and Expr2 evaluates to Number2, returns the bitwise exclusive or of Number1 and Number2.

min(Expr1,Expr2)  Evaluable Functor (ISO)
    If Expr1 evaluates to Number1, and Expr2 evaluates to Number2, returns the minimum of the two.

max(Expr1,Expr2)  Evaluable Functor (ISO)
    If Expr1 evaluates to Number1, and Expr2 evaluates to Number2, returns the maximum of the two.

ceiling(Expr)  Evaluable Functor (ISO)
    If Expr evaluates to Number, ceiling(Number) returns the integer ceiling of Number if Number is a float, and Number itself if Number is an integer.

float(Expr)  Evaluable Functor (ISO)
    If Expr evaluates to Number, float(Number) converts Number to a float if Number is an integer, and returns Number itself if Number is a float.

ground(Expr)  Evaluable Functor (ISO)
    If Expr evaluates to Number, floor(Number) returns the integer floor of Number if Number is a float, and Number itself if Number is an integer.
mod(+Expr1,+Expr2)  
Evaluable Functor (ISO)  
If +Expr1 evaluates to Number1 and Expr2 evaluates to Number2 where Number2 is not 0, mod(Number1,Number2) returns  
\[\text{Number1} - (\lfloor (\text{Number1}/\text{Number2})\rfloor) \times \text{Number2}\]

rem(+Expr1,+Expr2)  
Evaluable Functor (ISO)  
If +Expr1 evaluates to Number1 and Expr2 evaluates to Number2 where Number2 is not 0, rem(Number1,Number2) returns  
\[\text{Number1} - (\text{Number1}//\text{Number2} \times \text{Number2})\]

Example:

| ?- X is 5 mod 2.  
X = 1  
yes  
| ?- X is 5 rem 2.  
X = 1  
yes  
| ?- X is 5 mod -2.  
X = -1  
yes  
| ?- X is 5 rem -2.  
X = 1  
yes

round(+Expr)  
Evaluable Functor (ISO)  
If +Expr evaluates to Number, round(Number) returns the nearest integer to Number if Number is a float, and Number itself if Number is an integer.

^/2  
Evaluable Functor (ISO)  
If Expr1 and Expr2 both evaluate to numbers, the infix function ^/2 raises Expr1 to the Expr2 power. If Expr1 and Expr2 both evaluate to integers, an integer is returned; otherwise a float is returned.
CHAPTER 6. STANDARD AND GENERAL PREDICATES

'+Expr1,+Expr2'+
Evaluable Functor (ISO)
If Expr1 and Expr2 both evaluate to numbers, the infix function **/2 raises Expr1 to the Expr2 power. A floating-point number is always returned.

sqrt(+Expr)
Evaluable Functor (ISO)
If +Expr evaluates to Number, sqrt(Number) returns the square root of Number.

truncate(+Expr)
Evaluable Functor (ISO)
If +Expr evaluates to Number, truncate(Number) truncates Number if Number is a float, and returns Number itself if Number is an integer.

sign(+Expr)
Evaluable Functor (ISO)
If +Expr evaluates to Number, sign(Number) returns 1 if Number is greater than 0, 0 if Number is equal to 0, and -1 if Number is less than 0.

pi
Evaluable Functor (ISO)
Evaluates to $\pi$ within an arithmetic expression.

e
Evaluable Functor
Evaluates to $e$, the base of the natural logarithm, within an arithmetic expression. (Use exp(1) for ISO compatibility.)

epsilon
Evaluable Functor
Evaluates to epsilon, the difference between the float 1.0 and the first larger floating point number.

Mathematical Functions from math.h

XSB also allows as evaluable functors, many of the functions from the C library math.h. Functions included in XSB Version 3.6 are cos/1 (ISO), sin/1 (ISO), tan/1 (ISO), acos/1 (ISO), asin/1 (ISO), atan/1 (ISO), log/1 (natural logarithm) (ISO), log10/1, and atan2/2 (ISO) (also available as atan2/2). For their semantics, see documentation to math.h.

6.4 Convenience

These predicates are standard and often self-explanatory, so they are described only briefly.
CHAPTER 6. STANDARD AND GENERAL PREDICATES

true
  Always succeeds.

otherwise
  Same as true/0.

call/1
  Always fails.

false
  Same as fail/0.

6.5 Negation and Control

'!/0
  Cut (discard) all choice points made since the parent goal started execution.
  Cuts across tabled predicates are not valid. The compiler checks for such cuts,
  although whether the scope of a cut includes a tabled predicate is undecidable
  in the presence of meta-predicates like call/1. Further discussion of conditions
  allowing cuts and of their actions can be found in Section 5.1.

\+ +P
  If the goal \( P \) has a solution, fails, otherwise it succeeds. Equivalently, it is true
  iff call(\( P \)) (see Section 6.11) is false. Argument \( P \) must be ground for sound
  negation as failure, although no runtime checks are made.

  Error Cases

  instantiation_error \( P \) is not instantiated.
  type_error(callable,\( P \)) \( P \) is not callable.

fail_if(+P)

not +P
  Like \+/1 and provided for compatibility with legacy code. Compilation of
  \+/1 and fail_if/1 is optimized by XSB’s compiler, while that of not/1 is not
  – therefore the first two syntactical forms are preferred in terms of efficiency,
  while \+/1 is preferred in terms of portability.

  All error cases are the same as call/1 (see Section 6.11).
tnot(+P)  
Tabling
The semantics of \texttt{tnot/1} allows for correct execution of programs with according to the well-founded semantics. \( P \) must be a tabled predicate, For a detailed description of the actions of tabled negation for in XSB Version 3.6 see [62, 64]. Chapter 5 contains further discussion of the functionality of \texttt{tnot/1}.

\textbf{Error Cases}

- \( P \) is not ground (floundering occurs)
  - \texttt{instantiation\_error}
- \( P \) is not callable
  - \texttt{type\_error(callable,P)}
- \( P \) is not a call to a tabled predicate
  - \texttt{table\_error}

\texttt{not\_exists(+P)}  
Tabling
If \(+P\) is a tabled predicate, \texttt{not\_exists/1} acts as \texttt{tnot/1} but permits variables in its subgoal argument The semantics in the case of unbound variables is as follows:

\[
\ldots :- \ldots, \text{not\_exists}(p(X)), \ldots
\]

is equivalent to

\[
\ldots :- \ldots, \text{tnot}(pp), \ldots
\]

\[
pp :- p(X).
\]

where \texttt{pp} is a new proposition. Thus, the unbound variable \( X \) is treated as \texttt{tnot(\exists X(p(X)))}.

If \(+P\) is a non-tabled predicate \texttt{not\_exists/1} ensures that \(+P\) is ground and called via a tabled predicate so that \texttt{not\_exists/1} can be used with non-tabled predicates as well, regardless of whether \(+P\) is ground or not \footnote{In previous versions of XSB, \texttt{not\_exists/1} was called \texttt{sk\_not/1}}.

\texttt{not\_exists/1} uses an auxiliary tabled predicate, \texttt{tunnumcall/1} in its execution. Therefore to reclaim space at the predicate or call level (e.g. using \texttt{abolish\_table\_pred/1} or similar predicates), \texttt{tunnumcall/1} must be explicitly abolished.

\textbf{Error Cases}
• \( P \) is not instantiated
  – instantiation_error
• \( P \) is not callable
  – type_error(callable,\( P \))

\texttt{u\_not(+P)}

\textit{module: tables}

If \( P \) is ground (or cyclic), \( u\_not(P) \) is equivalent to \( tnot(P) \); but \( u\_not/1 \) provides a different semantics than \( tnot/1 \) or \( not\_exists/1 \) if \( P \) is non-ground. In this latter case, \( u\_not(P) \) applies SLG delay to the goal \( P \), explicitly indicating that the default negation of \( P \) is floundered. This action is safe because any answer that relies on \( not \ P \) will be undefined, rather than true or false. A current limitation of \( u\_not/1 \) is that while floundering correctly causes a literal to be delayed, no simplification is ever performed if the delayed literal ever becomes ground (see the example below). \( u\_not/1 \) thus provides an informationally sound but incomplete semantics for floundering.

Thus, the use of \( tnot/1, not\_exists/1, or u\_not/1 \) depends on two conditions. \( not\_exists/1 \) is the only one of these predicates that allows \( P \) to be a nontabled predicate. However as mentioned, their main difference is in handling non-ground negative subgoals. If an error should be thrown for a non-ground negative subgoal, \( tnot/1 \) should be used; if it is semantically correct to skolemize if \( P \) is not ground, \( not\_exists/1 \) should be used; if it is semantically correct to treat the truth value of the negative subgoal as undefined, \( u\_not/1 \) should be used. From the perspective of performance, \( tnot/1 \) is fastest followed by \( u\_not/1 \) and then \( not\_exists/1 \).

The following examples should clarify the behavior of \( u\_not/1 \). For the program fragment:

\begin{verbatim}
:- table p/1,q/1.
p(1):- u\_not(q(X)).
q(1).
\end{verbatim}

the goal \( p(V) \) returns

\( V = 1 \) undefined

Examining this answer shows the following:

\begin{verbatim}
| ?- get\_residual(p(1),Res).
Res = [floundered(q(_h258))].
\end{verbatim}
The program fragment

```prolog
:- table r/1,q/1.
r(1):- u_not(q(X)),s(X).
q(1).
s(1).
```

shows a limitation in the current implementation of `u_not/1`. The goal `r(V)` returns

\[ V = 1 \text{ undefined} \]

as before. However, examining the answer shows

\[
| ?- \text{get_residual(r(1),Res)}
| \text{Res} = [\text{floundered(q(1))}]
\]

Note that the binding \( X=1 \) is propagated to the delayed literal after the resolution of `s(X)`. However, the call `tnot(q(1))` is not made once \( X \) is bound, so that the delayed literal does not fail.

**Error Cases** are the same as for `tnot/1`.

\[
P \rightarrow Q ; R
\]

**ISO**

Analogous to if `P` then `Q` else `R`, i.e. defined as if by

\[
(P \rightarrow Q ; R) :- P, \neg, Q.
(P \rightarrow Q ; R) :- R.
\]

\[
P \rightarrow Q
\]

**ISO**

When occurring other than as one of the alternatives of a disjunction, is equivalent to:

\[
P \rightarrow Q ; \text{fail}.
\]

**repeat**

Generates an infinite sequence of choice points (in other words it provides a very convenient way of executing a loop). It is defined by the clauses:

\[
\text{repeat}.
\text{repeat} :- \text{repeat}.
\]
between(+L,+U,B)

For L and U integers, with L less than or equal to U, successive calls to \texttt{between/3} unify B with all integers between L and U inclusively. If L is less than U the predicate fails.

\textbf{Error Cases:}

- L (or U) is a not an integer
  - \texttt{type\_error(integer,L)}

\section{Unification and Comparison of Terms}

The predicates described in this section allow unification and comparison of terms.\footnote{Arithmetic comparison predicates that may evaluate terms before comparing them are described in Section 6.3.1.}

Like most Prologs, default unification in XSB does not perform a so-called \textit{occurs check} — it does not handle situations where a variable X may be bound to a structure containing X as a proper subterm. For instance, in the goal

\[
X = f(X) \quad \% \text{incorrect!}
\]

X is bound to f(X) creating a term that is either cyclic or infinite, depending on one’s point of view. Prologs in general perform unification without occurs check since without occurs check unification is linear in the size of the largest term to be unified, while unification with occurs check may be exponential in the size of the largest term to be unified. Most Prolog programmers will rarely need to concern themselves with cyclic terms or unification with occurs check. However, unification with occurs check can be important for certain applications, in particular when Prolog is used to implement theorem provers or sophisticated constraint handlers. As a result XSB provides an ISO-style implementation of the predicate \texttt{unify\_with\_occurs\_check/2} described below, as well as a Prolog flag \texttt{unify\_with\_occurs\_check} that changes the behavior of unification in XSB’s engine.

As opposed to unification predicates, term comparison predicates described below take into account a standard total ordering of terms, which has as follows:

\begin{itemize}
  \item \texttt{variables} \texttt{@} < \texttt{floating point numbers} \texttt{@} < \texttt{integers} \texttt{@} < \texttt{atoms} \texttt{@} < \texttt{compound terms}
\end{itemize}

Within each one of the categories, the ordering is as follows:

- ordering of variables is based on their address within the SLG-WAM — the order is \textit{not} related to the names of variables. Thus note that two variables are
identical only if they share the same address – only if they have been unified or are the same variable to begin with. As a corollary, note that two anonymous variables will not have the same address and so will not be considered identical terms. As with most WAM-based Prologs, the order of variables may change as variables become bound to one another. If the order is expected to be invariant across variable bindings, other mechanisms, such as attributed variables, should be used.

- floating point numbers and integers are put in numeric order, from $-\infty$ to $+\infty$. Note that a floating point number is always less than an integer, regardless of their numerical values. If comparison is needed, a conversion should be performed (e.g. through float/1).
- atoms are put in alphabetical (i.e. UTF-8) order;
- compound terms are ordered first by arity, then by the name of their principal functor and then by their arguments (in a left-to-right order).
- lists are compared as ordinary compound terms having arity 2 and functor ‘.’.

For example, here is a list of terms sorted in increasing standard order:

\[
\text{[ X, 3.14, -9, fie, foe, fum(X), [X], X = Y, fie(0,2), fie(1,1) ]}
\]

The basic predicates for unification and comparison of arbitrary terms are:

\[
\begin{align*}
X = Y & \quad \text{ISO} \\
& \quad \text{Unifies X and Y without occur check.} \\
\text{unify_with_occurs_check(One,Two)} & \\
& \quad \text{Unifies One and Two using an occur check, and failing if One is a proper subterm of Two or if Two is a proper subterm of One.}
\end{align*}
\]

Example:

\[
| \text{?- unify_with_occurs_check(f(1,X),f(1,a(X))).} \\
\text{no} \\
| \text{?- unify_with_occurs_check(f(1,X),f(1,a(Y))).} \\
X = a(_h165) \\
Y = _h165 \\
\text{yes}
\]
| ?- unify_with_occurs_check(f(1,a(X)),f(1,a(X))).

X = _h165

yes

\[ T_1 == T_2 \] ISO
Tests if the terms currently instantiating \( T_1 \) and \( T_2 \) are literally identical (in particular, variables in equivalent positions in the two terms must be identical).
For example, the goal:
| ?- X == Y.
fails (answers no) because \( X \) and \( Y \) are distinct variables. However, the question
| ?- X = Y, X == Y.
succeeds because the first goal unifies the two variables.

\[ X \not= Y \] ISO
Succeeds if \( X \) and \( Y \) are not unifiable, fails if \( X \) and \( Y \) are unifiable. It is thus equivalent to \( \neg\neg(X = Y) \).

\[ T_1 \not\leq T_2 \] ISO
Succeeds if the terms currently instantiating \( T_1 \) and \( T_2 \) are not literally identical.

\[ \text{Term1 } \neq \text{Term2} \]
Succeeds if the equality of \( \text{Term1} \) and \( \text{Term2} \) can be compared safely, i.e. whether the result of \( \text{Term1} = \text{Term2} \) can change due to further instantiation of either term. It is specified as by \( \neq(A,B) :- (A==B ; A \neq B) \).

unifiable(X, Y, -Unifier) module: constraintLib
If \( X \) and \( Y \) can unify, succeeds unifying \texttt{Unifier} with a list of terms of the form \texttt{Var = Value} representing a most general unifier of \( X \) and \( Y \). \texttt{unifiable/3} can handle cyclic terms. Attributed variables are handles as normal variables. Associated hooks are not executed 9.

\[ T_1 \preceq T_2 \] ISO
Succeeds if term \( T_1 \) is before term \( T_2 \) in the standard order.

---

9In Version 3.6, \texttt{unifiable/3} is written as a Prolog predicate and so is slower than many of the predicates in this section.
CHAPTER 6.  STANDARD AND GENERAL PREDICATES

T1 @> T2  
    Succeeds if term T1 is after term T2 in the standard order.

T1 @=< T2  
    Succeeds if term T1 is not after term T2 in the standard order.

T1 @>= T2  
    Succeeds if term T1 is not before term T2 in the standard order.

T1 @= T2  
    Succeeds if T1 and T2 are identical variables, or if the main structure symbols of T1 and T2 are identical.

compare(?Op, +T1, +T2)  
    Succeeds if the result of comparing terms T1 and T2 is Op, where the possible values for Op are:

    ‘=’ if T1 is identical to T2,
    ‘<’ if T1 is before T2 in the standard order,
    ‘>’ if T1 is after T2 in the standard order.

Thus compare(=, T1, T2) is equivalent to T1=<T2. Predicate compare/3 has no associated error conditions.

ground(+X)  
    Succeeds if X is currently instantiated to a term that is completely bound (has no uninstantiated variables in it); otherwise it fails. While ground/1 has no associated error conditions, it is not safe for cyclic terms: if cyclic terms may be an issue use ground_or_cyclic/1.

ground_and_acyclic(+X)  

ground_or_cyclic(+X)  
    ground_or_cyclic/1 succeeds if X is currently instantiated to a term that is completely bound (has no uninstantiated variables in it) or is a cyclic term; otherwise it fails. Alternately, ground_and_acyclic/1 succeeds if X is currently instantiated to an acyclic term that is completely bound (has no uninstantiated variables in it). Neither predicate has no associated error conditions.

Both predicates are written to be as efficient as possible, and each requires a single traversal of a term, regardless of whether the term is ground, nonground or cyclic. However, due to the nature of checking for cyclicity, these predicates are somewhat slower than the unsafe ground/1.
CHAPTER 6. STANDARD AND GENERAL PREDICATES

subsumes(?Term1, +Term2)  
Term subsumption is a sort of one-way unification. Term Term1 and Term2 unify if they have a common instance, and unification in Prolog instantiates both terms to that (most general) common instance. Term1 subsumes Term2 if Term2 is already an instance of Term1. For our purposes, Term2 is an instance of Term1 if there is a substitution that leaves Term2 unchanged and makes Term1 identical to Term2. Predicate subsumes/2 does not work as described if Term1 and Term2 share common variables.

subsumes_chk(+Term1, +Term2)  
The subsumes_chk/2 predicate is true when Term1 subsumes Term2; that is, when Term2 is already an instance of Term1. This predicate simply checks for subsumption and does not bind any variables either in Term1 or in Term2. Term1 and Term2 should not share any variables.

Examples:

| ?- subsumes_chk(a(X,f,Y,X),a(U,V,b,S)).  
no
| ?- subsumes_chk(a(X,Y,X),a(b,b,b)).  
X = _595684  
Y = _595624

variant(?Term1, ?Term2)  
This predicate is true when Term1 and Term2 are alphabetic variants. That is, you could imagine that variant/2 as being defined like:

\[
\text{variant}(\text{Term1}, \text{Term2}) : - \\
\text{subsumes_chk}(\text{Term1}, \text{Term2}), \\
\text{subsumes_chk}(\text{Term2}, \text{Term1}).
\]

but the actual implementation of variant/2 is considerably more efficient. However, in general, it does not work for terms that share variables; an assumption that holds for most (reasonable) uses of variant/2.

check_variant(?Term1)  
check_variant(+Term1,+DontCares)  
check_variant/[1,2] provide efficient means of checking whether the variant
of a term has been asserted to a trie indexed predicate. A call \texttt{?- check\_variant(Term)} thus succeeds if a variant of \texttt{Term} has been trie indexed and asserted, and fails otherwise; the check performs no unification, and no backtracking is possible.

\texttt{check\_variant/2} allows the user to specify that the last \(n\) arguments of \texttt{Term} are not to be checked for variance. This \texttt{check\_variant(Term,N)} succeeds of there is a trie indexed term whose first \(\text{arity} - n\) arguments are variants of those in \texttt{Term}.

These predicates exploit the trie data structure to obtain their efficiency; as a result our implementation does not allow don’t care arguments apart from the final \(n\) arguments. \textbf{More importantly, for efficiency, no check is made to determine whether a predicate has been trie-indexed.} If unsure, the user should call \texttt{current\_index/2}.

\textbf{Example 6.6.1} \texttt{?- import check\_variant/1 from tables.}

\begin{verbatim}
yes
?- index(cmp/3, trie).
yes
| ?- assert(cmp(a,b,c)),assertcmp(d,e,f)).
yes
| ?- check\_variant(cmp(a,b,c)).
yes
| ?- check\_variant(cmp(a,b,1)).
no
| ?- check\_variant(cmp(a,b,X)).
no
| ?- check\_variant(cmp(a,b,X),1).
X = _h183
\end{verbatim}

\textbf{Error Cases}

\texttt{type\_error} Argument 1 of \texttt{check\_variant/[1,2]} is not a callable structure.
\texttt{type\_error} Argument 2 of \texttt{check\_variant/[2]} is not an integer
6.6.1 Sorting of Terms

Sorting routines compare and order terms without instantiating them. Users should be careful when comparing the value of uninstantiated variables. The actual order of uninstantiated variables may change in the course of program evaluation due to variable aliasing, garbage collection, or other reasons.

sort(+L1, ?L2)
ISO
The elements of the list L1 are sorted into the standard order, and any identical (i.e. ‘==’) elements are merged, yielding the list L2. The time to perform the sorting is $O(n\log n)$ where $n$ is the length of list L1.

Examples:

| ?- sort([3.14,X,a(X),a,2,a,X,a], L).
| L = [X,3.14,2,a,a(X)];
| no

Exceptions:

instantiation_error Argument 1 of sort/2 is a variable or is not a proper list.

type_error Argument 1 of sort/2 is a non-variable, non-list term.

keysort(+L1, ?L2)
ISO
The list L1 must consist of elements of the form Key-Value. These elements are sorted into order according to the value of Key, yielding the list L2. The elements of list L1 are scanned from left to right. Unlike sort/2, in keysort/2 no merging of multiple occurring elements takes place. The time to perform the sorting is $O(n \log n)$ where $n$ is the length of list L1. Note that the elements of L1 are sorted only according to the value of Key, not according to the value of Value. The sorting of elements in L1 is not guaranteed to be stable in the presence of uninstantiated variables.

Example:

| ?- keysort([3-a,1-b,2-c,1-a,3-a], L).
| L = [1-b,1-a,2-c,3-a,3-a]
| yes
Error Cases:

- instantiation_error L1 is a variable or is not a proper list.
- type_error L1 is a non-variable, non-list term.
- domain_error(key_value_pair,Element) L1 contains an element Element that is not of the form Key-Value.

parsort(+L1, +SortSpec, +ElimDupl, ?L2) module: machine

depsort/4 is a very general sorting routine. The list L1 may consist of elements of any form. SortSpec is the atom asc, the atom desc, or a list of terms of the form asc(I) or desc(I) where I is an integer indicating a sort argument position. The elements of list L1 are sorted into order according to the sort specification. asc indicates ascending order based on the entire term; desc indicates descending order. For a sort specification that is a list, the individual elements indicate subfields of the source terms on which to sort. For example, a specification of [asc(1)] sorts the list in ascending order on the first subfields of the terms in the list. [desc(1),asc(2)] sorts into descending order on the first subfield and within equal first subfields into ascending order on the second subfield. The order is determined by the standard predicate compare. If ElimDupl is nonzero, merging of multiple occurring elements takes place (i.e., duplicate (whole) terms are eliminated in the output). If ElimDupl is zero, then no merging takes place. A SortSpec of [] is equivalent to “asc”. The time to perform the sorting is \( O(n\log n) \) where \( n \) is the length of list L1. The sorting of elements in L1 is not guaranteed to be stable. parsort/4 must be imported from module machine.

Example:

```
? paritysort([f(3,1),f(3,2),f(2,1),f(2,2),f(1,3),f(1,4),f(3,1)],
            [asc(1),desc(2)],1,L).
L = [f(1,4),f(1,3),f(2,2),f(2,1),f(3,2),f(3,1)];
no
```

Error Cases:

- instantiation_error L1 is a variable or is not a proper list.
6.7 Meta-Logical

To facilitate manipulation of terms as objects in themselves, XSB provides a number meta-logical predicates. These predicates include the standard meta-logical predicates of Prolog, along with their usual semantics. In addition are provided predicates which provide special operations on HiLog terms. For a full discussion of Prolog and HiLog terms see Section 4.1.

\textbf{var(?X)} \quad \text{ISO}

Succeeds if \(X\) is currently uninstantiated (i.e. is still a variable); otherwise it fails.

Term \(X\) is uninstantiated if it has not been bound to anything, except possibly another uninstantiated variable. Note in particular, that the HiLog term \(X(Y,Z)\) is considered to be instantiated. There is no distinction between a Prolog and a HiLog variable.

Examples:

\begin{verbatim}
| ?- var(X).
yes
| ?- var([X]).
no
| ?- var(X(Y,Z)).
no
| ?- var((X)).
yes
| ?- var((X)(Y)).
no
\end{verbatim}

\textbf{nonvar(?X)} \quad \text{ISO}

Succeeds if \(X\) is currently instantiated to a non-variable term; otherwise it fails.

This has exactly the opposite behaviour of \textit{var/1}.

\textbf{atom(?X)} \quad \text{ISO}

Succeeds only if the \(X\) is currently instantiated to an atom, that is to a Prolog or HiLog non-numeric constant.

Examples:

\begin{verbatim}
| ?- atom(HiLog).
no
| ?- atom(10).
no
\end{verbatim}
integer(?X)  
Succeeds if \( X \) is currently instantiated to an integer; otherwise it fails.

float(?X)  
float/1 Same as real/1. Succeeds if \( X \) is currently instantiated to a floating point number; otherwise it fails.

real(?X)  
Succeeds if \( X \) is currently instantiated to a floating point number; otherwise it fails. This predicate is included for compatibility with earlier versions of XSB.

number(?X)  
Succeeds if \( X \) is currently instantiated to either an integer or a floating point number (real); otherwise it fails.

atomic(?X)  
Succeeds if \( X \) is currently instantiated to an atom or a number; otherwise it fails.
Examples:

| ?- atomic(10).  
  yes  
| ?- atomic(p).  
  yes  
| ?- atomic(h).  
  yes  
| ?- atomic(h(X)).  
  no  
| ?- atomic("foo").  
  no  
| ?- atomic('foo').  
  yes  
| ?- atomic(X).  
  yes  
| ?- atom('HiLog').  
  yes  
| ?- atom(X(a,b)).  
  no  
| ?- atom(h).  
  yes  
| ?- atom(+).  
  yes  
| ?- atom([]).  
  yes
no
| ?- atomic(X((Y))).
no

compound(?X)
ISO
Succeeds if \(X\) is currently instantiated to a compound term (with arity greater that zero), i.e. to a non-variable term that is not atomic; otherwise it fails.
Examples:

| ?- compound(1).
no
| ?- compound(foo(1,2,3)).
yes
| ?- compound([foo, bar]).
yes
| ?- compound("foo").
yes
| ?- compound('foo').
no
| ?- compound(X(a,b)).
yes
| ?- compound((a,b)).
yes

structure(?X)

Same as compound/1. Its existence is only for compatibility with previous versions.

is_list(?X)

Succeeds if \(X\) is a proper list. In other words if it is either the atom \([\]\) or \([H|T]\) where \(H\) is any Prolog or HiLog term and \(T\) is a proper list; otherwise it fails.
Examples:

| ?- is_list([p(a,b,c), h(a,b)]).
yes
| ?- is_list([_,_]).
yes
| ?- is_list([a,b|X]).
no
| ?- is_list([a|b]).
no
is_charlist(+X)
Succeeds if X is a Prolog string, i.e., a list of characters. Examples:

| ?- is_charlist("abc").
  yes
| ?- is_charlist(abc).
  no

is_charlist(+X,-Size)
Works as above, but also returns the length of that string in the second argument, which must be a variable.

is_attv(+Term)
Succeeds is Term is an attributed variable, and fails otherwise.

is_most_general_term(?X)
Succeeds if X is compound term with all distinct variables as arguments, or if X is an atom. (It fails if X is a cons node.)

| ?- is_most_general_term(f(_,_,_,_)).
  yes
| ?- is_most_general_term(abc).
  yes
| ?- is_most_general_term(f(X,Y,Z,X)).
  no
| ?- is_most_general_term(f(X,Y,Z,a)).
  no
| ?- is_most_general_term([_|_]).
  no

is_number_atom(?X)
Succeeds if X is an atom (e.g. ‘123’) (as opposed to a number 123) which can be converted to a numeric atom (integer or float) and fails otherwise. In particular, if is_number_atom(X) succeeds, then

| ?- atom_codes(X,Codes),number_codes(N,Codes).

will succeed.
callable(?X)

Callable if \( X \) is currently instantiated to a term that standard predicate \texttt{call/1} could take as an argument and not give an instantiation or type error. Note that it only checks for errors of predicate \texttt{call/1}. In other words it succeeds if \( X \) is an atom or a compound term; otherwise it fails. Predicate \texttt{callable/1} has no associated error conditions.

Examples:

\[
\begin{align*}
| ?- callable(p). & \text{yes} \\
| ?- callable(p(1,2,3)). & \text{yes} \\
| ?- callable([_,_]). & \text{yes} \\
| ?- callable(a). & \text{yes} \\
| ?- callable(3.14). & \text{no}
\end{align*}
\]

proper_hilog(?X)

Succeeds if \( X \) is a proper HiLog term – i.e. a HiLog term that is not a Prolog term; otherwise the predicate fails.

Examples: (In this example and the rest of the examples of this section we assume that \( h \) is the only parameter symbol that has been declared a HiLog symbol).

\[
\begin{align*}
| ?- proper_hilog(X). & \text{no} \\
| ?- proper_hilog(foo(a,f(b),[A])). & \text{no} \\
| ?- proper_hilog(X(a,b,c)). & \text{yes} \\
| ?- proper_hilog(3.6(2,4)). & \text{yes} \\
| ?- proper_hilog(h). & \text{no} \\
| ?- proper_hilog([a, [d, e, X(a)], c]). & \text{yes} \\
| ?- proper_hilog(a(a(X(a)))). & \text{yes}
\end{align*}
\]
functor(?Term, ?Functor, ?Arity)

Succeeds if the functor of the Prolog term Term is Functor and the arity (number of arguments) of Term is Arity. Functor can be used in either the following two ways:

1. If Term is initially instantiated, then
   - If Term is a compound term, Functor and Arity are unified with the name and arity of its principal functor, respectively.
   - If Term is an atom or a number, Functor is unified with Term, and Arity is unified with 0.

2. If Term is initially uninstantiated, then either both Functor and Arity must be instantiated, or Functor is instantiated to a number, and
   - If Arity is an integer in the range 1..255, then Term becomes instantiated to the most general Prolog term having the specified Functor and Arity as principal functor and number of arguments, respectively. The variables appearing as arguments of Term are all distinct.
   - If Arity is 0, then Functor must be either an atom or a number and it is unified with Term.
   - If Arity is anything else, then functor/3 aborts.

Error Cases

atom_or_variable Functor is not an atom or variable.

instantiation_error Both Term, and either Functor, or Arity are uninstantiated.

Examples:

```
| ?- functor(p(f(a),b,t), F, A).
  F = p
  A = 3

| ?- functor(T, foo, 3).
  T = foo(_595708,_595712,_595716)

| ?- functor(T, 1.3, A).
  T = 1.3
  A = 0

| ?- functor(foo, F, 0).
  F = foo
```
hilog_functor(?Term, ?F, ?Arity)

The XSB standard predicate `hilog_functor/3` succeeds

• when `Term` is a Prolog term and the principal function symbol (`functor`) of `Term` is `F` and the `arity` (number of arguments) of `Term` is `Arity`, or

• when `Term` is a HiLog term, having `name` `F` and the number of arguments `F` is applied to, in the HiLog term, is `Arity`. 

\[ \text{\texttt{| ?- functor("foo", F, A).}} \]
\[ \text{F = .} \]
\[ \text{A = 2} \]

\[ \text{\texttt{| ?- functor([], [], A).}} \]
\[ \text{A = 0} \]

\[ \text{\texttt{| ?- functor([2,3,4], F, A).}} \]
\[ \text{F = .} \]
\[ \text{A = 2} \]

\[ \text{\texttt{| ?- functor(a+b, F, A).}} \]
\[ \text{F = +} \]
\[ \text{A = 2} \]

\[ \text{\texttt{| ?- functor(f(a,b,c), F, A).}} \]
\[ \text{F = f} \]
\[ \text{A = 3} \]

\[ \text{\texttt{| ?- functor(X(a,b,c), F, A).}} \]
\[ \text{F = apply} \]
\[ \text{A = 4} \]

\[ \text{\texttt{| ?- functor(map(P)(a,b), F, A).}} \]
\[ \text{F = apply} \]
\[ \text{A = 3} \]

\[ \text{\texttt{| ?- functor(T, foo(a), 1).}} \]
\[ \text{++Error: Wrong type in argument 2 of functor/3} \]
\[ \text{Aborting...} \]

\[ \text{\texttt{| ?- functor(T, F, 3).}} \]
\[ \text{++Error: Uninstantiated argument 2 of functor/3} \]
\[ \text{Aborting...} \]

\[ \text{\texttt{| ?- functor(T, foo, A).}} \]
\[ \text{++Error: Uninstantiated argument 3 of functor/3} \]
\[ \text{Aborting...} \]
The first of these cases corresponds to the “usual” behaviour of Prolog’s `functor/3`, while the second is the extension of `functor/3` to handle HiLog terms. Like the Prolog’s `functor/3` predicate, `hilog_functor/3` can be used in either of the following two ways:

1. If `Term` is initially instantiated, then
   - If `Term` is a Prolog compound term, `F` and `Arity` are unified with the name and arity of its principal functor, respectively.
   - If `Term` is an atom or a number, `F` is unified with `Term`, and `Arity` is unified with 0.
   - If `Term` is any other HiLog term, `F` and `Arity` are unified with the name and the number of arguments that `F` is applied to. Note that in this case `F` may still be uninstantiated.

2. If `Term` is initially uninstantiated, then at least `Arity` must be instantiated, and
   - If `Arity` is an integer in the range 1..255, then `Term` becomes instantiated to the most general Prolog or HiLog term having the specified `F` and `Arity` as name and number of arguments `F` is applied to, respectively. The variables appearing as arguments are all unique.
   - If `Arity` is 0, then `F` must be a Prolog or HiLog constant, and it is unified with `Term`. Note that in this case `F` cannot be a compound term.
   - If `Arity` is anything else, then `hilog_functor/3` aborts.

In other words, the standard predicate `hilog_functor/3` either decomposes a given HiLog term into its name and arity, or given an arity —and possibly a name— constructs the corresponding HiLog term creating new uninstantiated variables for its arguments. As happens with `functor/3` all constants can be their own principal function symbols.

Examples:

```
| ?- hilog_functor(f(a,b,c), F, A).
  F = f
  A = 3

| ?- hilog_functor(X(a,b,c), F, A).
  X = _595836
  F = _595836
  A = 3
```
arg(+Index, +Term, ?Argument)  
ISO
Unifies Argument with the Index\textsuperscript{th} argument of Term, where the index is taken to start at 1. In accordance with ISO semantics, Index must be instantiated to a non-negative integer, and Term to a compound term, otherwise an error is thrown as described below. If Index is 0 or a number greater than the arity of Term, the predicate quietly fails.
Examples:

\begin{verbatim}
| ?- arg(2, p(a,b), A).
A = b

| ?- arg(2, h(a,b), A).
A = a

| ?- arg(0, foo, A).
no

| ?- arg(2, [a,b,c], A).
A = [b, c]
\end{verbatim}
CHAPTER 6. STANDARD AND GENERAL PREDICATES

| ?- arg(2, "HiLog", A).  
| A = [105,108,111,103] 

| ?- arg(2, a+b+c, A).  
| A = c 

| ?- arg(3, X(a,b,c), A).  
| X = _595820  
| A = b 

| ?- arg(2, map(f)(a,b), A).  
| A = a 

| ?- arg(1, map(f)(a,b), A).  
| A = map(f) 

| ?- arg(1, (a+b)(foo,bar), A).  
| A = a+b 

Error Cases

- Index is a variable
  - instantiation_error
- Index neither a variable nor an integer
  - type_error(integer,Index)
- Index is less than 0
  - domain_error(not_less_than_zero,Index)
- Term is a variable
  - instantiation_error
- Term neither a variable nor a compound term
  - type_error(integer,Index)

arg0(+Index, +Term, ?Argument)
Unifies Argument with the Index\textsuperscript{th} argument of Term if Index > 0, or with the functor of Term if Index = 0.

hilog_arg(+Index, +Term, ?Argument)
HiLog
If Term is a Prolog term, it has the same behaviour as arg/3, but if Term is a proper HiLog term, hilog_arg/3 unifies Argument with the (Index + 1)\textsuperscript{th}
argument of the Prolog representation of Term. Semantically, Argument is the \( \text{Index}^{th} \) argument to which the \text{HiLog functor} of Term is applied. The arguments of the Term are numbered from 1 upwards. An atomic term is taken to have 0 arguments.

Initially, Index must be instantiated to a positive integer and Term to any non-variable Prolog or HiLog term. If the initial conditions are not satisfied or \( I \) is out of range, the call quietly fails. Note that like \text{arg}/3 this predicate does not succeed for \text{Index}=0.

Examples:

\[
| \text{?- hilog_arg}(2, p(a,b), A). \quad A = b \\
| \text{?- hilog_arg}(2, h(a,b), A). \quad A = b \\
| \text{?- hilog_arg}(3, X(a,b,c), A). \quad X = _{595820} \quad A = c \\
| \text{?- hilog_arg}(1, \text{map}(f)(a,b), A). \quad A = a \\
| \text{?- hilog_arg}(2, \text{map}(f)(a,b), A). \quad A = b \\
| \text{?- hilog_arg}(1, (a+b)(\text{foo},\text{bar}), A). \quad A = \text{foo} \\
| \text{?- hilog_arg}(1, \text{apply}(\text{foo}), A). \quad A = \text{foo} \\
| \text{?- hilog_arg}(1, \text{apply}(\text{foo},\text{bar}), A). \quad A = \text{bar} \\
\]

Note the difference between the last two examples. The difference is due to the fact that apply/1 is a Prolog term, while apply/2 is a proper HiLog term.

\text{?Term =.. ?List} \quad \text{ISO}

Given proper instantiation of the arguments, =../2 (pronounced \textit{univ}) succeeds when (1) Term unifies with a compound Prolog or HiLog term and List unifies with a list whose head is the functor of Term and whose tail is a list of the arguments of Term; or (2) when Term unifies with an atomic term and List unifies with a list whose only element is Term. More precisely,
• If initially Term is uninstantiated, then List must be instantiated either to a proper list (list of determinate length) whose head is an atom, or to a list of length 1 whose head is a number.

• If the arguments of =../2 are both uninstantiated, or if either of them is not what is expected, =../2 throws the appropriate error message.

Examples:

<table>
<thead>
<tr>
<th>?- X - 1 =.. L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = _h112</td>
</tr>
<tr>
<td>L = [-, _h112, 1]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>?- p(a,b,c) =.. L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L = [p, a, b, c]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>?- h(a,b,c) =.. L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L = [apply, h, a, b, c]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>?- map(p)(a,b) =.. L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L = [apply, map(p), a, b]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>?- T =.. [foo].</th>
</tr>
</thead>
<tbody>
<tr>
<td>T = foo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>?- T =.. [apply, X, a, b].</th>
</tr>
</thead>
<tbody>
<tr>
<td>T = apply(X, a, b)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>?- T =.. [1,2].</th>
</tr>
</thead>
<tbody>
<tr>
<td>++Error[XSB/Runtime/P]: [Type (1 in place of atomic)] in arg 2 of predicate =../2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>?- T =.. [a+b,2].</th>
</tr>
</thead>
<tbody>
<tr>
<td>++Error[XSB/Runtime/P]: [Type (a + b in place of atomic)] in arg 2 of predicate =../2</td>
</tr>
</tbody>
</table>

| ?- X =.. [foo|Y]. |
|-------------------|
| ++Error[XSB/Runtime/P]: [Instantiation] in arg 2 of predicate =../2 |

Error Cases

• Term is a variable and List is a variable, a partial list, a or a list whose head is a variable
  – instantiation_error

• List is neither a variable nor a non-empty list
  – type_error(list, H)
• **List** is a list whose head $H$ is neither an atom nor a variable, and whose tail is not the empty list
  - `type_error(atomic, H)`

• **Term** is a variable and the tail of **List** has a length greater than XSB’s maximum arity for terms (256)
  - `representation_error(max arity)`

$\text{Term} \not=.. [?F |?ArgList]$  

**HiLog** When **Term** is a Prolog term, this predicate behaves exactly like the Prolog `=../2`. However when **Term** is a proper HiLog term, `^=../2` succeeds unifying $F$ to its HiLog functor and $\text{ArgList}$ to the list of the arguments to which this HiLog functor is applied. Like `=../2`, the use of `^=../2` can nearly always be avoided by using the more efficient predicates `hilog_functor/3` and `hilog_arg/3`. The behaviour of `^=../2`, on HiLog terms is as follows:

- If initially **Term** is uninstantiated, then the list in the second argument of `^=../2` must be instantiated to a *proper list* (list of determinate length) whose head can be any Prolog or HiLog term.
- If the arguments of `^=../2` are both uninstantiated, or if the second of them is not what is expected, `^=../2` aborts, producing an appropriate error message.

**Examples:**

```
| ?- p(a,b,c) ^=.. L.  
L = [p,a,b,c]  
| ?- h(a,b,c) ^=.. L.  
L = [h,a,b,c]  
| ?- map(p)(a,b) ^=.. L.  
L = [map(p),a,b]  
| ?- T ^=.. [X,a,b].  
T = apply(X,a,b)  
| ?- T ^=.. [2,2].  
T = apply(2,2)  
| ?- T ^=.. [a+b,2].  
T = apply(a+b,2)  
| ?- T ^=.. [3|X].
```
++Error: Argument 2 of ^=../2 is not a proper list
Aborting...

Error Cases

instantiation_error Argument 2 of ^=../2 is not a proper list.

copy_term(+Term, -Copy)
   ISO
   Makes a Copy of Term in which all variables have been replaced by brand new variables which occur nowhere else. Variable attributes are also copied. It can be very handy when writing (meta-)interpreters for logic-based languages. The version of copy_term/2 provided is space efficient in the sense that it never copies ground terms. Predicate copy_term/2 has no associated errors or exceptions.
   Examples:

   | ?- copy_term(X, Y).
   X = _598948
   Y = _598904

   rr | ?- copy_term(f(a,X), Y).
   X = _598892
   Y = f(a,_599112)

copy_term_nat(+Term, -Copy)
   module: machine
   Behaves as copy_term/2, however it replaces attributed variables with non-attributed variables in the copy.\textsuperscript{10}

term_variables(+Term,-VariableList)
   ISO
   Collects the variables in Term into the list VariableList. The variables are in the order of their first occurrences in a depth-first traversal of Term.

term_depth(+Term, -Depth)
   term_depth/2 provides an efficient way to find the maximal depth of a term. Term depth is defined recursively as follows:
   • The depth of a structure is defined as 1 + the maximal depth of any argument of that structure.

\textsuperscript{10}The name of this predicate was chosen for consistency with SWI Prolog, and stands for copy_term no attributes.
• The depth of an attributed variable is the depth of the attribute structure associated with that variable.

• The depth of a list \([H|T]\) is defined as \(1 + \) the maximal depth of \(H\) and \(T\).

• The depth of any other element is 1.

Note that according to this definition, the depth of the list \([a,b]\) is 3, since the list is equivalent to the structure \(.\cdot(a,.\cdot(b,[]))\) whose depth is 3.

\texttt{term_depth/2} does not check for cyclic structures, so it must be ensured that \texttt{Term} is acyclic.

\texttt{term_size(+Term, -Size)}

\texttt{term_size/2} provides an efficient way to find the total number of constituents of a term. Term size is defined recursively as follows:

• The size of an attributed variable is \(1\) (the variable size) + the size of the attribute structure.

• The size of a non-compound term is 1.

• The size of a compound term is defined as \(1 + \) the sum of the sizes of all arguments of that term.

• The size of a list \([H|T]\) is defined as the size of the term \().'\cdot(H,T)\).

\texttt{term_size/2} does not check for cyclic structures, so it must be ensured that \texttt{Term} is acyclic.

\texttt{intern_term(+Term, -InternedTerm)} \hspace{1cm} \text{module: machine}

\texttt{intern_term} makes an “interned” version of its first argument and returns that term in its second argument. The terms are equal terms (i.e., \texttt{Term == InternedTerm} would succeed.) The interned term has all its ground subterms represented (uniquely) in a global space. Subterms that contain variables are not copied but remain on the heap. The interned representation of ground terms can save space and/or time in some situations. Note that already interned subterms or \texttt{Term} do not need to be traversed in this operation.
6.8 Cyclic Terms

6.8.1 Unification with and without Occurs Check

Cyclic terms are created when Prolog unifies two terms whose variables have not been standardized apart: for instance

\[ X = f(X) \]

will produce the cyclic term \( f(f(f(f(f(...)))))) \) – in other words, a term with an “infinite” depth. Note that according to the mathematical definition of unification, \( X \) should not unify with a term containing itself. There are two reasons why XSB (along with virtually all other Prologs) has this default behavior.

- The default unification algorithm, when it unifies a variable \( V \) with a term \( T \), does not check for the occurrence of \( V \) in \( T \), in other words it does not perform an occurs check. Unification without an occurs check is linear in the size of the smaller of the terms to be unified. Unification with occurs check is (essentially) linear in the size of the larger term. Since unification is often used to assign a value to a variable, it is important in a programming language that assignment be constant time, and not linear in the size of the term being assigned.

- Some programs purposefully construct cyclic terms: this occurs with various constraint libraries such as CHR. These libraries do not perform as expected when a mathematically correct unification algorithm is used.

XSB provides two mechanisms for overriding this default behavior for unification.

- First, there is a Prolog flag `unify_with_occurs_check` which when set to `on` ensures that all unification is mathematically correct. Care should be taken when using this flag, for the above two reasons.

- For more detailed usages, the ISO predicate `unify_with_occurs_check/2` can be used syntactically rather than Prolog’s default unification operator `=/2`.

6.8.2 Cyclic Terms

Fortunately, the creation of cyclic terms is uncommon for most types of programming; even when cyclic terms arise they can often be avoided by the proper use of
copy_term/2 or other predicates. Nevertheless cyclic terms do arise when XSB is used for meta-programming or if XSB is used as the basis of a high-level knowledge representation language such as Flora-2 or Silk. It is important that XSB's behavior be cycle-safe in the sense that the creation of cyclic terms per se will not create infinite loops in XSB's tabling or XSB's builtins. Like some other Prologs, XSB supports unification of cyclic terms. In addition, most predicates like functor/3, or =../2 that either take non-compound terms or that do not require term traversal are cycle-safe. A few builtins that require term-traversal are “safe” for cyclic terms. For instance writing in XSB is subject to a depth check, which terminates for cyclic terms. Most importantly, the XSB heap garbage collector is guaranteed to be safe for cyclic terms.

Variant tabling can also handle cyclic terms if the proper flags are set. These flags are max_table_subgoal_depth which determines the maximal “reasonable” depth of a subgoal; and max_table_answer_depth, max_table_answer_list_depth which determine the maximal “reasonable” depth for non-list terms or lists (respectively) in answers. These last two flags also determine a “reasonable” depth for interned tries. Each of these depth flags have an associated answer flag: max_table_subgoal_action, max_table_answer_action and max_table_answer_list_action respectively. The actions can be of three types: error which throws an error if a term with a certain depth is encountered as a tabled subgoal or answer (regardless of whether that term is tabled); failure which causes failure for these cases; and fail_on_cycles which fails on cyclic terms, and otherwise throws an error for a term of a certain depth.

While the above operations cycle-safe, cyclic terms can cause problems in XSB for builtins or predicates that require term traversal. For instance the library predicates length/2 and append/2 currently go into infinite loops with cyclic terms; unless otherwise specified it is the user’s responsibility to check library predicates (as opposed to standard builtins) for acyclicity using is_acyclic/1 or is_cyclic/1. In addition the following XSB builtins are not cycle-safe:

- bagof/3, copy_term/2, ground/1 numbervars/[1,3,4], setof/3, subsumes/2, subsumes_chk/2, term_depth/2, term_size/2, term_to_atom/[2,3], term_to_codes/[2,3], term_variables/2, unifiable/2 and variant/2.

- Various table inspection builtins based on get_call/2 or similar routines (including get_residual/2).

Arguably, programs should not intentionally create cyclic terms, and the above flags, as well as the following predicates, can help debug when cyclic terms are created.

11We hope to efficiently integrate cycle checking into XSB’s subsumptive tabling in the reasonably near future.
12The predicate ground_or_cyclic/1 is safe for cyclic terms.
is_cyclic(?X)
   Succeeds if X is a cyclic term.

is_acyclic(?X)

acyclic_term(?X) ISO
   Succeeds if X is not a cyclic term.

6.9 Manipulation of Atomic Terms

This section lists some of XSB’s standard predicates for manipulating atomic terms. See also in Volume 2, Section 1.5 for other library predicates. Section 7 for wildcard matching, and Section 8 for an interface to the PCRE library.

atom_codes(?Atom, ?CharCodeList) ISO
   The standard predicate atom_codes/2 performs the conversion between an atom and its character list representation. If Atom is supplied (and is an atom), CharList is unified with a list of UTF-8 codes representing the “name” of that atom. In that case, CharList is exactly the list of UTF-8 character codes that appear in the printed representation of Atom. If on the other hand Atom is a variable, then CharList must be a proper list of UTF-8 character codes. In that case, Atom is instantiated to an atom containing exactly those characters, even if the characters look like the printed representation of a number.

Examples:

| ?- atom_codes('Foo', L).  
| L = [70,111,111]          

| ?- atom_codes([], L).     
| L = [91,93]               

| ?- atom_codes(X, [102,111,111]).  
| X = foo                   

| ?- atom_codes(X, []).      
| X = ''                    

| ?- atom_codes(X, "Foo").  
| X = 'Foo'                 

| ?- atom_codes(X, [52,51,49]).
\[ X = '431' \]

\[ \begin{align*} | ?- \text{atom_codes}(X, [52,51,49]), \text{integer}(X). & \text{no} \\
| ?- \text{atom_codes}(X, [52,Y,49]). & \text{++Error[XSB/Runtime/P]: [Instantiation] in arg 2 of predicate atom_codes/2} \\
& \text{Forward Continuation...} \\
| ?- \text{atom_codes}(431, L). & \text{++Error[XSB/Runtime/P]: [Type (431 in place of atom)] in arg 1 of predicate atom_codes/2} \\
& \text{Forward Continuation...} \\
| ?- \text{atom_codes}(X, [52,300,49]). & \text{[Representation (300 is not character code)] in arg 2 of predicate atom_codes/2} \\
& \text{Forward Continuation...} \\
\end{align*} \]

Error Cases

- **Atom** is a variable and **CharCodeList** is a partial list or a list with an element which is a variable
  - \text{instantiation_error}
- **Atom** is neither a variable nor an atom
  - \text{type_error(atom, Atom)}
- **Atom** is a variable and **CharCodeList** is neither a list nor a partial list
  - \text{type_error(list, CharCodeList)}
- **Atom** is a variable and an element E of **CharCodeList** is neither a variable nor a character code
  - \text{representation_error(character_code, E)}

\text{number_codes(?Number, ?CharCodeList)}

The standard predicate \text{number_codes/2} performs the conversion between a number and its character list representation. If \text{Number} is supplied (and is a number), \text{CharList} is unified with a list of UTF-8 (= ASCII) codes comprising the printed representation of that \text{Number}. If on the other hand \text{Number} is a variable, then \text{CharList} must be a proper list of UTF-8 (ASCII) character codes that corresponds to the correct syntax of a number (either integer or float) In that case, \text{Number} is instantiated to that number, otherwise \text{number_codes/2} will simply fail.
Examples:

```
| ?- number_codes(123, L).
L = [49,50,51];

| ?- number_codes(N, [49,50,51]), integer(N).
N = 123

| ?- number_codes(31.4e+10, L).
L = [51,46,49,51,57,57,57,55,69,43,49,48]

| ?- number_codes(N, "314e+8").
N = 3.14e+10

| ?- number_codes(foo, L).
++Error[XSB/Runtime/P]: [Type (foo in place of number) in arg 1 of predicate number_codes
Forward Continuation...
```

Error Cases

- **Number** is a variable and **CharCodeList** is a partial list or a list with an element which is a variable
  - instantiation_error
- **Number** is neither a variable nor a number
  - type_error(number, Number)
- **Number** is a variable and **CharCodeList** is neither a list nor a partial list
  - type_error(list, CharCodeList)
- **Number** is a variable and an element **E** of **CharCodeList** is neither a variable nor a character code
  - representation_error(character_code, E)

**name(?Constant, ?CharList)**

The standard predicate **name/2** performs the conversion between a constant and its character list representation. If **Constant** is supplied (and is any atom or number), **CharList** is unified with a list of UTF-8 codes representing the “name” of the constant. In that case, **CharList** is exactly the list of UTF-8 character codes that appear in the printed representation of **Constant**. If on the other hand **Constant** is a variable, then **CharList** must be a proper list
of UTF-8 character codes. In that case, `name/2` will convert a list of UTF-8 characters that can represent a number to a number rather than to a character string. As a consequence of this, there are some atoms (for example `18`) which cannot be constructed by using `name/2`. If conversion to an atom is preferred in these cases, the standard predicate `atom_codes/2` should be used instead. The syntax for numbers that is accepted by `name/2` is exactly the one which `read/1` accepts.

Examples:

```
| ?- name('Foo', L).
L = [70,111,111]

| ?- name([], L).
L = [91,93]

| ?- name(431, L).
L = [52,51,49]

| ?- name(X, [102,111,111]).
X = foo

| ?- name(X, []).
X = ''

| ?- name(X, "Foo").
X = 'Foo'

| ?- name(X, [52,51,49]).
X = 431

| ?- name(X, [45,48,50,49,51]), integer(X).
X = -213

| ?- name(3.14, L).
++Error[XSB/Runtime/P]: [Miscellaneous] Predicate name/2 for reals is not implemented
Aborting...
```

- **Constant** is a variable and **CharCodeList** is a partial list or a list with an element which is a variable
  - `instantiation_error`
- **Constant** is neither a variable nor atomic
  - `type_error(atomic, Constant)`
• **Constant** is a variable and **CharCodeList** is neither a list nor a partial list
  - `type_error(list, CharCodeList)`

• **Constant** is a variable and an element **E** of **CharCodeList** is neither a variable nor a character code
  - `representation_error(character_code, E)`

```atom_chars(?Number, ?CharList)```

ISO

Like `atom_codes/2`, but the list returned (or input) is a list of characters as *atoms* rather than UTF-8 codes. For instance, `atom_chars(abc,X)` binds **X** to the list `[a,b,c]` instead of `[97,98,99]`.

Error Cases

• **Atom** is a variable and **CharCodeList** is a partial list or a list with an element which is a variable
  - `instantiation_error`

• **Atom** is neither a variable nor an atom
  - `type_error(atom, Atom)`

• **Atom** is a variable and **CharCodeList** is neither a list nor a partial list
  - `type_error(list, CharCodeList)`

• An element **E** of **CharCodeList** is not a single-character atom
  - `type_error(character, E)`

• **Atom** is a variable and an element **E** of **CharCodeList** is not a single-character atom
  - `representation_error(character, E)`

```number_chars(?Number, ?CharList)```

ISO

Like `number_codes/2`, but the list returned (or input) is a list of characters as *atoms* rather than codes. For instance, `number_chars(123,X)` binds **X** to the list `['1','2','3']` instead of `[49,50,51]`.

Error Cases

• **Number** is a variable and **CharCodeList** is a partial list or a list with an element which is a variable
  - `instantiation_error`

• **Number** is neither a variable nor a number
- `type_error(number, Number)`

- **Number** is a variable and **CharList** is neither a list nor a partial list
  - `type_error(list, CharList)`

- **An element** `E` of **CharList** is not a single-character atom
  - `type_error(character, E)`

- **CharList** is a list of single-character atoms but is not parsable as a number (by XSB)
  - `syntax_error(CharList)`

**number_digits(?Number, ?DigitList)**

Like `number_codes/2`, but the list returned (or input) is a list of digits as numbers rather than UTF-8 codes (for floats, the atom ‘,’ ‘+’ or ‘-’, and ‘e’ will also be present in the list). For instance, `number_digits(123,X)` binds `X` to the list `[1,2,3]` instead of `[‘1’, ‘2’, ‘3’]`, and `number_digits(123.45,X)` binds `X` to `[1,.,2,3,4,5,0,0,e,+0,2]`.

Error cases are the same as `number_chars/2`.

**char_code(?Character, ?Code)**

The standard predicate `char_code/2` is true if **Code** is the current code for **Character**. In XSB it is defined as `atom_codes(Character,[Code])`.

**atom_length(+Atom1,?Length)**

This standard predicate succeeds if **Length** unifies with the length of (the name of) **Atom**.

Example

```
|?- atom_length(trilobyte,L).

L = 9
```

Error Cases

- **Atom** is a variable
  - `instantiation_error`

- **Atom** is neither a variable nor an atom
  - `type_error(atom,Atom)`

- **Length** is neither a variable nor an integer
– type_error(integer,Length)

call_atom(+AtomList,?Atom) module: string

If Atom is a variable, then AtomList must be a list structure containing atoms, integers and/or floats. This predicate flattens AtomList and concatenates the atoms and integers into a single atom, returned in Atom. Integers and floats are converted to character strings using number_codes/2.

If Atom is an atom, then AtomList must be a list containing atoms, and/or variables. In this case atom_codes binds the variables in the list to atoms in such a way that the atoms of AtomList concatenate to the atom Atom. For example, call_atom([X,abb,Y,cc],aabbabdefcc) will succeed twice, first binding X to a and Y to abdef, and then binding X to aabb and Y to def.

This is a somewhat more general predicate than the ISO atom_concat/2 described below, and can be more efficient if numerous atoms are to be concatenated together.

call_atom(+AtomList,+Sep,?Atom) module: string

AtomList must be a list containing atoms, integers and/or floats, and Sep must be an atom. This predicate concatenates the atoms and integers into a single atom, separating each by Sep, return the resulting atom in Atom. Integers and floats are converted to character strings using number_codes/2.

This is a somewhat more general predicate than the ISO atom_concat/2 described below, and can be more efficient if numerous atoms are to be concatenated together.

atom_concat(Atom1,Atom2,Atom3) ISO

• Usage: atom_concat(?Atom,?Atom,+Atom)

• Usage: atom_concat(+Atom,+Atom,-Atom)

Succeeds if Atom12 is the concatenation of Atom1 and Atom2.

Examples

| ?- atom_concat(hello,world,F).

F = hello world

| ?- atom_concat(X,Y,'hello world').
\begin{verbatim}
X =
Y = hello world;
X = h
Y = ello world
\end{verbatim}

The last query will re-succeed for all combinations of atoms that produce \texttt{hello world}.

\textbf{Error Cases}

- \texttt{Atom1} and \texttt{Atom3} are both variables
  - \texttt{instantiation_error}
- \texttt{Atom2} and \texttt{Atom3} are both variables
  - \texttt{instantiation_error}
- \texttt{Atom1} is neither a variable nor an atom
  - \texttt{type_error(atom,Atom1)}
- \texttt{Atom2} is neither a variable nor an atom
  - \texttt{type_error(atom,Atom2)}
- \texttt{Atom3} is neither a variable nor an atom
  - \texttt{type_error(atom,Atom3)}

\texttt{sub_atom(+Atom,?LeftLength,?CenterLength,?RightLength,?CenterAtom ISO
Succeeds if Atom can be broken into three pieces: A left atom of length \texttt{LeftLength}, a center atom \texttt{CenterAtom} of length \texttt{CenterLength} and a right atom of length \texttt{RightLength}. If sufficient arguments are uninstantiated to produce \texttt{CenterAtom} in non-deterministic starting positions, the predicate will backtrack through all center atoms for which the left atom length is the smallest, up to those whose left atom length is greatest (see examples below).

\textbf{Examples}

\begin{verbatim}
| ?- sub_atom(trilobyte,5,4,RL,CA).

RL = 0
CA = byte
| ?- sub_atom(trilobyte,1,CL,2,CA).

CL = 6
\end{verbatim}
CA = riloby
| ?- sub_atom(trilobyte,LL,6,RL,riloby).

LL = 1
RL = 2
| ?- sub_atom(trilobyte,RL,4,LL,CA).

RL = 0
LL = 5
CA = tril;

RL = 1
LL = 4
CA = rilo;

RL = 2
CL = 3
CA = ilob
| ?- sub_atom(trilobyte,LL,CL,RL,CA).

LL = 0
CL = 0
RL = 9
CA = ;

LL = 0
CL = 1
RL = 8
CA = t;

LL = 0
CL = 2
RL = 7
CA = tr;

: /* after more backtracking */

LL = 0
CL = 9
RL = 0
CA = trilobyte;
Error Cases

- Atom is a variable
  - instantiation_error
- Atom is neither a variable nor an atom
  - type_error(atom, Atom)
- CenterAtom is neither a variable nor an atom
  - type_error(atom, CenterAtom)
- LeftLength is neither a variable nor an integer
  - type_error(integer, LeftLength)
- CenterLength is neither a variable nor an integer
  - type_error(integer, CenterLength)
- RightLength is neither a variable nor an integer
  - type_error(integer, RightLength)
- LeftLength is an integer that is less than zero
  - domain_error(not_less_than_zero, LeftLength)
- CenterLength is an integer that is less than zero
  - domain_error(not_less_than_zero, CenterLength)
- RightLength is an integer that is less than zero
  - domain_error(not_less_than_zero, RightLength)

string_substitute(+InpStr, +SubstrList, +SubstitutionList, -OutStr) module: string

InputStr can an atom or a list of characters. SubstrList must be a list of terms of the form s(BegOffset, EndOffset), where the name of the functor
is immaterial. The meaning of the offsets is the same as for substring/4. (In particular, negative offsets represent offsets from the first character past the end of String.) Each such term specifies a substring (between BegOffset and EndOffset; negative EndOffset stands for the end of string) to be replaced. SubstitutionList must be a list of atoms or character lists.

Offsets start from 0, as in C/Java.

This predicate replaces the substrings specified in SubstrList with the corresponding strings from SubstitutionList. The result is returned in OutStr. OutStr is a list of characters, if so is InputStr; otherwise, it is an atom.

If SubstitutionList is shorter than SubstrList then the last string in SubstitutionList is used for substituting the extra substrings specified in SubstitutionList. As a special case, this makes it possible to replace all specified substrings with a single string.

As in the case of re_substring/4, if OutStr is an atom, it is not interned. The user should either intern this string or convert it into a list, as explained previously.

The string_substitute/4 predicate always succeeds.

Here are some examples:

| ?- string_substitute('qaddf', [s(2,4)], ['123'], L).

L = qa123f

| ?- string_substitute('qaddf', [s(2,-1)], ['123'], L).

L = qa123

| ?- string_substitute("abcdefg", [s(4,-1)], ["123"], L).

L = [97,98,99,100,49,50,51]

| ?- string_substitute('1234567890123', [f(1,5),f(5,7),f(9,-2)], ["pppp", 111], X).

X = 1pppp11189111

| ?- string_substitute('1234567890123', [f(1,5),f(6,7),f(9,-2)], ["---"], X).

X = 1---6---89---
term_to_atom(+Term,-Atom,+Options)

Converts +Term to an atomic form according to a list of write options, Options, that are similar to those used by write_term/[2,3]. The various options of term_to_atom/[2,3] are especially useful for the interface from C to XSB (see Calling XSB from C in Volume 2 of this manual).

• quoted(+Bool). If Bool = true, then atoms and functors that can’t be read back by read/1 are quoted, if Bool = false, each atom and functor is written as its unquoted name. Default value is false.

• ignore_ops(+Bool). If Bool = true each compound term is output in functional notation; list braces are ignored, as are all explicitly defined operators. If Bool = canonical, bracketed list notation is used. Default value is canonical. The corresponding value of false, that would enable operator precedence, is not yet implemented.

• numbervars(+Bool). If Bool = true, a term of the form ’$VAR’(N) where N is an integer, is output as a variable name consisting of a capital letter possibly followed by an integer. A term of the form ’$VAR’(Atom) where Atom is an atom, is output as itself (without quotes). Finally, a term of the form ’$VAR’(String) where String is a character string, is output as the atom corresponding to this character string. If bool is false these cases are not treated in any special way. Default value is false.

Error Cases

• Options is a variable
  – instantiation_error

• Options neither a variable nor a list
  – type_error(list,Options)

• Options contains a variable element, O
  – instantiation_error

• Options contains an element O that is neither a variable nor a write option.
  – domain_error(write_option,O)

Examples:

| ?- term_to_atom(f(a,1,X,[’3cpio’,d(3),’$VAR’(”Foo”)],F,[]).
X = _h131
F = f(a,1,_h0,[3cpio,d(3),$VAR([70,111,111])])}
CHAPTER 6. STANDARD AND GENERAL PREDICATES

yes
| ?- term_to_atom(f(a,1,X,[‘3cpio’,d(3),’$VAR’("Foo")]),F,[numbervars(true)]).

X = _h131
F = f(a,1,_h0,[3cpio,d(3),Foo])

yes
| ?- term_to_atom(f(a,1,X,[‘3cpio’,d(3),’$VAR’("Foo")]),F,[numbervars(true),quoted(true)]).

X = _h131
F = f(a,1,_h0,[‘3cpio’,d(3),Foo])

yes
| ?- term_to_atom(f(a,1,X,[‘3cpio’,d(3),’$VAR’("Foo")]),F,[numbervars(true),quoted(true),ignore_ops(true)]).

X = _h131
F = f(a,1,_h0,’.’(‘3cpio’,’.’(d(3),’.’(Foo,[]))))

yes

term_to_atom(+Term,-Atom) module: string
This predicate converts an arbitrary Prolog term Term into an atom, putting the result in Atom. It is defined using the default options for term_to_atom/3, e.g. ignore_ops(canonical), quoted(false), and numbervars(false).

term_to_codes(+Term,-CodeList,+OptionList) module: string
This predicate is used in the definition of term_to_atom/3 but only converts a term into a list of UTF-8 codes, and does not intern the list as an atom. Allowed values for OptionList and error cases are the same as in term_to_atom/3.

term_to_codes(+Term,-CodeList) module: string
This predicate converts a term to a list of UTF-8 codes. It is defined using the default options for term_to_atom/3, e.g. ignore_ops(canonical), quoted(false), and numbervars(false).

gc_atoms
Explicitly invokes the garbage collector for atomsthatarecreated, but no longer needed. By default, gc_atoms/1 is called automatically, unless the Prolog_flag atom_garbage_collection is set to false, or if more than one thread is active. However there are reasons why a user may need to invoke atom table garbage collection. First, in Version 3.6, if atom table garbage collection is invoked automatically, it occurs periodically on heap garbage collection, or if numerous asserts and retracts have taken place. These heuristics overlook certain cases
where numerous atoms may be created without invoking the garbage collector – e.g. through repeated uses of \texttt{format_write_string/3}. In addition if user-defined C code contains pointers to XSB’s atom table, atom table garbage collection will be unsafe, as Version 3.6 of XSB does not detect such pointers in external code. In such cases, atom table garbage collection should be turned off via the Prolog flag \texttt{atom_garbage_collection}, and reinvoked at a point where the external pointers are no longer used.

### 6.10 All Solutions and Aggregate Predicates

Often there are many solutions to a problem and it is necessary somehow to compare these solutions with one another. The most general way of doing this is to collect all the solutions into a list, which may then be processed in any way desired. So XSB provides ISO-standard predicates such as \texttt{setof/3}, \texttt{bagof/3}, and \texttt{findall/3} to collect solutions into lists. Sometimes however, one wants simply to perform some aggregate operation over the set of solutions, for example to find the maximum or minimum of the set of solutions. XSB uses answer subsumption to produce a powerful aggregation facility as discussed in Section 5.4

\texttt{setof(\?Template, +Goal, \?Set)}

This predicate may be read as “\texttt{Set} is the set of all instances of \texttt{Template} such that \texttt{Goal} is provable”. If \texttt{Goal} is not provable, \texttt{setof/3} fails. The term \texttt{Goal} specifies a goal or goals as in \texttt{call(Goal)}. \texttt{Set} is a set of terms represented as a list of those terms, without duplicates, in the standard order for terms (see Section 6.6). If there are uninstantiated variables in \texttt{Goal} which do not also appear in \texttt{Template}, then a call to this evaluable predicate may backtrack, generating alternative values for \texttt{Set} corresponding to different instantiations of the free variables of \texttt{Goal}. Variables occurring in \texttt{Goal} will not be treated as free if they are explicitly bound within \texttt{Goal} by an existential quantifier. An existential quantification can be specified as:

\[ Y \ ^{\exists} \ G \]

meaning there exists a \texttt{Y} such that \texttt{G} is true, where \texttt{Y} is some Prolog term (usually, a variable).

Error cases are the same as predicate \texttt{call/1} (see Section 6.11).

Example: Consider the following predicate:
The goal ?- setof(Color,Height^Val^p(Color,Height,Val),List) returns a single solution:

    Color = _h73
    Height = _h87
    Val = _h101
    L = [black,blue,green,red]

If Height is removed from the sequence of existential variables, so that the goal becomes:

?- setof(Color,Val^p(Color,Height,Val),List)

the first solution is:

    Color = _h73
    Val = _h87
    Height = high
    L = [black,blue,red];

upon backtracking, a second solution is produced:

    Color = _h73
    Val = _h87
    Height = low
    L = [black,green]

\[\text{bagof(?Template, +Goal, ?Bag)}\]  ISO

This predicate has the same semantics as setof/3 except that the third argument returns an unsorted list that may contain duplicates.

Error Cases are the same as predicate call/1 (see Section 6.11).

Example: For the predicate \(p/3\) in the example for setof/3, the goal

?\(-\) bagof(Color,Height^Val^p(Color,Height,Val),L) returns the single solution:
CHAPTER 6. STANDARD AND GENERAL PREDICATES

\[ \text{Color} = _{h73} \]
\[ \text{Height} = _{h87} \]
\[ \text{Val} = _{h101} \]
\[ \text{L} = [\text{red,green,blue,black,black}]; \]

If \text{Height} is removed from the sequence of existential variables, so that the goal becomes: \( \text{?- bagof(Color,Val}^\text{p(Color,Height,Val),List}) \), the first solution is:

\[ \text{Color} = _{h73} \]
\[ \text{Val} = _{h87} \]
\[ \text{Height} = \text{high} \]
\[ \text{L} = [\text{red,blue,black}]; \]

upon backtracking, a second solution is produced:

\[ \text{Color} = _{h73} \]
\[ \text{Val} = _{h87} \]
\[ \text{Height} = \text{low} \]
\[ \text{L} = [\text{green,black}]; \]

\text{findall(?Template, +Goal, ?List)} \quad \text{ISO}

Similar to predicate \text{bagof/3}, except that variables in \text{Goal} that do not occur in \text{Template} are treated as existential, and alternative lists are not returned for different bindings of such variables. Note that this means that \text{Goal} should not contain existential variables. This makes \text{findall/3} deterministic (non-backtrackable). Unlike \text{setof/3} and \text{bagof/3}, if \text{Goal} is unsatisfiable, \text{findall/3} succeeds binding \text{List} to the empty list.

Error cases are the same as \text{call/1} (see Section 6.11).

Example: For the predicate \text{p/3} in the example for \text{setof/3}, the goal \text{findall(Color,p(Color,Height,Val),L)} returns a single solution:

\[ \text{Color} = _{h73} \]
\[ \text{Height} = _{h107} \]
\[ \text{Val} = _{h121} \]
\[ \text{F} = [\text{red,green,blue,black,black}] \]

\text{findall(?Template, +Goal, ?List,?Tail)}

Acts as \text{findall/3}, but returns the result as the difference-list \text{Bag-Tail}. In fact, the 3-argument version is defined in terms of the 4-argument version:
findall(Templ, Goal, Bag) :- findall(Templ, Goal, Bag, [])

Error cases are the same as findall/3 (or call/1).

tfindall(?Template, +Goal, ?List)  Tabling

Like findall/3, tfindall/3 treats all variables in Goal that do not occur in Template as existential. However, in tfindall/3, the Goal must be a call to a single tabled predicate.

tfindall/3 allows findall functionality to be used safely with tabling by throwing an error if it is called recursively. Its use can be seen by considering the following series of programs.

\[ p1(X) :\neg \text{findall}(Y, p1(Y), X). \]

When executing the goal \( p(X) \), XSB will throw an error when it reaches the maximum number of recursive invocations of findall.

Next, consider the program

\[
\text{:- table t/1.}
\text{t(X)} :\neg \text{findall}(Y, t(Y), X).
\text{t(a).}
\]

The query \( t(X) \) will terminate without error, but will return two answers: \( X = [\] \) and \( X = a \). These answers are hard to defend semantically, since there is an implicit domain closure axiom in findall-like predicates. On the other hand, for the program

\[
\text{:- table t2/1.}
\text{t2(X)} :\neg \text{tfindall}(Y, t2(Y), X).
\text{t2(a).}
\]

the query \( t2(X) \) will throw a table error, indicating that a call to tfindall/3 is apparently non-stratified footnote. Detection of non-stratification is based on the approximate detection of dependencies among subgoals maintained by XSB. This approximation is quite close for local evaluation, but is less close for batched evaluation. Other behavior for tabled aggregation is provided by answer subsumption as discussed in Section 5.4.

Other differences between predicates findall/3 and tfindall/3 can be seen from the following example:
| ?- [user].
| [Compiling user]
| :- table p/1.
p(a).
p(b).
| [user compiled, cpu time used: 0.639 seconds]
| [user loaded]

yes
| ?- p(X), findall(Y, p(Y), L).

X = a
Y = _922928
L = [a];

X = b
Y = _922820
L = [a,b];

no
| ?- abolish_all_tables.

yes
| ?- p(X), tfindall(Y, p(Y), L).

X = b
Y = _922820
L = [b,a];

X = a
Y = _922820
L = [b,a];

no

Error cases include those of findall/3 (see above), along with

**table_error** Upon execution Goal is not a subgoal of a tabled predicate.

**table_error** A call to tfindall/3 is apparently non-stratified

\[ X \subseteq \text{Goal} \]  ISO

Within setof/3, bagof/3 and the like, the \( \subseteq \) operator means there exists an \( X \) such that Goal is true.

excess_vars(+Term, +ExistVarTerm, +AddVarList, -VarList) module: setof

Returns in VarList the list of (free) variables found in Term concatenated to the
end of AddVarList. (In normal usage AddVarList is passed in as an empty list.) ExistVarTerm is a term containing variables assumed to be quantified in Term so none of these variables are returned in the resulting list (unless they are in AddVarList.) Subterms of Term of the form (VarTerm ^ SubTerm) are treated specially: all variables in VarTerm are assumed to be quantified in SubTerm, and so no occurrence of these variables in SubTerm is collected into the resulting list.

Error Cases

type_error AddVarList is not a list of variables
memory Not enough memory to collect the variables.

find_n(+N,?Template, +Goal, ?List) module: setof
Acts as findall/3 but returns only the first N bindings of Template to List.

6.11 Meta-Predicates

call(#X) ISO
If X is a non-variable term in the program text, then it is executed exactly as if X appeared in the program text instead of call(X), e.g.

... p(a), call( (q(X), r(Y)) ), s(X), ...

is equivalent to

... p(a), q(X), r(Y), s(X), ...

However, if X is a variable in the program text, then if at runtime X is instantiated to a term which would be acceptable as the body of a clause, the goal call(X) is executed as if that term appeared textually in place of the call(X), except that any cut (!) occurring in X will remove only those choice points in X. If X is not instantiated as described above, an error message is printed and call/1 fails.

Error Cases

instantiation_error X is a variable
type_error(callable,X) X is not callable.
#X
(where X is a variable) executes exactly the same as call(X). However, the explicit use of call/1 is considered better programming practice. The use of a top level variable subgoal elicits a warning from the compiler.

call(Goal, Arg,...) ISO call(Goal, Arg)
where Goal is an N-ary callable term first constructs a new N+1-ary term NewGoal with the same functor and first N arguments as Goal and with Arg as its N+1th argument, and then calls NewGoal. As an example,
call(member(X), [a, b, c])
is equivalent to call(member(X, [a, b, c]). Goal must be a callable term, but can be prepended by a module name using the :/2 symbol. call(Goal, Arg1, Arg2,...) will act similarly. Note that Goal should usually be atomic – if the outer functor of Goal is, say, ,/2, call/[2-10] will try to add the extra argument(s) to the comma functor, which is generally not the intended behavior.

While meta calls are generally fast in XSB, the extra term manipulation of call/[2-10] makes it somewhat slower than call/1.

call_tv(#Goal, -TV)
Tabling
calls Goal just as with call/1, and if Goal does not fail, instantiates TV with either true or undefined, depending on the truth value of Goal in the well-founded semantics. Goal need not be tabled itself. Note that Goal might succeed with truth value undefined before succeeding with truth value true, so that to avoid confusion, call_tv/2 should usually be invoked as a top-level goal.

Since call_tv/2 actually calls Goal, if Goal fails, call_tv/2 will fail; if Goal succeeds unconditionally (is true in the well-founded semantics), call_tv/2 will succeed unconditionally; and if if Goal succeeds conditionally (is neither true nor false in the well-founded semantics) call_tv/2 will itself succeed unconditionally.

An alternative approach would be to call get_residual(Goal, DelayList): Goal is true at a given state of an evaluation iff get_residual(Goal, DelayList) succeeds with DelayList = []. This approach either fails or succeeds with truth value true; however if Goal is not completed, calling get_residual(Goal, DelayList) may lead to confusing results.

Error cases are the same as call/1.

timed_call(#Goal, +Interval, #Handler, +Option)
timed_call(#Goal,+Interval,#Handler)

This predicate calls Goal and, if Goal is still being evaluated after Interval milliseconds, Goal will be interrupted and Handler executed. In the case where Handler succeeds or fails, the execution of Goal will be continued; if Handler throws an uncaught exception the execution of Goal may be aborted. In this way timed_call/3 can be used enforce a time-out on Goal.

Interval can be either a positive integer or the term repeating(Int) where Int is a non-negative integer. In this latter case, Goal is interrupted every Int milliseconds until it terminates (whether by normal termination or by Handler throwing an exception). In the case of repeated interrupts, the time taken to execute Handler is not counted as part of Interval milliseconds.

Nested calls to timed_call/3 are not allowed unless Option is set to nesting.

- If Option is not equal to nesting, or if timed_call/3 is used, then a nested call to timed_call/4 will throw a permission error.
- If Option is equal to nesting then the nested timed call is simply treated as a call to Goal: in other words the interval and handler for the nested call is ignored.

As an implementation note, timed_call/[3,4] is based on XSB’s internal interrupt mechanism, used for attributed variable handlers and thread signalling. As such, the ability to execute complex actions upon interrupt and to resume is very robust. However, checks for interrupts are only made whenever XSB’s SLG-WAM engine is executing. Because of this, if XSB is suspended on I/O, calling a C or java function, in a C-implemented builtin, or otherwise outside of its virtual machine, the interrupt will not be executed until computation is back within XSB’s virtual machine.

timed_call/3 is not yet implemented for the multi-threaded engine but its functionality is easily duplicated using thread signalling (Section 7.5).

Examples Consider the simple (and non-tabled) program fragment

\[\text{loop} :- \text{loop}.\]

which goes into an infinite loop on the query \(-\text{loop}\). However, the query

\[\text{timed_call(loop,repeating(500),abort)}.\]

will interrupt loop and abort its computation after 500 milliseconds. Alternatively, the query
timed_call(loop,500,statistics).

will interrupt the computation after 500 seconds, print out statistics, and resume
the computation where it left off.

timed_call(loop,repeating(500),statistics).

will interrupt the computation every 500 milliconds tp print statistics. More
sophisticated interrupt handlers could introspect a computation (e.g., using
statistics/2 or table_dump/[1,3]) and possibly modify parameters of the
computation when possible (e.g., by changing one form of tabling to another,
when permitted).

**Error Cases** Error cases are the same as in call/1 for the first and third
arguments of timed_call/3, along with these other errors.

- **Interval** is not an integer
  - type_error(integer,Interval)
- **Interval** is not a positive integer
  - domain_error(positive_integer,Interval)

A call C to timed_call/3 is made within the scope of some other call to
timed_call/3

- permission_error(nested_call,predicate,Goal)

**bounded_call**/3 is called from the multi-threaded engine

- misc_error

bounded_call(#Goal,+MaxMemory,+MaxCPU,#Handler) module: standard
bounded_call(#Goal,+MaxMemory,+MaxCPU) module: standard

These predicates call Goal and check once per second whether the total CPU
time to execute Goal is greater than MaxCPU seconds, and whether the total
memory taken by XSB is greater than MaxMemory bytes. Under bounded_call/4
if either of these conditions arise, Handler is called; under bounded_call/3 a
resource exception is thrown for memory or CPU time.

These predicates are implemented directly using timed_call/3 and inherit the
advantages and limitations of that predicate. As an advantage, the ability to
execute complex actions upon interrupt and to resume is very robust. How-
ever, checks for interrupts are only made whenever XSB’s SLG-WAM engine is
executing. Because of this, if XSB is suspended on I/O, calling a C or java func-
tion, in a C-implemented builtin, or otherwise outside of its virtual machine, the
interrupt will not be executed until computation is back within XSB’s virtual
machine.

Handler cannot cause \texttt{timed\_call/3} to be executed as a subgoal; but otherwise
Handler has no restrictions.

\texttt{bounded\_call/[3,4]} is not yet implemented for the multi-threaded engine but
its functionality is easily duplicated using thread signalling (Section\textsuperscript{7.5}).

**Error Cases** Error cases are the same as in \texttt{call/1} for the first argument of
\texttt{bounded\_call/3}, and are the same as that of \texttt{timed\_call} for Handler.

MaxCPU or MaxMemory is not an integer
- \texttt{type\_error(integer)}

MaxCPU or MaxMemory is not a positive integer
- \texttt{domain\_error(positive\_integer)}

\texttt{once(#X)} \hspace{1cm} \textit{ISO}
\texttt{once/1} is defined as \texttt{once(X):- call(X),!}. \texttt{once/1} should be used with care
in tabled programs. The compiler can not determine whether a tabled predicate
is called in the scope of \texttt{once/1}, and such a call may lead to runtime errors. If a
tabled predicate may occur in the scope of \texttt{once/1}, use \texttt{table\_once/1} instead.

Error cases are the same as \texttt{call/1}.

\texttt{forall(Generate, Test)}\hspace{1cm} \textit{Tabling}
\texttt{forall(Generate, Test)} is true iff for all possible bindings of Generate, the
goal Test is true. Procedurally, abstracting error checking, the predicate shall
behave as being defined by \texttt{\(+\) (call(Generator), \(+\) call(Test))}.

Error cases are the same as \texttt{call/1}.

\texttt{table\_once(#X)}
\texttt{table\_once/1} is a weaker form of \texttt{once/1}, suitable for situations in which a single
solution is desired for a subcomputation that may involve a call to a tabled
predicate. \texttt{table\_once(?Pred)} succeeds only once even if there are many solu-
tions to the subgoal \texttt{Pred}. However, it does not “cut over” the subcomputation
started by the subgoal \texttt{Pred}, thereby ensuring the correct evaluation of tabled
subgoals.

\texttt{call\_cleanup(#Goal,#Handler)}
\texttt{call\_cleanup(Goal, Cleanup)} calls Goal just as if it were called via \texttt{call/1},
but it is ensures that Handler will be called after Goal finishes execution.
\texttt{call\_cleanup/2} is thus useful when Goal uses a resource, (such as a stream,
mutex, database cursor, etc.) that should be released when Goal finishes execution.

More precisely, Goal finishes execution either 1) by failure, 2) by determining that the success of Goal is deterministic, 3) when an error is thrown and not handled by Goal or one of its subgoals; or 4) when Goal is cut over. In all of these cases, Handler will be called and will succeed non-deterministically. We illustrate these cases through examples.

- **Failure of Goal:**
  
  \[
  \text{?- call_cleanup(fail,writeln(failed(Goal))).}
  \]
  
  In this case, Goal has no solutions, and the handler is invoked when the engine backtracks out of Goal.

- **Deterministic success of Goal.** Assume that p(1) and p(2) have been asserted. Then
  
  \[
  \text{?- call_cleanup((p(X),writeln(got(p(X)))),writeln(handled(p(X)))),}\n  \]
  
  \[
  \text{got(p(1))}\n  \]
  
  \[
  X = 1;\n  \]
  
  \[
  \text{got(p(2))}\n  \]
  
  \[
  \text{handled(p(2))}\n  \]
  
  \[
  X = 2;\n  \]
  
  \[
  \text{no}\n  \]

  Note that Handler is called only after the last solution of the goal p(X) has been obtained. XSB decides to call Handler only when it can be determined that the success of Goal has left no choice points. In such a case, the final solution has been obtained for Goal. Of course, it may be that a solution S to Goal leaves a choice point but the choice point will produce no further solutions for Goal. XSB will not call Handler in this case, rather it will wait until there are no choice points left for Goal.

- **An uncaught error E is thrown out of Goal.** In this case, Handler will be called, and then, if E is uncaught, E will be rethrown. This is illustrated in the following example (Error handling is discussed further in Section 12.3.2):
  
  \[
  \text{?- catch(call_cleanup(throw(my_error),writeln(invoking_handler))),Ball,write(Ball).}\n  \]
  
  \[
  \text{invoking_handler}\n  \]
my_error
yes

Of course, Handler itself can be wrapped in a catch/3 so that any errors will be caught by call_cleanup/2.

• Choice points for Goal are removed via a cut. Consider an example in which p/1 has the same extension as above (p(1),p(2):

\[
\text{call_cleanup(p(X),writeln(handled_1)),!}.
\]
handled_1

X = 1

yes

The handler is invoked immediately when the choice point laid down by p(X) is cut over – before returning to the command line. If a cut cuts over more than goal to be cleaned, more than one handler will be executed:

\[
?-\text{call_cleanup(p(X),writeln(handled_4_1))},
\text{call_cleanup(p(Y),writeln(handled_4_2))},
\text{call_cleanup(p(Z),writeln(handled_4_3))},
!.
\]
handled_4_3
handled_4_2
handled_4_1

X = 1
Y = 1
Z = 1

call_cleanup/2 is thus an extremely powerful and flexible mechanism when used in a simple manner. While Handler is “guaranteed” to be invoked whenever Goal finishes execution\(^{13}\), it may be difficult to predict when Handler will be invoked, as Handler may be invoked because of deeply non-local cuts over Goal, and even when such cuts are not present, the invocation depends on XSB determining when the last solution for Goal has been obtained. Baroque usages, such as invoking call_cleanup/2 and cuts in the handler are supported, but may lead to code that is difficult to debug, since handlers may be invoked based on the state of XSB’s choice point stack.

\(^{13}\text{In fact we don’t guarantee anything, see XSB’s license.}\)
Error Cases

Goal is a variable
  • instantiation error

Goal is neither a variable nor a callable term
  • type error(callable, Goal)

Handler is a variable
  • instantiation error

Handler is neither a variable nor a callable term
  • type error(callable, Handler)

6.12 Information about the System State

Various aspects of the state of an instance of XSB — information about what predicates, modules, or dynamic clauses have been loaded, their object files, along with other information can be inspected in ways similar to many Prolog systems. However, because the atom-based module system of XSB may associate structures with particular modules, predicates are provided to inspect these elements as well. The following descriptions of state predicates use the terms predicate indicator, term indicator and current module to mean the following:

• By predicate indicator we mean a compound term of the form M:F/A or simply F/A. When the predicate indicator is fully instantiated, M and F are atoms representing the module name and the functor of the predicate respectively and A is a non negative integer representing its arity.
  Example: usermod:append/3

• By term indicator we mean a predicate or function symbol of arity N followed by a sequence of N variables (enclosed in parentheses if N is greater than zero). A term indicator may optionally be prefixed by the module name, thus it can be of the form M:Term.
  Example: usermod:append(_,_,_)

• A module M becomes a current (i.e. “known”) module as soon as it is loaded in the system or when another module that is loaded in the system imports some predicates from module M.
Note that due to the dynamic loading of XSB, a module can be current even if it has not been loaded, and that some predicates of that module may not be defined. In fact, a module can be current even if it does not exist. This situation occurs when a predicate is improperly imported from a non-existent module. Despite this, a module can never lose the property of being *current*.

**current_input(?Stream)**  
ISO  
Succeeds iff stream $\text{Stream}$ is the current input stream, or procedurally unifies $\text{Stream}$ with the current input stream.  

**Error Cases**  
- $\text{Stream}$ is neither a variable nor a stream identifier  
  - domain_error(stream_or_variable,Stream))

**current_output(?Stream)**  
ISO  
current_output/1 Succeeds iff stream $\text{Stream}$ is the current output stream, or procedurally unifies $\text{Stream}$ with the current output stream.  

**Error Cases**  
- $\text{Stream}$ is neither a variable nor a stream identifier  
  - domain_error(stream_or_variable,Stream))

**ISO Compatibility Note:** In XSB current_input/1 does not throw an error if $\text{Stream}$ is not a current input stream, but quietly fails instead.

**current_prolog_flag(?Flag_Name, ?Value)**  
ISO  
current_prolog_flag/2 allows the user to examine both dynamic aspects of XSB along with certain non-changeable ISO flags and non-changeable Prolog-commons flags. Calls to current_prolog_flag/2 will unify against ISO, Prolog-commons, and XSB-specific flags.  

ISO and Prolog-commons flags are as follows:

- **bounded** Indicates whether integers in XSB are bounded. This flag always has the value *true*
- **min_integer, max_integer** The minimum integer available in the current XSB configuration (differs between 32- and 64-bits).
- **max arity** Indicates the maximum arity of terms in XSB. This flag always has the value 255
- **integer_rounding_function** This flag always has the value *toward_zero*
• **debug** Indicates whether trace or debugging is turned on or off

• **unknown** Indicates the behavior taken when calling an unknown predicate. Values can be set to **fail**, **warning**, or **error**, indicating that calls to unknown predicates fail, produce a warning message to **user_warning** or throw an existence error. The default setting is **error**.

• **double_quotes** Indicates that double-quoted terms in XSB represent lists of character codes. Value is **codes**

• **dialect** indicates the implementation of Prolog that is running. Using this flag, applications intended to run on more than one Prolog can take actions that conditional on the executing Prolog. The value is **xsb**.

• **version_data** indicates the version of XSB that is running. Using this flag, applications intended to run on more than one Prolog can take actions that conditional on the executing Prolog. The value is **xsb**.

**ISO Compatibility Note:** The ISO flags **char_conversion** is not available – XSB does not use character conversion. XSB reads double quoted strings as lists of character codes, so that the value of the flag **double_quotes** is always **codes**, and this flag is not settable.

Non-standard flag names may be specific to XSB or may be common to XSB and certain other Prolog. These flag names are:

• **backtrace_on_error** on iff system-handled errors automatically print out the trace of the execution stack where the error arose, **off** otherwise. Default is **on**. In the multi-threaded engine, this flag is thread-specific and controls whether the backtrace for a current execution will be printed to **STDERR**.

• **dcg_style** the DCG style currently used; **xsb** or **standard** (standard is used in Quintus, SICSTUS, etc.). See Section 11.4 for more details. Default is **xsb**. This flag affects all threads in the process.

• **heap_garbage_collection** **indirection**, **none**, **sliding**, or **copying** depending on the heap garbage collection strategy that is currently being employed (see also Section 3.7). Default is **indirection**. This flag is private to each thread.

• **heap_margin** Specifies the size in **bytes** of the margin used to determine whether to perform heap garbage collection or reallocation of the environment stack. The default is 8192 (8K) bytes for 32-bit platforms 16384 (16K) for 64-bit platforms. Setting this field to a large value (e.g. in the
megabyte range) can cause XSB to be more aggressive in terms of expanding heap and local stack and to do fewer heap garbage collections than with the default value. However heap_margin should not be set lower than its default, as this may prevent XSB from properly creating large terms on the heap.

- clause_garbage_collection on if garbage collection for retracted clauses is allowed, and off otherwise. Default is on. This flag is private to each thread.

- atom_garbage_collection on if garbage collection for atomic constants is allowed, and off otherwise. Default is on. This flag is global for all threads (currently, string garbage collection will only be invoked if there is a single active thread.)

- table_gc_action The setting abolish_tables_transitively causes predicates or subgoals that depend on a conditional answer of an abolished table to be abolished automatically; the setting abolish_tables_singly not does not cause this action. The distinction is important, since if table $T_1$ depends on table $T_2$, and $T_2$ is abolished but $T_1$ is not, then predicates that introspect the dependencies of $T_1$ could cause memory violations (e.g., get_residual/2). Default is abolish_tables_transitively. This flag affects all threads in the process.

- goal the goal passed to XSB on command line with the ‘-e’ switch; ‘true.’ if nothing is passed. This flag may be examined, but not set.

- tracing on iff trace mode is on; off otherwise. This flag affects all threads in the process.

- write_depth The depth to which a term is written by write-like predicates. Default is 64. This flag affects all threads in the process.

- warning_action The action to take on warnings: the default value print_warning prints a warning message to the XSB STDWARN stream when warning/1 is called; silent_warning silently succeeds when warning/1 is called; and error_warning/1 throws a miscellaneous exception.

- write_attributes Determines the action to take by write/1 when it writes an attributed variable. By default write/1 portrays attributed variables using module-specific routines (cf. Volume 2 of this manual) as Variable{Module : PA_Output} where PA_Output is the output of the portray_attributes/2 clause for Module. However the value ignore causes an attributed variable to be written simply as a variable; and dots causes Variable{< module_name > : ...} to be written. Finally, the value
write causes a variables attribute to be written as a term \(^{14}\). The default behavior is set to the value portray.

- **max_table_subgoal_action** The action to take when a tabled subgoal of maximum depth is encountered. To understand the use of this flag, consider that if a predicate such as

\[
p(X) :- p(f(X)) .
\]

is tabled, it can (semantically) create subgoals of infinite depth. When the maximum subgoal depth is reached, XSB can either throw a miscellaneous error (the default action); or XSB can fail – an action that may be valid for certain programs. The action is set to fail by the value failure while the action of throwing an error can be (re-)set using the value error.

- **max_table_subgoal_depth** The maximum depth of a subgoal argument that can be added to a table: when the depth is reached, an action is taken as indicated for the previous flag. The default value is maximum_integer.

- **max_table_answer_action** The action to take when a tabled answer of maximum depth is encountered. To understand the use of this flag, consider the program fragment:

\[
:- table p/1 .
p(f(X)) :- p(X) .
p(a) .
\]

is tabled, the model for the goal \(?- p(X)\) is infinite, so that this program will not terminate. When the maximum answer depth is reached, XSB can either

1. Throw a miscellaneous error, set using the value error. This is the default action.
2. Apply bounded rationality by abstracting the answer with a truth value of undefined; settable using the value bounded_rationality.
3. Issue a warning, settable using the value warning.
4. Fail the answer addition, an action that may be valid for certain programs, settable using the value fail.

*Note that this flag affects only structures that are not lists (since large lists are more common than other large structures).*

- **max_table_answer_depth** The maximum depth of an answer argument that can be added to a table: when the depth is reached, an action is taken as indicated for the previous flag. The default value is maximum_integer.

\[^{14}\text{When writing an attribute, any attributed variables in the attribute itself are written just as variables with their attributes ignored.}\]
Note that this flag affects only structures that are not lists (since large lists are more common than other large structures).

- **max_table_answer_list_action** The action to take when a tabled answer of maximum list depth is encountered. To understand the use of this flag, consider the program fragment:

  ```prolog
  :- table l/1.
  l([a|X]):- l(X). l([a]).
  ```

  is tabled, the model for the goal `?- l(X)` is infinite, so that this program will not terminate. When the maximum answer list depth is reached, XSB can either throw a miscellaneous error (the default action); emit a warning; or XSB can fail – an action that may be valid for certain programs. The action is set to fail by the value `failure` while the action of throwing an error can be (re-)set using the value `error`, and the action of warning is set by the value `warning`.

- **max_table_answer_list_depth** The maximum list depth of an answer argument that can be added to a table: when the depth is reached, an action is taken as indicated for the previous flag. The default value is `maximum_integer`.

  Note that this flag affects only structures that are lists (since large lists are more common than other large structures).

- **max_memory** The maximum amount of memory in kilobytes that an XSB thread (in the single-threaded engine) or all XSB threads (in the multi-threaded engine) can use for their combined execution stacks, program space, tables, or any other purpose. If a query exceeds this amount, XSB will abort the query with a resource exception and then try to reclaim space used by the query. As with other flags, this flag can be set during an XSB session. The value of 0 effectively disables the flag, allowing XSB to allocate as much memory as the underlying OS will grant. The default value is 0, so that the flag is disabled.

- **unify_with_occurs_check** If set to on, perform all unification using an occurs check, which makes unification mathematically correct, but computationally complex. Without the occurs check, the unification

\[ X = f(X) \]

\[ ^{15} \text{Failure in this case can be seen as an implicit form of answer abstraction.} \]
will produce a cyclic term \( X = f(f(f(f(\ldots)))) \); with the occurs check this unification will fail. Setting the flag to on may slow down programs, perhaps drastically, and may be incompatible with some constraint libraries such as CHR. An alternate to this flag is the ISO predicate unify_with_occurs_check/2: see Section 6.8 for further discussion. The default for this flag is off.

- **character_set** If set to utf_8, interprets input/output byte sequences as UTF-8 encodings of unicode code points; if set to cp1252 then interprets bytes using the Windows Code Page 1252; if set to latin_1, then input/output bytes are interpreted as directly representing unicode code points. Default for UNIX-style systems is utf_8 and for Windows-style systems is cp1252, but the flag (and character sets) may be changed at any time. (See section “Character Sets in XSB” in XSB User Manual Volume 2 for more details.)

- **errors_with_position** If set to on, then the Prolog read predicates, when they encounter a syntax error in the term being read, will throw a syntax error which contains a pair ErrorMessage-ErrorPosition. ErrorPosition is an integer indicating the position in the file at which the syntax error was detected. If set to off, then the read predicates will simply throw the syntax error message.

- **exception_action** If set to iso then ISO-style exceptions will be thrown whenever an error condition arises. However, if exception_action is set to undefined_truth_value then certain goals will succeed with an undefined truth value rather than throwing an error. When this occurs, a literal is added to the delay list of the current evaluation. Later, it can be determined if an undefined answer depends on an exceptional condition through explain_u_val/[3,6], get_residual_sccs/[3,5] or via a justification system that depends on these predicates. The default for this flag is iso.

- **exception_pre_action** If set to print_incomplete_tables, then the predicate print_incomplete_tables/0 is called before throwing an exception. The execution of print_incomplete_tables/0 causes the stack of incomplete tables to be printed to a temporary file in $XSBDIR/etc. The file can be obtained via the predicate get_scc_dumpfile/1; information in the file can be used to help understand the context in which the exception arose. The file will be created only if an exception is thrown over at least one incomplete table. The default for this flag is off.

- **max_tab_usage** If set to on, maintains the maximal table usage (in bytes) for display in statistics/[0,1]. This information can be useful if a
program performs various types of table abolishes. Setting this flag to on may slightly slow down computation. Default is off.

The following flags affect only the multi-threaded engine.

- **thread_gls** In the multi-threaded engine, the initial size, in kbytes, of the global and local stack area of a newly created thread if no such option is explicitly passed. By default this is 768 (or 1536 for 64-bit configurations), or whatever was passed in if the command-line option `-m` was used, but that value may be modified at any time by resetting the flag. This flag affects a thread created by any thread in the process.

- **thread_tcpsize** In the multi-threaded engine, the initial size, in kbytes, of the trail and choice point area of a newly created thread if no such option is explicitly passed. By default this is 768 (or 1536 for 64-bit configurations), or whatever was passed in if the command-line option `-c` was used, but that value may be modified at any time by resetting the flag. This flag affects a thread created by any thread in the process.

- **thread_complsize** In the multi-threaded engine, the initial size, in kbytes, of the completion stack area of a newly created thread if no such option is explicitly passed. By default this is 64 (or 128 for 64-bit configurations), or whatever was passed in if the command-line option `-0` was used, but that value may be modified at any time by resetting the flag. This flag affects a thread created by any thread in the process.

- **thread_pdlsize** In the multi-threaded engine, the initial size, in kbytes, of the unification stack area of a newly created thread if no such option is explicitly passed. By default this is 64 (or 128 for 64-bit configurations), or whatever was passed in if the command-line option `-m` was used, but that value may be modified at any time by resetting the flag. This flag affects a thread created by any thread in the process.

- **thread_detached** In the multi-threaded engine, this specifies whether threads are to be created as detached or joinable if no explicit option is passed. A value of `true` indicates that threads are to be created as detached, and `false` as joinable. If this flag is not set, its default is `false`.

- **max_threads** In the multi-threaded engine, the maximum number of valid threads. By default this is 1024 and this value may not be reset at runtime, but it may be set by the command-line option `-max_threads`. This option is settable only by a command-line argument, and has no effect in the single-threaded engine.
• **max_queue_size** In the multi-threaded engine, the default maximum number of terms a message queue contains before writes to the message queue block. By default this is 1000. If set to 0, queues by default will be unbounded. This option has no effect in the single-threaded engine.

• **shared_predicates** In the multi-threaded engine, indicates whether predicates are considered thread-shared by default – that is, whether tables or dynamic predicates are shared among threads. By default this is false, and predicates are considered thread-private by default. This option is settable only by a command-line argument, and has no effect in the single-threaded engine.

Note that the non-standard flags are used only for dynamic XSB settings, *i.e.*, settings that might change between sessions (via command line arguments) or within the same session (via modifiable flags). For static configuration information, the predicate `xsb_configuration/2` is used. `xsb_configuration/2`.

**Error Cases**

- **Flag_Name** is neither a variable nor an atom.
  - `domain_error(atom_or_variable,Flag_Name)`

```
set_prolog_flag(?Flag_Name, ?Value)
```

ISO

`set_prolog_flag/2` allows the user to change settable prolog flags. Currently the only settable ISO flag is the `unknown` flag. Setting the flag `unknown` to `fail` results in calls to undefined predicates to quietly fail. Setting it to `warning` causes calls to undefined predicates to generate a warning (to `STDWARN`) and then fail. Setting it to `error` (the default) causes calls to undefined predicates to throw an existence error.

Dynamic XSB settings can also be changed, as described in `current_prolog_flag/2`.

**Error Cases**

- **Flag_Name** or **Value** is a variable.
  - `instantiation_error`
  - **Flag_Name** is not the name of a recognized Prolog flag.
  - `domain_error(prolog_flag,Flag_Name)`

```
current_predicate(?Predicate_Indicator)
```

ISO

`current_predicate/1` can be used to backtrack through indicators for loaded user or system predicates. If `Predicate_Indicator` unifies with `Module:F/A` all loaded predicates unifying with this indicator is returned. If `Predicate_indicator`
is F/A. current_predicate/1 behaves as if it were called with the form usermod:F/A. Unlike current_functor/1, current_predicate/1 does not return indicators for predicates that have been imported but not actually loaded into code space. For more detailed analysis of predicate properties, the predicate predicate_property/2 can be used.

As an example to backtrack through all of the predicates defined and loaded in module blah, regardless of whether blah is a system or a user defined module, use:

\[
| ?- current_predicate(blah:Predicate).
\]

In this case Predicate will have the form: Functor/Arity.

To backtrack through all predicates defined and loaded in any current module, use:

\[
| ?- current_predicate(Module:Functor/Arity).
\]

This succeeds once for every predicate that is loaded in XSB’s database.

To find the predicates having arity 3 that are loaded in usermod, use:

\[
| ?- current_predicate(usermod:Functor/3).
\]

while to find all predicates loaded in the global modules of the system regardless of their arity, use:

\[
| ?- current_predicate(usermod:Predicate).
\]

**Error Cases**

- Predicate_indicator is neither a variable nor a predicate indicator
  - type_error(predicate_indicator,Predicate_indicator))

**ISO Compatibility Note:** In XSB, current_predicate will backtrack through system predicates as well as user predicates.

**current_module(??Module)**

The standard predicate current_module/1 allows the user to check whether a given module is *current* or to generate (through backtracking) all currently known modules. Succeeds iff Module is one of the modules in the database. This includes both user modules and system modules. For more detailed analysis of module properties, the predicate module_property/2 can be used.

Note that predicate current_module/1 succeeds for a given module even if that module does not export any predicates. There are no error conditions
associated with this predicate; if its argument does not unify with one of the current modules, `current_module/1` simply fails.

`current_module(?Module, ?ObjectFile)`
Predicate `current_module/2` gives the relationship between the modules and their associated object file names. The file name `ObjectFile` must be absolute and end with the object file extension for the system (by default, `.xwam`). It is possible for a current module to have no associated file name (as is the case for "usermod"), or for the system to be unable to determine the file name of a current module. In both cases, predicate `current_module/1` will succeed for this module, while `current_module/2` will fail. The system is unable to determine the file name of a given module if that module is not in one of the directories of the search path (see Section 3.6). Once again, there are no error conditions associated with this predicate; if the arguments of `current_module/2` are not correct, or `Module` has no associated `File`, the predicate will simply fail.

`current_functor(?Predicate_Indicator)`
`current_predicate/1` can be used to backtrack through indicators for all non-atomic terms occurring in loaded modules. If `Predicate_Indicator` unifies with `Module:F/A` all term indicators unifying with `F/A` in a module unifying with `Module` are returned. If `Predicate_Indicator` is `F/A`, `current_predicate/1` behaves as if it were called with the form `usermod:F/A`. Unlike `current_predicate/1`, `current_functor/1` returns not only structures occurring in predicates but predicates that are imported into loaded modules but are not yet themselves loaded.

As an example, to backtrack through all of the functors of positive arity (function and predicate symbols) that appear in the global modules of the system regardless of whether they are system or a user defined, use:

```
| ?- current_functor(Functor/Arity), Arity > 0.
```

There are no error conditions associated with this predicate; if its argument is not a predicate indicator the predicate simply fails.

`current_index(Functor/Arity,IndexSpec)`
XSB has a variety of ways to index dynamic predicate including alternate argument indexing, multiple argument indexing, star-indexing, and tries, as discussed in Section 6.14. In addition XSB allows a choice of which argument to index for compiled predicates as well. `current_index/2` returns the index specification for each functor/arity pair unifying with `Functor/Arity` and visible from the calling context of `current_index/2`.  

---

**CHAPTER 6. STANDARD AND GENERAL PREDICATES**

---

245
current_atom(?Atom_Indicator)
Generates (through backtracking) all currently known atoms, and unifies each
in turn with Atom_Indicator.

predicate_property(?Term_Indicator, ?Property)
The standard predicate predicate_property/2 can be used to find the prop-
eties of any predicate that is visible to a particular module. Succeeds iff
Term_Indicator is a term indicator for a current predicate whose principal
functor is a predicate having Property as one of its properties. Or procedu-
really, Property is unified with the currently known properties of the predicate
having Term_Indicator as its skeletal specification.

A brief description of predicate_property/2 is as follows:

- If Term_Indicator is not a variable, and is a structure or atom, then
  Property is successively unified with the various properties associated
  with Term_Indicator. If Term_Indicator is not a known to the sys-
tem, the call succeeds with Property successively unified to exported and
  unclassified. These properties can be considered as a default for any
  structure or atom.
- If Property is bound to a valid predicate property, then predicate_property/2
  successively unifies Term_Indicator with the skeletal specifications of all
  predicates known to the system having the specified Property.
- If Term_Indicator is a variable, then it is unified (successively through
  backtracking) with the most general term for a predicate whose known
  properties are unified with Property.
- If Term_Indicator is not a term indicator, or if Property is not a valid
  predicate property, the call fails.

For example, all the loaded predicate skeletal specifications in module "usermod"
may be enumerated using:

    | ?- predicate_property(Pred, loaded).

Also the following query finds all predicate skeletal specifications that are ex-
ported by module blah:

    | ?- predicate_property(blah:Pred, exported).

Currently, the following properties are associated with predicates either implicit-
ly or by declaration. Double lines show property categories, and a predicate
can have at most one property of each category.

- Execution Type which is one of
– unclassified The predicate symbol is not yet classified according to this
category. This property has various meanings. Usually for exported
predicate symbols in system or user defined modules it means that the
predicate is yet unloaded (because it has not been used). In usermod
it usually means that the predicate is either a function symbol, or an
unloaded predicate symbol (including constants).

– dynamic The predicate is dynamic.

– loaded The predicate (including internal predicates) is a Prolog pred-
icate loaded into the module in question; this is always the case for
predicates in usermod.

– unloaded The predicate is yet unloaded into the module in question.

– foreign The predicate is a foreign predicate. This implies that the
predicate is already loaded in the system, because currently there is
no way for XSB to know that a predicate is a foreign predicate until
it is loaded in the system.

• Visibility Type which can be one of

  – exported The predicate symbol is exported by the module in question;
in other words the predicate symbol is visible to any other module in
the system.

  – local The predicate symbol is local to the module in question.

  – imported_from(Mod) The predicate symbol is imported into the mod-
ule in question from module Mod.

• Tabling Call Behavior which can be one of

  – tabled(variant) The predicate has been declared tabled and to use
call variance.

  – tabled(subsumptive) The predicate has been declared tabled and to
use call subsumption

  – tabled(default) The predicate has been declared tabled and to use
the default tabling strategy of the session, which can be either call
variance or call subsumption.

• Incremental Tabling Behavior which can be one of

  – incremental The predicate was declared as either incremental dy-
namic or as incremental tabled; or

  – opaque The predicate was declared as opaque to incremental updates.

• spied The predicate symbol has been declared spied (either conditionally
or unconditionally).
• **shared** The predicate has been declared shared in the multi-threaded engine. This means that any dynamic code or tables for this predicate will be shared among threads, but it does not affect static, non-tabled code.

• **built_in** The predicate symbol has the same Functor and Arity as one of XSB’s standard predicates, and is available to the user without needing to load a file or import the predicate from a module.

• **meta_predicate(Template)** The predicate is a meta-predicate. This property provides compatibility with other Prolog compilers and with forthcoming ISO Prolog standards.

Finally, since **dynamic** is usually declared as an operator with precedence greater than 999, writing the following:

```prolog
| ?- predicate_property(X, dynamic).
```

will cause a syntax error. The way to achieve the desired result is to parenthesize the operator like in:

```prolog
| ?- predicate_property(X, (dynamic)).
```

**module_property(?Module, ?Property)**

The standard predicate **module_property/2** can be used to find the properties of any current module. Succeeds iff **Module** is the name of a current module having **Property** as one of its properties. Or procedurally, **Property** is unified with the currently known properties of the module having **Module** as its name.

Currently, the following properties are associated with modules implicitly:

<table>
<thead>
<tr>
<th><strong>Property</strong></th>
<th><strong>Explanation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>unloaded</td>
<td>The module (including system modules) though it is current, is yet unloaded in the system.</td>
</tr>
<tr>
<td>loaded</td>
<td>The module (including system modules) is loaded in the system; this is always the case for <strong>usermod</strong>.</td>
</tr>
</tbody>
</table>

**listing**

Lists in the current output stream the clauses for all dynamic predicates found in module **usermod**. Note that **listing/0** does not list any compiled predicates unless they have the **dynamic** property (see **predicate_property/2**). A predicate gets the **dynamic** property when it is explicitly declared as **dynamic**, or automatically acquires it when some clauses for that predicate are asserted in the database. In cases where a predicate was compiled but converted to **dynamic** by asserting additional clauses for that predicate, **listing/0** will just display...
an indication that there exist compiled clauses for that predicate and only the
dynamically created clauses of the predicate will be listed. For example:

```
| ?- [user].
[Compiling user]
a(X) :- b(X).
a(1).
[user compiled, cpu time used: 0.3 seconds]
[user loaded]

yes
| ?- assert(a(3)).

yes
| ?- listing.

a(A) :-
 $compiled.
a(3).

yes
```

Predicate listing/0 always succeeds. The query:
```
| ?- listing.
```

is just a notational shorthand for the query:
```
| ?- listing(X).
```

`listing(+Predicate_Indicator)`
If `Predicate_Indicator` is a variable then `listing/1` is equivalent to `listing/0`. If `Predicate_Indicator` is an atom, then `listing/1` lists the dynamic clauses for all predicates of that name found in module `usermod` of the database. The argument `Predicate_Indicator` can also be a predicate indicator of the form `Name/Arity` in which case only the clauses for the specified predicate are listed. Finally, it is possible for `Predicate_Indicator` to be a list of predicate indicators and/or atoms; e.g.
```
| ?- listing([foo/2, bar, blah/4]).
```

If `Predicate_Indicator` is not a variable, an atom or a predicate indicator (or list of predicate indicators) of the form `Name/Arity`, predicate `listing/1` will simply fail.
In future releases of XSB, we intend to allow the user to specify a predicate indicator of the form \texttt{Module:Name/Arity} as argument of \texttt{listing/1}.

\texttt{xsb\_configuration(Feature\_Name, \textquote{Value})}

Succeeds iff the current value of the XSB feature \texttt{Feature\_Name} is \texttt{Value}.

This predicate provides information on a wide variety of features related to how XSB was built, including the compiler used, the compiler and loader flags, the machine and OS on which XSB was built, the release number, the various directories that XSB uses to find its libraries, etc.

To find all features and their values, ask the following query:

\begin{verbatim}
| ?- xsb\_configuration(FeatureName, Value), fail.
\end{verbatim}

Here is how \texttt{xsb\_configuration} might look like:

\begin{verbatim}
xsb\_configuration(architecture, \textquote{i386-apple-darwin8.9.1}).
%% configuration is usually the same as architecture, but it can also
%% contain special tags, \{it e.g.\}, i386-apple-darwin8.9.1-dbg, for a version
%% built with debugging enabled.
xsb\_configuration(configuration, \textquote{i386-apple-darwin8.9.1-dbg}).
xsb\_configuration(host\_os, \textquote{darwin8.9.1}).
xsb\_configuration(os\_version, \textquote{8.9.1}).
xsb\_configuration(os\_type, \textquote{darwin}).
xsb\_configuration(host\_vendor, \textquote{apple}).
xsb\_configuration(host\_cpu, \textquote{i386}).
xsb\_configuration(compiler, \textquote{gcc}).
xsb\_configuration(compiler\_flags, \textquote{-faltivec -fPOC -Wall -pipe -g}).
xsb\_configuration(loader\_flags, \textquote{-g -lm '}).
xsb\_configuration(compile\_mode, \textquote{debug}).
%% The type of XSB engine configured.
xsb\_configuration(scheduling\_strategy, \textquote{(local)}).
xsb\_configuration(engine\_mode, \textquote{slg-wam}).
xsb\_configuration(word\_size, \textquote{32}).
%% The following is XSB release information
xsb\_configuration(major\_version, \textquote{3}).
xsb\_configuration(minor\_version, \textquote{3}).
xsb\_configuration(patch\_version, \textquote{1}).
xsb\_configuration(beta\_version, \textquote{''}).
xsb\_configuration(version, \textquote{3.3.1}).
xsb\_configuration(codename, \textquote{Pignoletto}).
xsb\_configuration(release\_date, date(2011, 04, 12)).
%% Support for other languages
\end{verbatim}
xsb_configuration(perl_support, 'yes').
xsb_configuration(perl_archlib, '/usr/lib/perl5/i386-linux/5.00404').
xsb_configuration(perl_cc_compiler, 'cc').
xsb_configuration(perl_ccflags, '-Dbool=char -DHAS_BOOL -I/usr/local/include').
xsb_configuration(perl_libs, '-lnsl -lndbm -lgdbm -ldeb -ldl -lm -lc -lposix -lcrypt').
xsb_configuration(javac, '/usr/bin/javac').
/* Tells where XSB is currently residing; can be moved */
xsb_configuration(install_dir, InstallDir) :- .../* User home directory. Usually HOME. If that is null, then it would
be the directory where XSB is currently residing.
This is where we expect to find the .xsb directory */
xsb_configuration(user_home, Home) :- .../* Where XSB invocation script is residing */
xsb_configuration(scriptdir, ScriptDir) :- .../* where are cmplib, syslib, lib, packages, etc live */
xsb_configuration(cmplibdir, CmplibDir) :- ...xsb_configuration(libdir, LibDir) :- ...
xsb_configuration(syslibdir, SyslibDir) :- ...
xsb_configuration(packagesdir, PackDir) :- ...
xsb_configuration(etcdir, EtcDir) :- ...
/* architecture and configuration specific directories */
xsb_configuration(config_dir, ConfigDir) :- ...
xsb_configuration(config_libdir, ConfigLibdir) :- ...
/* site-specific directories */
xsb_configuration(site_dir, '/usr/local/XSB/site').
xsb_configuration(site_libdir, SiteLibdir) :- ...
/* site and configuration-specific directories */
xsb_configuration(site_config_dir, SiteConfigDir) :- ...
xsb_configuration(site_config_libdir, SiteConfigLibdir) :- ...
/* Where user’s arch-specific libraries are found by default. */
xsb_configuration(user_config_libdir, UserConfigLibdir) :- ...

hilog_symbol(?Symbol)
Succeeds iff Symbol has been declared as a HiLog symbol, or procedurally uni-
ifies Symbol with one of the currently known (because of a prior declaration)
HiLog symbols. The HiLog symbols are always atoms, but if the argument of
hilog_symbol, though instantiated, is not an atom the predicate simply fails.
So, one can enumerate all the HiLog symbols by using the following query:
| ?- hilog_symbol(X).

current_op(?Precedence, ?Specifier, ?Name)
ISO
This predicate is used to examine the set of operators currently in force. It
succeeds when the atom Name is currently an operator of type Specifier and precedence Precedence. None of the arguments of current_op/3 need to be instantiated at the time of the call, but if they are, they must be of the following types:

Precedence must be an integer in the range from 1 to 1200.
Specifier must be one of the atoms:
   xfx xfy yfx fx fy hx hy xf yf
Name it must be an atom.

Error Cases

• Precedence is neither a variable nor an integer in the range from 1 to 1200.
  – domain_error(operator_priority,Precedence)
• Specifier is neither a variable nor an operator specifier of the types above.
  – domain_error(operator_specifier,Specifier)
• Name is neither a variable nor an atom.
  – domain_error(atom_or_variable,Name)

hilog_op(?Precedence, ?Type, ?Name)
This predicate has exactly the same behaviour as current_op/3 with the only difference that Type can only have the values hx and hy.

6.13 Execution State

break
Causes the current execution to be suspended at the beginning of the next call. The interpreter then enters break level 1 and is ready to accept input as if it were at top level. If another call to break/0 is encountered, it moves up to break level 2, and so on. While execution is done at break level $n > 0$ the prompt changes to $n:$ ?-.

To close a break level and resume the suspended execution, the user can type the the atom end_of_file or the end-of-file character applicable on the system (usually CTRL-d on UNIX systems). Predicate break/0 then succeeds (note in the following example that the calls to break/0 do not succeed), and the
execution of the interrupted program is resumed. Alternatively, the suspended execution can be abandoned by calling the standard predicate \texttt{abort/0}, which causes a return to the top level.

An example of \texttt{break/0} ’s use is the following:

```
| ?- break.
[ Break (level 1) ]
1: ?- break.
[ Break (level 2) ]
[ End break (level 2) ]
```

```
yes
1: ?-
```

Entering a break closes all incomplete tables (those which may not have a complete set of answers). Closed tables are unaffected, even if the tables were created during the computation for which the break was entered.

\textbf{halt}

\texttt{halt/0} Exits the XSB session regardless of the break level. On exiting the system cpu and elapsed time information is displayed.

\textbf{halt(Code)}

\texttt{halt/1} Exits the XSB session regardless of the break level, sending the integer \texttt{Code} to the parent process. Normally 0 is considered to indicate normal termination, while other exit codes are used to report various degrees of abnormality.

\textbf{Error Cases}

- \texttt{Code} is not an integer
  - \texttt{type_error(Integer,Code)}

\textbf{prompt(+NewPrompt, ?OldPrompt)}

Sets the prompt of the top level interpreter to \texttt{NewPrompt} and returns the old prompt in \texttt{OldPrompt}.

An example of \texttt{prompt/2} ’s use is the following:
| ?- prompt('Yes master > ', P).

P = | ?- ;

no
Yes master > fail.

no
Yes master >

\textbf{trimcore} \hspace{1cm} \texttt{module: machine}

A call to \texttt{trimcore/0} reallocates an XSB thread’s execution stacks (and some tabling stacks) to their initial allocation size, the action affecting only the memory areas for the calling thread. When XSB is called in standalone or server mode, \texttt{trimcore/0} is automatically called when the top interpreter level is reached. When XSB is embedded in a process, \texttt{trimcore/0} is called at the top interpreter level for any thread created through \texttt{xsb_ccall_thread_create()} (see Volume 2, Chapter 3 \textit{Embedding XSB in a Process}).

\textbf{gc_heap}

Explicitly invokes the garbage collector for a thread’s heap. By default, heap garbage collection is called automatically for each thread upon stack expansion, unless the Prolog flag \texttt{heap_garbage_collection} is set to \texttt{none}. Automatic heap garbage collection should rarely need to be turned off, and should rarely need to be invoked manually.

\textbf{statistics}

Displays usage information on the current output stream, including:

- Process-level information about allocated memory excluding execution stacks but including:
  - \texttt{atoms} Space used to maintain global information about predicates and structures.
  - \texttt{string} Space used to maintain information about atomic constants in XSB.
  - \texttt{asserted} Space allocated for dynamic code.
  - \texttt{asserted} Space allocated for static code.
  - \texttt{foreign} Space allocated for foreign predicates.
  - \texttt{table} Space allocated for XSB’s tables.
- **findall** Space allocated for buffers to support findall/3 and similar predicates.
- **mt-private** Private space used by threads.
- **profiling** Space used to maintain profiling information, if XSB is called with profiling on.
- **gc temp** Temporary space for used for heap garbage collector.
- **interprolog** Space allocated for the Interprolog XSB/Java interface.
- **thread** Space allocated for the thread table
- The space occupied by subgoal and answer tables (in the form of tries) [58, 18, 37]. In the multi-threaded configuration process level table space includes shared tables but not private tables.

- Thread-specific information about allocation of memory for the calling thread including the
  - Global stack (heap) and local (environment) stack (see e.g. [1]) for the calling thread. Memory for these two WAM stacks is allocated as a single unit so that each stack grows together; information is provided on the current allocation for the stacks as well as on the stack sizes themselves. (See Section 3.7 for the memory re-allocation algorithm).
  - Trail and choice point stack (see e.g. [1]) for the calling thread. Memory for these two WAM stacks is allocated as a single unit so that each stack grows together; information is provided on the current allocation for the stacks as well as on the stack sizes themselves. The (re-)allocation follows the algorithm sketched in Section 3.7. (See Section 3.7 for the memory re-allocation algorithm).
  - SLG unification stack for the calling thread. This stack is used as a space to copy terms from the execution stacks into table space, or back out. This stack will not be reallocated unless extremely large terms are tabled.
  - SLG completion stack for the calling thread. The completion stack is used to perform incremental completion for sets of mutually dependent tabled subgoals. One completion stack frame is allocated per tabled subgoal [62] but the size of these frames is version-dependent.
  - The space occupied by private subgoal and answer tables for the calling thread.

In XSB’s single-threaded configuration, maximum space used by each of will be output if the ’-s’ command-line option is used

- Information about the number of tabling operations performed in the session by any thread. Note that the statistics are divided up between calls to
predicates that use variant tabling and those that use (call) subsumptive tabling (see Section 5.2.1 and [37]).

- Call Subsumption Subgoal Operations. For predicates that use subsumptive tabling, the total number of subsumptive subgoal calls is given, as is the number of new calls (producers) and the number of repeated calls to non-completed tables (variants). Furthermore, the number of properly subsumed calls to incomplete tables is given, along with the number of subsumed calls to completed tables. Finally, the total number of subsumptive table entries overall is given, including all producer and consumer calls.

- Call Subsumption Answer Operations. In call subsumptive tabling, answer lists are copied from producer subgoals to subsumed consumer subgoals (this operation is not required in variant tabling). The number of answer ident operations represents the number of times this copy is done. In addition, the number of consumptions performed by all consuming subsumptive table entries is also given.

- Call Variance Subgoal Operations. For call variance the number of subgoal check/insert operations is given along with the unique number of subgoals encountered (generator) and the number of redundant consumer encountered (consumer).

- Total Answer Operations. For both variant and subsumptive tables, the number of answer check insert operations is given along with the number of answers actually inserted into the table and the number of redundant answers derived.

- Garbage Collection Information. Time spent garbage collecting by the calling thread and number of heap cells collected.

- Information about process CPU and clock time, as well as the number of active threads.

As mentioned above, if XSB is configured with the single-threaded engine and is invoked with the ’-s’ option (see Section 3.7), additional information is printed out about maximum use of each execution stack and table space. However, the ’-s’ option can substantially slow down the emulator so benchmarks of time should be performed separately from benchmarks of space.

Example: The following printout shows how the statistics/0 output looks if it is invoked with the ’-s’ option (without it the Maximum stack used, and Maximum table space used lines are not shown). Information about the allocation size is provided since the sizes can be changed through emulator options (see Section 3.7).
\section*{CHAPTER 6. STANDARD AND GENERAL PREDICATES}

| ?- statistics.

Memory (total) 2429504 bytes: 726696 in use, 1702808 free
permanent space 645520 bytes: 645520 in use, 0 free
atom 120328
string 156872
asserted 3184
compiled 358216
other 6920
glob/loc space 786432 bytes: 652 in use, 785780 free
global 456 bytes
local 196 bytes
trail/cp space 786432 bytes: 476 in use, 785956 free
trail 88 bytes
choice point 388 bytes
SLG unific. space 65536 bytes: 0 in use, 65536 free
SLG completion 65536 bytes: 0 in use, 65536 free
SLG table space 80048 bytes: 80048 in use, 0 free

Maximum stack used: global 436724, local 14780, trail 27304, cp 20292,
SLG completion 0 (0 subgoals)
Maximum table space used: 0 bytes

Tabling Operations
0 subsumptive call check/insert ops: 0 producers, 0 variants,
0 properly subsumed (0 table entries), 0 used completed table.
0 relevant answer ident ops. 0 consumptions via answer list.
0 variant call check/insert ops: 0 producers, 0 variants.
0 answer check/insert ops: 0 unique inserts, 0 redundant.

0 heap ( 0 string) garbage collections by copying: collected 0 cells in 0.000000 secs

Time: 0.190 sec. cputime, 13.921 sec. elapsetime

\texttt{statistics(+Key)}
\texttt{statistics/1} allows the user to output detailed statistical information about
the atom and symbol tables, as well as about table space. The following calls
to \texttt{statistics/1} are supported:

\begin{itemize}
  \item \texttt{statistics(reset)} Resets the CPU time as well as counts for various
        tabling operations.
  \item \texttt{statistics(atom)} Outputs statistics about both the atom and symbol
        tables. An example is:
        | ?- statistics(atom).
        Symbol table statistics:
------------------------
Table Size: 8191
Total Symbols: 1188
  used buckets: 1088 (range: [0, 8174])
  unused buckets: 7103
  maximum bucket size: 3 (#: 18)

String table statistics:
------------------------
Table Size: 16381
Total Strings: 1702
  used buckets: 1598 (range: [0, 16373])
  unused buckets: 14783
  maximum bucket size: 3 (#: 2318)

• statistics(table) Outputs very detailed statistics about table space, including breakdowns into variant and subsumptive call- and answer-trie nodes and hash tables; answer return list nodes, and structures for conditional answers (cf. [62, 58, 37, 17]). In the multi-threaded engine, these data structures are reported both for shared tables and for private tables of the calling thread.

While this option is intended primarily for developers, it can also provide valuable information for the serious user of tabling.

Error Cases

• Key not a valid atom for input to statistics/1
  – domain_error(statisticsInputDomain,Key))

statistics(?Key,-Result)

statistics/2 allows a user to determine information about resources used by XSB. Currently statistics/2 unifies Key with

• runtime, which instantiates Result to the structure [TotalCPU,IncrCPU] where TotalCPU is the total (process-level) CPU time at the time of call, and IncrCPU is the CPU time taken since the last call to statistics/2.
  Times are measured in seconds. The process-level CPU time includes time taken for system calls, as well as time taken for garbage collection and stack-shifting. Note that in the multi-threaded engine, statistics/2 measures the time for all threads.

• walltime, which instantiates Result to the list [TotalTime,IncrTime] where TotalTime is the total elapsed time at the time of call, and IncrTime is the elapsed time taken since the last call to statistics/2. Times are measured in seconds.
• **total_memory** which instantiates Result to the list \([\text{Alloc,Used}]\). In the single-threaded engine, \text{Alloc} is the total table space allocated and \text{Used} is the total table space used, both in bytes. In the multi-threaded engine, both refer to table space *private* to the calling thread.

• **tablespace** which instantiates Result to the list \([\text{Alloc,Used}]\). In the single-threaded engine, \text{Alloc} is the total table space allocated and \text{Used} is the total table space used, both in bytes. In the multi-threaded engine, both refer to table space *private* to the calling thread.

• **shared_tablespace** which instantiates Result to the list \([\text{Alloc,Used}]\). In the multi-threaded engine, \text{Alloc} is the total space allocated for *shared* tables and \text{Used} is the total table space used, both in bytes. An error is thrown if this option is called by the single-threaded engine.

• **trie_assert** which instantiates Result to the list \([\text{Alloc,Used}]\). In the single-threaded engine, \text{Alloc} is the total space allocated for trie-asserted facts and interned tries; \text{Used} is the total space used for these purposes, both in bytes.

• **heap** which instantiates Result to the total number of bytes used by XSB’s heap. In the multi-threaded engine, the number refers only to the heap of the calling thread.

• **local** which instantiates Result to the total number of bytes used by XSB’s local (environment) stack. In the multi-threaded engine, the number refers only to the local stack of the calling thread.

• **trail** which instantiates Result to the total number of bytes used by XSB’s trail stack. In the multi-threaded engine, the number refers only to the trail stack of the calling thread.

• **choice_point** which instantiates Result to the total number of bytes used by XSB’s choice point stack. In the multi-threaded engine, the number refers only to the choice point stack of the calling thread.

• **incomplete_tables** which instantiates Result to the number of incomplete tables in XSB’s completion stack. In the multi-threaded engine, this number refers to the completion stack of the calling thread, which may contain both thread-private and thread-shared tables.

• **atoms** which instantiates Result to the number of bytes taken by atoms in the atom table.

• **idg** which instantiates Result to a structure \(\text{idg(+Nodes,+Edges)}\) where \text{Nodes} is the number of nodes currently in the incremental dependency graph (IDG) and \text{Edges} the number of edges.
Example An example of using statistics/2 to check CPU time is as follows:

?- statistics(runtime,[BeforeCumu,BeforeIncr]),spin(100000000),
    statistics(runtime,[AfterCumu,AfterIncr]).

BeforeCumu = 5.0167
BeforeIncr = 5.0167
AfterCumu = 9.6498
AfterIncr = 4.6331

Note that statistics/2 can provide either cumulative or incremental times; here

\[ AfterCumu - BeforeCumu = AfterIncr \]

Checking wall time is done similarly.

?- statistics(walltime,Before),sleep(1),statistics(walltime,After).

Before = [35.0651,35.0651]
After = [36.0652,1.0001]

Error Cases

- Key not a valid atom for input to statistics/1
  - domain_error(statisticsInputDomain,Key))

\texttt{time(+Goal)}

Prints both the CPU time and wall time taken by the execution of \texttt{Goal}. Any choice-points of \texttt{Goal} are discarded. The definition of predicate is based on the SWI-Prolog definition (minus reporting the number of inferences, which XSB does not currently support). This predicate is also found on other Prolog compilers such as YAP.

### 6.14 Asserting, Retracting, and Other Database Modifications

XSB provides an array of features for modifying the dynamic database. As a default, using \texttt{assert/1}, clauses can be asserted using first-argument indexing in a manner
that is now standard to Prolog implementations. However, a variety of other behaviors can be specified using the (executable) directives \texttt{index/3} and \texttt{index/2}. For instance, dynamic clauses can be declared to have multiple or joint indexes, and this indexing can be either hash-based as is typical in Prolog systems or based on \texttt{tries}. No matter what kind of indexing is used, space is dynamically allocated when a new clause is asserted and, unless specified otherwise, released after it is retracted. Furthermore, the size of any index table expands dynamically as clauses are asserted.

All dynamic predicates are compiled into SLG-WAM code, however the manner of their compilation may differ, and the differences in compilation affect the semantics for the predicate. If a dynamic predicate $P/n$ is given an indexing directive of \texttt{trie}, clauses for $P/n$ will be compiled using trie instructions; otherwise clauses for $P/n$ will be compiled into SLG-WAM instructions along the lines of static predicates.

Consider first dynamic predicates that use any indexing other than \texttt{trie} – including multiple or joint indices and star indexing. XSB asserts WAM code for such clauses so that that the execution time of dynamic code is similar to compiled code for unit and binary clauses. Furthermore, tabling can be used by explicitly declaring a predicate to be both dynamic and tabled. In Version 3.6, when the clause of a dynamic predicate is asserted as WAM code, the “immediate semantics” rather than the ISO Semantics of assert/retract [46]. The immediate semantics allows assert and retract to be fast and spatially efficient, but requires that significant care must be taken when modifying the definition of a predicate which is currently being executed.

If a dynamic predicate is given an indexing directive of \texttt{trie}, clauses of the predicate are compiled (upon a call \texttt{assert/1}) using trie instructions as described in [58]. Creation of trie-based dynamic code is significantly faster than creation of other dynamic code, and execution time may also be faster. However, trie-based predicates can only be used for unit clauses where a relation is viewed as a set, and where the order of the facts is not important.

XSB does not at this time fully support dynamic predicates defined within compiled code. The only way to generate dynamic code is by explicitly asserting it, or by using the standard predicate \texttt{load_dyn/1} to read clauses from a file and assert them (see the section \texttt{Asserting Dynamic Code} in Volume 2). There is a \texttt{dynamic/1} predicate (see page 269) that declares a predicate within the system so that if the predicate is called when no clauses are presently defining it, the call will quietly fail instead of issuing an “Undefined predicate” error message.

\texttt{asserta(+Clause)} \quad \text{ISO}

If the index specification for the predicate is not \texttt{trie}, this predicate adds a dynamic clause, \texttt{Clause}, to the database \texttt{before} any other clauses for the same
predicate currently in the database. If the index specification for the predicate is \texttt{trie}, the clause is asserted arbitrarily within the trie, and a warning message sent to \texttt{stderr}.

Note that because of the precedence of \texttt{:/-2}, asserting a clause containing this operator requires an extra set of parentheses: \texttt{assert((Head :- Body))}.

\textbf{Error Cases}

- Clause is not instantiated
  - \texttt{instantiation\_error}
- Clause is not a callable clause.
  - \texttt{domain\_error(callable,Clause)}
- Clause has a head that is a static built-in
  - \texttt{permission\_error(modify,builtin,Clause)}
- Clause has a head that is a static user predicate
  - \texttt{permission\_error(modify,static,Clause)}

\texttt{assertz(+Clause)} \hspace{1cm} \texttt{ISO}

If the index specification for the predicate is not \texttt{trie}, this predicate adds a dynamic clause, \texttt{Clause}, to the database after any other clauses for the same predicate currently in the database. If the index specification for the predicate is \texttt{trie}, the clause is asserted arbitrarily within the trie, and a warning message sent to \texttt{stderr}. Error cases are as with \texttt{asserta/1}.

Note that because of the precedence of \texttt{:/-2}, asserting a clause containing this operator requires an extra set of parentheses: \texttt{assert((Head :- Body))}.

\texttt{assert(+Clause)}

If the index specification for the predicate is not \texttt{trie}, this predicate adds a dynamic clause, \texttt{Clause}, to the database after any other clauses for the same predicate currently in the database (acting as \texttt{assertz/1}). If the index specification for the predicate is \texttt{trie}, the clause is asserted arbitrarily within the trie. Error cases are as with \texttt{assertz/1}.

Note that because of the precedence of \texttt{:/-2}, asserting a clause containing this operator requires an extra set of parentheses: \texttt{assert((Head :- Body))}.

\texttt{assert(+Clause,+AorZandVar,+Index)}

This is a lower-level interface to (non-trie-indexed) \texttt{assert}. It is normally not needed except in one particular situation, when \texttt{assert} aborts because it needs
too many registers. In this case, this lower-level assert may allow the offending clause to be correctly asserted.

The default implementation of non-trie-indexed assert generates code with a single pass through the asserted term. Because of this, it cannot know when it has encountered the final occurrence of a variable, and thus it can never release (and thus re-use) registers that are used to refer to variables. Since there is a limit of 255 registers in the XSB virtual machine, asserting a clause with more than this many distinct variables results in an error. There is an alternative implementation of assert that initially traverses the clause to determine the number of occurrences of each variable and thus allows better use of registers during code generation.

Clause is the clause to assert. AorZandVar is an integer whose lower 2 bits are used: The low-order bit is 0 if the clause is to be added as the first clause, and 1 if it is to be added as the last clause. If the second bit (2) is on, then the clause is traversed to count variable occurrences and so improve register allocation for variables; if it is 0, the default one-pass code-generation is done. So, for example, if AorZandVar is 3, then the clause will be asserted as the last one in the predicate and the better register allocation will be used. Index indicates the argument(s) on which to index.

retract(+Clause)  
ISO
Removes through backtracking all clauses in the database that match with Clause. Clause must be of one of the forms: Head or Head :- Body. Note, that because of the precedence of :-/2, using the second form requires an extra set of parentheses: retract((Head :- Body)).

The technical details on space reclamation are as follows. When retract is called, a check is made to determine whether it is safe to reclaim space for that clause. Safety is ensured when:

- A check is made of the choice point stack indicating that no choice point will backtrack into space that is being reclaimed; AND
  - The predicate is thread-private; OR
  - there is a single active thread
- AND if the predicate is tabled, there is no incomplete table for that predicate.

If it is safe to reclaim space for the clause, space is reclaimed immediately. Otherwise the clause is marked so that its space may later be reclaimed through garbage collection. (See gc_dynamic/1).
### Error Cases

- **Clause** is not instantiated
  - instantiation_error
- **Clause** is not a callable clause.
  - domain_error(callable,Clause)
- **Clause** has a head that is a static built-in
  - permission_error(modify,builtin,Clause)
- **Clause** has a head that is a static user predicate
  - permission_error(modify,static,Clause)

#### retractall(+Head)

ISO

removes every clause in the database whose head matches with Head. The predicate whose clauses have been retracted retains the dynamic property (contrast this behavior with that of predicates abolish/[1,2] below). Predicate retractall/1 is determinate and always succeeds. The term Head is not further instantiated by this call. Conditions for space reclamation and error cases are as with retract/1.

#### abolish(+PredSpec)

ISO

Removes all information about the specified predicate. PredSpec is of the form Pred/Arity. Everything about the abolished predicate is completely forgotten by the system (including the dynamic or static property, whether the predicate is tabled, and whether the predicate is thread-shared or thread-private)\(^\text{16}\). Any completed tables for the predicate are also removed.

It is an error to abolish a predicate when there is more than 1 active thread, regardless of whether the predicate is thread-private or thread-shared. The reason for this is that, even if PredInd denotes a thread-private predicate, one thread may be making use of PredInd as another thread abolishes it. abolish/1 throws an error in such a case to prevent such a semantic inconsistency. Similarly, if there is a non-completed table for PredInd, an error is thrown to prevent incompleteness in the tabled computation.

**ISO Compatibility Note:** Version 3.6 of XSB allows static predicates to be abolished and their space reclaimed. Such space is reclaimed immediately, and unlike the case for abolished static code, no check is made to ensure that XSB’s

---

\(^{16}\text{For compatibility with older Prologs, there is also an}\ \text{abolish/2 which takes} \ Pred \text{and} \ Arity \text{as its two arguments.}
choice point stack is free of choice points for the abolished static predicate. Abolishing static code is thus dangerous and should be avoided unless a user is certain it is safe to use.

Error Cases

- **PredInd, Pred or Arity** is not instantiated
  - instantiation_error
- **Arity** is not in the range 0..255 (max arity)
  - domain_error(arity_indicator,Arity)
- **PredInd** indicates a static built-in
  - permission_error(modify,builtin,Predind)
- **abolish/1** is called when there is more than 1 active thread.
  - misc_error
- **PredInd** has a non-completed table in the current thread.
  - table_error
- There are active backtrack points to a (dynamic) clause for **PredInd**.  
  - misc_error

clause(+Head,?Body)

ISO

Returns through backtracking all dynamic clauses in the database whose head matches **Head** and Body matches **Body**. For facts the Body is true. **clause/2** works properly for all dynamically asserted clauses, even if they are trie-indexed; however **clause/2** does not access trie-inserted terms. In the multi-threaded engine, when a thread *T* calls **clause/2** it accesses both thread-shared dynamic code and thread-private dynamic code for *T*.

Error Cases

- **Head** is not instantiated
  - instantiation_error
- **Head** (or Body) is not a callable clause.
  - domain_error(callable,Head)
- **Head** is a static built-in

---

17XSB throws an error in this case because garbage collection for abolished predicates has not been implemented (unlike for retract(all) and various table abolishes). Besides, you shouldn’t be abolishing a predicate that you could backtrack into. What were you thinking?
CHAPTER 6. STANDARD AND GENERAL PREDICATES

- permission_error(access,builtin,Head)
  - Head is a static user predicate
- permission_error(access,static,Clause)

gc_dynamic(-N)
Invokes the garbage collector for dynamic clauses that have been retracted, or whose predicate has been abolished. When called with more than 1 active thread, gc_dynamic/1 will always perform garbage collection for that thread’s private retracted clauses; however in Version 3.6, it will only perform garbage collection for retracted thread-shared clauses if there is a single active thread. N is the number or shared and/or private frames left to be collected – if N is unified to 0, then all possible garbage collecting has been performed. N is unified to -1 garbage collection was not attempted (due to multiple active threads).
By default, gc_dynamic/1 is called automatically at the top level of the XSB interpreter, when abolishing a predicate, and when calling retractall for an “open” term containing no variable bindings.

index(+PredSpec, +IndexSpec)
In index(PredSpec, IndexSpec), PredSpec is a predicate indicator or term indicator, and IndexSpec is a form of index specification as described below.
In general, XSB supports hash-based indexing on various arguments of clauses, on combinations of arguments, as well as within the arguments of a clause. The availability of various kinds of indexing depends on whether code is static (e.g. compiled) or dynamic (e.g. asserted, loaded with load_dyn/1 and so on). Index directives can be given to the compiler as part of source code or executed during program execution (analogously to op/3). When executed during program execution, index/2 does not re-index an already existing predicate; however for dynamic predicates index/2 does affect the index for clauses asserted after the directive has been given.

- Hash-based Indexing
  - Static Predicates In this case IndexSpec must be a non-negative integer which indicates the argument on which an index is to be constructed. If IndexSpec is 0, then no index is kept (possibly an efficient strategy for predicates with only one or two clauses.)
  - Dynamic Predicates For a dynamic predicate, (to which no clauses have yet been asserted), a wide variety of indexing techniques are possible. We discuss their syntax first, and then their semantics. For dynamic predicates then, IndexSpec can be either an indexing element
or a list of indexing elements. Each indexing element defines a separate index and specifies an argument or group of arguments that make up the search key of that index. Thus an indexing element consists of one or more argument indicators joined together by +/2. An argument indicator is may be an integer (ArgNo) indicating an argument number (starting from 1) to use in the index, or it may have the form *(ArgNo). If ArgNo is an integer, only the main functor symbol of argument ArgNo will participate in the index. When annotated with the asterisk, the first 5 fields of argument ArgNo (in a depth-first traversal of the term) will be used in the index. If there are fewer than 5, they all will be used. If any of the first 5 is a variable, then the index cannot be used. An index is usually on a single argument, in which case the indexing element consists of a single argument indicator. If an indexing element contains more than one argument specifier, then a joint index is specified i.e. an index will be constructed so that the values of each argument indicator are to be concatenated to create the search key of the index.

Examples help clarify this. \texttt{index(p/3, [2,1])} indicates that clauses asserted for the predicate \texttt{p/3} should be indexed on both the second and the first argument. A query \texttt{Q} to \texttt{p/3} will first use the second argument index to \texttt{p/3} if the second argument of \texttt{Q} is non-variable, and will use the main functor of the second argument. Otherwise, if the second argument of \texttt{Q} is a variable, but not the first argument, the first argument index of \texttt{p/3} will be used. If both arguments in \texttt{Q} are variables, no index will be used and \texttt{Q} will backtrack through all clauses for \texttt{p/3}.

\texttt{index(p/3, [(2)*1])} would result in similar behavior as the previous example, but the first index to be tried (on the second argument) would be built using more of the term value in that second argument position (not just the main functor symbol.)

As another example, one could specify: \texttt{index(p/5, [1+2,1,4])}. After clauses are asserted to it, a call to \texttt{p/5} would first check to see if both the first and second arguments are non-variable and if so, use an index based on both those values. Otherwise, it would see if the first argument is non-variable and if so, use an index based on it. Otherwise, it would see if the fourth argument is non-variable and if so use an index based on it. As a last resort, it would use no index but backtrack through all the clauses in the predicate. In each of these cases, the indexes are built using only the main functor symbol in the indicated argument position. (Notice that it may well make sense to
include an argument that appears in a joint specification later alone, as 1 in this example, but it never makes sense forcing the single argument to appear earlier. In that case the joint index would never be used.) If we want to use similar indexing on p/5 of the previous example, except say argument 1 takes on complex term values and we want to index on more of those terms, we might specify the index as index(p/5, [*(1)+2, *(1), 4]).

- Trie-based Indexing If Predspec is dynamic, the executable directive \texttt{index(Predspec, trie)} causes clauses for Predspec to be asserted using tries (see [58], which is available through the XSB web page). The name trie indexing is something of a misnomer since the trie itself both indexes the term and represents it. In XSB, a trie index is formed using a left-to-right traversal of the unit clauses. These indexes can be very effective if discriminating information lies deep within a term, and if there is sharing of left-prefixes of a term, trie indexing can reduce the space needed to represent terms. Furthermore, asserting a unit clause as a trie is much faster than asserting it using default WAM code. Despite these advantages, representing terms as tries leads to semantic differences from asserted code, of which the user should be aware. First, the order of clauses within a trie is arbitrary: using \texttt{asserta/1} or \texttt{assertz} for a predicate currently using trie indexing will give the same behavior as using \texttt{assert}. Also, the current version of XSB only allows trie indexing for unit clauses.

If in doubt what indexing is being used for a predicate, a call to \texttt{current_index/2} can be made.

Error Cases

- PredSpec or IndexSpec is a variable
  - instantiation_error
- PredSpec is neither a variable, a predicate indicator, nor a callable term.
  - type_error(predicate_indicator_or_callable, PredSpec)
- IndexSpec is not ground
  - instantiation_error
- IndexSpec is neither a properly formed indexing element nor a list of indexing elements
  - domain_error(indexing_element, IndexSpec)
• **IndexSpec** is a list containing an element **IndexElt** that not a properly formed indexing element
  
  - `domain_error(indexing_element,IndexElt)`

• **PredSpec** represents a predicate that has been previously defined to be static

  - `permission_error(modify,static_predicate)`

**dynamic(+Operations)**

**ISO**

dynamic/1 can be used either as a compiler declaration or as an executable directive. Used as a compiler declaration, it indicates that all clauses for each predicate denoted by the command are dynamic – clauses for these predicates can be asserted or retracted. Without this declaration compiled clauses will be treated as static. Executed as a directive in a state of execution where no clauses exist for each denoted predicate **dynamic/1** ensures clauses for the affected predicates are to be treated as dynamic. If **PredSpec** contains a predicate that is defined as static or as foreign code, a permission error will be thrown. **Operations** can take one of two forms:

1. **Operations** is a predicate indicator, a callable term, or a comma-list of predicate indicators or callable terms.

2. **Operations** has the form **Predspec as Options** where

   - **PredSpec** is a predicate indicator, a callable term, or comma-list of predicate indicators or callable terms.

   - **Options** is either a dynamic_option or a list of dynamic_options. These dynamic options control the attributes of a dynamic predicate. In Version 3.6, the following dynamic options are supported

     - `intern` which causes every clause for this predicate, before being asserted, to force all its ground subterms to be interned into a global table.

     - `tabled` which causes the dynamic predicate to be tabled. The declaration/directive `dynamic p/n as tabled` has the same effect as `table p/n as dynamic`.

     - `variant` which causes the table evaluation method of the predicate(s) to use call variance.

     - `incremental` which allows (incremental) tables that are based on the dynamic predicate to be automatically updated when clauses are asserted or retracted.
- **Opaque.** This option is essentially the same as non-incremental dynamic code, *except* that opaque predicates can be made incremental by a later `dynamic/1` directive, and incremental predicates can be made opaque by a `dynamic/1` directive.

- **Private** which causes the predicate(s) to be treated as thread private.

- **Shared** which causes the predicate(s) to be treated as thread shared.

If the directive

```
dynamic p/n.
```

is executed, its behavior is as follows:

- If `p/n` is already dynamic, the directive has no effect, regardless of whether `p/n` is tabled, incremental or opaque, private or shared.

- If `p/n` has *not* already been defined, the directive makes `p/n` non-tabled, non-incremental, and to use the default thread sharing strategy (`private` unless XSB is called with `-shared_predicates`).

If the directive

```
dynamic PredList as Options.
```

is executed, various checks are performed on `Options`. These checks are (mostly) performed before any predicates are declared as dynamic or options changed, and reduce the possibility of leaving some `p/n` in `PredList` with inconsistent attributes.

- If a dynamic predicate in `PredList` is declared as `incremental` it may be changed to `opaque` at any time; similarly, a dynamic predicate that is `opaque` may be changed to `incremental`.

- Otherwise, an attempt to change an attribute of `p/n` in `PredList` — i.e. whether `p/n` is tabled or not, incremental/opaque or not, and thread-private or thread-shared — will throw a permission error.

In addition, regardless of the state of predicates in `PredList`, if options contains an inconsistent set of declarations, a domain error will be thrown. `Options` is inconsistent in the following cases:

- `Options` contains `tabled` or `variant` and `opaque` or `incremental`. Tabled dynamic incremental code is not yet supported in XSB.

- `Options` contains both `private` and `shared`
• Options contains both incremental and opaque
• Options contains intern and (dynamic or subsumptive or incremental or opaque)

Error Cases

Error cases are summarized as follows. Let Operations be of the form PredSpec or PredSpec as Options. Then if

• PredSpec or is a variable or a comma list containing a variable
  – instantiation_error
• An element of PredSpec is neither a variable nor a comma list
  – type_error(callable,PredSpec)
• A predicate in PredSpec has been previously defined to be static or foreign
  – permission_error(modify,static_predicate)
• Options is a variable or a list containing a variable
  – instantiation_error
• Options contains an element Option that isn’t a dynamic option (as described above)
  – domain_error(dynamic_option,Option)
• Options contains inconsistent elements (as described above)
  – table_error
• An option in Options would modify a predicate in predspec in a manner that is not allowed (as described above)
  – permission_error

In addition, if a predicate p/n was declared to be dynamic and a file containing clauses for p/n is later consulted, a permission error will be thrown.

6.14.1 Reading Dynamic Code from Files

Several built-in predicates are available that can assert the contents of a file into XSB’s database. These predicates are useful when code needs to be dynamic, or when the they contain a large number of clauses or facts. Configured properly, files
containing millions of facts can be read and asserted into memory in under a minute, making XSB suitable for certain kinds of in-memory database operations.

Each of the predicates in this section allow loading from files with proper prolog extensions, and makes use of the XSB library paths. See Sections 3.6 and 3.3 for details.

load_dyn(+FileName)

Asserts the contents of file FileName into the database. All existing clauses of the predicates in the file that already appear in the database, are retracted, unless there is a multifile/1 declaration for them. An indexing declaration of a predicate p/n in FileName will be observed as long as the declarations occur before the first clause of p/n. file will be observed as Clauses in FileName must be in a format that read/1 will process. So, for example, operators are permitted. As usual, clauses of predicates are not retracted if they are compiled instead of dynamically asserted. All predicates are loaded into usermod. Module declarations such as :- export are ignored and a warning is issued.

Dynamically loaded files can be filtered through the XSB preprocessor. To do this, put the following in the source file:

:- compiler_options([xpp_on]).

Of course, the name compiler_options might seem like a misnomer here (since the file is not being compiled), but it is convenient to use the same directive both for compiling and loading, in case the same source file is used both ways.

Error Cases

- FileName is a variable
  - instantiation_error
- FileName is not an atom.
  - type_error(atom,Filename)
- FileName has been loaded previously in the session and there is more than one active thread.
  - misc_error

\[^{18}\text{In Version 3.6, loading code dynamically can also be useful when the clauses contain atoms whose length is more than 255 that cannot be handled by the XSB compiler.}\]
load_dyn(+FileName,+Dir)

 Asserts the contents of file FileName into the database. Dir indicates whether assertz or asserta is to be used. If Dir is z, then assertz is used and the behavior of load_dyn(FileName) is obtained. If Dir is a, then asserta is used to add the clauses to the database, and clauses will be in the reverse order of their appearance in the input file. asserta is faster than assertz for predicates such that their indexing and data result in many hash collisions. Dir is ignored for facts in FileName that are trie-indexed.

Error Cases

- FileName is a variable
  - instantiation_error
- FileName is not an atom:
  - type_error(atom,FileName)
- Dir is not equal to a or z\(^{19}\):
  - domain_error(a_or_z,Dir)
- FileName has been loaded previously in the session and there is more than one active thread.
  - misc_error

load_dync(+FileName)

Acts as load_dyn/1, but assumes that facts are in “canonical” format and is much faster as a result. In XSB, a term is in canonical format if it does not use any operators other than list notation and comma-list notation. This is the format produced by the predicate write_canonical/1. (See cvt_canonical/2 to convert a file from the usual read/1 format to read_canonical format.) As usual, clauses of predicates are not retracted if they are compiled instead of dynamically asserted. All predicates are loaded into usermod. :- export declarations are ignored and a warning is issued.

Notice that this predicate can be used to load files of Datalog facts (since they will be in canonical format). This predicate is significantly faster than load_dyn/1 and should be used when speed is important. (See load_dyn/2 below for further efficiency considerations.) A file that is to be dynamically loaded often but not often modified by hand should be loaded with this predicate.

\(^{19}\)For backward compatibility, 0 and 1 are also allowed.
As with \texttt{load_dyn/1}, the source file can be filtered through the C preprocessor. However, since all clauses in such a file must be in canonical form, the \texttt{compiler_options/1} directive should look as follows:

\begin{verbatim}
:- (compiler_options(’.’(xpp_on,[]))).
\end{verbatim}

\textbf{Error Cases}

- \texttt{FileName} is a variable
  - \texttt{instantiation_error}
- \texttt{FileName} is not an atom.
  - \texttt{type_error(atom,FileName)}
- \texttt{FileName} has been loaded previously in the session \textit{and} there is more than one active thread.
  - \texttt{misc_error}

\texttt{load_dyn(+FileName,+Dir)}

Acts as \texttt{load_dyn/2}, but assumes that facts are in “canonical” format. \texttt{Dir} is ignored for trie-asserted code, but otherwise indicates whether \texttt{assertz} or \texttt{asserta} is to be used. If \texttt{Dir} is \texttt{z}, then \texttt{assertz} is used and the exact behavior of \texttt{load_dyn(FileName)} is obtained. If \texttt{Dir} is \texttt{a}, then \texttt{asserta} is used to add the clauses to the database, and clauses will end up in the reverse order of their appearance in the input file.

Setting \texttt{Dir} to \texttt{a} for non trie-asserted code can sometimes be \textit{much} faster than the default of \texttt{z}. The reason has to do with how indexes on dynamic code are represented. Indexes use hash tables with bucket chains. No pointers are kept to the ends of bucket chains, so when adding a new clause to the end of a bucket (as in \texttt{assertz}), the entire chain must be run. Notice that in the limiting case of only one populated bucket (e.g., when all clauses have the same index term), this makes assertz-ing a sequence of clauses quadratic. However, when using \texttt{asserta}, the new clause is added to the beginning of its hash bucket, and this can be done in constant time, resulting in linear behavior for asserta-ing a sequence of clauses.

\textbf{Error Cases}

- \texttt{FileName} is a variable
  - \texttt{instantiation_error}
• FileName is not an atom:
   - type_error(atom,FileName)
• Dir is not instantiated to a or z \(^{20}\):
   - domain_error(a_or_z,Dir)
• FileName has been loaded previously in the session and there is more than one active thread.
   - misc_error

ensure_loaded(+FileName,+Action)
This predicate does nothing if FileName has been loaded or consulted into XSB, and has not changed since it was loaded or consulted. Otherwise

• If Action is instantiated to dyn the behavior is as load_dyn/1 (or load_dyn(FileName,z)).
• If Action is instantiated to dyna the behavior is as load_dyn(FileName,a).
• If Action is instantiated to dync the behavior is as load_dync/1 (or load_dync(FileName,z)).
• If Action is instantiated to dynca the behavior is as load_dync(FileName,a).
• If Action is instantiated to consult, FileName is consulted (action is the same as ensure_loaded/1).

Error Cases

• FileName is not instantiated:
   - instantiation_error
• FileName is not an atom:
   - type_error(atom,FileName)
• Action is not a valid load action as described above
   - domain_error(loadAction,Action)

cvt_canonical(+FileName1,+FileName2) \quad \text{module: consult}
Converts a file from standard term format to “canonical” format. The input file name is FileName1; the converted file is put in FileName2. This predicate can be used to convert a file in standard Prolog format to one loadable by load_dync/1.

\(^{20}\)For backward compatibility, 0 and 1 are also allowed.
6.14.2 The storage Module: Associative Arrays and Backtrackable Updates

XSB provides a high-level interface that allows the creation of “objects” that efficiently manage the storage of facts or of associations between keys and values. Of course, facts and associative arrays can be easily managed in Prolog itself, but the storage module is highly efficient and supports the semantics of backtrackable updates as defined by Transaction logic [6] in addition to immediate updates. The semantics of backtrackable updates means that an update made by the storage module may be provisional until the update is committed. Otherwise, if a subgoal calling the update fails, the change is undone. The commit itself may be made either by the predicate storage_commit/1, or less cleanly by cutting over the update itself.

A storage object \( O \) is referred to by a name, which must be a Prolog atom. \( O \) can be associated either with a set of facts or a set of key-value pairs. Within a given storage object each key is associated with a unique value: however since keys and values can be arbitrary Prolog terms, this constraint need not be a practical restriction. A storage object \( O \) is created on demand, simply by calling (a backtrackable or non-backtrackable) update predicate that refers to \( O \). However to reclaim \( O \)’s space within a running thread, the predicate storage_reclaim_space/1 must be called. Both backtrackable and non-backtrackable updates can be made to the same storage object, although doing so may not always be a good programming practice.

If multiple threads are used, each storage object is private to a thread, and space for a storage object is reclaimed upon a thread’s exit. Thread-shared storage objects may be supported in future versions.

All the predicates described in this section must be imported from module storage.

Non-backtrackable Storage

storage_insert_keypair(+StorageName,+Key, +Value, ?Inserted)

Insert the given Key-Value pair into StorageName. If the pair is new, then Inserted unifies with 1. If the pair is already in StorageName, then Inserted unifies with 0. If StorageName already contains a pair with the given key that is associated with a different value, then Inserted unifies with -1. The first argument, StorageName, must be an atom naming the storage to be used. Different names denote different storages. In all cases the predicate succeeds.

storage_delete_keypair(+StorageName, +Key, ?Deleted)

Delete the key-value pair with the given key from StorageName. If the pair was
in StorageName then Deleted unifies with 1. If it was not in StorageNames then Deleted unifies with 0. The first argument, StorageName, must be an atom naming the storage object to be used. Different names denote different storages. In both cases the predicate succeeds.

\texttt{storage\_find\_keypair(+StorageName, +Key, ?Value)}

If StorageName has a key pair with the given key, then Value unifies with the value stored in StorageName. If no such pair exists in the database, then the goal fails.

Note that this predicate works with non-backtrackable associative arrays described above as well as with the backtrackable ones, described below.

\texttt{storage\_insert\_fact(+StorageName, +Fact, ?Inserted)}

Similar to keypair insertion, but this primitive inserts facts rather than key pairs.

\texttt{storage\_delete\_fact(+StorageName, +Fact, ?Inserted)}

Similar to key-pair deletion, but this primitive deletes facts rather than key pairs.

\texttt{storage\_find\_fact(+StorageName, +Fact)}

Similar to key-pair finding, but this primitive finds facts facts rather than key pairs.

\textbf{Backtrackable Updates}

\texttt{storage\_insert\_keypair\_bt(+StorageName, +Key, +Value, ?Inserted)}

A call to this predicate inserts a key pair into StorageName as does \texttt{storage\_insert\_keypair/4}, and the key-value pair may be queried via \texttt{storage\_find\_keypair/3}, just as with the non-backtrackable updates described above. In addition, the key-value pair can be removed from StorageName by explicit deletion. However, the key pair will be removed from StorageName upon failing over the insertion goal unless a commit is made to StorageName through the goal \texttt{storage\_commit(StorageName)}. The exact semantics is defined by Transaction Logic [6].

Note it is the update itself that is backtrackable, not the key-value pair. Hence, a key-pair may be (provisionally) inserted by a backtrackable update and deleted by a non-backtrackable update, or inserted by a non-backtrackable update and (provisionally) deleted by a backtrackable update. Of course, whether such a mixture makes sense would depend on a given application.
storage_delete_keypair_bt(+StorageName, +Key, ?Deleted)
   Like storage_delete_keypair/3, but backtrackable as described for storage_insert_keypair_bt/4.

storage_insert_fact_bt(+StorageName, +Goal)
   Like storage_insert_fact/2, but backtrackable.

storage_delete_fact_bt(+StorageName, +Goal)
   This is a backtrackable version of storage_delete_fact/2.

storage_commit(+StorageName)
   Commits to StorageName any backtrackable updates since the last commit, or since initialization if no commit has been made to StorageName. If StorageName does not exist, the predicate silently fails.

Reclaiming Space

storage_reclaim_space(+StorageName)
   This is similar to reclaim_space/1 for assert and retract, but it is used for storage managed by the primitives defined in the storage module. As with reclaim_space/1, this goal is typically called just before returning to the top level.

6.15 Tabling Declarations and Builtin

In XSB, tables are designed so that they can be used transparently by computations. However, it is necessary to first inform the system of which predicates should be evaluated using tabled resolution (Section 3.10.2) along with the properties to be used, such as call variance or call subsumption (Chapter 5). Further, it is often useful to be able to explicitly inspect a table, or to alter its state. The predicates described in this section are provided for these purposes. In order to ground the discussion of these predicates, we continue our overview of tables and table creation from Chapter 5. For a detailed description of the implementation of table access routines in XSB, the reader is referred to [58, 37, 18, 76] and other papers listed in the bibliography.

Tables and Table Entries

Abstractly, at a subgoal-level table can be seen as a set of entry triples \((S, \mathcal{A}, Status)\) where \(S\) is a subgoal, \(\mathcal{A}\) is its associated answer set, and \(Status\) its status — whether
it is complete or incomplete, along with tabling properties it uses (e.g., incremental or non-incremental, cf. Chapter 5 for a discussion of tabling properties). However, we also refer to a predicate-level table, which contains the set of all subgoal-level tables for some (tabled) predicate \( p/n \). At execution time, invocation of a tabled subgoal \( S \) leads to the classification \( S \) according to the properties associated with its predicate, as well as its possible creation of a table for \( S \). Each occurrence of a subgoal that is not yet completely evaluated can be classified as either (a) a generator, of answers or (b) a consumer of those answers.

**Answers, Returns, and Return Templates**

Given a table entry \((S, A, \text{Status})\), the set of variables in \( S \) is sometimes called the substitution factor of \( S \). The order of arguments in the substitution factor corresponds to the order of distinct variables in a left-to-right traversal of \( S \). Each answer in \( A \) substitutes values for the variables in the substitution factor of \( S \); this substitution is sometimes called an answer substitution. The table inspection predicates allow access to substitution factors and answer substitutions through a family of terms called return templates and whose principle functors have the form \( \text{ret}/n \), where \( n \) is the size of the substitution factor.

**Example 6.15.1** Let \( S = p(X, f(Y)) \) be a generator subgoal. Using a return template, the substitution factor can be depicted as \( \text{ret}(X, Y) \), while the answer substitution \( \{X=a, Y=b\} \) is depicted as \( \text{ret}(a, b) \). Note that the application of the answer substitution to the generator subgoal yields the answer \( p(a, f(b)) \).

To take a slightly more complex example, consider the subgoal \( q(X) \) where \( X \) is an attributed variable whose attribute is \( f(Z, Y, Y) \). In this case the substitution factor is \( \text{ret}(X, Z, Y) \).

In a similar manner, XSB maintains substitutions between generator subgoals and consuming subgoals when call subsumption is used. The return template for a consuming subgoal is a substitution that maps variables of its generator to subterms of the consuming subgoal. This template can then be used to select answers from the generator that unify with the consuming call.

**Example 6.15.2** Let \( p/2 \) of the previous example be evaluated using call subsumption and let the subgoal \( S = p(A, f(B)) \) be present in its table. Further, let \( S_1: p(A, f(B)) \) and \( S_2: p(g(Z), f(b)) \) be two consuming subgoals of \( S \). Then the return template of \( S_1 \) is \( \text{ret}(A, B) \) and that of \( S_2 \) is \( \text{ret}(g(Z), b) \). \( S_1 \), being a variant of \( S \),
selects answers for $S$ such that \{X=A, Y=B\}; i.e., all answers of $S$. $S_2$, on the other hand, selects only relevant answers of $S$, those that satisfy \{X=g(Z), Y=b\}.

Skeletons and Predicate Specifications

A skeleton for a functor $f/n$ is a structure of the form $f(Arg_1, \ldots, Arg_n)$ where each $Arg_i$ is a distinct variable. Similarly the skeleton of a term is the skeleton formed from the principal functor of the term, so that skeletons from the terms $f(1,2)$ and $f(A,B)$ are the same. A return skeleton is a specific application of this notion to answers. From it, one may discern the size of the template for a given subgoal. Finally, we assume that a predicate specification for a predicate $p$ and arity $n$, represented as $\text{PredSpec}$ below, can be given either using the notation $p/n$ or as a skeleton, $p(t_1, \ldots, t_n)$.

6.15.1 Declaring and Modifying Tabled Predicates

table(+Operations)  

$\text{table}/1$ can be used either as a compiler declaration or as an executable directive. Used as a compiler declaration, it indicates that each predicate denoted by the command is to be compiled using (a particular form of) tabling, and may indicate that the predicate itself is dynamic or thread-shared or thread-private. Executed as a directive in a state of execution where no clauses exist for each denoted predicate $\text{table}/1$ ensures that any clauses asserted for each predicate use tabling and may indicate the mode of tabling to be used. The parameter $\text{Operations}$ can take one of three forms:

1. $\text{Operations}$ is a predicate indicator, a skeleton, or a comma-list or list of predicate indicators or skeletons.

2. $\text{Operations}$ is a term indicating that a predicate is to be tabled with a particular form of answer subsumption (cf. Section 5.4).

3. $\text{Operations}$ has the form $\text{PredSpec as Options}$ where
   - $\text{PredSpec}$ is a predicate indicator, a skeleton, or a comma-list or list of predicate indicators or skeletons.
   - $\text{Options}$ is either a table option or a list of table options. In Version 3.6, the following table options are supported
     - dynamic or dyn which causes the predicate(s) to be treated as dynamic in addition to being tabled, and is equivalent to ?- dynamic
PredSpec

- subsumptive which causes the table evaluation method of the predicate(s) to use call subsumption.
- variant which causes the table evaluation method of the predicate(s) to use call variance.
- intern which causes all ground subterms of subgoals and answers entered into the table for the predicate(s) to be interned.
- incremental which causes the table evaluation method of the predicate(s) to be incremental.
- opaque which indicates that the tables predicate is used in the definition of an incremental table, but are not to be incrementally maintained themselves.
- private which causes the predicate(s) to be treated as thread private in addition to being tabled.
- shared which causes the predicate(s) to be treated as thread shared in addition to being tabled.
- subgoal_abstract(n) which enables depth-n subgoal abstraction for the predicate(s).
- answer_abstract(n) which enables depth-n answer abstraction for the predicate(s).

If the directive

**table PredList as Options.**

is executed, various checks are performed on Options. These checks are (mostly) performed before any predicates are declared as dynamic or options changed, and reduce the possibility of leaving some p/n in PredList with inconsistent attributes, which could cause an error to be thrown during program execution.

- If a predicate in Predlist has been declared as incremental it may be changed to opaque at any time; similarly, a predicate that is opaque may be changed to incremental.
- If a predicate in Predlist has been declared to use call variance it may be changed to use call subsumption at any time; similarly, a predicate that uses call subsumption may be changed to use call variance.

---

21Because dynamic is an operator, the declaration requires parentheses, e.g.: table p/n as (dynamic).
• Otherwise, an attempt to change an attribute of \( p/n \) in \( \text{PredList} \) – i.e. whether \( p/n \) is tabled or not, dynamic or not and thread-private or thread-shared – will throw a permission error.

In addition, regardless of the state of predicates in \( \text{PredList} \), if options contains an unsupported set of declarations, a permission error will be thrown (see Table 5.1 for a list of supported and non-supported combinations of tabling modes and predicate properties). \( \text{Options} \) throws a table error in the following cases:

- \( \text{Options} \) contains \text{dynamic} and either \text{opaque} or \text{incremental}. Tabled dynamic incremental code is not yet supported in XSB.
- \( \text{Options} \) contains \text{(incremental or opaque)} and \text{(subsumptive or shared)}
- \( \text{Options} \) contains \text{subsumptive} and \text{(variant or shared or subgoal_abstract/1 or answer_abstract/1)}
- \( \text{Options} \) contains \text{intern} and \text{(dynamic or subsumptive or approximate or incremental or opaque or answer_abstract or subgoal_abstract)}
- \( \text{Options} \) contains both \text{private} and \text{shared}
- \( \text{Options} \) contains both \text{incremental} and \text{opaque}

**Error Cases**

Error cases are summarized as follows. Let \( \text{Operations} \) be of the form \( \text{PredSpec} \) or \( \text{PredSpec} \) as \( \text{Options} \). Then if

- \( \text{PredSpec} \) or is a variable or a comma list containing a variable
  - \text{instantiation_error}
- An element of \( \text{PredSpec} \) is neither a variable nor a predicate indicator, nor a skeleton.
  - \text{type_error} (callable, \text{PredSpec})
- A predicate in \( \text{PredSpec} \) has been previously defined to be static or foreign and \( \text{Options} \) contains \text{dynamic} or \text{dyn}
  - \text{permission_error} (modify, static_predicate)
- \( \text{Options} \) is a variable or a list containing a variable
  - \text{instantiation_error}
- \( \text{Options} \) contains an element \( \text{Option} \) that isn’t a table option (as described above)
  - \text{domain_error} (table_option, \text{Option})
CHAPTER 6.  STANDARD AND GENERAL PREDICATES

- **Options** contains a non-supported combination of elements (as described above)
  - permission_error

- An option in *Options* would modify a predicate in *PredSpec* in a manner that is not allowed (as described above)
  - permission_error

### 6.15.2 Predicates for Table Inspection

The user should be aware that skeletons that are dynamically created (e.g., by `functor/3`) are located in `usermod` (refer to Section 3.4). In such a case, the tabling predicates below may not behave in the desired manner if the tabled predicates themselves have not been imported into `usermod`.

We maintain two running examples in this section for explanatory purposes. One uses tabling based on call variance:

<table>
<thead>
<tr>
<th>Call Variance Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Program</strong></td>
</tr>
<tr>
<td>: table p/2 as variant.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

and the other uses tabling based on call subsumption::

```
Call Subsumption Example

<table>
<thead>
<tr>
<th>Program</th>
<th>Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>:- table q/2 as subsumptive.</td>
<td>Subgoal</td>
</tr>
<tr>
<td>q(a,b).</td>
<td>q(X,Y)</td>
</tr>
<tr>
<td>q(b,c).</td>
<td></td>
</tr>
<tr>
<td>q(a,c).</td>
<td>q(a,Y)</td>
</tr>
<tr>
<td></td>
<td>q(X,c)</td>
</tr>
</tbody>
</table>

Note that in the call subsumption example, the subgoals \(q(a,Y)\) and \(q(X,c)\) are subsumed by, and hence obtain their answers from, the subgoal \(q(X,Y)\).

get_call(+CallTerm,-TableEntryHandle,-ReturnTemplate) Tabling

If call variance is used for the predicate corresponding to CallTerm, then this predicate searches the table for an entry whose subgoal is a variant of CallTerm. If subsumption is used, then this predicate searches for some entry that subsumes (properly or not) CallTerm. In either case, should the entry exist, then the handle to this entry is assigned to the second argument, while in the third, its return template is constructed. These latter two arguments must be uninstantiated at call time.

Error Cases

- CallTerm is not a callable term
  - type_error(callable_term,CallTerm)
- CallTerm does not correspond to a tabled predicate
  - permission_error(table access,non-tabled predicate,CallTerm)
Example 6.15.3

### Variant Predicate

<table>
<thead>
<tr>
<th>?- get_call(p(X,Y),Ent,Ret).</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
</tr>
<tr>
<td>?- get_call(p(1,Y),Ent,Ret).</td>
</tr>
<tr>
<td>Y = _h92</td>
</tr>
<tr>
<td>Ent = 136039108</td>
</tr>
<tr>
<td>Ret = ret(_h92);</td>
</tr>
<tr>
<td>no</td>
</tr>
<tr>
<td>?- get_call(p(X,3),Ent,Ret).</td>
</tr>
<tr>
<td>X = _h84</td>
</tr>
<tr>
<td>Ent = 136039156</td>
</tr>
<tr>
<td>Ret = ret(_h84);</td>
</tr>
<tr>
<td>no</td>
</tr>
<tr>
<td>?- get_call(p(1,3),Ent,Ret).</td>
</tr>
<tr>
<td>no</td>
</tr>
</tbody>
</table>

### Subsumptive Predicate

<table>
<thead>
<tr>
<th>?- get_call(q(X,Y),Ent,Ret).</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = _h80</td>
</tr>
<tr>
<td>Y = _h94</td>
</tr>
<tr>
<td>Ent = 136043988</td>
</tr>
<tr>
<td>Ret = ret(_h80,_h94);</td>
</tr>
<tr>
<td>no</td>
</tr>
<tr>
<td>?- get_call(q(a,Y),Ent,Ret).</td>
</tr>
<tr>
<td>Y = _h88</td>
</tr>
<tr>
<td>Ent = 136069412</td>
</tr>
<tr>
<td>Ret = ret(a,_h88);</td>
</tr>
<tr>
<td>no</td>
</tr>
<tr>
<td>?- get_call(q(X,c),Ent,Ret).</td>
</tr>
<tr>
<td>X = _h80</td>
</tr>
<tr>
<td>Ent = 136069444</td>
</tr>
<tr>
<td>Ret = ret(_h80,c);</td>
</tr>
<tr>
<td>no</td>
</tr>
</tbody>
</table>

get_calls(#CallTerm,-TableEntryHandle,-ReturnTemplate)  

tabling  

Identifies through backtracking each subgoal in the table which unifies with CallTerm. For those that do, the handle to the table entry is assigned to the second argument, and its return template is constructed in the third. These latter two arguments must be uninstantiated at call time. The error terms are the same as for get_calls/1.
Example 6.15.4

**Variant Predicate**

<table>
<thead>
<tr>
<th>?- get_calls(p(X,Y),Ent,Ret).</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = _h80</td>
</tr>
<tr>
<td>Y = 3</td>
</tr>
<tr>
<td>Ent = 136039156</td>
</tr>
<tr>
<td>Ret = ret(_h80);</td>
</tr>
</tbody>
</table>

X = 1
Y = _h94
Ent = 136039108
Ret = ret(_h94);

no

<table>
<thead>
<tr>
<th>?- get_calls(p(X,3),Ent,Ret).</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = _h80</td>
</tr>
<tr>
<td>Ent = 136039156</td>
</tr>
<tr>
<td>Ret = ret(_h80);</td>
</tr>
</tbody>
</table>

X = 1
Ent = 136039108
Ret = ret(3);

no

<table>
<thead>
<tr>
<th>?- get_calls(p(1,3),Ent,Ret).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ent = 136039156</td>
</tr>
<tr>
<td>Ret = ret(1);</td>
</tr>
</tbody>
</table>

Ent = 136039108
Ret = ret(3);

no

**Subsumptive Predicate**

<table>
<thead>
<tr>
<th>?- get_calls(q(X,Y),Ent,Ret).</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = a</td>
</tr>
<tr>
<td>Y = _h94</td>
</tr>
<tr>
<td>Ent = 136069412</td>
</tr>
<tr>
<td>Ret = ret(a,_h94);</td>
</tr>
</tbody>
</table>

X = _h80
Y = _h94
Ent = 136069444
Ret = ret(_h80,c);

X = _h80
Y = c
Ent = 136069444
Ret = ret(_h80,c);

no

<table>
<thead>
<tr>
<th>?- get_calls(q(a,Y),Ent,Ret).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y = _h88</td>
</tr>
<tr>
<td>Ent = 136069412</td>
</tr>
<tr>
<td>Ret = ret(a,_h88);</td>
</tr>
</tbody>
</table>

Y = _h88
Ent = 136069412
Ret = ret(a,_h88);

no

<table>
<thead>
<tr>
<th>?- get_calls(q(1,Y),Ent,Ret).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y = _h88</td>
</tr>
<tr>
<td>Ent = 136069412</td>
</tr>
<tr>
<td>Ret = ret(a,c);</td>
</tr>
</tbody>
</table>

Y = _h88
Ent = 136069444
Ret = ret(a,c);

no

no

**get_calls_for_table(+PredSpec,?Call)**

Identifies through backtracking all the subgoals whose predicate is that of *PredSpec* and which unify with *Call*. *PredSpec* is left unchanged while *Call* contains the unified result.
Example 6.15.5

<table>
<thead>
<tr>
<th><strong>Variant Predicate</strong></th>
<th><strong>Subsumptive Predicate</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><code>?- get_calls_for_table(p(1,3),Call).</code></td>
<td><code>?- get_calls_for_table(q(X,Y),Call).</code></td>
</tr>
</tbody>
</table>
| Call = p(_h142,3); | X = _h80  
| Call = p(1,_h143); | Y = _h94  
| no | Call = q(a,_h167); |
| `?- get_calls_for_table(p/2,Call).` | X = _h80  
| Call = p(_h137,3); | Y = _h94  
| Call = p(1,_h138); | Call = q(_h166,c); |
| no | X = _h80  
| | Y = _h94  
| | Call = q(_h166,_h167); |

get_returns(+TableEntryHandle,#ReturnTemplate) **Tabling**

Backtracks through the answers for the subgoal whose table entry is referenced through the first argument, TableEntryHandle, and instantiates ReturnTemplate with the variable bindings corresponding to the answer.

The supplied values for the entry handle and return skeleton should be obtained from some previous invocation of a table-inspection predicate such as get_call/3 or get_calls/3.
Example 6.15.6

|
| **Variant Predicate**
| \[- get\_calls(p(X,3),\text{Ent},\text{Ret}),
| \text{get\_returns(Ent,Ret)}.
| X = 2
| \text{Ent} = 136039156 \quad \% p(X,3)
| \text{Ret} = \text{ret}(2);
| X = 1
| \text{Ent} = 136039156
| \text{Ret} = \text{ret}(1);
| X = 1
| \text{Ent} = 136039108 \quad \% p(1,Y)
| \text{Ret} = \text{ret}(3);
| X = 1
| \text{Ent} = 136039108
| \text{Ret} = \text{ret}(3);
| no

| **Subsumptive Predicate**
| \[- get\_calls(q(a,c),\text{Ent},\text{Ret}),
| \text{get\_returns(Ent,Ret)}.
| \text{Ent} = 136069412 \quad \% q(a,Y)
| \text{Ret} = \text{ret}(a,c);
| \text{Ent} = 136069444 \quad \% q(X,c)
| \text{Ret} = \text{ret}(a,c);
| \text{Ent} = 136043988 \quad \% q(X,Y)
| \text{Ret} = \text{ret}(a,c);
| no
| \[- get\_calls(q(c,a),\text{Ent},\text{Ret}),
| \text{get\_returns(Ent,Ret)}.
| no

\text{get\_returns\_and\_tvs}(\text{+TableEntryHandle},\text{#ReturnTemplate},-\text{TruthValue}) \quad \text{Tabling}

Identical to \text{get\_returns/2}, but also obtains the truth value of a given answer, setting \text{TruthValue} to \text{t} if the answer is unconditional and to \text{u} if it is conditional. If a conditional answer has multiple delay lists, this predicate will succeed only once, so that using this predicate may be more efficient than \text{get\_residual/2} (although less informative).

\text{get\_returns}(\text{+TableEntryHandle},\text{#ReturnSkeleton},-\text{ReturnHandle}) \quad \text{Tabling}

Functions identically to \text{get\_returns/2}, but also obtains a handle to the answer given in the second argument.

\text{get\_returns\_for\_call}(\text{+CallTerm},?\text{AnswerTerm}) \quad \text{Tabling}

Succeeds through backtracking for each answer of the subgoal \text{CallTerm} which unifies with \text{AnswerTerm}. Fails if \text{CallTerm} is not a subgoal in the table or \text{AnswerTerm} does not unify with any of its answers or the answer set is empty. The answer is created in its entirety, including fresh variables; the call is \textit{not}
further instantiated. Of course the user may unify the call with its answer if desired.

**Example 6.15.7**

<table>
<thead>
<tr>
<th>Variant Predicate</th>
<th>Subsumptive Predicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>! - get_returns_for_call(p(1,Y),</td>
<td>! - get_returns_for_call(q(a,Y),</td>
</tr>
<tr>
<td>AnsTerm).</td>
<td>AnsTerm).</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Y = _h88</td>
<td>Y = _h88</td>
</tr>
<tr>
<td>AnsTerm = p(1, _h161);</td>
<td>AnsTerm = q(a,c);</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Y = _h88</td>
<td>Y = _h88</td>
</tr>
<tr>
<td>AnsTerm = p(1,3);</td>
<td>AnsTerm = q(a,b);</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Y = _h88</td>
<td>no</td>
</tr>
<tr>
<td>AnsTerm = p(1,2);</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td></td>
</tr>
<tr>
<td>! - get_returns_for_call(p(X,Y),</td>
<td></td>
</tr>
<tr>
<td>AnsTerm).</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>X = _h80</td>
<td>X = _h80</td>
</tr>
<tr>
<td>AnsTerm = q(b,c);</td>
<td>AnsTerm = q(a,c);</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

get_residual(#CallTerm,?DelayList) Tableing

variant_get_residual(#CallTerm,?DelayList) Tableing

call_residual/2 backtracks through the answers to each completed subgoal in the table that unifies with CallTerm. With each successful unification, this argument is further instantiated as well as DelayList.

**Example 6.15.8** For the following program and table

<table>
<thead>
<tr>
<th>Subgoal</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(1,X)</td>
<td>p(1,2)</td>
</tr>
<tr>
<td></td>
<td>p(1,3):- tnot(p(2,3))</td>
</tr>
<tr>
<td>p(1,3)</td>
<td>p(1,3):- tnot(p(2,3))</td>
</tr>
<tr>
<td>p(1,3)</td>
<td>p(1,3):- tnot(p(2,3))</td>
</tr>
<tr>
<td>p(2,3)</td>
<td>p(2,3):- tnot(p(1,3))</td>
</tr>
<tr>
<td>p(2,3)</td>
<td>p(2,3):- tnot(p(1,3))</td>
</tr>
</tbody>
</table>

:- table p/2.
p(1,2).
p(1,3) :- tnot(p(2,3)).
p(2,3) :- tnot(p(1,3)).
the completed subgoals are \( p(1,X), p(1,3), \) and \( p(2,3) \). Calls to \texttt{get_residual/2} will act as follows

\[
\text{\texttt{?- get_residual(p(X,Y),List).}}
\]

\[
X = 1 \quad \text{\% from subgoal } p(1,X) \\
Y = 2 \\
\text{List} = [];
\]

\[
X = 1 \quad \text{\% from subgoal } p(1,X) \\
Y = 3 \\
\text{List} = [\text{tnot} (p(2,3))];
\]

\[
X = 1 \quad \text{\% from subgoal } p(1,3) \\
Y = 3 \\
\text{List} = [\text{tnot} (p(2,3))];
\]

\[
X = 2 \quad \text{\% from subgoal } p(2,3) \\
Y = 3 \\
\text{List} = [\text{tnot} (p(1,3))];
\]

no

Since the delay list of an answer consists of those literals whose truth value is unknown in the well-founded model of the program (see Chapter 5) \texttt{get_residual/2} is useful to examine the residual program (e.g. for XASP).

For other purposes, it may be desired to examine the answers for a particular subgoal, rather than for all subgoals that unify with \texttt{CallTerm}. In this case, \texttt{variant_get_residual/2} can be used, which backtracks through all answers for \texttt{CallTerm} if \texttt{CallTerm} is a tabled subgoal with answers, and fails otherwise. For the above example, \texttt{variant_get_residual/2} behaves as follows:

\[
\text{\texttt{?- variant_get_residual(p(X,Y),List).}}
\]

\[
\text{no}
\]

\[
\text{\texttt{?- variant_get_residual(p(1,Y),List).}}
\]

\[
X = 1 \quad \text{\% from subgoal } p(1,X) \\
Y = 2 \\
\text{List} = [];
\]
CHAPTER 6. STANDARD AND GENERAL PREDICATES

X = 1 % from subgoal p(1,X)
Y = 3
List = [\texttt{tnot(p(2,3))}];

no

Error Cases

- CallTerm is not a callable term
  - \texttt{type_error\texttt{(callable\_term,CallTerm)}}
- CallTerm does not correspond to a tabled predicate
  - \texttt{permission\_error\texttt{(table\_access,non\_tabled\_predicate,CallTerm)}}

\texttt{table\_state(+CallTerm,?PredType,?CallType,?AnsSetStatus)} \texttt{Tabling}
\texttt{table\_state(+TableEntryHandle,?PredType,?CallType,?AnsSetStatus)} \texttt{Tabling}

Succeeds whenever CallTerm is a subgoal in the table, or TableEntryHandle is a valid reference to a table entry, and its predicate type, the type of the call, and the status of its answer set, unify with arguments 2 through 4, respectively.

XSB defines three sets of atomic constants, one for each output argument. Taken together, they provide a detailed description of state of the given subgoal within an evaluation. The valid combinations and their specific meaning is given in the following table. Notice that not only can these combinations describe the characteristics of a subgoal in the table, but they are also equipped to predict how CallTerm would have been treated had it been called at that moment.
<table>
<thead>
<tr>
<th>PredType</th>
<th>CallType</th>
<th>AnsSetStatus</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>variant</td>
<td>producer</td>
<td>complete</td>
<td>Self explanatory.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>incremental_needs_reeval</td>
<td>An incremental table that has been invalidated, and is therefore inconsistent with a KB and needs recomputation (which will be lazily done).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>incomplete</td>
<td>Self explanatory.</td>
</tr>
<tr>
<td>variant</td>
<td>no_entry</td>
<td>undefined</td>
<td>The call does not appear in the table.</td>
</tr>
<tr>
<td></td>
<td>producer</td>
<td>complete</td>
<td>Self explanatory.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>incomplete</td>
<td>Self explanatory.</td>
</tr>
<tr>
<td></td>
<td>subsumed</td>
<td>complete</td>
<td>The call is in the table and is properly subsumed by a completed producer.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>incomplete</td>
<td>The call is in the table and is properly subsumed by an incomplete producer.</td>
</tr>
<tr>
<td></td>
<td>no_entry</td>
<td>complete</td>
<td>The call is not in the table, but if it were to be called, it would consume from a completed producer.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>incomplete</td>
<td>The call is not in the table, but if it had been called at this moment, it would consume from an incomplete producer.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>undefined</td>
<td>The call is not in the table, but if it had been called at this moment, it would be a producer.</td>
</tr>
<tr>
<td></td>
<td>undefined</td>
<td>undefined</td>
<td>The given predicate is not tabled.</td>
</tr>
</tbody>
</table>

**table_dump(+OptionList)**

**table_dump(+Stream,#Term,+OptionList)**

`table_dump/[2,3]` provides an easy method to view subgoals and answers that are present in a table. Given an input `Term`, `table_dump/[2,3]` provides information about all tabled subgoals that are subsumed by `Term`; if `Term` is a variable, information is provided about all tables.

The information can be provided at three levels of aggregation, and the form of the information is determined by the options in `OptionList`.

- If the option `summary(true)` is set, the aggregate sum of subgoals and answers that are subsumed by `Term` is collected, along with the aggregate sum of calls to these subgoals. If `Term` is a variable this information is broken down by tabled predicates.
  - If `details(answers)` is set, a list is collected of every tabled subgoal `S` such that `S` is subsumed by `Term` along with the number of answers for each `S` along with a list of those answers and the truth value of...
each answer \( t \) if true and \( u \) if undefined). If Term is a variable this information is broken down by tabled predicates.

- If \texttt{details(subgoals)} is set, a list is collected of all subgoals \( S \) such that \( S \) is subsumed by Term along with the number of answers for each \( S \). However, unlike the action for \texttt{details(answers)} the actual list of answers for \( S \) is not returned. If Term is a variable this information is broken down by tabled predicates.

- If \texttt{details(false)} is set, no detail information is provided for the actual subgoals or their answers.

• If \texttt{OptionsList} contains the option \texttt{results(X)} for some variable \( X \), \( X \) will be instantiated upon backtracking to all information collected about the tables.

• If the option \texttt{output(true)} is set, the information is written to Stream or to \texttt{userout} in Prolog-readable form.

If not otherwise specified the default options are \texttt{summary(true)}, \texttt{details(false)}, \texttt{output(true)}.

\textbf{Example} Consider the program:

\begin{verbatim}
:- table p/2.
p(1,a).
p(1,b) :- p(2,b).
p(2,b) :- p(1,a).
p(3,X) :- q(X).

:- table q/1.
q(1). q(2).

:- table r/1.
r(a).

:- table s/2.
s(1,a). s(2,b). s(1,a1). s(2,b1).
\end{verbatim}

and suppose the top-level query \(?- p(X,Y)\) has been made. Then \texttt{table_dump/2} provides the following information (reformatted for readability):

\begin{verbatim}
| ?- table_dump(_X,[summary(true)]).
\end{verbatim}
As the above example shows, each line of the summary has the form:

\[
\text{summary} = \text{Pred/Goal} - \text{subgoals}(N_{\text{subgoals}}) - \text{total\_times\_called}(N_{\text{called}}) - \text{total\_answers}(N_{\text{answers}})
\]

where

- \text{Pred/Goal} is either a term indicator, if the Term argument of table_dump/[2,3] was a variable (to indicate there should be no filtering of tabled calls); or Term itself.
- \(N_{\text{subgoals}}\) are the total number tabled subgoals that are subsumed by \text{Pred/Goal} (perhaps including \text{Pred/Goal} itself).
• $N_{called}$ is the total number of times all subgoals subsumed by $Pred/Goal$ have been called.
• $N_{answers}$ is the total number of answers currently derived by all subgoals subsumed by $Pred/Goal$.

Each line of details has the form:

$$Details = \text{Pred/Goal} - \text{subgoals}(N_{subgoals}) - \text{details(List)}$$

where $Pred/Goal$ and $N_{subgoals}$ are as above. If $\text{details(answers)}$ was an input option

$$List = \text{Subgoal} - \text{times\_called}(N_{called}) - \text{answers}(N_{answers}) - \text{List\_of\_Answers} - \text{Status}$$

for each $Subgoal$ in the table subsumed by $Pred/Goal$. $N_{called}$ and $N_{answers}$ are as above, while $\text{List\_of\_Answers}$ contains $A - TV$ for each answer $A$ with truth value $TV$ that is currently derived for $Subgoal$. On the other hand, if $\text{details(subgoals)}$ was an input option

$$List = \text{Subgoal} - \text{times\_called}(N_{called}) - \text{answers}(N_{answers}) - \text{Status}$$

where all elements are as before. Finally $Status$ is

• $\text{completed}$ if $Subgoal$ has been completed; and
• $\text{scc}(N_{SCC})$ if $Subgoal$ is incomplete. $N_{SCC}$ is relative: if $N_{SCC}$ is greater than $M_{SCC}$ then $N_{SCC}$ is a descendent of $M_{SCC}$: i.e., subgoals in SCC $M_{SCC}$ depend on subgoals in SCC $N_{SCC}$. However, these numbers should only be used relatively: at a given state in the computation there may be fewer than $M_{SCC}$ Sccs $^{22}$.

**Error Cases**

• $\text{OptionList}$ is a variable, or contains a variable as an element
  - $\text{instantiation\_error}$

$^{22}$XSB keeps track of SCCs through an algorithm similar to depth-first search: the numbers associated with subgoals are the depth-first numbers of the minimal back-dependency of a subgoal (cf. [62])
• OptionList is not a list
  – type_error(list,OptionList)
• OptionList contains an element, 0, that is not a valid table_dump_option.
  – domain_error(table_dump_option,0)

print_incomplete_tables  module: tables
These predicates, which can be useful for debugging, print out each incomplete subgoal in the current state of the evaluation, followed by the ordinal number of the SCC to which that subgoal belongs. This information describes the dependencies among tabled predicates. In local evaluation (the default evaluation method for XSB) all subgoals in SCC \( m \) depend on all subgoals in SCC \( n \) if \( m < n \). Furthermore, all subgoals in a given SCC depend on one another.\(^{23}\)
As its name implies, `print_incomplete_tables/[0,1]` print out SCC information only for incomplete tables; for full information about the SCCs of a computation, forest logging must be used (cf. Section 10.3).

In `print_incomplete_tables/0`, the information is output to `stdout`.

**Example:** For the program

```prolog
:- table q/2.
q(0,_):- !,print_incomplete_tables.
q(3,A):- q(5,A).
q(N,A):- N1 is N - 1,q(N1,A).
```

the goal `?- q(5,foo)` will produce the output

```
q(5, foo)- scc(1).
q(4, foo)- scc(1).
q(3, foo)- scc(1).
q(2, foo)- scc(2).
q(1, foo)- scc(3).
q(0, foo)- scc(4).
```

get_scc_dumpfile(-Filename)  module: tables
If the Prolog flag `exception_pre_action` is set to `print_incomplete_tables` (its default setting is `none`), then when an exception is thrown, incomplete tables

\(^{23}\)This assumes that there is no early completion, which can remove dependencies. In batched evaluation, the dependencies are less exact – see [62] for details, as SCCs represent a dag of dependencies rather than a chain as in local evaluation.
and their SCC information are printed to a “SCC dumpfile” via \texttt{print_incomplete_tables/1}. Note that the information is output for the state of execution where the error was thrown, and so is more informative than calling \texttt{print_incomplete_tables/} when the error is caught. (No file is generated unless the exception is thrown over at least one incomplete table.)

This predicate returns the name of the last such file generated and fails if there is no such file. Files are written to the \$XSBDIR/etc directory with the prefix \texttt{scc_dump}. Users are responsible for removing these files.

Note that XSB backtraces (Section 12.5) provide information about the context in which an exception is thrown, but the SCC dumpfile provides explicit SCC information along with parameter values for tabled predicates.

\textbf{Error Cases}

- \textit{Filename} is a not a variable
  - \texttt{instantiation\_error}

\begin{verbatim}
get_residual_sccs(+Subgoal,+Answer,-SCCList) module: tables
get_residual_sccs(+Subgoal,+Answer,-SCCList,-DepList,-SignList) module: tables
\end{verbatim}

As discussed in Section 5.3.3, answers that are undefined in the well-founded semantics are stored in XSB along with their delay lists, forming a residual program. This residual program can be materialized through the various predicates discussed above, in particular \texttt{get_residual/2} and \texttt{variant\_get\_residual/2}.

At times it can be useful to view the residual program as a directed graph, for instance in order to understand why a given answer might be undefined. In a manner somewhat analogous to the incremental dependency graph (Section 5.6) the \textit{residual dependency graph} is a directed graph whose nodes are subgoal/atom pairs and whose edges are labelled with: 1) a sign indicating whether the edge is positive or negative; and 2) the label \textit{depends on or affects}.

\textbf{Example 6.15.9} Consider the program

\begin{verbatim}
:- table p/2.
p(1,2).
p(1,3):- tnot(p(2,3)).
p(2,3):- tnot(p(1,3)).
p(2,3):- r(a).
r(a):- tnot(r(b))
r(b):- tnot(r(a)).
\end{verbatim}
to which the query ?- p(1,X) was made, generating the tables:

<table>
<thead>
<tr>
<th>Subgoal</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(1,X)</td>
<td>p(1,2)</td>
</tr>
<tr>
<td></td>
<td>p(1,3):- tn(p(2,3))</td>
</tr>
<tr>
<td>p(1,3)</td>
<td>p(1,3):- tn(p(2,3))</td>
</tr>
<tr>
<td>p(2,3)</td>
<td>p(2,3):- tn(p(1,3))</td>
</tr>
<tr>
<td></td>
<td>p(2,3):- tn(r(a))</td>
</tr>
<tr>
<td>r(a)</td>
<td>r(a):- tn(r(b))</td>
</tr>
<tr>
<td>r(b)</td>
<td>r(b):- tn(r(a))</td>
</tr>
</tbody>
</table>

The residual dependency graph for this program and query would have a node for each subgoal/answer combination with an undefined truth value, and a dependency edge for nodes $S_1/A_1$ and $S_2/A_2$ if $A_2$ occurs in a literal in the delay list for $S_1/A_1$, and the original subgoal for $A_2$ was $S_2$ in the subcomputation for $S_1$. The edge also has a sign indicating whether $A_2$ occurs positively or negatively in the delay list for $A_1$. In this example, the residual dependency graph could be represented as

depends_on(p(1,X),p(1,3),p(2,3),p(2,3),neg).
depends_on(p(1,3),p(1,3),p(2,3),p(2,3),neg).
depends_on(p(2,3),p(2,3),p(1,3),p(1,3),neg).
depends_on(p(2,3),p(2,3),r(a),r(a),pos).
depends_on(r(a),r(a),r(b),r(b),neg).
depends_on(r(b),r(b),r(a),r(a),neg).

Using the residual dependency graph, a user may be able to determine why an answer $A$ to a subgoal $S$ was unexpectedly undefined either because $S/A$ was involved in or depended on a loop through negation; or because $S/A$ depended on some other answer that was undefined because of the use of bounded rationality (Section 5.5) or because of floundering and the use of u_not/1.

The residual dependency graph can be constructed in a straightforward way from variant_get_residual/2. get_residual_sccs/[3,5] provides an alternate view that is slightly higher-level and much faster. Given a subgoal/answer pair as input, each of these predicates constructs SCC-based information about the residual dependency graph via structures of the form:

ret(Subgoal,Answer,SCCIndex).
where SCC is a numerical index for the SCCs of Subgoal. Two subgoals are in the same SCC iff they have the same index, however no other dependency information can be otherwise directly inferred from the index.\footnote{The actual number for each SCC index depends on how the residual dependency graph happens to be traversed; as a result it is best to rely on the index only as a “generated” name for each SCC.}

To obtain dependency information, get_residual_sccs/5 also returns a list indicating the direct dependencies among the SCCs, along with a list indicating whether given SCCs contain a negative edge. For the example above, the SCC information would have a form such as:

\[
[ \text{ret}(\text{p}(1,X),\text{p}(1,3),1), \text{ret}(\text{p}(1,3),\text{p}(1,3),2), \text{ret}(\text{p}(2,3),\text{p}(2,3),2), \\
\text{ret}(\text{r}(a),\text{r}(a),3), \text{ret}(\text{r}(b),\text{r}(b),3) ]
\]

The dependency list would have a form such as:

\[
[ \text{depends}(1,2), \text{depends}(2,3) ]
\]

while the sign list would have a form such as:

\[
[ \text{sign}(1,\text{no_neg}), \text{sign}(2,\text{neg}), \text{sign}(3,\text{neg}) ]
\]

If it is necessary to know which subgoal(s) in SCC1 directly depends on which subgoal(s) in SCC2, the information can be easily reconstructed from the output of get_residual_sccs/[4,5] using variant_get_residual/2. A similar approach can be used to determine the actual edges within a given SCC.

SCC detection is implemented using Tarjan’s algorithm in C working directly on XSB’s data structures. The algorithm is $O(|V| + |E|)$ where $|V|$ is the number of vertices and $|E|$ the number of edges in the dependency graph. As a result, get_residual_sccs/3 provides an efficient means to materialize the high-level topography of the dependency graph.\footnote{Currently, the materialization of dependency information between SCCs is implemented in a naïve manner, so that get_residual_sccs/6 is $O(|V|^2)$.}

\begin{verbatim}
explain_u_val(+Subgoal,+Answer,-Reason) module: tables
explain_u_val(+Subgoal,+Answer,-Sccs,-Deps,-Signs,-Reason) module: tables

The XSB predicate explain_u_val(+Subgoal,+Answer,?Reason) can be used to query why Answer is undefined when derived in an evaluation of Subgoal. Reason may be
\end{verbatim}
• **negative_loops(cycle)** if the derivation of **Answer** involves a loop through though negation that includes **Answer** itself.

• **negative_loops(dependent)** if the derivation of **Answer** depends on an atom that is involved in a loop through though negation.

• **unsafe_negation** if the derivation of **Answer** depends on a negative subgoal that is non-ground (XSB does not automatically perform subgoal reordering). The action of making a non-ground subgoal undefined is performed by **u_not/1**.

• **bounded_rationality** if the derivation of answer depends on bounded rationality based on radial restraint [32].

These reasons are not exclusive, and complex derivations may well involve several of the above reasons.

**Example 6.15.10** After executing the query **p** to the program

```prolog
:- table p/0, q/0, r/0, s/1.
p:- q, tnot p. p:- s(f(f(f(f(0))))).
q:- tnot r. r:- tnot q.
s(f(X)):- s(X). s(0).
```

where the bounded rationality depth has been set to 3. The query **explain_u_value(p,P,Reason)** will bind **Reason** to **negative_loops(cycle)**, to **negative_loops(dependent)**, and to **bounded_rationality** (this ordering is not guaranteed).

### 6.15.3 Abolishing Tables and Table Components

The following predicates are used to *abolish* tables: to ensure that they are not used by new computations and to reclaim their space when it is safe to do so. The use of the word “tables” in this section is rather unspecific. For
the purpose of deletion a table can either refer to a single subgoal and its answers, or to all subgoals and answers for a tabled predicate. Predicates are provided to abolish tables not only for particular predicates and subgoals, but for all tabled predicates, all tabled predicates in a module, and in the multi-threaded engine all thread-private tabled predicates or all thread-shared tabled predicates. Overall, these predicates share similar characteristics.

Abolishing a Table that is being Computed  An incomplete tabled subgoal $S$ may not be directly abolished by the user. This restriction is made since if $S$ is incomplete there may be pointers to $S$ from various elements of the current execution environment, and removing all of these pointers may be difficult to do. (Not to mention that abolishing an incomplete table has a murky semantics.) Accordingly, calling an abolish_xxx predicate on an incomplete table raises an error.

However, note that incomplete tables may be abolished automatically by XSB on exceptions, and when the interpreter level is resumed. Because tabled computation is more complex than Prolog computation, error handling must be correspondingly more complex. Suppose an exception is thrown over some incomplete table, so that the system looks for some catch/3 or similar call that will catch the error. In order to ensure safe space reclamation, XSB looks for the catcher $C$ that is nearest to the throw, but is also between SCCs. Both XSB’s command line interpreter and the interpreter XSB uses when embedded in a process use a top-level catch/3 goal, which is considered to be “between” SCCs, so that a thrown error will eventually be caught.

Because of the complexity of error handling in tabled computations, it is usually best to ensure that user-level catches are close to where an exception may be thrown so that there is no goal to an incomplete table between the thrower and catcher. In such a case XSB’s error handling mechanism conforms to the ISO standard for Prolog.

View Consistency and Table Garbage Collection  If one of the table abolish predicates is called when the current execution environment contains a failure continuation (i.e., a choice point) to an answer $A$ in a completed table $T$, space for $T$ is not immediately reclaimed. Rather the space for $T$ will be reclaimed by the table garbage collector at a later point. More precisely, if the current global tabling environment (including suspended states) has either

- a choice point that points to an answer $A$ in $T$;
or a (heap) delay list that points to a subgoal \( S \) in \( T \)

we say that \( T \) is *active*. Also, since tables can be abolished and rederived during the course of an evaluation, the table deletion system marks the tables with versions. Accordingly, if a tabled predicate \( P_{\text{version}} \) or subgoal \( S_{\text{version}} \) to be abolished is active in the current environment, reclamation of space for that version of \( P \) or \( S \) will be delayed until no answers for \( P_{\text{version}} \) or \( S_{\text{version}} \) are active. Meanwhile the older version of table will be available for backtracking, ensuring view consistency for the choice points. New calls to \( P \) or \( S \), however, will force rederivation of a new table version, rather than using the abolished information.

**Maintenance of the Residual Program** When conditional answers are present, abolishing a specific table or call may lead to semantic or implementational complications. Consider the conditional answer \( r(a,b) :- \text{undef} \) from Figure 6.1. If the predicate \( r/2 \) (or subgoal \( r(a,X) \)) is abolished and later rederived, the rederivation of \( r(a,X) \) might have different semantics than the original derivation (e.g. if \( \text{undef} \) depended on a database predicate whose definition has changed). From an implementation perspective, if space for \( r(a,X) \) is reclaimed, then the call \( \text{get\_residual}(p(a,X),Y) \) may core dump, even if there are no choice points for completed tables anywhere in the choice point stack. To address this problem, by default abolishing a subgoal \( S \) (predicate \( P \)) will abolish all subgoals (predicates) that (transitively) depend on \( S \) (\( P \)). In this case the goal \( \text{abolish\_table\_call}(r(a,X)) \) would cause the deletion of \( p(a,X) \) while the goal \( \text{abolish\_table\_pred}(r/2) \) would cause the deletion of \( p/2 \), since there are tabled subgoals of \( p/2 \) that depend on \( r/2 \). Only dependencies from subgoals or answers to the answers that are conditional on them are taken into account for table deletion: thus the deletion \( r(a,X) \) deletes \( p(a,X) \), but not \( \text{undef} \).

Users with programs that give rise to conditional answers in completed tables are encouraged to maintain this default behavior. However the default behavior may be changed either by setting a Prolog flag:

\[
?- \text{set\_prolog\_flag(table\_gc\_action,abolish\_tables\_singly)}.\]

or by calling a 2-ary abolish command with \texttt{abolish\_tables\_singly} in the options list.

\[26\text{Dao Tran Minh contributed to implementing this functionality.}\]
Abolishing Incremental Tables In XSB, incremental tables react to changes in underlying dynamic predicates and/or external events (cf. Section 5.6). To support this, XSB maintains an Incremental Dependency Graph (IDG) among incrementally tabled subgoals and incremental dynamic predicates. When an incremental table $T$ is abolished, the IDG needs to be restructured. Fortunately, with lazy incremental tabling as used by XSB, the only maintenance needed for the IDG outside of $T$ is to delete direct links between other IDG tables and $T$. In addition, all tables that depend on $T$ are incrementally invalidated. As a result, if some $T'$ which had previously depended on $T$ is called after $T$ was abolished, $T'$ will be seen to be incrementally invalid and will be recomputed. This recomputation will re-insert $T$ into the IDG in a manner that reflects the new state of the program.

Multiple Threads In the multi-threaded engine abolishing tables private to a thread behaves exactly as in the sequential engine, regardless of whether the tables are complete or incomplete, or contain conditional answers. In addition, when a thread $T$ exits (by normal termination or via an exception), tables private to $T$ are abolished automatically and their space reclaimed, as are any
incomplete shared tables owned by \( T \) in local evaluation. Shared tables can be abolished by the user at any time, but their space will not be reclaimed until there is a single active thread.

Table Deletion Predicates

\[ \text{abolish_table_pred}(+\text{Pred}) \]

Tabling

Invalidates all tabled subgoals for the predicate denoted by the predicate or term indicator \( \text{Pred} \). If any subgoal for \( \text{Pred} \) contains an answer \( A \) that is active in the current environment, \( \text{Pred} \) space reclamation for the \( \text{Pred} \) tables will be delayed until \( A \) is no longer active; otherwise the space for the \( \text{Pred} \) tables will be reclaimed immediately.

If \( \text{Pred} \) has a subgoal that contains a conditional answer, the default behavior will be to transitively abolish any tabled predicates with subgoals having answers that depend on any conditional answers of \( S \). This default may be changed either by setting a Prolog flag:

?- set_xsb_flag(table_gc_action,abolish_tables_singly).

or by calling \( \text{abolish_table_pred}/2 \) with the appropriate option. If the transitive abolishers are turned off, and \( \text{Pred} \) contains a conditional answer, the warning

\[ \text{abolish_table_pred}/[1,2] \text{ is deleting a table with conditional answers: delay dependencies may be corrupted.} \]

will be issued.

In the multi-threaded engine, if \( \text{Pred} \) is shared, reclamation for \( \text{Pred} \) will be delayed until there is a single active thread and no answer in \( \text{Pred} \) is active in the current execution environment. Otherwise, the behavior of \( \text{abolish_table_pred}/1 \) is the same as in the sequential engine.

Finally, \( \text{abolish_table_pred}/1 \) will throw an error if the predicate to be abolished is incremental. Until \( \text{abolish_table_pred}/[1,2] \) is extended to support incremental tables, use \( \text{abolish_table_call}/[1,2] \) or \( \text{abolish_all_tables}/0 \).

Error Cases

- \( \text{Pred} \) is not instantiated
  - instantiation_error

- \( \text{PredSpec} \) is not a predicate_indicator or a term_indicator
  - domain_error(predicate_or_term_indicator,Pred)
• PredSpec does not indicate a tabled predicate
  – table_error
• PredSpec indicates an incrementally tabled predicate.
  – permission_error
• There is currently an incomplete table for an atomic subgoal of Pred.
  – permission_error

abolish_table_pred(+CallTerm,+Options)  Tableing
Behaves as abolish_table_pred/1, but allows the default table_gc_action
to be over-ridden with a flag, which can be either abolish_tables_transitively
or abolish_tables_singly.

Error Cases Error cases are the same as abolish_table_pred/1 but
with the additions:
• Options is a variable, or contains a variable as an element
  – instantiation_error
• Options is not a list
  – type_error(list,Options)
• Options contains an option O that is not a table abolish option.
  – domain_error([abolish_tables_transitively, abolish_tables_singly,O])

abolish_table_subgoals(+Subgoal)  Tableing
Invalidates the table for any subgoal that unifies with Subgoal. If a subgoal
S unifying with Subgoal contains an answer A that is active in the current
environment, the table entry for S will not be reclaimed until A is no longer
active; otherwise the space for S will be reclaimed immediately.
If S contains a conditional answer, the default behavior will be to transi-
tively abolish any subgoals that depend on any conditional answers of S.
This default may be changed either by setting an XSB flag:

?- set_xsb_flag(table_gc_action,abolish_tables_singly).

or by calling abolish_table_call/2 with the appropriate option. If the
transitive abolishes are turned off, and S contains a conditional answer,
the warning
abolish_table_call/1 is deleting a table with conditional answers:
delay dependencies may be corrupted.
will be issued.
In the multi-threaded engine, if S is a subgoal for a predicate that is shared,
reclamation for S will be delayed until there is a single active thread and
no answer in $S$ is active in the current execution environment. Otherwise, the behavior of `abolish_table_call/1` is the same as in the sequential engine on tabled predicates that are thread-private.

For incremental tables, `abolish_table_call/[1,2]` not only deletes the table structures for `Subgoal`, but pointers to `Subgoal` in the Incremental Dependency Graph (IDG), after invalidating all subgoals that depend on `Subgoal`. The node and edges for `Subgoal` will be reinserted into the IDG when `Subgoal` is re-evaluated, either lazily or by an explicit update command.

**Error Cases**

- The term spec `Subgoal` does not correspond to a tabled predicate:
  - `table_error`
- The term spec `Subgoal` unifies with a tabled subgoal that is incomplete:
  - `permission_error`
- The term spec `Subgoal` is a cyclic term:
  - `table_error`

`abolish_table_subgoals(+Subgoal,+Options)`

Behaves as `abolish_table_subgoals/1`, but allows the default `table_gc_action` to be over-ridden with a flag, which can be either `abolish_tables_transitively` or `abolish_tables_singly`.

**Error Cases** Error cases are the same as `abolish_table_call/1` but with the additions:

- `Options` is a variable, or contains a variable as an element
  - `instantiation_error`
- `Options` is not a list
  - `type_error(list,Options)`
- `Options` contains an option $O$ that is not a table abolish option.
  - `domain_error([abolish_tables_transitively, abolish_tables_singly,$O$])`

`abolish_table_subgoal(+Subgoal)`

`abolish_table_subgoal(+Subgoal,+Options)`

These predicates behave as `abolish_table_subgoals/[1,2]`. However rather than abolishing all tables whose subgoal unifies with `Subgoal` they only abolish the table whose subgoal is a variant of `Subgoal`, if such a table exists.
abolish_all_tables

In the single-threaded engine, removes all tables presently in the system and frees all the memory held by XSB for these structures. Predicates that have been declared tabled remain so, but information in their table is deleted. `abolish_all_tables/0` works directly on the memory structures allocated for table space. This makes it very fast for abolishing a large amount of tables, and to maintain its speed it throws an error if the current execution environment contains any incomplete tables, or any active completed tables. `abolish_all_tables/0` can be used regardless of whether there are incremental tables, or tables that use call or answer subsumption.

In the multi-threaded engine `abolish_all_tables/0` additionally raises an error unless it is called when there is a single active thread. In that case, all shared tables are abolished as well as all private tables for the main thread.

Error Cases

- There are incomplete tables at the time of the predicate’s call;
  - `permission_error`
- The current execution environment has an active completed table \( T \)
  - `permission_error`
- (Multi-threaded engine only) More than one thread is active:
  - `table_error`

abolish_nonincremental_tables

Abolishes all tabled calls for predicates that are not declared to be incremental. \(^{27}\) This predicate allows XSB to function in a manner similar to that of a deductive database: incremental tables will be automatically updated when the data they depends on changes; while non-incremental tables, which may have become invalid, can be abolished.

As currently implemented, `abolish_nonincremental_tables/1` traverses through each nonincremental tabled predicate, `Pred`, and if `Pred` has any incomplete subgoals, a permission error will be thrown. However, unlike with `abolish_all_tables/0` no errors will be thrown if there are active completed tables: rather these tables will be marked for deletion and their space later garbage collected. In addition, no error will be thrown if there are incomplete incremental subgoals.

Error Cases

\(^{27}\)Calls for predicates that are declared as opaque are considered to be non-incremental.
• There are incomplete nonincremental tables at the time of the predicate’s call;
  – permission_error

abolish_all_private_tables

Tabling

In the multi-threaded engine, removes all tables private to the thread and frees all the memory held by XSB for these structures, including space for conditional answers. Predicates that have been declared tabled remain so, but information in their table is deleted. Like abolish_all_tables/0, abolish_all_private_tables/0 works directly on the memory structures allocated for table space. This makes it very fast for abolishing a large amount of tables, and to maintain its speed it throws an error if the current execution environment contains any incomplete tables, or any active completed tables. abolish_all_private_tables/0 can be used regardless of whether there are incremental tables, or tables that use call or answer subsumption.

Error Cases

• There are incomplete tables at the time of the predicate’s call;
  – permission_error
• The current execution environment for the thread has an active private table T for the current thread
  – table_error

abolish_all_shared_tables

Tabling

In the multi-threaded engine, removes all tables private to the thread and frees all the memory held by XSB for these structures, including space for conditional answers. Predicates that have been declared tabled remain so, but information in their table is deleted. abolish_all_private_tables/0 works directly on the memory structures allocated for table space. This makes it very fast for abolishing a large amount of tables, and to maintain its speed it throws an error if the current execution environment contains any incomplete tables, or any active completed tables. abolish_all_private_tables/0 can be used regardless of whether there are incremental tables, or tables that use call or answer subsumption. In addition, abolish_all_shared_tables/0 raises an error unless it is called when there is a single active thread. If called with a single active thread, all shared tables are abolished, but private tables for the main thread are unaffected.

Error Cases

• There are incomplete tables at the time of the predicate’s call;
permission_error

- The current execution environment has an active table $T$
  - permission_error
- More than one thread is active:
  - table_error

abolish_module_tables(+Module)

Tabling

Given a module name (or the default module, usermod), this predicate abolishes all tables for each tabled predicate in Module. It is implemented using a series of calls to abolish_table_pred/1 and so inherits the behavior of that predicate.

gc_tables(-Number)

Tabling

When a tabled subgoal or predicate is abolished, reclamation of its space may be postponed if the subgoal or predicate has an answer that is active in the current environment. A garbage collection routine is called at various points in execution to check which answers are active in the current environment, and to reclaim the space for subgoals and predicates with no active answers. In particular, space for all abolished tables is reclaimed whenever the engine re-executes the main command-line or C thread interpreter code. However in rare situations this strategy may not be adequate. For this reason, the user can explicitly call the table garbage collector to reclaim space for any deleted tabled predicates or subgoals that no longer have active answers.

gc_tables/1 always succeeds, unifying Number to −1 if garbage collection was not attempted (due to multiple active threads) and otherwise to the number of tables still unreclaimed at the end of garbage collection.

Error Cases

- Number is not a variable
  - type_error(variable)

delete_return(+TableEntryHandle,+ReturnHandle)

Tabling

Removes the answer indicated by ReturnHandle from the table entry referenced by TableEntryHandle. The value of each argument should be obtained from some previous invocation of a table-inspection predicate.

This predicate is low-level so no error checking is done. In Version 3.6, this predicate does not reclaim space for deleted returns, but simply marks the returns as invalid.

Warning: While useful for purposes such as tabled aggregation, delete_return/2 can be difficult to use, both from an implementation and semantic perspective.
Chapter 7

Multi-Threaded Programming in XSB

id with Version 3.0, XSB supports the use of POSIX threads to perform separable computations, and in certain cases to parallelize them. POSIX threads have a simple and clear API, and are available on all Unixes and by using open-source libraries, on Windows as well (see Section 7.8 to configure under Windows). This chapter introduces how to program with threads in XSB through a series of examples sections discuss performance aspects of our implementation as well as describing relevant predicates. A general knowledge of multi-threaded programming is assumed, such as can be found in [45, 8].

7.1 Getting Started with Multi-Threading

In Version 3.6 the default configuration of XSB does not include multi-threading. This is partly because multi-threading is new, and despite our efforts, the multi-threaded engine may contain bugs not present in the single-threaded engine. However the main reason is because in Version 3.6, not all libraries and packages have yet been made thread-safe so that not all configurations are supported with multi-threading. Both the XSB-calling-C and the C-calling-XSB interfaces are supported in the multi-threaded engine. All XSB libraries have been ported to the multi-threaded engine except the profiling library and the string library (which is not yet thread-safe). The packages ODBC and CHR, FLORA-2, and regmatch are supported by the multi-threaded engine, but the packages dbdrivers, xpath, interprolog, smodels, perlmatch, libwww and posix are not yet fully supported. We note, however that all basic/ISO
Prolog functionality is thread-safe (at least, as far as we know :-).

With this in mind, making the multi-threaded engine is simple: configure and make XSB as in Chapter 3, but include the command \texttt{-enable-mt}. When you invoke the newly made configuration of XSB you should see \texttt{engine: multi-threading} in the configuration list below the banner rather than \texttt{engine: slg-wam} as in the sequential engine.

**Hello World for Beginners** We naturally start with a program to print “hello world”. Within the multi-threaded engine, import \texttt{thread_create/2} from the module \texttt{thread}, and type the command

```prolog
?- thread_create(writeln('hello world'),Id)
```

you should see something like

```
Id = 1hello world
```

while the output is a little ugly, the “hello world” program does illustrate simple multi-threading at work. The calling thread (i.e. the thread controlling the command-line interpreter which we call \( T_{prompt} \)) executes the predicate \texttt{thread_create/2} which creates a thread \( T_{child} \) and immediately returns with the \texttt{XSB thread id} of the created thread. Meanwhile, \( T_{child} \) initializes its stacks and other memory areas and executes the goal \texttt{writeln('hello world')}. \( T_{child} \) and \( T_{prompt} \) share most of their process-level information: in particular they share a common I/O stream for standard output, leading to the output above. What is happening may be seen a little more easily by executing the command

```prolog
?- thread_create((sleep(1),writeln('hello world')),Id)
```

In this case the interpreter reports that \( F \) is bound to a thread id, then about a second later \texttt{writeln/1} is executed.

The simple “hello world” program illustrates a couple of points. First, it is easy to create a thread in XSB and have that thread do work. Second, it can be tricky to coordinate actions among threads. We’ll explore these two themes in more detail, but first suppose we are determined to extend our multi-threaded program so that it produces good output. One way to do this is to \texttt{join} \( T_{prompt} \) and \( T_{child} \) as follows

```prolog
?- thread_create(writeln('hello world'),Id),
```
thread_join(Id,ExitCode).

hello world

Id = 1
ExitCode = true

In this case, as soon as $T_{\text{prompt}}$ has issued a command to create $T_{\text{child}}$, it executes thread_join/2. This latter predicate makes a system call to the underlying operating system to suspend $T_{\text{prompt}}$ until $T_{\text{child}}$ has exited. thread_join/2 returns a status term indicating whether the goal to thread Id succeeded, failed, exited with an error term, or was cancelled (in this case Id succeeded).

So far, we've introduced a few concepts that have not been fully discussed. First is the concept of an XSB thread id: XSB manages up to $M$ active threads using XSB thread ids. The default for $M$ in Version 3.6 is 1024, but $M$ can be reset via the max_threads command line option to XSB (cf. Section 3.7). Once XSB is initialized, the maximum number of threads for an XSB session can be obtained at runtime via the Prolog flag max_threads (cf. Section 6.12). It should be noted that the XSB thread id of a thread is different from the identifier of the underlying Pthread. An XSB thread id is a Prolog term, and unlike POSIX thread ids, XSB thread ids can be compared for equality using unification. The actual form of an XSB thread id, however, is subject to change between versions, so programs should not make use of the exact form of an XSB thread id. In the multi-threaded engine, the XSB thread id of any thread can be queried using the predicate thread_self/1.

### 7.2 Communication among Threads

**Example 7.2.1** Consider the program fragment

```prolog
:- dynamic p/1.

test:- thread_create(assert(p(1)),_X).
```

*If you type the goal ?- test and then the goal ?- p(X), the call p(X) will fail.*

This illustrates an important point about dynamic and tabled predicates in the multi-threaded engine: by default clauses for a dynamic predicate p/n are private to the thread that asserts them; and by default tables created in an evaluation of a goal for p/n are private to the thread that evaluates the goal. This behavior contrasts to
that of static code which is always shared between threads. In the example above, to allow \( p(1) \) to be visible to various threads, \( p/1 \) must be declared to be shared with the following declaration.

\[
:=- \text{table } p/1 \text{ as shared.}
\]

or

\[
:=- \text{dynamic } p/1 \text{ as shared.}
\]

Alternately, dynamic and tabled predicates can be made thread-shared by default by invoking XSB with the command-line argument \(-\text{shared_predicates}\), in which case a predicate may be declared thread-private through the declaration

\[
:=- \text{table } p/1 \text{ as private.}
\]

or

\[
:=- \text{dynamic } p/1 \text{ as private.}
\]

The ability to share dynamic code between predicates provides an extremely powerful mechanism for threads to communicate. So why does XSB make dynamic predicates thread-private by default? The main reason for this is that if dozens or hundreds of threads are running concurrently, shared dynamic code becomes an expensive synchronization point. Code for shared predicates must be more heavily mutexed than code for private predicates. In the case of dynamic code, XSB does not always immediately reclaim the space of retracted clause, to avoid the possibility of some computation backtracking into a clause that has been reclaimed. Rather, (like most Prologs), XSB may decide to garbage collect the space of the retracted clauses at a later time. While clause garbage collection is simple enough to implement for a single thread, garbage collecting clauses for shared dynamic predicates is difficult to do when multiple threads are active. Accordingly, in Version 3.6, space for shared dynamic clauses is not reclaimed until there is a single active thread. However for \textit{thread-private} dynamic predicates, there is no problem in reclaiming space when multiple threads are active: from the engine’s perspective garbage collection is no different than in the sequential case. Thus one set of reasons for making dynamic predicates private by default are based on efficiency \textsuperscript{1}.

\textsuperscript{1}Future versions may offer more powerful garbage collectors for shared predicates.
The second reason for making dynamic predicates thread-private by default is semantic. Suppose thread $T_1$ starts a tabled computation that depends on the dynamic shared predicate $p/1$. While $T_1$ is computing the table, thread $T_2$ asserts a clause to $p/1$. $T_1$’s table is likely to be inconsistent, leading to the problem of read consistency of any table that depends on thread-shared dynamic predicates. In Version 3.6, users are responsible for ensuring read consistency of any tables that depend on shared dynamic data. Future versions of XSB are intended to allow more sophisticated mechanisms for read consistency.

Not only can tables depend on thread-shared or thread-private dynamic data, but the tables themselves may be thread-shared or thread-private. Like dynamic code, the declaration `table Predspec as shared` allows sharing of tables for a predicate evaluated with call-variance to be shared among threads. To some extent, tabling considerations for making a predicate thread-shared or thread-private are like those of dynamic code. Thread-private tables require fewer synchronization points overall. The situation for reclaiming space for abolished tables is analogous to reclaiming space for retracted dynamic clauses: the garbage collector treats abolished tables for thread-private predicates as in the sequential case, while space for shared tables is not reclaimed until there is a single active thread. However the precise semantics of how tabling information is shared depends on whether the multi-threaded engine is configured with the default local evaluation or with batched evaluation. As discussed in Chapter 5, local evaluation is so-named because computation always takes place in the SCC most recently created, and no answer is returned outside of an SCC until the SCC has been completely evaluated. Within this scheduling strategy it is not often useful to share answers between tables that have not been completed – as local evaluation would allow these answers to be returned only if the tables were in the same SCC. This leads to a concurrency semantics called Shared Completed Tables [49, 50, 52]. Shared Completed Tables can in fact be supported by a relatively simple algorithm for optimistic concurrency control. If goals to two mutually dependent tables $Table_a$ and $Table_b$ are called concurrently by two different threads, $Thread_a$ and $Thread_b$, nothing is done until it is detected that $Table_a$ and $Table_b$ are both incomplete and are contained in the same SCC of the table dependency graph. At that time, one of the threads (e.g. $Thread_a$) takes over recomputation of all tables in the SCC, and when the SCC is completed, any remaining answers are returned to other threads that had invoked goals in the SCC. While $Thread_a$ is completing this computation, $Thread_b$ suspends until the SCC is complete. Thus the semantics of Shared Completed Tables supports concurrency for the well-founded semantics, but only supports the most coarse-grained parallelism.

---

In Version 3.6, tabled predicates using call-subsumption are always private; an attempt to make such a predicate thread-shared throws an exception.
Batched evaluation, on the other hand, allows answers to be returned outside of an SCC before that SCC has been completed. Concurrency control for batched evaluation is similar to that for local evaluation, except in the following case. Assume as before that Table\(_a\), first called by Thread\(_a\), and Table\(_b\) first called by Thread\(_b\) are determined to be in the same SCC, and that Thread\(_a\) takes over computation of subgoals in the SCC. Now, Thread\(_b\), rather than suspending, may continue work. In particular, Thread\(_b\) can return any answers in Table\(_b\) that it finds whenever it finds them, regardless of whether they have been produced by Thread\(_b\) (before Thread\(_a\) took over the SCC) or by Thread\(_a\) (afterwards). We call this type of concurrency semantics, Table Parallelism. Table Parallelism can be used to program producer-consumer examples, as well as to implement Or- and And- parallelism. Table Parallelism was first introduced in [26], but the mechanism now used for implementing Table Parallelism differs significantly from what was described there. In Version 3.6 of XSB, the implementation of Table Parallelism is experimental: in particular, it does not yet support tabled negation.

As mentioned, for either semantics of shared tables, in Version 3.6, users of thread-shared tables are responsible for ensuring read consistency. Note that, in principle, thread-shared tables may depend on thread-private tables and vice-versa. Either type of table may depend on thread-private or thread-shared dynamic code. In addition, a predicate may be both dynamic and tabled, and its clauses and tables may be either thread-private or thread-shared.

### 7.3 Thread Statuses: Joinable and Detached Threads

So far we have assumed that the goal called in thread_create/2 terminates normally — by success or failure. But what if a thread throws an error while executing a goal? How long should error information for a thread persist, and how can it be checked?

Our approach relies on the semantics of Pthreads, which can be either joinable or detached. Within this framework, we consider a thread to be valid if it has not yet terminated, or if it is joinable and has not yet been joined. After a joinable Pthread T\(_{dead}\) has terminated, status information about T\(_{dead}\) persists until some other thread joins it — at which time the information is removed. On the other hand, if T\(_{dead}\) is detached, status information is removed as soon as T\(_{dead}\) terminates. Reclamation of thread status information may be contrasted to that of thread-specific data structures such as stacks. Upon normal or exceptional termination of T\(_{dead}\), any memory automatically allocated in the process of initializing T\(_{dead}\)'s, or executing its goal – including stacks, private dynamic code, private tables is reclaimed. In addition,
any mutexes held by $T_{\text{dead}}$, are released. On the other hand, XSB-specific status information about threads follows the Pthread model: by default, error information is available when joining a joinable thread, but not otherwise.

**Example 7.3.1** Suppose the goal

```prolog
?- thread_create(functor(X,Y,Z),F).
```

is executed. By default, this will produce the result

\[
\begin{align*}
X &= \_h113 \\
Y &= \_h127 \\
Z &= \_h141 \\
F &= 1++Error[XSB/Runtime/P]: [\text{Instantiation}] \text{ in arg 2 of predicate functor/3}
\end{align*}
\]

In fact, the variable bindings are output to **STDOUT**, while the error message

\[
++Error[XSB/Runtime/P]: [\text{Instantiation}] \text{ in arg 2 of predicate functor/3}
\]

is output to **STDERR**, and may be redirected. The call

```prolog
?- thread_join(2,Error).
```

returns

\[
\text{Error} = \text{exception(error(instantiation\_error, in arg 2 of predicate functor/3,}
\]

\[
[[\text{Forward Continuation...},..., \text{standard:call/1},..., \text{standard:catch/3}],\]
\]

\[
\text{Backward Continuation...})]]
\]

In other words, **Error** is instantiated to a **exception/1** structure, containing a standard XSB error term (including backtrace).

The error term in the above example is one example of a **thread status** term. In XSB, these thread statuses are as follows.

- **running** The thread is still executing

---

3This behavior can, of course, be overridden by embedding goals within catch/3 and handling errors separately, or simply by adding a default user error handler: see Chapter 12 for details.
• **true** The thread has exited and successfully evaluated its goal.

• **false** The thread has exited and failed its goal.

• **exception(Exception)** The thread has been terminated due to an uncaught exception, represented by the term `Exception` which is a standard XSB error term.

• **cancelled(Exception)** The thread has been terminated due to a thread cancellation, represented by the term `Exception` which is a standard XSB error term.

• **exited(ExitTerm)** The thread has been terminated using the predicate `thread exit/1` with `ExitTerm` as its argument.

Any of these statuses except **running** may be returned by `thread_join/2`. In Prolog, the statuses of exited threads provide much more information than C exit codes.

As with pthreads, XSB threads are created as joinable by default, but can be created as detached using an option in `thread_create/3`. Alternatively, a thread created as joinable can be made detached by `thread_detach/1`. All of the predicates mentioned in this section are fully described in Section 7.9.

### 7.4 Prolog Message Queues

While Prolog predicates can communicate through shared dynamic code and tables, message queues provide a useful mechanism for one thread to pass a command to another or to synchronize on the return of data. A Prolog message queue contains an arbitrary Prolog Term, and unification may be used to obtain a term from a queue. More specifically, when a producer writes `Term` into a queue, the term is copied into the queue so that no binding are shared between `Term` and the producer’s stacks. `Term` may include structures or lists and need not be bound, and any variable bindings within `Term` are preserved. When a consumer `T_cons` accesses the queue it provides a goal `G` and traverses the queue until it finds a term in the queue that unifies with `G`. If `T_cons` finds a term in the queue that unifies with `G`, it removes it from the queue and continues in its computation. If there is no term in the queue that unifies with `G`, `T_cons` will suspend until at least one other term is added to the queue. When it awakens it will retraverse the queue from the beginning to find a term that unifies with `G`.

\footnote{Note that this traversal is necessary since the position of `T_cons` may in the queue may not be valid due to the addition and deletion of terms by other threads.}

Because of the behavior of message queues, it is usually
good programming practice to ensure that terms written into the queue will unify with the goals of consumers. This can usually be done by abstracting a consumers goal (say to a variable, X) or by splitting one “multiplexed” queue into two separate queues.

A Prolog message queue can be public or private: a public message queue can have any number of readers and writers. In addition, each thread T also has a private message queue Q_T: any thread can write to Q_T but only T can read from it. The following example illustrates how to use private message queues:

test_private:-
  thread_id(Tid),
  thread_create(child(Tid),Id),
  thread_get_message('Mom Im home'(ChildId)),
  thread_send_message(ChildId,'Im in the kitchen'),
  thread_join(Id,_).

call(Parent):-
  thread_self(Id),
  thread_send_message(Parent,'Mom, Im home'(Id)),
  thread_get_message('Im in the kitchen').

If ?- test is called by T_parent, it will obtain its own thread id, create a new thread T_child to execute call/1, wait for a message that T_child is operational using thread_get_message/1, send a message to T_child using thread_send_message/2 and then wait for T_child to terminate. When it is created, T_child immediately sends a message to its parent, waits for a message back from its parent, and terminates.

It is illustrative to compare

test_public:-
  message_queue_create(Qid)
  thread_create(child(Qid),Id),
  thread_get_message(Qid,'Mom Im home'(ChildQ)),
  thread_send_message(ChildQ,'Im in the kitchen'),
  thread_join(Id,_),
  message_queue_destroy(Qid).

call(ParentQ):-
  message_queue_create(Qid),
  thread_send_message(ParentQ,'Mom, Im home'(Qid)),
  thread_get_message(Qid,'Im in the kitchen').
test_public is essentially the same program as test_private, but uses public message queues, rather than private queues. The public queues must be explicitly created and destroyed, and they are referred to via a queue id (or alias) rather than via a thread id (or alias). Like thread ids, queue ids in XSB are integers, but a user should not depend on their precise form: aliases should be used if a user wants control of queue or thread identifiers.

Thus, apart from who can read from them, private and public message queues have essentially the same behavior. In addition, any queue can be created with a bound, size on the number of messages (terms) it contains. If size is 0, the queue is taken to be unbounded. If a bounded queue already contains size elements, the producer will suspend until one or more elements are removed from the queue. For public queues, a size argument can be passed using the predicate message_queue_create/2 (See Section 7.9). For private queues, and for public queues created with message_queue_create/1, the value for size is taken from the settable Prolog flag max_queue_terms. The default value for max_queue_terms is currently 100.

7.5 Thread Cancellation and Signalling

There may be a number of situations in which it is useful to give one thread the ability to cancel the execution of another thread. Within the semantics of pthreads, this is called thread cancellation. At the C level, thread cancellation can be tricky, as mutexes must be released, allocated memory freed, and so on. Accordingly, the predicate thread_cancel/1 cancels XSB threads by acting purely within the SLG-WAM engine. When thread $T_1$ interrupts thread $T_2$, $T_1$ writes to the thread-specific XSB interrupt vector in $T_2$. Later, when $T_2$ checks its interrupt vector, it throws a cancellation error, which causes it to clean up its mutexes, memory, private tables and dynamic code, and then exit.

Thread cancellation is just a special case of Prolog thread signalling, in which one thread can signal another thread to interrupt what it is doing and execute a goal. The following code provides an example of thread signalling.

test_signal:-
thread_self(Tid),
thread_create(child(Tid),T1,[]),
thread_get_message('Im alive'),
thread_signal(T1,writeln('Excuse me, but did you just kick me?')),
thread_join(T1,_Ball),
writeln(test5_ok).

child(Tid):-
    thread_send_message(Tid,'Im alive'),
    loop.

loop:- loop.

test_signal begins like test_private, but rather than waiting for a signal from its parent, the child goes into an infinite loop. The signal interrupts the child, which writes out a message and returns to the infinite loop.

Thread signals may be any callable Prolog term. As with private message queues, each thread is created with its own private signal queue (there are no public signal queues). In XSB, threads handle Prolog signal interrupts (including cancellation messages) at the same time as attributed variable interruptions. This means that Prolog signal interrupts will be handled very quickly if SLG-WAM code is being executed. On the other hand, if a thread executing a builtin to, e.g. waiting on a mutex, the thread may be immediately awakened to process the signal, but not always: if a thread is waiting for input on a stream or socket, the thread may not handle the signal interrupt until the input is received. Furthermore, in a very few critical sections of code, thread signal handling may be distabled. However, the thread is guarenteed to handle the signal interrupt or cancellation message very shortly after it finishes the builtin.

So, while thread cancellation and signalling is useful, it must be used with a certain amount of care. Any thread can signal any other thread, and any thread can cancel any other thread, with the exception that the main thread, which controls the console (or interface to C or interprolog) cannot be cancelled. The main thread always has XSB thread id 0 in both the single-threaded and multi-threaded systems, and has the thread alias main.
7.6 Performance and other Considerations

For running programs that do not use multiple threads, the multi-threaded engine has a minimal overhead compared to the single-threaded engine. Times for single-threaded execution of Prolog or tabled programs range from about 10–20% slower to 10–20% faster for the multi-threaded engine compared to the single-threaded engine. Speedups for running multiple threads on multiple processors depends heavily on the applications run and on the underlying operating system.

The size of a given thread may be a consideration for multi-threaded applications, especially on a 32-bit platform (the multi-threaded engine has been tested on both 32-bit and 64-bit platforms). Each thread has an area of thread-private variables that are “global” to its own virtual machine. This area, called the thread context, which accounts for about 4 Kbytes of space. Much larger are the various stacks used by the threads for tabled and Prolog execution. Almost all of XSB’s memory areas are fully expandable, and the initial size of the execution stacks may be set explicitly as options in thread_create/3. Explicitly setting a default thread stack size for an XSB thread to be smaller than the default process stack size may be useful for applications that have a large number of concurrently running threads.

Other performance considerations involve the contention by threads for shared resources. As discussed above, contention may arise when creating or abolishing tables, or when asserting or retracting dynamic code — however in either case thread-private predicates give rise to less contention than thread-shared predicates. In terms of I/O, each XSB stream up to the maximum number of file descriptors has its own mutex; as a result threads writing to different streams will not contend for I/O. Thus, in multi-threaded applications, it may be more efficient to open and close streams and access these streams explicitly, than to redirect standard input or standard output through see/1 and tell/1.

7.7 Examples of Multi-Threaded Programs in XSB

Figure 7.1 shows an example of a multi-threaded goal server in XSB, which makes use of XSB’s socket library (see Volume 2 of this manual) \(^6\). The server listens for requests from clients using socket_accept/2 and spawns a thread to handle each request via the goal accept_client/2 which actually calls the goals. The goals executed by the server could be tabled and take advantage of the shared table implementation, shared dynamic code, or any other mechanism in XSB. Halting of the server is done by the

---

\(^6\)Material in this section is based on [49].
thread cancellation mechanism, and a shared dynamic predicate is used to make the
server’s thread identifier known to the other threads. Note that this is the reason a
specific thread was created to execute server_loop, as the main thread cannot be
canceled.

Figure 7.2!la uses a multi-threaded execution model to compute a series of prime
numbers in parallel ⁷. The master thread partitions the work and creates two worker
threads. The worker threads each compute its portion of the interval and return their
results to the master through a message queue.

Notice how the primes/2 predicate uses difference lists to avoid the use of the
append predicate ⁸, and while threads don’t share variables, the bindings of the terms
in the messages are correctly handled, allowing Prolog’s unification to assume its full
power. Although only two threads are used, the program could easily be extended to
use an arbitrary number of threads.

7.8 Configuring the Multi-threaded Engine under
Windows

Libraries for pthreads are included on most versions of Unix and Linux. Windows also
supports multi-threading, but with a somewhat different semantics and API than that
of pthreads. To run multi-threaded XSB under Windows, a library must be included
to translate the Pthread library, used by XSB, to the native thread API of Windows.

Different libraries are available for this purpose. Internally, the multi-threaded en-
gine has been tested using the Win32 pthreads interface, available via http://sourceware.org/pthreads-win32,
but other libraries may also work, including Pthread library included with Cygwin.
To install the sourceware library, let $XSBENV be the parent directory of $XSBDIR the
root directory of XSB – i.e. $XSBENV is the directory into which XSB is installed.

- Download a version such as pthreads-2005-01-25.exe or later, and extract it into
  $XSBENV
  pthreads. Add $XSBENV\pthreads\Pre-built\lib to your system path
- To configure with windows enter the commands:

  ⁷This example was inspired by a similar example for multi-threaded computation of primes in
  from Logtalk [53]
  ⁸For a description on how to program with difference lists see a Prolog programming text, such
  as[70]).
:- dynamic server_id/1 as shared.

server :-
    socket(SockFD),
    socket_set_option( SockFD, linger, SOCK_NOLINGER ),
    xsb_port(XSBport),
    socket_bind(SockFD, XSBport),
    socket_listen(SockFD, Q_LENGTH),
    thread_create( server_loop(SockFD), Id, [] ),
    assert( server_id(Iden) ),
    thread_join( Iden ).

server_loop(SockFD) :-
    socket_accept(SockFD, SockClient),
    thread_create( attend_client(SockClient) ),
    server_loop(SockFD).

attend_client(SockClient) :-
    socket_recv_term(SockClient, Goal),
    ( Goal == stop ->
        retract(server_id( Server )),
        thread_cancel( Server ),
        socket_close( SockClient ),
        thread_exit
        ; true
    ),
    ( is_valid(Goal) ->
        call(Goal),
        socket_send_term(SockClient, Goal),
        fail,
        ; socket_send_term(SockClient, invalid_goal(Goal))
    ),
    socket_send_term(SockClient, end),
    socket_close(SockClient).

Figure 7.1: A multi-threaded goal server in XSB
prime(P, I) :- I < sqrt(P), !.
prime(P, I) :- Rem is P mod I, Rem = 0, !, fail.
prime(P, I) :- I₁ is I - 1, prime(P, I₁).

prime(P) :- I is P - 1, prime(P, I).

list_of_primes(I, F, Tail, Tail) :- I > F, !.
list_of_primes(I, F, [I|List], Tail) :-
    prime(I), !,
    I₁ is I + 1, list_of_primes(I₁, F, List, Tail).
list_of_primes(I, F, List, Tail) :-
    I₁ is I + 1, list_of_primes(I₁, F, List, Tail).

partition_space(N, H, H₁) :-
    H is N // 2, H₁ is H + 1.

worker(Q, Iden, I, F, List, Tail) :-
    list_of_primes(I, F, List, Tail),
    thread_send_message(Q, primes(Iden, List, Tail)).

master(N, L) :-
    partition_space(N, H, H₁),
    message_queue_create(Q),
    thread_create(worker(Q, p1, 1, H, L, L₁)),
    thread_create(worker(Q, p2, H₁, N, L₁, [])),
    thread_get_message(Q, primes(p1, L, L₁)),
    thread_get_message(Q, primes(p2, L₁, [])).

Figure 7.2: A multi-threaded program to calculate prime numbers in XSB
sh configure --enable-mt --with-wind \
--with-includes='c:\XSBSYS\XSBENV\threads\Pre-built\include' \
--with-static-libraries='c:\XSBSYS\XSBENV\threads\Pre-built\lib'

makexsb_wind

Note that the Unix sh shell must be available in order to reconfigure.

- To configure with cygwin enter the commands:

sh configure --enable-mt \
--with-includes='/cygdrive/c/XSBSYS/XSBENV/threads/Pre-built/include' \
--with-static-libraries='/cygdrive/c/XSBSYS/XSBENV/threads/Pre-built/lib'

sh makexsb --config-tag=mt

7.9 Predicates for Multi-Threading

The predicates described in this section do not address tabling or dynamic code. With only a few minor deviations the provisional working standard described in [36] is supported. As a result, these predicates are substantially the same as those in SWI, YAP, and other Prologs. In the single-threaded engine, semantically correct calls to these predicates will give a miscellaneous error.

thread_create(+Goal,Threadid,+OptionsList)

When called from thread T, this predicate creates a new XSB thread T_{new} to execute Goal. When goal either succeeds, throws an unhandled error, exits, or fails, T_{new} exits, but thread_create/2 will succeed immediately, binding Threadid to the XSB thread id of T_{new}. Goal must be callable, but need not be fully instantiated. No bindings from Goal are passed back from T to T_{new}, so communication between T_{new} and T must be through tables, asserted code, message queues or other side effects.

OptionList allows optional parameters in the configuration for the initial size of XSB stacks, for aliases, and to indicate whether T_{new} is to be created as detached. Note that XSB threads allow automatic stack allocation, so that the size options may be most useful for (32-bit) applications with very large numbers of threads. In this case, setting initial stack sizes to be small may allow more threads to be created on a given hardware platform. Also note that
only XSB stacks are affected, the stack size of the underlying Pthread remains unaltered.

- \texttt{glsize(N)}: create thread with global (heap) plus local stack size initially set to $N$ kbytes. If not specified, the default size is used. The default size can be set at the command line (cf. Section 3.7), and altered at run time by the Prolog flag \texttt{thread_glsize} (cf. Section 6.12).

- \texttt{tcpsize(N)}: create thread with trail plus choice point stack size initially set to $N$ kbytes. If not specified, the default size is used (cf. Section 3.7). The default size can be set at the command line (cf. Section 3.7), and altered at run time by the Prolog flag \texttt{thread_tcpsize} (cf. Section 6.12).

- \texttt{complsize(N)}: create thread with completion stack size initially set to $N$ kbytes. If not specified, the default size is used (cf. Section 3.7). The default size can be set at the command line (cf. Section 3.7), and altered at run time by the Prolog flag \texttt{thread_complsize} (cf. Section 6.12).

- \texttt{pdlsize(N)}: create thread with $N$ kbytes of unification stack. If not specified, the default size is used (cf. Section 3.7). The default size can be set at the command line (cf. Section 3.7), and altered at run time by the Prolog flag \texttt{thread_pdlsize} (cf. Section 6.12).

- \texttt{detached(Boolean)}: if \texttt{Boolean} is true, creates detached thread. If \texttt{Boolean} is false, the thread created will be joinable, while if no option is given the default will be used. In Version 3.6 threads are created joinable by default, but this default can be altered at run time by the Prolog flag \texttt{thread_default} (cf. Section 6.12).

- \texttt{on_exit(Handler)}: Ensures that \texttt{Handler} is called whenever the thread exits: whether that exit arises from success of \texttt{Goal}, failure, throwing an error that is unhandled in the user’s program, or an explicit call to \texttt{thread_exit/1}.

- \texttt{alias(Alias)}: Allow thread \texttt{ThreadId} to be referred to via \texttt{Alias} in all standard thread predicates. \texttt{Alias} remains active for \texttt{ThreadId} until it is joined. Note that the main XSB thread has alias \texttt{main}.

Finally, each thread is created with a signal queue and a private message queue, so these queues do not need to be explicitly created. Their size is obtained through the settable Prolog flag \texttt{max_queue_terms}.

**Error Cases**

- \texttt{Goal} is a variable
– instantiation_error.

• Goal is not callable
  – type_error(callable, Goal).

• ThreadId is not a variable
  – type_error(variable, ThreadId)

• OptionList is a partial list or contains an option that is a variable
  – instantiation_error

• OptionList is neither a list nor a partial list
  – type_error(list, OptionList)

• OptionList contains an option, Option not described above
  – domain_error(thread_option, Option)

• An element of OptionList is alias(A) and A is already associated with an existing thread, queue, mutex or stream
  – permission_error(create, alias, A)

• An element of OptionList is alias(A) and A is not an atom
  – type_error(atom, A)

• An element of OptionList is on_exit(Handler) and Handler is not callable
  – type_error(callable, Handler).

• No more system threads are available (EAGAIN)
  – resource_error(system threads)

thread_create(+Goal, -ThreadId)
Acts as thread_create(Goal, ThreadId, []).

thread_create(+Goal)
Acts as thread_create(Goal, _, [detached(true)]).

thread_join(+Threads_or_aliases, -ExitDesignators)
When thread_join/2 is called by thread T, Threads_or_aliases must be instantiated to either 1) an XSB thread id or alias; or 2) a list where each element is an XSB thread id or an alias; ExitDesignators must be uninstantiated. The action of the predicate is to suspend T until all of the threads denoted by Threads_or_aliases have exited. At this time, any remaining resources for the threads in ThreadIds will have been reclaimed. Upon success ExitDesignators
is either a the thread status of the associated thread (see page 316) or a list of such elements.

**Error Cases**

- **Thread_or_Aliases** is not instantiated  
  - instantiation_error
- **Threads_or_aliases** is not a list of XSB thread ids or aliases  
  - domain_error(listof(thread_or_alias),ThreadIds)
- **ExitDesignators** is not a variable  
  - type_error(variable,ExitDesignatorst)
- **ThreadId** does not correspond to a valid thread  
  - existence_error(valid_thread,ThreadId)
- **ThreadId** does not correspond to a joinable thread (i.e. **ThreadId** is detached).  
  - permission_error(join,non_joinable_thread,ThreadId)

**thread_exit(+ExitTerm)**
Exits a thread $T$ with **ExitTerm** after releasing any mutexes held by $T$, freeing any thread-specific memory allocated for $T$ (we hope), as well as calling any exit handlers for $T$. **ExitTerm** will be used if the caller of $T$ joins to $T$, but will be ignored in other cases. There is no need to call this routine on normal termination of a thread as it is called implicitly on success or (final) failure of a thread’s goal.

**Error Cases**

- **ExitCode** is a variable  
  - instantiation_error

**thread_self(?ThreadId_or_Alias)**
If **ThreadId** is an atom, unifies **ThreadId_or_Alias** with an alias of the calling thread. Otherwise, unifies **ThreadId_or_Alias** with the XSB thread id of the calling thread. There are no error conditions.

**thread_detach(+Thread_or_Alias)**
Detaches a joinable thread denoted by **Thread_or_Alias** so that all resources will be reclaimed upon its exit. The thread denoted by **ThreadId** will no longer be joinable, once it is detached. If **Thread_or_Alias** has already exited, all resources used by **Thread_or_Alias** are removed from the system.

**Error Cases**
• **Thread_or_Alias** is a variable
  - **instantiation_error**

• **Thread_or_Alias** is not a thread id or alias
  - **domain_error(thread_or_alias,Thread_or_Alias)**

• **Thread_or_Alias** does not correspond to a valid thread
  - **existence_error(valid_thread,Thread_or_Alias)**

• **Thread_or_Alias** is active but not joinable
  - **permission_error(thread_detach,thread,Thread_or_Alias)**

**thread_cancel(+Thread_or_Alias)**
Cancels the XSB thread denoted by **Thread_or_Alias**. The cancellation does not use Pthread cancellation mechanisms, rather it uses XSB’s interrupt mechanism to set **Thread_or_Alias**’s interrupt vector \(^9\). When this interrupt vector is checked, **Thread_or_Alias** will throw a thread cancellation error, which can be caught within **Thread_or_Alias** like any other error. However, the default behavior is for **Thread_or_Alias** to exit with an exit ball indicating that it has been cancelled.

As noted above, an executing thread that is cancelled will exit very shortly after the **thread_cancel/1** predicate is called. Blocked threads, however, are not always guaranteed to exit when cancelled. Currently a blocked thread may be cancelled

• when it is waiting to read or write a message on a queue

• when it is executing **thread_sleep/1**

On the other hand, a blocked thread may not be cancelled while it is waiting to read from a stream or waiting for a mutex.

During critical operations a thread may want to prevent itself from being cancelled. This can be done by If \( \leftarrow \text{thread_cancel}(T) \) is called for a thread **T** for which cancelling has been disabled, **T** will be cancelled immediately after **T** re-enables cancellation through calling the predicate **thread_enable_cancel/0**.

The main XSB thread cannot be cancelled; apart from that any thread can cancel any other thread.

**Error Cases**

\(^9\)This interrupt vector is checked upon every it is checked on every SLG-WAM call and **execute** instruction.
• Thread_or_Alias is not instantiated
  – instantiation_error
• Thread_or_Alias is not a thread id or alias
  – domain_error(thread_or_alias,Thread_or_Alias)
• Thread_or_Alias does not correspond to valid thread
  – existence_error(valid_thread,Thread_or_Alias)
• Thread_or_Alias denotes the main thread.
  – permission_error(cancel,main_thread,Thread_or_Alias)

thread_signal(Thread_or_Alias,Goal)

thread_signal(ThreadOrAlias, Goal) interrupts thread ThreadOrAlias so that it executes Goal at the first opportunity. Specifically, once Goal is placed onto the signal queue of ThreadOrAlias and the interrupt vector of ThreadOrAlias is adjusted, thread_signal/2 succeeds. ThreadOrAlias handles the interrupt asynchronously, and if the interrupt is handled while ThreadOrAlias is executing a goal with continuation $C$, all solutions for Goal will be obtained, and the failure continuation of Goal will be $C$. If Goal throws an exception $E$, the continuation will be the handler for $E$.

For blocked threads, signalling works much like cancellation (described above), and a blocked thread will handle a signal whenever it can be cancelled. However, the thread does not return to the blocking operation after the signal – rather it will execute the signal and then execute the continuation to be taken after the blocking operation.

**Error Cases**

• Thread_or_Alias is not instantiated
  – instantiation_error
• Thread_or_Alias is not a thread id or alias
  – domain_error(thread_or_alias,Thread_or_Alias)
• Thread_or_Alias does not correspond to valid thread
  – existence_error(valid_thread,Thread_or_Alias)
• Goal is not instantiated
  – instantiation_error
• Goal is not callable
  – type_error(callable,Goal)
thread_disable_cancel  
Disables the calling thread from being cancelled, so that it can be ensured that critical operations can run to completion. This predicate always succeeds.

thread_enable_cancel  
Enables the calling thread to be cancelled. By default, threads may be cancelled, so this predicate needs to be called if thread_disable_cancel/0 has been previously called. This predicate always succeeds.

thread_yield  
Make the calling thread ready to be run after other threads of the same priority. This predicate relies on the real-time extensions to pthreads specified in POSIX 1b, and may not be available on all platforms.

Error Cases

- The current platform does not support POSIX real-time extensions
  - misc_error

thread_property(?ThreadOrAlias,?Property)
If ThreadOrAlias is instantiated, unifies Property with current properties of the thread that unify with Property; if ThreadOrAlias is a variable, backtracks through all the current threads whose properties unify with Property. Note that there is no guarantee that that the information returned will be valid, due to concurrency issues.

Currently Property can have the form

- detached(Bool): if Bool is true the thread is detached, otherwise it is joinable.
- alias(Alias): if the thread has an alias Alias
- status(Status): see Section 7.3 for thread statuses that are currently supported.

Example: The following predicate may be used to clear resources from the thread table, although due to concurrency reasons, non-running threads may remain in the thread table after this predicate terminates.

clear_thread_table:-
  thread_property(Tid,status(S)),
  \+ (S = running),
  thread_join(Tid),

fail.
clear_thread_table.

Error Cases

- ThreadOrAlias is neither a variable nor an XSB thread id nor an alias
  - domain_error(thread_or_alias, ThreadOrAlias)
- ThreadOrAlias is not associated with a valid thread
  - existence_error(thread, ThreadOrAlias)

thread_sleep(+Seconds)
Causes the calling thread to sleep approximately Seconds before resuming. A thread may be cancelled while sleeping. However, a sleeping thread that is signaled will execute the signaled goal and resume execution without returning to sleep.

Error Cases

- Seconds is a variable
  - instantiation_error.
- Seconds is not a number
  - type_error(number, Seconds).

7.9.1 Predicates for Thread Synchronization and Communication

Threads can communicate to some extent through shared tables and dynamic code. However, it is often useful to use message queues as a synchronizable form of communication. Similarly, while the XSB engine itself is thread-safe, thread synchronization may be needed when calling a package that is not itself thread safe (see the beginning of this chapter for a list of which packages are and are not thread-safe). Synchronization may also be needed to protect data accessed by foreign function calls, or to coordinate responses to external events.

Prolog Message Queues

As described previously, each thread is created with a private message queue that is readable only by itself. The following predicates are used to communicate using private and public message queues.
message_queue_create(-Queue,+Options)

Creates a new public message queue with identifier Queue. Options allows optional parameters to be passed for the maximum number of terms in the queue, and for aliases of the queue.

- **max_terms(N)**: create queue so that it can contain at most N terms before writes to the queue block. If not specified, the default size is used. This default can be queried and altered at run time via the Prolog flag `queue_max_terms`. (cf. Section 6.12). If the flag `queue_max_terms` is set to 0, the queue size will be bounded only by available memory.

- **alias(Alias)**: Allow queue Queue to be referred to via Alias in all standard queue predicates. Alias remains active for Queue until it is destroyed.

**Error Cases**

- Queue is not a variable
  - `type_error(variable,Queue)`
- Options is a partial list or a list with an element that is a variable
  - instantiation error
- Options is neither a partial list or a list
  - `type_error(list, Options)`
- Options contains an option, Option not described above
  - `domain_error(queue_option,Option)`
- An element of Options is alias(A) and A is already associated with an existing thread, queue, mutex or stream
  - `permission_error(create,alias, A)`
- An element of Options is alias(A) and A is not an atom
  - `type_error(atom,A)`

message_queue_destroy(+Queue_or_Alias)

Destroys a public message queue with alias or id Queue_or_Alias, as created by `message_queue_create/[1,2]`. If any threads are currently waiting on Queue_or_Alias to read or write a term, they will be awakened and will throw an existence error.

**Error Cases**

- Queue_or_Alias is a variable
– instantiation_error
• Queue_or_Alias is not a queue id or alias
  – domain_error(queue_or_alias,Queue_or_Alias)
• Queue_or_Alias denotes a private message queue or signal queue rather than a public message queue
  – permission_error(destroy,private_signal_or_message_queue,Queue_or_Alias)
• Queue_or_alias is not the queue name or alias of a public message queue.
  – existence_error(message_queue, Queue_or_Alias)

thread_send_message(+Queue_or_Alias,#Message)
Queue_or_alias may either be a queue id or alias, or a thread id or alias in which latter case the private queue for a thread is used. If there are fewer terms on Queue_or_Alias than the queue’s maximum allowed number thread_send_message/2 puts Message onto Queue_or_Alias, and returns immediately. Otherwise, the calling thread suspends until there are fewer elements on Queue_or_Alias than the queue’s maximum allowed number, when the thread will be awakened to put Message onto the queue.

Error Cases
• Queue_or_Alias is a variable
  – instantiation_error
• Queue_or_Alias is not a queue id, queue alias, thread id, or thread alias.
  – domain_error(queue_or_alias,Queue_or_Alias)

thread_get_message(+Queue_or_Alias,?Message)
If there are terms on Queue_or_Alias thread_get_message/2 traverses Queue_or_Alias to obtain the first term $T$ that unifies with Message. If $T$ exists, the predicate returns with Message bound to the most general unifier of Message and $T$. If there are no terms on Queue_or_Alias or if no terms unify with Message, the calling thread suspends until at least one term is added to Queue_or_Alias. When the thread awakes, it will recheck Queue from its beginning for a term that unifies with Message.

Error Cases
• Queue_or_Alias is a variable
  – instantiation_error
• Queue_or_Alias is not a queue id or alias
- domain_error(queue_or_alias,Queue_or_Alias)
- existence_error(queue, Queue_or_Alias)

\textbf{thread\_get\_message(?Message)}
\hspace{1em}Acts as \texttt{thread\_get\_message/2}, but on a thread’s private queue.

\textbf{thread\_peek\_message(+Queue_or_Alias,?Message)}
\hspace{1em}If there are terms on \texttt{Queue\_or\_Alias} \texttt{thread\_peek\_message/2} traverses \texttt{Queue\_or\_Alias} to obtain the first term \( T \) that unifies with \texttt{Message}. If \( T \) exists, the predicate returns with \texttt{Message} bound to the most general unifier of \texttt{Message} and \( T \). If there are no terms on \texttt{Queue\_or\_Alias} or if no terms unify with \texttt{Message}, the predicate fails.

\textbf{Error Cases}
\begin{itemize}
\item \texttt{Queue\_or\_Alias} is a variable
  \hspace{1em} \texttt{instantiation\_error}
\item \texttt{Queue\_or\_Alias} is not a queue id or alias
  \hspace{1em} \texttt{domain\_error(queue\_or\_alias,Queue\_or\_Alias)}
\item \texttt{Queue\_or\_Alias} is not associated with a current queue
  \hspace{1em} \texttt{existence\_error(queue, Queue\_or\_Alias)}
\end{itemize}

\textbf{thread\_peek\_message(?Message)}
\hspace{1em}Acts as \texttt{thread\_peek\_message/2}, but on a thread’s private queue.

\textbf{User-defined Mutexes}

Usually, running multi-threaded evaluations does not require a user to set any mutexes – necessary mutexes are handled by XSB itself (we hope), and programs can often be written so that user-level locking is unnecessary. However, under certain conditions, locking is useful or even necessary: for instance, a user may need to set a lock so that a set of shared dynamic facts cannot be accessed when it is updated.

One of the simplest and most powerful primitives for locking are mutexes. The mutexes provided by the following predicates are \textit{recursive}: if a thread \( T \) locks a recursive mutex \( M \), any calls to \texttt{mutex\_lock(M)} made by \( T \) will immediately succeed without suspending while \( M \) is locked. Other threads that attempt to lock \( M \) will suspend until \( M \) is unlocked. To unlock \( M \) after \( n \) calls to \texttt{mutex\_lock(M)}, \( T \) must make \( n \) calls to \texttt{mutex\_unlock(M)}. 
When using mutexes in XSB, programmers must not only avoid explicitly creating deadlocks, but must also ensure that a mutex is unlocked when leaving a critical area, and destroyed when it is no longer needed. Making sure that this happens for successful goals, for failed goals and for goals that raise exceptions can sometimes be complicated. The predicate \texttt{with_mutex/2} handles all of these cases. We recommend using it if possible, and making use of lower-level calls to \texttt{mutex_lock/1}, \texttt{mutex_unlock/1} and \texttt{mutex_trylock/1} only in rare cases when \texttt{with_mutex/2} is not applicable.

\texttt{with_mutex(+Mutex,?Goal)}

Locks a current mutex or alias \texttt{Mutex}, executes \texttt{Goal} deterministically, then unlocks \texttt{Mutex}. If \texttt{Goal} leaves choice-points, these are destroyed. \texttt{Mutex} is unlocked regardless of whether \texttt{Goal} succeeds, fails or raises an exception. Any exception thrown by \texttt{Goal} is re-thrown after the mutex has been successfully unlocked.

\textbf{Error Cases}

- \texttt{Mutex} is a variable
  - \texttt{instantiation_error}
- \texttt{Mutex} is not a mutex id or alias
  - \texttt{domain_error(mutex_or_alias,Mutex_or_Alias)}
- \texttt{Mutex} is not associated with a current mutex.
  - \texttt{existence_error(mutex,Mutex)}
- Locking \texttt{Mutex} would give rise to a deadlock\footnote{This error case handles the EDEADLK return code on MacOS X, and other platforms.}
  - \texttt{permission_error(mutex,lock,Mutex)}
- \texttt{Goal} is a variable
  - \texttt{instantiation error}
- \texttt{Goal} is neither a variable nor a callable term
  - \texttt{type error(callable, Goal)}

\texttt{mutex_create(?Mutex)}

Creates a new recursive user mutex with identifier \texttt{Mutex}. \texttt{Options} allows optional parameters to be passed, currently only for aliases of the mutex.

- \texttt{alias(Mutex)}: Allow queue \texttt{Mutex} to be referred to via \texttt{Mutex} in all standard queue predicates. \texttt{Mutex} remains active for \texttt{Mutex} until it is destroyed.
Error Cases

- Mutex is not a variable
  - type_error(variable,Mutex)
- Options is a partial list or a list with an element that is a variable
  - instantiation error
- Options is neither a partial list or a list
  - type_error(list, Options)
- Options contains an option, Option not described above
  - domain_error(mutex_option,Option)
- An element of Options is alias(A) and A is already associated with an existing thread, queue, mutex or stream
  - permission_error(create,alias, A)
- An element of Options is alias(A) and A is not an atom
  - type_error(atom,A)

mutex_destroy(+Mutex)
Destroys a current unlocked mutex with alias or id Mutex along with any memory it uses.

Error Cases

- Mutex is a variable
  - instantiation_error
- Mutex is not a mutex id or alias
  - domain_error(mutex_or_alias,Mutex_or_Alias)
- Mutex is not associated with a current mutex.
  - existence_error(mutex,Mutex)
- Mutex is locked
  - permission_error(mutex,destroy,Mutex)

mutex_lock(+Mutex)
mutex_lock(Mutex) locks a (recursive) mutex with alias or id Mutex. Locking and unlocking mutexes should be paired carefully in order to avoid deadlocks. In particular, a programmer needs to ensure that mutexes are properly unlocked even if the protected code fails or raises an exception.

Error Cases
• Mutex is a variable
  – instantiation_error
• Mutex is not a mutex id or alias
  – domain_error(mutex_or_alias,Mutex_or_Alias)
• Mutex is not associated with a current mutex.
  – existence_error(mutex,Mutex)
• Locking Mutex would give rise to a deadlock
  – permission_error(mutex,lock,Mutex)

mutex_trylock(+Mutex)
Works as mutex_lock/1 but fails immediately if Mutex is held by another
thread, rather than suspending the calling thread.

Error Cases
• Mutex is a variable
  – instantiation_error
• Mutex is not a mutex id or alias
  – domain_error(mutex_or_alias,Mutex_or_Alias)
• Mutex is not associated with a current mutex.
  – existence_error(mutex,Mutex)

mutex_unlock(+Mutex)
Unlocks the mutex with alias or id Mutex when called by the same thread that
locked Mutex.

Error Cases
• Mutex is a variable
  – instantiation_error
• Mutex is not a mutex id or alias
  – domain_error(mutex_or_alias,Mutex_or_Alias)
• Mutex is not associated with a current mutex.
  – existence_error(mutex,Mutex)
• Mutex is not held by the calling thread

\[11\] This error case handles the EDEADLK return code on MacOS X, and other platforms.
permission_error(unlock,mutex,Mutex)

mutex_unlock_all
mutex_unlock_all/0 unlocks all user mutexes owned by the current thread. It has no error cases.

mutex_property(?MutexOrAlias,?Property)
If MutexOrAlias is instantiated, unifies Property with current properties of the mutex; if MutexOrAlias is a variable, backtracks through all the current mutexes whose properties unify with Property. Note that there is no guarantee that the information returned will be valid, due to concurrency issues.

Currently Property can have the form

- alias(Alias): if the mutex has an alias Alias
- status(Status). If the mutex is locked, Status will be a term of the form locked(ThreadId,NumLocks) where ThreadId is the thread id of the owner of the lock, and NumLocks is the number of times the mutex has been locked by the current owner (recall that user-defined mutexes are recursive and must be unlocked as many times as they have been locked in order to be freed). If the mutex is unlocked, Status will be a term of the form unlocked.

Example: The query

?- mutex_property(M,status(_)).

can be used to enumerate all active user-defined mutexes.

Error Cases

- MutexOrAlias is neither a variable nor an XSB mutex id nor an alias
  - domain_error(mutex_or_alias, MutexOrAlias)
- MutexOrAlias is not associated with an active mutex
  - existence_error(mutex, MutexOrAlias)
- Property is neither a variable nor a valid mutex property
  - domain_error(mutex_property, Property)
Chapter 8

Storing Facts in Tries

XSB offers a mechanism by which large numbers of facts can be directly stored and manipulated in tries, which can either be private to a thread or shared among threads. The mechanism described in this chapter is in some ways similar to trie-indexed asserted code as described in Section 6.14, but allows creation of tries that are shared between threads, and of associative tries that support efficient memory management\(^1\).

When stored in a trie, facts are compiled into trie-instructions similar to those used for XSB’s tables. For instance set of facts

\[
\{ \text{rt}(a,f(a,b),a), \text{rt}(a,f(a,X),Y), \text{rt}(b,V,d) \}
\]

would be stored in a trie as shown in Figure 8, where each node corresponds to an instruction in XSB’s virtual machine. Using a trie for storage has the advantage that discrimination can be made on a position anywhere in a fact, and directly inserting into or deleting from a trie is 4-5x faster than with standard dynamic code. In addition, in trie-dynamic code, there is no distinction between the index and the code itself, so for many sets of facts trie storage can use much less space than standard dynamic code. For instance, Figure 8 shows how the prefix \text{rt}(a,f(a,...)\ is shared for the first two facts. However, trie storage comes with tradeoffs: first, only facts can be stored in a trie; second, unlike standard dynamic code, no ordering is preserved among the facts; and third, duplicate facts are not supported.

In Version 3.6 of XSB, tries that store facts may have the following forms:

\(^1\)For nearly all purposes, the predicates in this chapter replace the low-level API for interned tries in previous versions, which included \text{trieIntern}, \text{trieUnintern}, \text{trieInterned} etc. However that API continues to be supported for low-level systems programming.
• *Private, general* tries allow arbitrary terms to be inserted in a trie. These tries are thread-private so that inserting a term in a trie $Tr$ in one thread will not be visible to another thread. Although such tries are general, they have limitations in memory reclamation in Version 3.6 of XSB. If a term is deleted from $Tr$, memory will be reclaimed if it is safe to do so at the time of deletion; otherwise the space will not be reclaimed until all terms in $Tr$ are removed by truncating $Tr$ or until the thread exits.

• *Private, associative* associative tries are more restricted than general tries: an associative trie combines a *key* which can be any ground term, with a *value* which can be any term. Memory for deleted key-value pairs in an associative trie is always immediately reclaimed, and insert or delete operations can be faster for an associative trie than for a general trie. These tries are private to a thread, and in addition to reclaiming memory when a term is deleted, memory is reclaimed when the trie is truncated or dropped, and when the thread exits.

• *Shared, associative* tries are associative tries that are shared among threads. Memory for deleted key-value pairs is always immediately reclaimed, and when the trie is truncated or dropped.

---

2That is, if no choice points are around that may cause backtracking into $Tr$. 

---

Figure 8.1: Terms Stored as a Trie
8.1 Examples of Using Tries

A handle for a trie can be obtained using the `trie_create/2` predicate. Terms can then be inserted into or deleted from that trie, and terms can be unified with information in the trie, as shown in the following example:

**Example 8.1.1** First, we create a private general trie:

```
| ?- trie_create(X, [type(prge)]).
X = 1
yes
```

Next, we insert some terms into the trie

```
| ?- trie_insert(1, f(a,b)), trie_insert(1, [a, dog, walks]).
yes
```

Now we can make arbitrary queries against the trie

```
| ?- trie_unify(1, X).
X = [a, dog, walks];
X = f(a, b);
no
```

Above, a general query was made, but the query could have been any Prolog term. Now we delete a term, and see what’s left.

```
| ?- trie_delete(1, f(X, B)).
X = a
B = b
yes
| ?- trie_unify(1, X).
X = [a, dog, walks];
no
```
The behavior of general tries can be contrasted with that of associative tries as seen in the next example.

**Example 8.1.2** Now we start by creating a shared associative trie, with abbreviation `shas` using the multi-threaded engine

```
?- trie_create(X,[type(shas),alias(foo)]).
X = 1048577
```

This time we used an alias so now we can use `foo` to refer to insert a couple of key-value pairs into the trie (we could also use the trie handle itself)

```
?- trie_insert(foo,pair(sentence(1),[a,dog,walks])),
   trie_insert(foo,pair(sentence(2),[a,man,snores])).
```

Yes

However, inserting a general term into an associative trie throws an error

```
?- trie_insert(foo,f(a,b)).
++Error[XSB/Runtime/P]: [Domain (f(a,b) not in domain pair/2)]
in arg 2 of predicate trie_insert/2
(Inserted term must be key-value pair in trie 1048577)
```

Finally, in an associative trie, if we insert a value for a key that is already in the trie, it will update the value for that key.

```
?- trie_insert(foo,pair(sentence(1),[a,dog,snores])).
```

Yes

```
?- trie_unify(foo,pair(sentence(1),X)).
X = [a,dog,snores]
```

8.2 Space Management for Tries

When creating or adding terms to an interned trie, XSB manages all space necessary for the terms and their indexes. However, when removing a term from a trie an
issue may arise if there is a possibility of backtracking into the term to be removed; this issue also arises for retracting dynamic code. In the sequential engine and in private tries XSB’s dynamic clause garbage collector handles space reclamation when terms are removed from a trie through `trie_delete/2` or similar low-level predicates. However, in the case of `trie_truncate/1` or `trie_drop/1`, an exception is thrown if there are active choice points to terms in a trie that is to be truncated or dropped.

In the multi-threaded engine the space reclamation problem becomes even more difficult for tries that can be shared among threads. In this case, no garbage collection is performed until there is a single active thread.

These space reclamation issues arise for non-associative tries only. Associative tries essentially contain key-value pairs, and so may have their space reclaimed upon deletion of a term, or upon truncation or dropping their trie, regardless of the number of active threads.

### 8.3 Predicates for Tries

The following subsections describe predicates for inserting terms into a trie, deleting terms from a trie, and unifying a term with terms in a trie, predicates for creating, dropping, and truncating tries, as well as predicates for bulk inserts into and deletes from a trie. These predicates can apply to any type of trie, and perform full error checking on their call arguments. As such, they are safer and more general than the lower-level trie predicates described in Chapter 1 of Volume 2 of this manual. Use of the predicates described here is recommended for applications unless the need for speed is paramount.

```prolog
trie_create(-TrieId,+OptionList)  module: intern  
```

`OptionList` allows optional parameters in the configuration of a trie to indicate its type and whether an alias should be used. In the present version, `OptionList` may contain the following terms

- `type(Type)` where `Type` can be one of
  - `prge` (private, general) maintains information that is accessible only to the calling thread. No other restrictions are made for accessing information in a private trie. In the single-threaded engine, tries are private by default.

---

3Future versions of XSB may extend garbage collection to handle trie truncation, trie dropping and better space reclamation in the multi-threaded engine.
– pras (private, associative) creates a private trie that maintains key-value pairs in a manner similar to an associative array, using the term \( \text{pair}(\text{Key}, \text{Value}) \). Each key must be ground, and there may be only one value per key.

– shas (shared associative) creates a shared trie that maintains key-value pairs in a manner similar to an associative array, using the term \( \text{pair}(\text{Key}, \text{Value}) \). Each key must be ground, and there may be only one value per key. This option is available only in the multi-threaded engine.

• alias(Alias): Allow trie \( \text{TrieId} \) to be referred to via \( \text{Alias} \) in all standard trie predicates. \( \text{Alias} \) remains active for \( \text{TrieId} \) until it is dropped.

• incremental: Allows tables that depend on trie \( \text{TrieId} \) to be automatically updated as information in \( \text{TrieId} \) changes (cf. Section 5.6.3).

• nonincremental: Specifies that tables that depend on trie \( \text{TrieId} \) should not be automatically updated as information in \( \text{TrieId} \) changes (cf. Section 5.6.3).

Error Cases

• \( \text{TrieId} \) is not a variable
  – type_error(variable,\( \text{TrieId} \))

• OptionList is a partial list or contains an option that is a variable
  – instantiation_error

• OptionList is neither a list nor a partial list
  – type_error(list,OptionList)

• OptionList contains an option, Option not described above
  – domain_error(trie_option,Option)

• An element of OptionList is alias(A) and A is already associated with an existing thread, queue, mutex or stream
  – permission_error(create,alias, A)

• An element of OptionList is alias(A) and A is not an atom
  – type_error(atom,A)

\text{trie_insert(+TrieIdOrAlias,Term)}

Inserts Term into the trie denoted by \( \text{TrieIdOrAlias} \). If \( \text{TrieIdOrAlias} \) denotes an associative trie, Term must be of the form \( \text{pair}(\text{Key}, \text{Value}) \) where Key
is ground. If TrieIdOrAlias is a general trie and already contains Term, the predicate fails (as the same term cannot be inserted multiple times in the same trie). Similarly, if TrieIdOrAlias is an associative trie and already contains a value for Key the predicate fails.

Insertion of tries can be controlled by the flags max_answer_term_depth, max_answer_list_depth, max_answer_term_action, and max_answer_list_action, which are also used to control additions of answers to tables. Using these flags, if a term to be inserted is cyclic and exceeds a stated depth, trie insertion may either fail or throw an error depending on the associated action: see pg. 236.

Error Cases

- TrieIdOrAlias is a variable
  - instantiation_error.
- TrieIdOrAlias is not a trie id or alias
  - domain_error(trie_id_or_alias,TrieIdOrAlias)
- TrieIdOrAlias denotes an associative array, and Term does not unify with pair(_,_)
  - domain_error(pair/2,Term)
- TrieIdOrAlias denotes an associative array, Term = pair(Key,Value) but Key is not ground
  - misc_error
- Key or Value is a cyclic term, or exceeds the depth
  - misc_error

trie_unify(+TrieIdOrAlias,Term) module: intern
Unifies Term with a term in the trie denoted by TrieIdOrAlias. If TrieIdOrAlias denotes a general trie, successive unifications will succeed upon backtracking. If TrieIdOrAlias denotes an associative trie, Term must be of the form pair(Key,Value) where Key is ground.

Error Cases

- TrieIdOrAlias is a variable
  - instantiation_error.
- TrieIdOrAlias is not a trie id or alias
  - domain_error(trie_id_or_alias,TrieIdOrAlias)
• *TrieIdOrAlias* denotes an associative array, and *Term* does not unify with pair(_,_)
  - domain_error(pair/2, *Term*)

• *TrieIdOrAlias* denotes an associative array, *Term* = pair(*Key*, *Value*) but *Key* is not ground
  - misc_error

`trie_delete(+TrieIdOrAlias, *Term*)`  
module: **intern**  
Deletes a term unifying with *Term* from the trie denoted by *TrieIdOrAlias*.  
*TrieIdOrAlias* denotes a general trie, all such terms can be deleted upon backtracking.  
If *TrieIdOrAlias* denotes an associative trie, *Term* must be of the form pair(*Key*, *Value*) where *Key* is ground.  
In either case, if *TrieIdOrAlias* does not contain a term unifying with *Term* the predicate fails.

**Error Cases**

• *TrieIdOrAlias* is a variable  
  - instantiation_error.

• *TrieIdOrAlias* is not a trie id or alias  
  - domain_error(trie_id_or_alias, *TrieIdOrAlias*)

• *TrieIdOrAlias* denotes an associative array, and *Term* does not unify with pair(_,_)
  - domain_error(pair/2, *Term*)

• *TrieIdOrAlias* denotes an associative array, *Term* = pair(*Key*, *Value*) but *Key* is not ground
  - misc_error

`trie_truncate(+TrieIdOrAlias)`  
module: **intern**  
Removes all terms from *TrieIdOrAlias*, but does not change any of its properties (e.g. the type of the trie or its aliases).

@@

**Error Cases**

• *TrieIdOrAlias* is a variable  
  - instantiation_error.

• *TrieIdOrAlias* is not a trie id or alias  
  - domain_error(trie_id_or_alias, *TrieIdOrAlias*)
• There are active failure continuations to terms in \texttt{TrieIdOrAlias}
  \begin{itemize}
    \item \texttt{miscellaneous_error}
  \end{itemize}

\texttt{trie_drop(+TrieIdOrAlias)} \hspace{1cm} \texttt{module: \texttt{intern}}

Drops \texttt{TrieIdOrAlias}. \texttt{trie_drop/1} not only removes all terms from \texttt{TrieIdOrAlias}, but also removes information about its type and any aliases the trie may have.

\textbf{Error Cases}

• \texttt{TrieIdOrAlias} is a variable
  \begin{itemize}
    \item \texttt{instantiation_error}.
  \end{itemize}

• \texttt{TrieIdOrAlias} is not a trie id or alias
  \begin{itemize}
    \item \texttt{domain_error(trie_id_or_alias,TrieIdOrAlias)}
  \end{itemize}

• There are active failure continuations to terms in \texttt{TrieIdOrAlias}
  \begin{itemize}
    \item \texttt{miscellaneous_error}
  \end{itemize}

\texttt{trie_bulk_insert(+TrieIdOrAlias,+Generator)} \hspace{1cm} \texttt{module: \texttt{intern}}

Used to insert multiple terms into the trie denoted by \texttt{TrieIdOrAlias}. \texttt{Generator} must be a callable term. Upon backtracking through \texttt{Generator} its first argument should successively be instantiated to the terms to be interned in \texttt{TrieIdOrAlias}. When inserting many terms into a general trie, \texttt{trie_bulk_insert/2} is faster than repeated calls to \texttt{trie_insert/2} as it does not need to make multiple checks that the choice point stack is free of failure continuations that point into the \texttt{TrieIdOrAlias} trie. For associative tries, \texttt{trie_bulk_insert/2} can also be faster as it needs to perform fewer error checks on the arguments of the insert.

\textbf{Example 8.3.1} Given the predicate

\begin{verbatim}
bulk_create(p(One,Two,Three),N):-
  for(One,1,N),
  for(Two,1,N),
  for(Three,1,N).
\end{verbatim}

and a general trie \texttt{Trie}, the goal

\begin{verbatim}
?- trie_bulk_insert(Trie,bulk_create(_Term,N))
\end{verbatim}

will add $N^3$ terms to \texttt{Trie}. 
Error Cases

- TrieIdOrAlias is a variable
  - instantiation_error.
- TrieIdOrAlias is not a trie id or alias
  - domain_error(trie_id_or_alias, TrieIdOrAlias)
- Generator is not a compound term
  - type_error(compound, Generator)
- TrieIdOrAlias denotes an associative array, and Generator does not unify with pair(_, _)
  - domain_error(pair/2, Term)
- TrieIdOrAlias denotes an associative array, and Generator succeeds with a term that unifies with pair(Key, Value) and Key is not ground
  - misc_error
- Key or Value is a cyclic term
  - misc_error

trie_bulk_delete(+TrieIdOrAlias, Term) module: intern
Deletes all terms that unify with Term from TrieIdOrAlias. If TrieIdOrAlias denotes an associative trie, the key of the key value pair need not be ground.

Example 8.3.2 For the trie in the previous example, the goal

?- trie_bulk_delete(Trie, p(1, _, _))

will delete the $N^2$ terms that unify with p(1, _, _) from TrieIdOrAlias.

Error Cases

- TrieIdOrAlias is a variable
  - instantiation_error.
- TrieIdOrAlias is not a trie id or alias
  - domain_error(trie_id_or_alias, TrieIdOrAlias)
trie_bulk_unify(+TrieOrAlias,#Term,-List) module: intern
Returns in List all terms in TrieOrAlias that unify with Term. If TrieOrAlias
denotes an associative trie, the key of the key value pair need not be ground.

This predicate is useful for two reasons. First, it provides a safe way to backtrack
through an associative trie while maintaining the memory management and
concurrency properties of associative tries. Second, it enforces read consistency
for TrieOrAlias, regardless of whether the trie is private or shared, general
or associative.

Example 8.3.3 Continuing from Example 8.3.2 the goal

?- trie_bulk_unify(Trie,X),List

will return the the \( N^3 - N^2 \) terms still in TrieOrAlias.

Error Cases

- TrieOrAlias is a variable
  - instantiation_error.
- TrieOrAlias is not a trie id or alias
  - domain_error(trie_id_or_alias,TrieOrAlias)
- List is not a variable
  - type_error(variable,List).

trie_property(?TrieOrAlias,?Property) module: intern
If TrieOrAlias is instantiated, unifies Property with current properties of the
trie; if TrieOrAlias is a variable, backtracks through all the current tries whose
properties unify with Property. In the MT engine, thread_property/2 ac-
cesses only tries private to the calling thread and shared tries; however note
that there is no guarantee that that the information returned about shared
tries will be valid, due to concurrency issues

Currently Property can have the form

- type(Type): where Type is the type of the trie.
- alias(Alias): if the trie has an alias Alias

Error Cases

\(^4\)trie_property/2 is not yet implemented for shared tries.
• TrieOrAlias is neither a variable nor an XSB trie id nor an alias
  - domain_error(trie, TrieOrAlias)
• TrieOrAlias is not associated with a valid trie
  - existence_error(trie, TrieOrAlias)

8.4 Low-level Trie Manipulation Utilities

The previous sections indicate how tries can be used as an efficient mechanism to store thread-private and thread-shared terms. In this section we describe lower-level trie manipulation predicates that are suitable for implementing XSB libraries. As with other tries, these utilities are suitable for storing terms rather than executable clauses, use a set based semantics, and do not maintain an ordering among these terms. In addition

• These predicates create and maintain thread-private, general tries.
• These predicates do not always perform error checking. If not explicitly specified in the description of the predicate, errors returned may be confusing, and calling with improper arguments may even cause memory violations.
• For historical reasons, the ordering of arguments in these predicates is not consistent.

Despite (and sometimes because of) these limitations, the trie manipulation facilities can be extremely fast, so that interning and uninterning terms in a trie may be much faster than assert and retract in XSB or in any other Prolog.

8.4.1 A Low-Level API for Interned Tries

\[
\begin{align*}
\text{new_trie} & (-\text{Root}) \\
\text{trie_intern} & (+\text{Term}, +\text{Root}) \\
\text{trie_intern} & (+\text{Term}, +\text{Root}, -\text{Leaf}, -\text{Flag}, -\text{Skel}) \\
\end{align*}
\]

\text{module: intern}  

**new_trie**(-Root)

Root is instantiated to a handle for a new private, general trie.

**trie_intern**(+Term,+Root)

module: intern

trie_intern(+Term,+Root,-Leaf,-Flag,-Skel)

module: intern

**trie_intern**/2 effectively asserts **Term** by interning into the trie designated by

\(^5\)Flora-2, XASP, XSB’s storage library and others use these predicates.
Root. If a variant of Term is already in Root the predicate succeeds, but a new copy of Term is not added to the trie.

trie_intern/5 acts as trie_intern/2 but returns additional information: Leaf is the handle for the interned Term in the trie. Flag is 1 if the term is “old” (already exists in the trie); it is 0, if the term is newly inserted. Skel represents the collection of all the variables in Term. It has the form \( \text{ret}(V_1,V_2,\ldots,V_N) \), exactly as in get_calls (see Vol. 1 of the XSB manual).

**Error Cases**

- **Root** is uninstantiated
  - instantiation_error
- **Root** is instantiated, but not an integer (trie handle)
  - type_error(integer,Root)

```
trie_interned(?Term,+Root)                    module: intern
```

```
trie_interned(?Term,+Root,+Leaf,-Skel)        module: intern
```

trie_interned/2 backtracks through the terms that unify with Term and that are interned into the trie represented by the handle Root. Term may be free, or partially bound.

If Leaf is a free variable, trie_interned/5 works as trie_interned/2: it backtracks through the terms that unify with Term and that are interned into the trie represented by the handle Root. In addition it returns Leaf as the handle for each such term and returns in Skel the collection of all the variables in Term using the form \( \text{ret}(V_1,\ldots,V_n) \). Otherwise, if Leaf is bound, trie_interned/5 will unify Term with the term in the trie designated by Leaf, returning a vector of variables in Skel.

**Error Cases**

- **Root** is uninstantiated
  - instantiation_error
- **Root** is instantiated, but not an integer (trie handle)
  - type_error(integer,Root)

```
trie_unintern(+Root,+Leaf)                   module: intern
```

```
trie_unintern_nr(+Root,+Leaf)               module: intern
```

trie_unintern(+Root,+Leaf) deletes a term from a trie using the handle Leaf, as obtained from trie_intern/[2,4] or trie_interned/[2,4]. Space is reclaimed for the term only if it is safe to do so – if there are no failure continuations that may consume the term (cf. Section 8.2).
trie_unintern_nr/2 does not perform space reclamation and as a result requires no garbage collection – it simply marks a term as “deleted”. This makes trie_unintern_nr/2 suitable if trie garbage collection may be an issue, and also allows it to be used in libraries that support backtrackable updates, such as XSB’s storage library.

**Error Cases**

- Root or Leaf is uninstantiated
  - instantiation_error
- Root or Leaf is instantiated, but not an integer (trie handle or trie leaf)
  - type_error(integer,Root) or type_error(integer,Leaf)

reclaim_uninterned_nr(+Root)  
module: intern

Runs through the chain of leaves of the trie Root and deletes the terms that have been marked for deletion by trie_unintern_nr/2. This can be viewed either as a garbage collection step or as a commit.

**Error Cases**

- Root is uninstantiated
  - instantiation_error
- Root is instantiated, but not an integer (trie handle)
  - type_error(integer,Root)

unmark_uninterned_nr(+Root,+Leaf)  
module: intern

The term pointed to by Leaf should have been previously marked for deletion using trie_unintern_nr/2. This term is then “unmarked” (or undeleted) and becomes again a normal interned term.

**Error Cases**

- Root or Leaf is uninstantiated
  - instantiation_error
- Root or Leaf is instantiated, but not an integer (trie handle or trie leaf)
  - type_error(integer,Root) or type_error(integer,Leaf)

delete_trie(+Root)  
module: intern

Deletes all the terms in the trie pointed to by Root. Garbage collection ensures that space reclamation is performed only if it is safe to do so.

**Error Cases**
• Root is uninstantiated
  – instantiation_error
• Root is instantiated, but not an integer (trie handle)
  – type_error(integer,Root)
• Failure continuations point to one or more nodes in the trie with root Root
  – misc_error
Chapter 9

Hooks

Sometimes it is useful to let the user application catch certain events that occur during XSB execution. For instance, when the user asserts or retracts a clause, etc. XSB has a general mechanism by which the user program can register hooks to handle certain supported events. All the predicates described below must be imported from xsb_hook.

9.1 Adding and Removing Hooks

A hook in XSB can be either a 0-ary predicate or a unary predicate. A 0-ary hook is called without parameters and unary hooks are called with one parameter. The nature of the parameter depends on the type of the hook, as described in the next subsection.

add_xsb_hook(+HookSpec) module: xsb_hook

This predicate registers a hook; it must be imported from xsb_hook. HookSpec has the following format:

hook-type(your-hook-predicate(_))

or, if it is a 0-ary hook:

hook-type(your-hook-predicate)

For instance,
:- add_xsb_hook(xsb_assert_hook(foobar(_))).

registers the hook foobar/1 as a hook to be called when XSB asserts a clause. Your program must include clauses that define foobar/1, or else an error will result.

The predicate that defines the hook type must be imported from xsb_hook:

    :- import xsb_assert_hook/1 from xsb_hook.

or add_xsb_hook/1 will issue an error.

remove_xsb_hook(+HookSpec)                module: xsb_hook

Unregisters the specified XSB hook; imported from xsb_hook. For instance,

    :- remove_xsb_hook(xsb_assert_hook(foobar(_))).

As before, the predicate that defines the hook type must be imported from xsb_hook.

### 9.2 Hooks Supported by XSB

The following predicates define the hook types supported by XSB. They must be imported from xsb_hook.

xsb_exit_hook(_)                         module: xsb_hook

These hooks are called just before XSB exits. You can register as many hooks as you want and all of them will be called on exit (but the order of the calls is not guaranteed). Exit hooks are all 0-ary and must be registered as such:

    :- add_xsb_hook(xsb_exit_hook(my_own_exit_hook)).

xsb_assert_hook(_)                       module: xsb_hook

These hooks are called whenever the program asserts a clause. An assert hook must be a unary predicate, which expects the clause being asserted as a parameter. For instance,
:- add_xsb_hook(xsb_assert_hook(my_assert_hook(_))).

registers my_assert_hook/1 as an assert hook. One can register several assert hooks and all of them will be called (but the order is not guaranteed).

\texttt{xsb_retract_hook(_)} \hspace{1cm} \textit{module: xsb\_hook}

These hooks are called whenever the program retracts a clause. A retract hook must be a unary predicate, which expects as a parameter a list of the form \texttt{[Head,Body]}, which represent the head and the body parts of the clause being retracted. As with assert hooks, any number of retract hooks can be registered and all of them will be called in some order.
Chapter 10

Debugging and Profiling

10.1 Prolog-style Tracing and Debugging

XSB supports a version of the Byrd four-port debugger for interactive debugging and tracing of Prolog code. In this release (Version 3.6), it does not work very well when debugging code involving tabled predicates. If one only creeps (see below), the tracing can provide some useful information. For programs that involve large amounts of tabling forest-view tracing can be used (Section 10.3). To turn on tracing, use trace/0, trace/1, or trace/2. To turn tracing off, use notrace/0.

```
trace
notrace
```

When tracing is on, the system will print a message each time a predicate is:

1. initially entered (Call),
2. successfully returned from (Exit),
3. failed back into (Redo), and
4. completely failed out of (Fail).

When debugging interactively, a message may be printed and tracer stopped and prompts for input. (See the predicates show/1 and leash/1 described below to modify what is traced and when the user is prompted.)

---

1The current version of XSB's Prolog debugger does not include exceptions as a debugging port.
In addition to single-step tracing, the user can set spy points to influence how the tracing/debugging works. A spy point is set using `spy/1`. Spy points can be used to cause the system to enter the tracer when a particular predicate is entered. Also the tracer allows “leaping” from spy point to spy point during the debugging process. The debugger also has profiling capabilities, which can measure the cpu time spent in each call. The cpu time is measured only down to 0.0001-th of a second. When the tracer prompts for input, the user may enter a return, or a single character followed by a return, with the following meanings:

- **c, <CR>:** Creep  Causes the system to single-step to the next port (i.e. either the entry to a traced predicate called by the executed clause, or the success or failure exit from that clause).
- **a:** Abort  Causes execution to abort and control to return to the top level interpreter.
- **b:** Break  Calls the evaluable predicate `break`, thus invoking recursively a new incarnation of the system interpreter. The command prompt at break level n is

  \[ n: \ ?- \]

  The user may return to the previous break level by entering the system end-of-file character (e.g. `ctrl-D`), or typing in the atom `end_of_file`; or to the top level interpreter by typing in `abort`.
- **f:** Fail  Causes execution to fail, thus transferring control to the Fail port of the current execution.
- **h:** Help  Displays the table of debugging options.
- **l:** Leap  Causes the system to resume running the program, only stopping when a spy-point is reached or the program terminates. This allows the user to follow the execution at a higher level than exhaustive tracing.
- **n:** Nodebug  Turns off debug mode.
- **r:** Retry (fail)  Transfers to the Call port of the current goal. Note, however, that side effects, such as database modifications etc., are not undone.
- **s:** Skip  Causes tracing to be turned off for the entire execution of the procedure. Thus, nothing is seen until control comes back to that procedure, either at the Success or the Failure port.
- **q: Quasi-skip** This is like Skip except that it does not mask out spy points.
CHAPTER 10. DEBUGGING AND PROFILING

- **S**: *Verbose skip*  Similar to *Skip* mode, but trace continues to be printed. The user is prompted again when the current call terminates with success or failure. This can be used to obtain a full trace to the point where an error occurred or for code profiling. (See more about profiling below.)

- **e**: *Exit*  Causes immediate exit from XSB back to the operating system.

```
trace(+Filename,+option)
```

`trace/2` is like `trace/0` except that it is non-interactive and dumps trace information into a log file, *Filename*. Currently the only supported option is `log`. However, the log is written in the form of Prolog facts, which can be loaded queried. The format of the facts is:

```
xsb_tracelog(CallId,CallNum,PortType,ParentCallNum,DepthOfCall,CurrentCall,Time)
```

where *CallId* is an identifier generated when XSB encounters a new top-level call. This identifier remains the same for all subgoals called while tracing that top-level call.

- **CallNum** is a generated number to show the nesting of the calls being traced. It is the same number that the user sees when tracing interactively.
- **PortType** is 'Call', 'Redo', 'Exit', or 'Fail'.
- **ParentCallNum** is the call number of the parent call.
- **DepthOfCall** is the nesting depth of the current call with respect to its ancestor calls.
- **CurrentCall** is the call being traced
- **Time** is the CPU time it took to execute *CurrentCall*. On 'Call' and 'Redo', Time is always 0 —— it has a meaningful value only for the 'Exit' and 'Fail' log entries.

It should be noted that when calls are delayed due to the well-founded negation computation of because of the *when/2* primitive, the parent call might be off in some cases. However, the parent property repairs itself for subsequent calls.

'The name of the predicate (`xsb_tracelog`) used for logging can be changed by asserting it into the predicate `debug_tracelog_predicate/1`, which should be imported from `usermod`. For instance,

```
:- import debug_tracelog_predicate/1 from usermod.
?- assert(debug_tracelog_predicate(foobar)).
```
spy(Preds)

where Preds is a spy specification or a list of such specifications, and must be instantiated. This predicate sets spy points (conditional or unconditional) on predicates. A spy specification can be of several forms. Most simply, it is a term of the form $P/N$, where $P$ is a predicate name and $N$ its arity. Optionally, only a predicate name can be provided, in which case it refers to all predicates of any arity currently defined in usermod. It may optionally be prefixed by a module name, e.g. $\text{ModName}:P/N$. (Again, if the arity is omitted, the specification refers to all predicates of any arity with the given name currently defined in the given module.) A spy specification may also indicate a conditional spy point. A conditional spy specification is a Prolog rule, the head indicating the predicate to spy, and the body indicating conditions under which to spy. For example, to spy the predicate $p/2$ when the first argument is not a variable, one would write: $\text{spy}(p(X,_) : \neg\text{nonvar}(X))$. (Notice that the parentheses around the rule are necessary). The body may be empty, i.e., the rule may just be a fact. The head of a rule may also be prefixed (using :) with a module name. One should not put both conditional and unconditional spy points on the same predicate.

nospy(Preds)

where Preds is a spy specification, or a list of such specifications, and must be instantiated at the time of call. What constitutes a spy specification is described above under spy. nospy removes spy points on the specified predicates. If a specification is given in the form of a fact, all conditional spy points whose heads match that fact are removed.

debug

Turns on debugging mode. This causes subsequent execution of predicates with trace or spy points to be traced, and is a no-op if there are no such predicates. The predicates $\text{trace/0, trace/1, trace/2}$, and $\text{spy/1}$ cause debugging mode to be turned on automatically.

nodebug

Turns off debugging mode. This causes trace and spy points to be ignored.

debugging

Displays information about whether debug mode is on or not, and lists predicates that have trace points or spy points set on them.

debug_ctl(option,value)

debug_ctl/2 performs debugger control functions as described below. These commands can be entered before starting a trace or inside the trace. The latter
can be done by responding with "b" at the prompt, which recursively invokes an XSB sub-session. At this point, you can enter the debugger control commands and type `end_of_file`. This returns XSB back to the debugger prompt, but with new settings.

1. `debug_ctl(prompt, off)` Set non-interactive mode globally. This means that trace will be printed from start to end, and the user will never be prompted during the trace.

2. `debug_ctl(prompt, on)` Make tracing/spying interactive.

3. `debug_ctl(profile, on)` Turns profiling on. This means that each time a call execution reaches the `Fail` or `Exit` port, CPU time spent in that call will be printed. The actual call can be identified by locating a `Call` prompt that has the same number as the "cpu time" message.

4. `debug_ctl(profile, off)` Turns profiling off.

5. `debug_ctl(redirect, +File)` Redirects debugging output to a file. This also includes program output, errors and warnings. Note that usually you cannot see the contents of `+File` until it is closed, i.e., until another redirect operation is performed (usually `debug_ctl(redirect, tty)`, see next).

6. `debug_ctl(redirect, tty)` Attaches the previously redirected debugging, error, program output, and warning streams back to the user terminal.

7. `debug_ctl(show, +PortList)` Allows the user to specify at which ports should trace messages be printed. `PortList` must be a list of port names, i.e., a sublist of `["Call", "Exit", "Redo", "Fail"]`.

8. `debug_ctl(leash, +PortList)` Allows the user to specify at which ports the tracer should stop and prompt the user for direction. `PortList` must be a list of port names, i.e., a sublist of `["Call", "Exit", "Redo", "Fail"]`. Only ports that are `show-n` can be `leash-ed`.

9. `debug_ctl(hide, +PredArityPairList)` The list must be of the form `[P1/A1, P2/A2, ...]`, i.e., each either must specify a predicate-arity pair. Each predicate on the list will become non-traceable. That is, during the trace, each such predicate will be treated as an black-box procedure, and trace will not go into it.

10. `debug_ctl(unhide, ?PredArityPairList)` If the list is a predicate-arity list, every predicate on that list will become traceable again. Items in the list can contain variables. For instance, `debug_ctl(unhide, [_/2])` will
make all 2-ary that were previously made untraceable traceable again. As a special case, if PredArityPairList is a variable, all predicates previously placed on the “untraceable”-list will be taken off.

11. debug_ctl(hidden, -List) This returns the list of predicates that the user said should not be traced.

10.2 Low-Level Tracing

XSB also provides a facility for low-level tracing of execution. This can be activated by invoking the emulator with the -T option (see Section 3.7), or through the predicate trace/0. It causes trace information to be printed out at every call (including those to system trap handlers). The volume of such trace information can very become large very quickly, so this method of tracing is not recommended in general.

XSB debugger also provides means for the low-level control of what must be traced. Normally, various standard predicates are masked out from the trace, since these predicates do not make sense to the application programmer. However, if tracing below the application level is needed, you can retract some of the facts specified in the file syslib/debugger_data.P (and in some cases assert into them). All these predicates are documented in the header of that file. Here we only mention the four predicates that an XSB developer is more likely to need. To get more trace, you should retract from the first three predicates and assert into the last one.

- hide_this_show(Pred,Arity): specifies calls (predicate name and arity) that the debugger should not show at the prompt. However, the evaluation of this hidden call is traced.
- hide_this_hide(Pred,Arity): specifies calls to hide. Trace remains off while evaluating those predicates. Once trace is off, there is no way to resume it until the hidden predicate exits or fails.
- show_this_hide(Pred,Arity): calls to show at the prompt. However, trace is switched off right after that.
- trace_standard_predicate(Pred,Arity): Normally trace doesn’t go inside standard predicates (i.e., those specified in syslib/std_xsb.P. If you need to trace some of those, you must assert into this predicate.
In principle, by retracting all facts from the first three predicates and asserting enough facts into the last one, it is possible to achieve the behavior that approximates the -T option. However, unlike -T, debugging can be done interactively. This does not obviate -T, however. First, it is easier to use -T than to issue multiple asserts and retracts. Second, -T can be used when the error occurs early on, before the moment when XSB shows its first prompt.

10.3 Analyzing the Execution of Tabled Programs

The tracing and debugging described in previous sections has proven useful for Prolog programs for 30 or more years. However, when tabling is added to Prolog, things change. First, as described in Chapter 5, tabling can be used to find the least fixed point of mutually recursive predicates. Operationally, this requires the ability to suspend one computation path and to resume another. Second, the addition of tabled negation for the well-founded semantics requires the ability to delay negative goals whose only proof may be involved in a loop through negation and to simplify these goals once their truth value has become known. Furthermore, a tabled subgoal has different states: it may be new; it may be incomplete so that new answers might be derived for it; or completed (completely evaluated) so that the answers may simply be read from the table. In short, tabling, which can execute much more general programs than Prolog and which can use the stronger well-founded semantics, requires a more complex set of operations than Prolog’s SLDNF. Accordingly, debugging and tracing is correspondingly more complex. Thus, while Prolog’s 4-port debugger may be useful for programs that involve just a few tabled predicates, it may not be useful for programs that heavily use tabling for complex recursions, non-monotonic reasoning or other purposes.

There is currently no standard approach to debugging tabled programs. One possible approach would be to extend the 4-port debugger to include other ports for tabling operations. Such extensions have not yet been explored, and whether the paradigm of n-port debugging can be extended to full tabling so that it can be useful to programmers is an open question. Another approach would be use the declarative approach of justification [33, 55] to explain why derivations were or were not made. XSB does in fact have a justification package but it is not currently robust enough to be recommended for general use. Below we present the logforest approach.
10.3.1 Tracing a tabled evaluation through forest logging

While the operations used for tabling are more complex than those of SLDNF, they have a clear formal operational semantics through SLG and the forest-of-trees model. We recall this model briefly below for a definite program but assume a background knowledge of tabled logic programming (see, for instance [78]).

Example 10.3.1 Figure 10.1 shows a program fragment along with an SLG forest for the query \texttt{?- reach(1,Y)} to the the right-recursive tabled predicate \texttt{reach/1}. An SLG forest consists of an SLG tree for each tabled subgoal \( S \): this tree has root \( S :- S \). In a definite program an SLG tree represents resolution of program clauses and answers to prove \( S \). In Figure 10.1 each non-root node of the form \( K.N \) where \( N = (S :- Goals)\theta \) is a clause in which the bindings to a subgoal \( S \) are maintained in \( S\theta \), the goals remaining to prove \( S \) are in \( Goals\theta \), and the order of creation of \( N \) within the tabled evaluation is represented by a number, \( K \) (local scheduling is used in this example). Children of a root node are obtained through resolution of a tabled subgoal against program clauses. Children of non-root nodes are obtained through answer clause resolution, if the left most selected literal is tabled (e.g. children of node 3 or 11 in the tree for \texttt{reach(1,Y)}), or through program clause resolution if the leftmost selected literal is not tabled (e.g. children of nodes 2 and 18 in the tree for \texttt{reach(1,Y)}). Nodes that have empty \( Goals \) are termed \textit{answers}. Note that the evaluation keeps track of each tabled subgoal \( S \) that it encounters. Later if \( S \) is selected again, resolution will use answers rather than program clauses; if no answers are available, the computation will \textit{suspend} at that point and the evaluation will backtrack to try to derive answers using some other computation path. Once more answers have been derived, the evaluation \textit{resumes} the suspended computation. Similarly, once the computation has backtracked through all answers available for \( S \) in the current state, the computation path will suspend, and resume after further answers are found. Thus a tabled evaluation is a fixed point computation for a set of interdependent subgoals. When it is determined that a (perhaps singleton) set of subgoals can produce no more answers, the subgoals are completed.

The forest logging approach (\texttt{logforest}) allows one to run a tabled query and produce a log that can be interpreted as (a partial image of) an SLG forest. The log can then used to analyze program correctness, to optimize performance and so on. Because \texttt{logforest} produces a log, it superficially resembles the non-interactive trace described earlier in this chapter. However,

- \texttt{trace/1} produces a Prolog-style trace that takes little account of tabling.
  \texttt{logforest} structures its output according to the forest-of-trees model, and
Figure 10.1: A program $P_{R_{ree}}$ and SLG forest for (local) evaluation of $?-\text{reach}(1,Y)$
takes little account of program clause resolution.

- **logforest** is implemented in C for efficiency, while **trace/1** is built on top of XSBs interactive debugger. Unlike **trace/1**, **logforest** can therefore to produce logs for very large evaluations with little overhead.

*We stress that the forest logging approach is under development and its features are subject to change.*

Currently, **logforest** captures the following actions.

- **A call to a tabled subgoal** If a positive call to a tabled subgoal $S_1$ is made from a tree for $S_2$ a Prolog-readable fact of the form $\text{tc}(S_1, S_2, \text{Stage}, \text{Counter})$ is logged, where *Counter* is the ordinal number of the fact, and *Stage* is
  - **new** if $S_1$ is a new subgoal
  - **cmp** if $S_1$ is not a new subgoal and has been completed
  - **incmp** if $S_1$ is not a new subgoal but has *not* been completed

If the call is negative a fact of the form $\text{nc}(S_1, S_2, \text{Stage}, \text{Counter})$ is logged, where all arguments are as above.

For instance, in the above example, node 3 would be represented as $\text{tc}(\text{reach}(2, Y), \text{reach}(1, Y), 2)$ (the reason for using the counter value of 2 rather than 3 is explained below). If $S_1$ is the first tabled subgoal in an evaluation, $S_2$ is the atom *null*.

- **Derivation of a new answer** When a new *unconditional* answer $A$ is derived for subgoal $S$ and added to the table (i.e. $A$ is not already an answer for $S$) a fact of the form $\text{na}(A, S, \text{Counter})$ is logged. In the above example, the answer node 9 would be represented as $\text{na}([2], \text{reach}(2, _\text{v1}), 4)$ where the first argument is a list of substitutions for the variables $_\text{v1}, \ldots, _\text{vn}$ in $S$.

  When a new *conditional* answer $A :- D$, with substitution $A$ and delayed literals $D$, is derived for subgoal $S$ and added to the table a fact of the form $\text{nda}(A, S, D, \text{Counter})$ is logged.

- **Return of an answer to a consuming subgoal** When an unconditional answer $A$ is returned to a consuming subgoal $S$ in a tree for $S_T$, a fact of the form $\text{ar}(A, S, S_T, \text{Counter})$ is logged. A log entry is made only if the table for $S$ is incomplete (see the explanation below).

  If the answer $A$ is conditional, the fact has the form $\text{dar}(A, S, S_T, \text{Counter})$, where each argument is as above.
• **Delaying a selected negative literal.** If a selected negative literal $L$ of a node $N$ is delayed, because it is involved in a loop through negation, and $N$ is in a tree for $S_T$, a fact of the form $\text{dly}(L,S_T,\text{Counter})$ is logged.

• **Subgoal completion**
  
  – When a set $S$ of subgoals is determined to be completely evaluated and is completed, a fact of the form $\text{cmp}(S,\text{SCCNum},\text{Counter})$ is logged for each $S \in S$. Here $\text{SCCNum}$ is simply a number giving an ordinal value that can be used to group subgoals into mutually dependent sets of subgoals (here called *Strongly Connected Components* or SCCs), i.e. the $\text{SCCNum}$ of each $S \in S$ has the same value, but that value is not used for a completion fact of any subgoal not in $S$.

  – When a subgoal $S$ is *early completed*, i.e. it is determined that no more answers for $S$ are possible or are desired a fact of the form $\text{cmp}(S,\text{ec},\text{Counter})$ is logged. If $S$ belonged to a larger mutually dependent set $S$ when it was early completed, $S$ will also be included in the completion facts for $S$.

• **Table Abolishes**
  
  – When a tabled subgoal $S$ is abolished, a fact of the form $\text{ta}(\text{subg}(S),\text{Counter})$ is logged.

  – When all tables for a predicate $p/n$ are abolished, a fact of the form $\text{ta}(\text{pred}(p/n),\text{Counter})$ is logged.

  – When all tables are abolished, a fact of the form $\text{ta}(\text{all},\text{Counter})$ is logged.

• **Location of errors** Whenever an error is thrown and the execution is in a tree for a subgoal $S$, a Prolog-readable fact of the form $\text{err}(S,\text{Counter})$ is logged, where $\text{Counter}$ is the ordinal number of the fact. The primary purpose of this fact is to indicate the nearest tabled call that gave rise to an uncaught error.

**logforest** does not contain

• Information about the occurrence of program clause resolution either when used to produce children of tabled predicates, or when it is used to produce children whose nodes have a selected literal that is non-tabled.

• Information about the return of answers from completed tables. XSB uses a so-called *completed table optimization* which treats answer return from completed tables in a manner akin to program clause resolution.
The inclusion of the above two features in `logforest` would significantly slow down execution of XSB. However, future versions of `logforest` may include expanded logging features for negation, for call and answer subsumption and for incremental tabling.\(^2\)

Example 10.3.2 *The forest for reach(1,Y) in the foregoing example has the log file as shown in Table 10.1.*

```prolog
log_forest(+Call)  % module: tables
log_forest(+Call,+Options)  % module: tables
These predicates turn on forest logging, call Call, then turn logging off when Call is finished. Options is a list of possible options.

- If Options contains the term `file(File)` then direct the logging to File; otherwise the log will be sent to standard output.

- If Options contains the term `level(Level)` where Level = full or partial then the level of forest logging is set to the respective value. The default level is full. The level partial does not output answer return operations, and so can reduce the size of the log for certain computations. The actions of partial logging and full logging are otherwise the same.

Error Cases

- Options is a variable, or contains a variable as an element
  - instantiation_error

- Options is not a list
  - type_error(list,Options)

- Options contains an option O that is not a forest logging option.
  - domain_error(forest_logging_option,O)

load_forest_log(+File)  % module: tables
The log produced by `log_forest/[1,2]` is a Prolog file that can be compiled and/or loaded dynamically just as any other Prolog file. However, for large logs (i.e. those of many megabytes) use of `load_dync/[1,2]` XSB commands can drastically reduce the time needed to load the file, while use of the proper `index/2` declarations can greatly improve query time. The simple predicate, `load_forest_log/1` loads a log file and indexes needed arguments.

\(^2\)Currently, attributes of attributed variables are not printed out.
<table>
<thead>
<tr>
<th>Log File</th>
<th>Forest</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>tc(reach(1, v0), null, new, 0)</td>
<td>node</td>
<td>created by program clause resol.</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>tc(reach(2, v0), reach(1, v0), new, 1)</td>
<td>node</td>
<td>created by program clause resol.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>node</td>
<td>created by program clause resol.</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>tc(reach(2, v0), reach(2, v0), incmp, 2)</td>
<td>node</td>
<td>repeated subgoal registered</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>node</td>
<td>created by program clause resol.</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>na([2], reach(2, v0), 3)</td>
<td>node</td>
<td>created by program clause resol.</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>ar([2], reach(2, v0), reach(2, v0), 4)</td>
<td>node</td>
<td>registered as answer</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>cmp(reach(2, v0), 2, 5)</td>
<td>node</td>
<td>created by answer resol.</td>
</tr>
<tr>
<td></td>
<td>9a</td>
<td>reach(2, v0) completed</td>
</tr>
<tr>
<td></td>
<td>node</td>
<td>created by return from completed table</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>na([2], reach(1, v0), 6)</td>
<td>node</td>
<td>created by program clause resol.</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>tc(reach(3, v0), reach(1, v0), new, 7)</td>
<td>node</td>
<td>created by program clause resol.</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>tc(reach(1, v0), reach(3, v0), incmp, 8)</td>
<td>node</td>
<td>repeated subgoal registered</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>ar([2], reach(1, v0), reach(3, v0), 9)</td>
<td>node</td>
<td>created by answer resol.</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>na([2], reach(3, v0), 10)</td>
<td>node</td>
<td>created by program clause resol.</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>na([1], reach(3, v0), 11)</td>
<td>node</td>
<td>created by program clause resol.</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>na([3], reach(1, v0), 12)</td>
<td>node</td>
<td>created by program clause resol.</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>ar([3], reach(1, v0), reach(3, v0), 13)</td>
<td>node</td>
<td>created by answer return</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>ar([3], reach(3, v0), 14)</td>
<td>node</td>
<td>created by answer resol.</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>ar([2], reach(3, v0), reach(1, v0), 15)</td>
<td>node</td>
<td>created by answer resol.</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>ar([1], reach(3, v0), reach(1, v0), 16)</td>
<td>node</td>
<td>created by answer resol.</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>na([3], reach(1, v0), reach(1, v0), 17)</td>
<td>node</td>
<td>created by answer resol.</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>ar([1], reach(3, v0), reach(1, v0), 18)</td>
<td>node</td>
<td>created by answer resol.</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>ar([1], reach(1, v0), reach(3, v0), 19)</td>
<td>node</td>
<td>created by answer resol.</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>cmp(reach(1, v0), 1, 20)</td>
<td>node</td>
<td>created by answer resol.</td>
</tr>
<tr>
<td>cmp(reach(3, v0), 1, 21)</td>
<td>node</td>
<td></td>
</tr>
</tbody>
</table>

Table 10.1: Log file for computation in Figure 10.1
10.3.2 Analyzing the log; seeing the forest through the trees

As previously described, forest logging is based on the formal operational semantics of SLG, and as a result the log can be analyzed to query any result that can be modelled by the theory. But despite the power of forest logging, it can be difficult to use. Not all users have the background to fully understand the operational semantics of SLG. Even those users with a formal background may find it difficult to write efficient analysis routines for logs of large computations\(^3\). Accordingly, XSB provides routines that analyze logs and display information about a computation. These routines can answer many questions about a computation and can provide the starting point for further exploration. We introduce these routines via an extended example.

Example 10.3.3 This example arises from the actual use of forest logging to understand a Flora-2 computation\(^8\), in which the Cyc reasoner (cf. http://www.cyc.com) was translated into Silk (cf. http://silk.semwebcentral.org) and used to answer various questions in biology. Silk itself compiles into Flora-2 which in turn compiles into XSB\(^4\). After translation, query answering took more resources than expected, and users wanted to determine why. Using the features of Version 3.6, the first step is to call statistics/0 at the end of the computation. The statistics indicated that the computation took about 30 seconds of CPU time and 300 megabytes of table space, while XSB’s trail had allocated over 1 gigabyte of space. The call to statistics/0 also showed the following information:

8678944 variant call check/insert ops: 615067 producers, 8063877 variants.
317346 answer check/insert ops: 304899 unique inserts, 12447 redundant.

In other words, there were nearly 10 million tabled subgoals that were called, indicating that this computation was heavily tabled (a characteristic of most Flora-2 computations). It also shows that the average number of answers per tabled subgoal is rather small.

This basic information leads to several questions. Why were there so many tabled subgoals? Did the tabling have anything to do with the large amount of choice-point/trail space that was allocated? Which tabled subgoals had answers? How many times did a given tabled predicate call another tabled predicate?

Some of these questions can be answered by table_dump/[2,3]: particularly, what tabled subgoals were called, and which had answers. However table_dump/[2,3] cannot provide other information, such as the dependencies of given tabled subgoals.

\(^3\)I find it difficult myself!
\(^4\)This example was run in 2012 using a 64-bit server with a large amount of RAM.
on other tabled subgoals or the order in which operations occurred. From a formal perspective, \texttt{table_dump/[2,3]} does not allow a user to analyze an entire SLG forest: only the “table”, i.e., the subgoals in the forest and the unordered set of its answers. The table omits any information about interior nodes or completion information, both of which are used to compute dependency information. Dependencies are useful in analyzing most computations, but is especially important in Flora-2 computations such as this one, that make heavy use of HiLog. This use of HiLog means that the dependencies of tabled predicates on one another is not at all obvious, and may not easily be determined by static analysis.

The next step, therefore, in analyzing this computation is to rerun it with forest logging. For this computation forest logging has no impact on memory usage, but increases the time of the computation from about 30 seconds to about 52 seconds — around 73% in this case. It is worthwhile noting that the actual overhead of forest logging varies depending on how heavily the computation is tabled. The log itself had slightly over 14 million entries which were loaded into XSB via \texttt{load_forest_log/1}. The log took about 140 seconds to load and about 7.8 Gbytes of space for the log facts and their multiple and trie indexes \(^5\).

The easiest way to start the analysis is to ask the query \(-\texttt{forest_log_overview}\), which for this example gives:

\begin{verbatim}
There were 613496 subgoals in 463330 (completed) SCCs.
93918 subgoals were early-completed.
0 subgoals were not completed in the log.
There were a total of 8670043 tabled subgoal calls:
  613496 were calls to new subgoals
  4467747 were calls to incomplete subgoals
  3588800 were calls to complete subgoals

Number of SCCs with 1 subgoals is 463322
Number of SCCs with 4 subgoals is 1
Number of SCCs with 7 subgoals is 1
Number of SCCs with 52 subgoals is 1
Number of SCCs with 110 subgoals is 4
Number of SCCs with 149671 subgoals is 1
\end{verbatim}

\(^{5}\)The load time for this example, about 100,000 facts/second is typical for 2012 CPUs; the size of the loaded code is larger than usual, due in part to the expansion in the size of terms caused by the HiLog encoding.
The overview extends the information shown by \texttt{statistics/0}. First, the total number of completed and non-completed SCCs is given along with a count of how many of the completed subgoals were early completed. Information about non-completed SCCs is useful, since the forest log may be analyzed for a computation that does not terminate. Since this computation did terminate, all subgoals in the log were completed \footnote{The slight difference between the number of subgoals shown here and the number shown by \texttt{statistics/0} is due to the use of tabling in the Flora compiler.}. Note that there is also a breakdown of calls to tabled subgoals that distinguishes whether the tabled subgoal was new, completed, or incomplete. Recall that calls to completed tabled subgoals essentially treat the answers in the table as facts, so that these calls are efficient. Making a call to an incomplete subgoals on the other hand means that the calling and called subgoals are mutually recursive \footnote{This statement is true in local evaluation but not in batched evaluation.} and execution of recursive sets of subgoals can be expensive, especially in terms of space.

Finally, the overview report provides the distributions of tabled subgoals across SCCs. While most of the SCCs were small there was a large one, with nearly 150,000 mutually dependent subgoals. Clearly the large SCC should be examined. The first step is to obtain its index. The query

\begin{verbatim}
get_scc_size(SCC,Index)), Index > 1000.
\end{verbatim}

returns the information that the index of the large SCC was 39. The query \texttt{analyze_an_scc(39,userout)} then provides the following information.

\begin{verbatim}
There are 149671 subgoals and 4461290 links (average of 30.8073 edges per subgoal) within the SCC
There are 2 subgoals in the SCC for the predicate backchainForbidden / 0
There are 2 subgoals in the SCC for the predicate http://www.cyc.com/silk/implementation/transformationPredicate / 0
: There are 15613 subgoals in the SCC for the predicate gpLookupSentence / 3
There are 15613 subgoals in the SCC for the predicate removalSentence / 3
There are 18770 subgoals in the SCC for the predicate forwardSentence / 3
There are 18771 subgoals in the SCC for the predicate lookupSentence / 3

Calls from assertedSentence/3 to lookupSentence/3 : 32
Calls from backchainForbidden/0 to 'http://www.cyc.com/silk/implementation/transformationPredicate' / 0 :
\end{verbatim}
Calls from transformationSentence/2 to sbhlSentence/3 : 5479
Calls from tvaSentence/3 to removalSentence/3 : 7695

It is evident from the first line in this report that the vast majority of the calls to incomplete tables during this computation occur in the SCC under investigation. Since information on incomplete tables is kept in XSB's choice point stack (cf. [62]), the evaluation of SCC 39 is the likely culprit behind the large amount of stack space required. The subgoals in the SCC are first broken out by their predicate name and arity, then the edges within the SCC are broken out by the predicates of their caller and called subgoals. At this point a programmer can review the various rules for lookupSentence/3, forwardSentence/3 and other predicates to determine whether the recursion is intended and if so, whether it can be simplified.

Using abstraction in the analysis

Within the SCC analysis, information about a given tabled subgoal $S$ was abstracted to the functor and arity of $S$. For this example, abstraction was necessary, as reporting 150,000 subgoals or 4,000,000+ would not provide useful information for a human being. However, it could be the case that seeing the tabled subgoals themselves would be useful for a smaller SCC. Even for an SCC of this size, different levels of abstraction could be useful: mode information or type information might be useful in a given circumstance.

**Example 10.3.4** Making the call `?- analyze_an_scc(39,userout,abstract_modes(_,_))` applies the predicate `abstract_modes/2` to each term, producing an output of the form:

There are 149671 subgoals and 4461290 links (average of 30.8073 edges per subgoal)
within the SCC

There are 3 subgoals in the SCC for the predicate backchainRequired(g,g)
There are 2 subgoals in the SCC for the predicate backchainForbidden(g,g):

There are 29254 subgoals in the SCC for the predicate gpLookupSentence(g,g)
There are 29254 subgoals in the SCC for the predicate removalSentence(g,g)

Calls from assertedSentence(g,g) to lookupSentence(g,g) : 10
Calls from assertedSentence(m,g) to lookupSentence(m,g) : 22:

: 
abstract_modes(In,Out) simply goes through each argument of In and unifies the corresponding argument of Out with a v if the argument is a variable, a g if the argument is ground, and m otherwise.

abstract_modes/2 is simply an example: any term-abstraction predicate may be passed into the last argument of analyze_an_scc/3.

Analyzing Negation

Many programs that use negation are stratified in such a way that they do not require the use of Delaying and Simplification operations, and the routines described in the previous section are sufficient for these programs. However if a program does not have a two-valued well-founded model, a user would often like to understand why. Even in a program that is two-valued, the heavy use of Delaying and Simplification can indicate that some rules may need to be optimized by having their literals reordered.

Example 10.3.5 Figure 10.2 shows a program with negation and illustrates SLG resolution for the query \( p(c) \) to the program. The nodes in Figure 10.2 have been annotated with the order in which they were created under local scheduling. In the formalism used by Figure 10.2, the symbol | in a node separates the unresolved goals to the right from the delayed goals to the left. In the evaluation state where nodes 1 through 10 have been created, \( p(b) \) has been completed, and \( p(a) \) and \( p(c) \) are in the same SCC. There are no more clauses or answers to resolve, but \( p(a) \) is involved in a loop through negation in node 5, and nodes 2 and 10 involve \( p(a) \) and \( p(c) \) in a negative loop.

In situations such as this, where all resolution has been performed for nodes in an SCC, an evaluation may have to apply a Delaying operation to a negative literal such as \( \text{not}(p(a)) \), in order to explore whether other literals to its right might fail. When multiple literals can be delayed, an arbitrary one is chosen to be delayed first. So the evaluation delays the selected literal of node 2 to generate node 12 producing

---

\[ \text{Example 10.3.4} \]

Because of the special representation of Flora-2 terms, abstraction was used to produce the output of Example 10.3.4, while a more sophisticated version of abstract_modes/2 was used in Example 10.3.4.

\[ \text{Example 10.3.5} \]

In this example, we ignore the effects of early completion which would complete \( p(b) \) immediately upon creation of node 8, obviating the need to create node 9.
a conditional answer – an answer with a non-empty delay list (cf. Section 5.3.2 for an overview of how XSB computes and allows inspection of delayed literals). Next, not p(a) in node 5 is delayed, failing that computation path, and not p(c) in node 10 is delayed to produce node 15 and failing the final computation path for p(a).

At this stage the SCC {p(a), p(c)} is completely evaluated meaning that there are no more operations applicable for goal literals (as opposed to delay literals). Since p(a) is completely evaluated with no answers, conditional or otherwise, the evaluation determines it to be failed and a SIMPLIFICATION operation can be applied to the conditional answer of node 12, leading to the unconditional answer in node 17 and success of the literal p(c).

As indicated previously, the forest log overview includes a total count of DELAYING and SIMPLIFICATION operations, as well as a count of conditional answers. In addition, SCC analysis counts negative as well as positive links within the SCC. The
current version of forest logging also provides a means to examine the causes of answers that have an undefined truth value. Recall from Example 10.3.5 that there are two types of causes of an undefined truth value: either 1) a negative literal explicitly undergoes a DELAYING operation; or 2) a conditional answer may be used to resolve a literal. It can be shown that in local evaluation, a conditional answer $A$ will never be returned out of an SCC if $A$ is successful or failed in the well-founded model of a program. This means that if an answer for $S$ is undefined, then it would be caused operationally by a DELAYING operation within the SCC of $S$ or within some other SCC on which $S$ depends. So to understand why an atom is undefined it can be useful understand the “root causes” of the delay: to examine SCCs in which DELAYING operations were executed and conditional answers were derived, but the answers could not be simplified.

Example 10.3.6 As a use case, logging was made of execution of a Flora-2 program that tested out a new defeasibility theory. The forest log overview indicated that the top-level query was undefined:

: There were a total of 55 negative delays
There were a total of 0 simplifications
There were a total of 695 unconditional answers derived:
There were a total of 66 conditional answers derived:

The analysis predicate three_valued_scc(List) produces a list of all SCC indices in which DELAYING caused the derivation of conditional answers. These SCCs can then be analyzed as discussed in the previous section.

10.3.3 Discussion

Using log forest imposes a relatively minimal overhead on most computations, considering the information it can provide, and loading and analysis is relatively quick. For this example, the top level analysis took around 10 seconds, and analysing SCC 39 took about 20 seconds in Example 10.3.3 and about 60 seconds in Example 10.3.4. For more information, see [75].

10.3.4 Predicates for Forest Logging

forest_log_overview module: tables
Provides an overview of subgoals, calls, and SCCs in the forest log as indicated in Section 10.3.2.
CHAPTER 10. DEBUGGING AND PROFILING

get_scc_size(?Index,?Size) module: tables

This simple predicate determines the indices of SCCs whose size is \texttt{Size}, for use with \texttt{analyze_an_scc/[2,3]}.

three_valued_sccs(List) module: tables

If there are any SCCs in the log where delay is performed, causing conditional answers to be added that were not simplified into unconditional answers, unifies \texttt{List} with the index of all such SCCs.

analyze_an_scc(+Index,+File) module: tables

analyze_an_scc(+Index,+File,+Abstraction) module: tables

These predicates can be used to analyze the SCC indexed by \texttt{Index} in a forest log, as explained in Section 10.3.2. The output is written to \texttt{File}; calling the predicate with \texttt{File} set to \texttt{userout} causes the output to be written to the console. In \texttt{analyze_an_scc/2}, tabled subgoals are abstracted to predicate indicators, in \texttt{analyze_an_scc/3}, a two-ary abstraction predicate in \texttt{usermod} is called.

Error conditions on \texttt{File} are the same as \texttt{tell/1}.

abstract_modes(Term,AbstractedTerm) module: usermod

\texttt{abstract_modes(In,Out)} simply goes through each argument of \texttt{Term} and unifies the corresponding argument of \texttt{Abstracted} with a \texttt{v} if the argument is a variable, a \texttt{g} if the argument is ground, and \texttt{m} otherwise.

set_forest_logging_for_pred(+PredSpec,+Mode) module: tables

If forest logging is active, this predicate allows any logging specific to the predicate or term indicator, \texttt{PredSpec}, to be turned on or off. Thus, for instance, tabled predicates in a pre-existing library need not clutter up the log.

Error Cases

- \texttt{PredSpec} is not a predicate or term indicator.
  - \texttt{type_error}
- \texttt{Mode} is not in the set \{\texttt{on},\texttt{off}\}
  - \texttt{domain_error}
Chapter 11

Definite Clause Grammars

11.1 General Description

Definite clause grammars (DCGs) are an extension of context free grammars that have proven useful for describing natural and formal languages, and that may be conveniently expressed and executed in Prolog. A Definite Clause Grammar rule is executable because it is just a notational variant of a logic rule that has the following general form:

\[ \text{Head} \rightarrow \text{Body}. \]

with the declarative interpretation that “a possible form for Head is Body”. The procedural interpretation of a grammar rule is that it takes an input sequence of symbols or character codes, analyses some initial portion of that list, and produces the remaining portion (possibly enlarged) as output for further analysis. In XSB, the exact form of this sequence is determined by whether XSB’s DCG mode is set to use tabling or not, as will be discussed below. In either case, the arguments required for the input and output lists are not written explicitly in the DCG rule, but are added when the rule is translated (expanded) into an ordinary normal rule during parsing. Extra conditions, in the form of explicit Prolog literals or control constructs such as if-then-elses (’\(\rightarrow\)'/2) or cuts (’!'/0), may be included in the Body of the DCG rule and they work exactly as one would expect.

The syntax of DCGs is orthogonal to whether tabling is used for DCGs or not. An overview of DCG syntax supported by XSB is as follows:

1. A non-terminal symbol may be any HiLog term other than a variable or a
number. A variable which appears in the body of a rule is equivalent to the appearance of a call to the standard predicate phrase/3 as it is described below.

2. A terminal symbol may be any HiLog term. In order to distinguish terminals from nonterminals, a sequence of one or more terminal symbols $\alpha, \beta, \gamma, \delta, \ldots$ is written within a grammar rule as a Prolog list $[\alpha, \beta, \gamma, \delta, \ldots]$, with the empty sequence written as the empty list $[]$. The list of terminals may contain variables but it has to be a proper list, or else an error message is sent to the standard error stream and the expansion of the grammar rule that contains this list will fail. If the terminal symbols are UTF-8 character codes, they can be written (as elsewhere) as strings.

3. Extra conditions, expressed in the form of Prolog predicate calls, can be included in the body (right-hand side) of a grammar rule by enclosing such conditions in curly brackets, '{' and '}'. For example, one can write:

$$\text{positive_integer}(N) \rightarrow [N], \{\text{integer}(N), N > 0\}. \quad (1)$$

4. The left hand side of a DCG rule must consist of a single non-terminal, possibly followed by a sequence of terminals (which must be written as a unique Prolog list). Thus in XSB, unlike SB-Prolog version 3.1, Semicontext (formerly called push-back lists) is supported.

5. The right hand side of a DCG rule may contain alternatives (written using the usual Prolog’s disjunction operator ‘;’ or using the usual BNF disjunction operator ‘|’).

6. The Prolog control primitives if-then-else (‘->’/2), nots (not/1, fail_if/1, ‘\+’/1 or tnot/1) and cut (‘!’/0) may also be included in the right hand side of a DCG rule. These symbols need not be enclosed in curly brackets. All other Prolog’s control primitives, such as repeat/0, must be enclosed explicitly within curly brackets if they are not meant to be interpreted as non-terminal grammar symbols.

---

1 A term like {foo} is just a syntactic-sugar for the term ‘{’(foo).
2 Readers familiar with Quintus Prolog may notice the difference in the treatment of the various kinds of not. For example, in Quintus Prolog a not/1 that is not enclosed within curly brackets is interpreted as a non-terminal grammar symbol.
11.2 Translation of Definite Clause Grammar rules

In this section we informally describe the translation of DCG rules into normal rules in XSB. Each grammar rule is translated into a Prolog clause as it is consulted or compiled. This is accomplished through a general mechanism of defining the hook predicate `term_expansion/2`, by means of which a user can specify any desired transformation to be done as clauses are read by the reader of XSB’s parser. This DCG term expansion is as follows:

A DCG rule such as:

```
p(X) --> q(X).
```

will be translated (expanded) into:

```
p(X, Li, Lo) :-
    q(X, Li, Lo).
```

If there is more than one non-terminal on the right-hand side, as in

```
p(X, Y) --> q(X), r(X, Y), s(Y).
```

the corresponding input and output arguments are identified, translating into:

```
p(X, Y, Li, Lo) :-
    q(X, Li, L1),
    r(X, Y, L1, L2),
    s(Y, L2, Lo).
```

Terminals are translated using the predicate `’C’/3` (See section 11.3 for its description). For instance:

```
p(X) --> [go, to], q(X), [stop].
```

is translated into:

```
p(X, S0, S) :-
    ’C’(S0, go, S1),
    ’C’(S1, to, S2),
    q(X, S2, S3),
    ’C’(S3, stop, S).
```

Extra conditions expressed as explicit procedure calls naturally translate into themselves. For example,

```
positive_number(X) -->
    [N], {integer(N), N > 0},
    fraction(F), {form_number(N, F, X)}.
```
translates to:

```
positive_number(X, Li, Lo) :-
  'C'(Li, N, L1),
  integer(N),
  N > 0,
  L1 = L2,
  fraction(F, L2, L3),
  form_number(N, F, N),
  L3 = Lo.
```

Similarly, a cut is translated literally.

Semicontext (or a push-back list, which is a proper list of terminals on the left-hand side of a DCG rule) translate into a sequence of 'C'/3 goals with the first and third arguments reversed. For example,

```
it_is(X), [is, not] --> [aint].
```

becomes

```
it_is(X, Li, Lo) :-
  'C'(Li, aint, L1),
  'C'(Lo, is, L2),
  'C'(L2, not, L1).
```

Disjunction has a fairly obvious translation. For example, the DCG clause:

```
expr(E) -->
  expr(X), "+", term(Y), \{E \equiv X+Y\}
| term(E).
```

translates to the Prolog rule:

```
expr(E, Li, Lo) :-
  ( expr(X, Li, L1),
    'C'(L1, 43, L2), % 0'+ = 43
    term(Y, L2, L3)
    E = X+Y,
    L3 = Lo
  ; term(E, Li, Lo)
).```

11.2.1 Definite Clause Grammars and Tabling

Tabling can be used in conjunction with Definite Clause Grammars to get the effect of a more complete parsing strategy. When Prolog is used to evaluate DCG’s, the resulting parsing algorithm is “recursive descent”. Recursive descent parsing, while efficiently implementable, is known to suffer from several deficiencies: 1) its time can be exponential in the size of the input, and 2) it may not terminate for certain context-free grammars (in particular, those that are left or doubly recursive). By appropriate use of tabling, both of these limitations can be overcome. With appropriate tabling, the resulting parsing algorithm is a variant of Earley’s algorithm and of chart parsing algorithms.

In the simplest cases, one needs only to add the directive :- auto_table (see Section 3.10.5) to the source file containing a DCG specification. This should generate any necessary table declarations so that infinite loops are avoided (for context-free grammars). That is, with a :- auto_table declaration, left-recursive grammars can be correctly processed. Of course, individual table directives may also be used, but note that the arity must be specified as two more than that shown in the DCG source, to account for the extra arguments added by the expansion. However, the efficiency of tabling for DCGs depends on the representation of the input and output sequences used, a topic to which we now turn.

Consider the expanded DCG rule from the previous section:

\[
p(X, S0, S) :-
\begin{align*}
  & 'C'(S0, go, S1), \\
  & 'C'(S1, to, S2), \\
  & q(X, S2, S3), \\
  & 'C'(S3, stop, S).
\end{align*}
\]

In a Prolog system, each input and output variable, such as S0 or S is bound to a variable or a difference list. In XSB, this is called list mode. Thus, to parse go to lunch stop the phrase would be presented to the DCG rule as a list of tokens [go, to, lunch, stop] via a call to phrase/3 such as:

\[
\text{phrase}(p(X), [go, to, lunch, stop]).
\]

or an explicit call to p/3, such as:

\[
p(X, [go, to, lunch, stop|X], X).
\]

Terminal elements of the sequence are consumed (or generated) via the predicate 'C'/3 which is defined for Prolog systems as:

\[
'C'([Token|Rest], Token, Rest).
\]
While such a definition would also work correctly if a DCG rule were tabled, the need to copy sequences into or out of a table can lead to behavior quadratic in the length of the input sequence (See Section 5.2.5). As an alternative, XSB allows a mode of DCGs that defines ‘C’/3 as a call to a Datalog predicate \texttt{word/3}:

\begin{verbatim}
'\texttt{C'}(\texttt{Pos},\texttt{Token},\texttt{Next_pos})\texttt{:}- \texttt{word(\texttt{Pos},\texttt{Token},\texttt{Next_pos})}.
\end{verbatim}

assuming that each token of the sequence has been asserted as a \texttt{word/3} fact, e.g:

\begin{verbatim}
\texttt{word(0,go,1)}.
\texttt{word(1,to,2)}.
\texttt{word(2,lunch,3)}.
\texttt{word(3,stop,4)}.
\end{verbatim}

The above mode of executing DCGs is called \textit{datalog mode}.

\texttt{word/3} facts are asserted via a call to the predicate \texttt{tphrase_set_string/1}. Afterwards, a grammar rule can be called either directly, or via a call to \texttt{tphrase/1}. To parse the list [\texttt{go,to,lunch,stop}] in datalog mode using the predicate \texttt{p/3} from above, the call

\begin{verbatim}
\texttt{tphrase_set_string([go,to,lunch,stop])}
\end{verbatim}

would be made, afterwards the sequence could be parsed via the goal:

\begin{verbatim}
\texttt{tphrase(p(X))}.
\end{verbatim}

or

\begin{verbatim}
\texttt{p(X,0,F)}.
\end{verbatim}

To summarize, DCGs in list mode have the same syntax as they do in datalog mode: they just use a different definition of ‘C’/3. Of course tabled and non-tabled DCGs can use either definition of ‘C’/3. Indeed, this property is necessary for tabled DCG predicates to be able to call non-tabled DCG predicates and vice-versa. At the same time, tabled DCG rules may execute faster in datalog mode, while non-tabled DCG rules may execute faster in list mode.

Finally, we note that the mode of DCG parsing is part of XSB’s state. XSB’s default mode is to use list mode: the mode is set to datalog mode via a call to \texttt{tphrase_set_string/3} and back to list mode by a call to \texttt{phrase/2} or by a call to \texttt{reset_dcg_mode/0}.

### 11.3 Definite Clause Grammar predicates

The library predicates of XSB that support DCGs are the following:
CHAPTER 11. DEFINITE CLAUSE GRAMMARS

phrase(+Phrase, ?List)
This predicate is true iff the list List can be parsed as a phrase (i.e. sequence of terminals) of type Phrase. Phrase can be any term which would be accepted as a nonterminal of the grammar (or in general, it can be any grammar rule body), and must be instantiated to a non-variable term at the time of the call; otherwise an error message is sent to the standard error stream and the predicate fails. This predicate is the usual way to commence execution of grammar rules.

If List is bound to a list of terminals by the time of the call, then the goal corresponds to parsing List as a phrase of type Phrase; otherwise if List is unbound, then the grammar is being used for generation.

tphrase(+Phrase)
This predicate succeeds if the current database of word/3 facts can be parsed via a call to the term expansion of +Phrase whose input argument is set to 0 and whose output argument is set to the largest N such that word(_,_,N) is currently true.

The database of word/3 facts is assumed to have been previously set up via a call to tphrase_set_string/1 (or variant). If the database of word/3 facts is empty, tphrase/1 will abort.

phrase(+Phrase, ?List, ?Rest)
This predicate is true iff the segment between the start of list List and the start of list Rest can be parsed as a phrase (i.e. sequence of terminals) of type Phrase. In other words, if the search for phrase Phrase is started at the beginning of list List, then Rest is what remains unparsed after Phrase has been found. Again, Phrase can be any term which would be accepted as a nonterminal of the grammar (or in general, any grammar rule body), and must be instantiated to a non-variable term at the time of the call; otherwise an error message is sent to the standard error stream and the predicate fails.

Predicate phrase/3 is the analogue of call/1 for grammar rule bodies, and provides a semantics for variables in the bodies of grammar rules. A variable X in a grammar rule body is treated as though phrase(X) appeared instead, X would expand into a call to phrase(X, L, R) for some lists L and R.

expand_term(+Term1, ?Term2)
This predicate is used to transform terms that appear in a Prolog program before the program is compiled or consulted. The default transformation performed by expand_term/2 is that when Term1 is a grammar rule, then Term2 is the corresponding Prolog clause; otherwise Term2 is simply Term1 unchanged. If
Term1 is not of the proper form, or Term2 does not unify with its clausal form, predicate expand_term/2 simply fails.

Users may augment the default transformations by asserting clauses for the predicate term_expansion/2 to usermod. After term_expansion(Term_a,Term_b) is asserted, then if a consulted file contains a clause that unifies with Term_a the clause will be transformed to Term_b before further compilation. (Term_b can be a list of clauses, so term_expansion can transform a single clause into a sequence of clauses.) expand_term/2 calls user clauses for term_expansion/2 first; if the expansion succeeds, the transformed term so obtained is used and the standard grammar rule expansion is not tried; otherwise, if Term1 is a grammar rule, then it is expanded using dcg/2; otherwise, Term1 is used as is.

Example: Suppose the following clause is asserted:

?- assert(term_expansion(foo(X),bar(X))).

and that the file te.P contains the clause foo(a) then the clause will automatically be expanded upon consulting the file:

| ?- [te].
[Compiling /Users/macuser/te]
[te compiled, cpu time used: 0.0170 seconds]
[te loaded]

yes
| ?- bar(X).

X = a

yes
| ?- foo(X).
++Error[XSB/Runtime/P]: [Existence (No procedure usermod : foo / 1 exists)] []
Forward Continuation...

However, read/[1,2] does not automatically perform term expansion

| ?- use_module(standard,[expand_term/2]).

yes
| ?- read(X),expand_term(X,Y).
CHAPTER 11. DEFINITE CLAUSE GRAMMARS

foo(a).

X = foo(a)
Y = bar(a)

yes

'C'(?L1, ?Terminal, ?L2)
This predicate generally is of no concern to the user. Rather it is used in the transformation of terminal symbols in grammar rules and expresses the fact that L1 is connected to L2 by the terminal Terminal. This predicate is needed to avoid problems due to source-level transformations in the presence of control primitives such as cuts ('!/0), or if-then-elses ('->'/2) and is defined by the single clause:

'C'([Token|Tokens], Token, Tokens).

The name 'C' was chosen for this predicate so that another useful name might not be preempted.

tphrase_set_string(+List)
This predicate
1. abolishes all tables;
2. retracts all word/3 facts from XSB’s store; and
3. asserts new word/3 facts corresponding to List as described in Section 11.2.1.

implicitly changing the DCG mode from list to datalog.

tphrase_set_string_keeping_tables(+List) module: dcg
This predicate is the same as tphrase_set_string, except it does not abolish any tables. When using this predicate, the user is responsible for explicitly abolishing the necessary tables.

tphrase_set_string_auto_abolish(+List) module: dcg
This predicate is the same as tphrase_set_string, except it abolishes tables that have been indicated as dcg-supported tables by a previous call to set_dcg_supported_table/1.
setting_dcg_supported_table(+TabSkel)

module: dgc

This predicate is used to indicate to the DCG subsystem that a particular
tabled predicate is part of a DCG grammar, and thus the contents of its table
depends on the string being parsed. TabSkel must be the skeleton of a tabled
predicate. When tphrase_set_string_auto_abolish/1 is called, all tables
that have been indicated as DCG-supported by a call to this predicate will be
abolished.

dcg(+DCG_Rule, ?Prolog_Clause)

module: dgc

Succeeds iff the DCG rule DCG_Rule translates to the Prolog clause Prolog_Clause.
At the time of call, DCG_Rule must be bound to a term whose principal functor
is '->'/2 or else the predicate fails. dcg/2 must be explicitly imported from
the module dgc.

11.4 Two differences with other Prologs

The DCG expansion provided by XSB is in certain cases different from the ones
provided by some other Prolog systems (e.g. Quintus Prolog, SICStus Prolog and
C-Prolog). The most important of these differences are:

1. XSB expands a DCG clause in such a way that when a '!'/0 is the last goal
   of the DCG clause, the expanded DCG clause is always steadfast.

That is, the DCG clause:
   a --> b, ! ; c.
gets expanded to the clause:
   a(A, B) :- b(A, C), !, C = B ; c(A, B).
and not to the clause:
   a(A, B) :- b(A, B), ! ; c(A, B).
as in Quintus, SICStus and C Prolog.

The latter expansion is not just optimized, but it can have a different (unintended) meaning if a/2 is called with its second argument bound.

However, to obtain the standard expansion provided by the other Prolog sys-
tems, the user can simply execute:
   set_dcg_style(standard).

To switch back to the XSB-style DCG’s, call
set_dcg_style(xsb).
This can be done anywhere in the program, or interactively. By default, XSB
starts with the XSB-style DCG’s. To change that, start XSB as follows:

xsb -e "set_dcg_style(standard)."

Problems of DCG expansion in the presence of cuts have been known for a long
time and almost all Prolog implementations expand a DCG clause with a !'/0
in its body in such a way that its expansion is steadfast, and has the intended
meaning when called with its second argument bound. For that reason almost
all Prologs translate the DCG clause:

a --> ! ; c.
to the clause:

a(A, B) :- !, B = A ; c(A, B).
But in our opinion this is just a special case of a !'/0 being the last goal in
the body of a DCG clause.
Finally, we note that the choice of DCG style is orthogonal to whether the DCG
mode is list or datalog.

2. Most of the control predicates of XSB need not be enclosed in curly brackets.
A difference with, say Quintus, is that predicates not/1, \+1/1, or fail_if/1
do not get expanded when encountered in a DCG clause. That is, the DCG
clause:

a --> (true -> X = f(a) ; not(p)).
gets expanded to the clause:

a(A, B) :- (true(A, C) -> =(X, f(a), C, B) ; not p(A, B))
and not to the clause:

a(A, B) :- (true(A, C) -> =(X, f(a), C, B) ; not(p, A, B))
that Quintus Prolog expands to.
However, note that all non-control but standard predicates (for example true/0
and '='/2) get expanded if they are not enclosed in curly brackets.
Chapter 12

Exception Handling

We define the term *exceptions* as errors in program execution that are handled by a non-local change in execution state. Exception handling in XSB is ISO-compatible, and has been extended to handle tabled evaluations.

12.1 The Mechanics of Exception Handling

We address the case of non-tabled evaluations before discussing the extensions for tabling.

12.1.1 Exception Handling in Non-Tabled Evaluations

The preferred mechanism for dealing with exceptions in XSB is to use the predicates `catch/3` and `default_user_error_handler/1` together with one of XSB’s error predicates (such as `misc_error/1`). These predicates are ISO-compatible, and their use can give a great deal of control to exception handling. At a high level, when an exception is encountered an error term $T$ is *thrown*. In a non-tabled Prolog program, throwing an error term $T$ causes XSB to examine its choice point stack until it finds a *catcher* that unifies with $T$. This catcher then calls a *handler*. If no explicit catcher for $T$ exists, a default handler is invoked, which usually results in an abort, and returns execution to the top-level of the interpreter, or to the calling C function.\footnote{Starting in Version 3.5.1, XSB uses the ISO compliant `error/2` for error terms, rather than `error/3` as in previous versions.}

A handler is set up when `catch(Goal,Catcher,Handler)` is called. At the time
of the call, a continuation is saved (i.e. a Prolog choice point), and Goal is called.
If no exceptions are encountered, answers for Goal are obtained as usual. However,
within the execution of Goal, an exception might be thrown by calling a Prolog pred-
icate in the error_handler module, or by executing a C-level error function. As
mentioned above, when an error is thrown in an environment Env, XSB searches for
an ancestor Env_{anc} of Env in which catch/3 was called, and in which the catcher
(second argument) unifies with Error. If such an ancestor is found, program exe-
cution reverts to the ancestor and all intervening choice points are removed. The
catcher’s Handler goal is called and the exception is thereby handled. On the other
hand, if no ancestor in the user’s program was called using catch/3 the exception
is handled via the handler associated with XSB’s goal interpreter at the top-level
command line or C API. This top-level handler checks whether a clause with head
default_user_error_handler(Term) has been asserted, such that Term unifies with
Error. If so, this handler is executed. If not, XSB’s default system error handler
in invoked an error message is output and execution returns to the top level of the
interpreter.

The following, somewhat fanciful, example helps clarify these concepts. Consider
the predicate userdiv/2 (Figure 12.1) which is designed to be called with the first
argument instantiated to a number. A second number is then read from a console,
and the first number is divided by the second, and unified with the second argument
of userdiv/2. By using catch/3 and throw/1 together the various types of errors
can be caught.

The behavior of this program on some representative inputs is shown below.

```
| ?- userdiv(p(1),F).
++Error[XSB/Runtime/P]: [Type (p(1) in place of number)] in arg 1 of predicate userdiv1/2
Forward Continuation...
... machine:xsb_backtrace/1
... error_handler:type_error/4
... standard:call/1
... x_interp:_$call/1
... x_interp:call_query/1
... standard:call/1
... standard:catch/3
... x_interp:interpreter/0
... loader:ll_code_call/3
... standard:call/1
```

\[A user-defined error type is desired, the Prolog predicate \texttt{throw/1} can also be called directly.\]

\[A code for this example can be found in $\texttt{XSBDIR/examples/exceptions.P}$.\]
import error_writeln/1 from standard.
import type_error/4 from error_handler.

userdiv(X,Ans):-
    catch(userdiv1(X,Ans),mydiv1(Y),handleUserdiv(Y,X)).

userdiv1(X,Ans):-
    (number(X) -> true; type_error(number,X,userdiv1/2,1)),
    write('Enter a number: '),read(Y),
    (number(Y) -> true ; throw(mydiv1(error1(Y)))),
    (Y < 0 -> throw(mydiv1(error2(Y))); true),
    (Y =:= 0 -> throw(error(zerodivision,userdiv/1,[],[])); true),
    Ans is X/Y.

handleUserdiv(error1(Y),_X):-
    error_writeln(['a non-numeric denominator was entered in userdiv/1: ',Y]),fail.
handleUserdiv(error2(Y),_X):-
    error_writeln(['a negative denominator was entered in userdiv/1: ',Y]),fail.

Figure 12.1: The userdiv/1 program
... standard:catch/3

no
| ?- userdiv(3,F).
Enter a number: foo.
a non-numeric denominator was entered in userdiv/1: foo

no
|| ?- userdiv(3,F).
Enter a number: -1.
a negative denominator was entered in userdiv/1: -1

no
| ?- userdiv(3,Y).
Enter a number: 2.

Y = 1.5000

yes

Note, however the following behavior.

| ?- userdiv(3,F).
Enter a number: 0.
++Error[XSB/Runtime/P] uncaught exception: error(zerodivision,userdiv / 1)
Aborting...

By examining the program above, it can be seen that if \(p(1)\) is entered, the predicate \texttt{type\_error/3}\ is called. \texttt{type\_error/3}\ is an XSB mechanism to throw a type error from Prolog. Error terms thrown by system predicates such as \texttt{type\_error/3}\ are in XSB’s \textit{standard error format}, which is ISO-compatible and which may encode useful information. For instance, the type error thrown in the above example is known to XSB’s default system error handler which prints out a message along with a \textit{backtrace} that indicates the calling context in which the error arose (this behavior can be controlled: see Section 12.5). Alternately, in the second case, when \(-1\) is entered, the (non-standard) error term \texttt{mydiv1(error2(-1))}\ is thrown, which is caught within \texttt{userdiv/2}\ and handled by \texttt{handleUserdiv/2}. Finally, when 0 is entered for the denominator, an error term of the form \texttt{error(zerodivision,userdiv/1)}\ is thrown, and this term does not unify with the second argument of the \texttt{catch/3} literal in the body of \texttt{userdiv/1}, or with any error in standard format. The error is instead caught
by XSB’s default system error handler which prints an uncaught exception message and aborts to the top level of the interpreter.

XSB has two default system error handlers: one used when XSB is called as a stand-alone process, and another when XSB is embedded in a process. Each recognizes the same error formats (see Section 12.2), and handles the rest as uncaught exceptions. However, there may be times when an application requires special default handling: perhaps the application calls XSB from through a socket, so that aborts are not practical. As another example, perhaps XSB is being called from a graphical user interface via Interprolog [9] or some other interface, so that in addition to a special abort handling, one would like to display an error window. In these cases it is convenient to make use of the dynamic predicate \texttt{default_user_error_handler/1}. 

\texttt{default_user_error_handler/1} is called immediately before the default system error handler, and after it is ascertained that no catcher for an error term is available via a \texttt{catch/3} ancestor.

It is important to note that the system error handlers catch errors only in the main thread, and do not affect errors thrown by goals executed by \texttt{thread_create/[2,3]}. Error terms thrown by goals executed by non-detached threads are stored internally, and can be obtained by \texttt{thread_join/2}. Error terms thrown by detached threads are lost when the thread exits, so that any error handling for a detached thread should be performed within the thread itself. See Chapter 7 for further information.

Accordingly, suppose the following clause is asserted into \texttt{usermod}:

\begin{verbatim}
?- assert((default_user_error_handler(error(zerodivision,Pred)): -
    error_writeln(['Aborting: division by 0 in: ',Pred])).
\end{verbatim}

The behavior will now be

\begin{verbatim}
| ?- userdiv(4,F).
Enter a number: 0.
Aborting: division by 0 in: userdiv / 1
\end{verbatim}

The actions of \texttt{catch/3} and \texttt{throw/1} resemble that of the Prolog cut in that they remove choice points that lie between a call to \texttt{throw/1} and the matching \texttt{catch/3} that serves as its ancestor.

The predicate \texttt{call_cleanup/2} (cf. Section 6.11) can be used with \texttt{catch/3}, since the goal \texttt{call_cleanup(Goal,Cleanup)} executes \texttt{Cleanup} whenever computation of \texttt{Goal} is completed, whether because \texttt{Goal} has thrown an exception, has failed, or has succeeded with its last answer. \texttt{call_cleanup/2} can thus be used to release resources created by \texttt{Goal} (such as streams, mutexes, database cursors, etc.). However, if \texttt{Goal}
CHAPTER 12. EXCEPTION HANDLING

395

throws an exception, \texttt{call_cleanup/2} will re-throw the exception after executing cleanup.

12.1.2 Exception Handling in Tabled Evaluation

The exception handling as previously described requires extensions in order to work well with tabled predicates. First, if an \textit{unhandled} exception is thrown during evaluation of a tabled subgoal \( S \) and \( S \) is not completed, the table for \( S \) is not meaningful and should be removed. (Tables that have been completed are not affected by exceptions.) Accordingly, the user will sometimes see the message:

\begin{verbatim}
Removing incomplete tables...
\end{verbatim}

written to standard feedback. But what about exceptions that are \textit{caught} during the computation of \( S \)?

The proper action to take in such a case is complicated by the scheduling mechanism of tabling which, as discussed in Chapter 5, is more complex than in Prolog. Rather than a simple depth-first search, as in Prolog, tabled evaluations effectively perform a series of fixed-point computations for various sets of mutually dependent subgoals, which are termed \textit{SCCs} \footnote{This term is used since sets of mutually dependent subgoals are formally modelled as (approximate) \textit{Strongly Connected Components} within a dependency graph.}. In fact, a tabled evaluation can be seen as a tree of SCCs (in batched evaluation) or a chain of SCCs (in local evaluation). In a tabled evaluation XSB’s throw mechanism searches for the nearest catcher \( C \) among its ancestors

- whose first argument unifies with the thrown error; and

- where \( C \) is between SCCs: that is where the set of subgoals that depend on \( C \) is disjoint from the set of subgoals upon which \( C \) depends. We term this the \textit{SCC restriction} for exception handling.

This behavior can be best understood by an example. Consider the query \texttt{a(X)} to the program in Figure 12.2 which has the following output:

\begin{verbatim}
| ?- a(X).
a_calling_b
b_calling_a
\end{verbatim}
:- table a/1, b/1, c/1, d/1.
a(X):- writeln(a_calling_b), b(X).
b(X):- writeln(b_calling_a), a(X).
b(X):- writeln(b_calling_c), catch(c(X),_,(writeln(handled_1),fail)).
c(X):- writeln(c_calling_d), d(X).
c(X):- writeln(c_aborting), abort.
d(X):- writeln(d_calling_c), catch(c(X),_,(writeln(handled_2),fail)).

Figure 12.2: A program to illustrate exception handling in tabled evaluations

Note that there are 2 SCCs, \{a(X), b(X)\} and \{c(X), d(X)\}. When the `abort` is called in the body of `c(X)` the catch in the body of `d(X)` is its nearest ancestor; however this catch is skipped over, and the catch in the body of `b(X)` takes effect. This catch is between the SCCs – the first SCC depends on it, but the second doesn’t. Due to the SCC restriction, the actual behavior of exception handling with tabling is thus somewhat less intuitive than in Prolog. If this restriction were lifted, there would be no guarantee that there existed a unique catch that was the closest ancestor of an exception.

While the above mechanism offers a great deal of flexibility, for many cases the best approach to exception handling is to keep it simple.

1. Use catches when there will be no tabled subgoal between an exception and its catcher. For instance, sometimes it may be annoying to have `atom_codes/2` throw an exception rather than failing, if given an integer in its first argument. This can be addressed by the predicate

\[
\text{my_atom_codes}(X,Y):= \\
\quad \text{catch}(\text{atom_codes}(1,B),\text{error}(\text{type_error}(A,B),C,D),\text{writeln}(E)).
\]
which, for a type error, does not interact with tabling in any way.

2. Similarly, if only subgoals to completed tables occur between an exception and its catcher, exception handling behaves just as in case 1).

3. Otherwise, abort the entire tabled computation and handle it from there. (Unless you really know what you’re doing!)

Obtaining Information about a Tabled Computation after an Exception is Thrown

XSB backtraces (Section 12.5) provide information about the context in which error is thrown, but in a tabled computation additional information is available. If the Prolog flag exception_pre_action is set to print_incomplete_tables (its default setting is none), then when an exception is thrown, incomplete tables and their SCC information at the time an exception is thrown are printed to a file via print_incomplete_tables/1. The file may be obtained through the predicate get_scc_dumpfile/1 in the module tables. No file is generated unless the exception is thrown over at least one incomplete table.

12.2 XSB’s Standard Format for Errors

All exceptions that occur during the execution of an XSB program can be caught. However, by structuring error terms in a consistent manner, different classes of errors can be handled much more easily by handlers, both system- and user-defined. This philosophy partly underlies the ISO Standard for defining classes of Prolog errors [34]. While the ISO standard defines various types of errors and how they should arise during execution of ISO Prolog predicates, it only partially defines the actual error terms a system should use. The ISO format can be represented as:

\[
\text{error}(\text{Tag}, \text{Context}),
\]

where \text{Tag} is specific to each class of error, while \text{Context} is implementation-dependent.\(^5\)

\(^5\)If a program catches errors itself, \text{error}/2 may need to be imported from \text{error_handler}.\)
12.2.1 Error Tags

In XSB, the ISO-compliant values for Tag are given below.

domain_error(Valid_type,Culprit) is the tag for an ISO domain error, where Valid_type is the domain expected and Culprit is the term observed. Various ISO predicates may have specific domains for input values; and in addition unlike types, domains can be user-defined.

evaluation_error(Flag) is the tag for an ISO evaluation error (e.g. overflow or underflow), and Flag is the type of evaluation error encountered (e.g., undefined, if an arithmetic function is undefined for a given input).

existence_error(Type,Culprit) is the tag for an ISO existence error, where Type is the type of a resource (e.g., a predicate, stream, attribute handler, etc.) and Culprit is the term observed.

instantiation_error is the tag for an ISO instantiation error.

permission_error(Op,Obj_type,Culprit) is the tag for an ISO permission error, when an operation Op was applied to an object of type Obj_type, but Culprit was observed.

representation_error(Flag) is the tag for an ISO representation error (e.g., the maximum arity of a predicate has been exceeded), and Flag is the type of representation error encountered.

resource_error(Flag) is the tag for an ISO resource error (e.g. allowed memory has been used, or too many files have been opened), and Flag is the type of resource error encountered.

syntax_error and syntax_error(Culprit) are alternate tags for an ISO syntax error, where Culprit denotes a syntactically-incorrect sequence of tokens.

system_error(Flag) is the tag for an ISO system error, and Flag is the type of system error encountered.

type_error(Valid_type,Culprit) is the tag for an ISO type error, where Valid_type is the type expected and Culprit is the term observed. As opposed to domain errors, type errors should be used for checks of Prolog types only (i.e. integers, floats, atoms, etc.)

In addition, XSB also makes use of two other classes of errors.
table_error and type_error(Subtype) are the tags for an error arising when using XSB’s tabling mechanism, when the condition giving rise to the error does not easily fit under one of the above classes.

misc_error is the tag for an error that is not otherwise classified.

error(thread_cancel,Id) is the format of an error ball for a thread that has been cancelled by XSB thread Id (See Chapter 7 for details on thread cancellation.)

In Version 3.6 of XSB, errors for ISO predicates usually, but not always ISO-compliant, although the number of such non-compliances will reduce over time. First, when XSB determines it is out of available system memory, recovering from such an error may be difficult at best. Accordingly the computation is aborted in the sequential engine, or XSB exits in the multi-threaded engine. Second, errors in XSB code sometimes arise as miscellaneous errors rather than as a designated ISO-error type.

12.2.2 XSB-Specific Information in Error Terms

XSB also encodes other information in error terms, which may vary with the error thrown, the form in which XSB was compiled, and the version of XSB. In addition, the specifics of how the information is represented may vary, so that this information should always be represented through the access methods described in this and the next section.

Message describes the error in human-readable format. Messages are present in all of XSB’s system errors, and can be obtained through xsb_error_get_message/2.

Goal represents tabled goal that is closest to the environment of the thrown error. It is present in some, but not all error terms thrown by XSB and can be obtained as a term through xsb_error_get_goal/2, and as an atom through xsb_error_get_goalatom/2.

Thread Id is an atom ‘th <tid>’ indicating the id of the thread that threw the error. Thread Id information is only present when using the multi-threaded version of XSB, and even in that version is not present in all error terms. The predicate xsb_error_get_tid/2 can be used to obtain this information if present.

6This does not include overflowing a memory limit specified by the flag max_memory.
Backtrace represents the stack of the forward continuations in the execution stack at the time the error was thrown. Backtraces are present by default in all XSB system error terms, and are described in Section 12.5. They may be obtained from an error term using the predicate `xsb_error_get_backtrace/2`.

## 12.3 Predicates to Throw and Handle Errors

### 12.3.1 Predicates to Throw Errors

XSB provides a variety of predicates that throw errors\(^7\). In general, we recommend the use of predicates such as `domain_error/4` over the direct use of `throw/1` when possible.

**throw(+ErrorTerm)**

ISO

Throws the error `ErrorTerm`. Execution traverses up the choice point stack until a goal of the form `catch(Goal,Term,Handler)` is found such that `Term` unifies with `ErrorTerm`. In this case, `Handler` is called. If no catcher is found in the main thread, the system looks for a clause of `default_user_error_handler(Term)` such that `Term` unifies with `ErrorTerm` — if no such clause is found the default system error handler is called. In a non-main joinable thread, the error term is stored internally and the thread exits; in a detached thread, the thread exits with no action taken. `throw/1` is most useful in conjunction with specialized handlers for new types of errors not already supported in XSB.

**domain_error(+Valid_type,-Culprit,+Predicate,+Arg)**

module: `error_handler`

Throws a domain error. Using the default system error handler (with the Prolog flag `backtrace_on_error` set to off) an example is

```
domain_error(posInt,-1,checkPosInt/3,3).
++Error[XSB/Runtime/P]: [Domain (-1 not in domain posInt)] in arg 3 of predicate checkPosInt/3
```

**evaluation_error(+Flag,+Predicate,+Arg)**

module: `error_handler`

Throws an evaluation error. Using the default system error handler (with the Prolog flag `backtrace_on_error` set to off) an example is

\(^7\) C functions for throwing terms and ISO-style errors are described in Volume 2, Chapter 3 Foreign Language Interface.
CHAPTER 12. EXCEPTION HANDLING

evaluation_error(zero_divisor,unidiv/1,2).
++Error[XSB/Runtime/P]: [Evaluation (zero_divisor)] in arg 2 of predicate unidiv/2

existence_error(+Object_type,?Culprit,+Predicate,+Arg) module: error_handler
Throws an existence error. Using the default system error handler (with the Prolog flag backtrace_on_error set to off) an example is

existence_error(file,'myfile.P','load_intensional_rules/2',2).
++Error[XSB/Runtime/P]: [Existence (No file myfile.P exists)] in arg 2 of predicate load_intensional_rules/2

instantiation_error(+Predicate,+Arg,+State) module: error_handler
Throws an instantiation error. Using the default system error handler, an example (with the Prolog flag backtrace_on_error set to off) is

?- instantiation_error(foo/1,1,nonvar).
++Error[XSB/Runtime/P]: [Instantiation] in arg 1 of predicate foo/1: must be nonvar

permission_error(+Op,+Obj_type,?Culprit,+Predicate) module: error_handler
Throws a permission error. Using the default system error handler, an example (with the Prolog flag backtrace_on_error set to off) is

| ?- permission_error(write,file,'myfile.P',foo/1).
++Error[XSB/Runtime/P]: [Permission (Operation) write on file: myfile.P] in foo/1

representation_error(+Flag,+Predicate,+Arg) module: error_handler
Throws a representation error. Using the default system error handler, an example (with the Prolog flag backtrace_on_error set to off) is

representation_error(max_arity,assert/1,1).
++Error[XSB/Runtime/P]: [Representation (max_arity)] in arg 1 of predicate assert/1

resource_error(+Flag,+Predicate) module: error_handler
Throws a resource error. Using the default system error handler (with the Prolog flag backtrace_on_error set to off) and example is

resource_error(open_files,open/3)
++Error[XSB/Runtime/P]: [Resource (open_files)] in predicate open/3
CHAPTER 12. EXCEPTION HANDLING

402

type_error(+Valid_type,-Culprit,+Predicate,+Arg)  module: error_handler

Throws a type error. Using the default system error handler, an example (with the Prolog flag backtrace_on_error set to off) is

?-> type_error(atom,f(1),foo/1,1).
++Error[XSB/Runtime/P]: [Type (f(1) in place of atom)] in arg 1 of predicate foo/1

misc_error(+Message)  module: error_handler

Throws a miscellaneous error that will be caught by the default system handler. For good programming practice miscellaneous errors should only be thrown when the cases above are not applicable, and the type of error is not of interest for structured error handling. Such situations occur can occur for instance in debugging, during program development, or for other reasons. Note that this misc_error/2 replaces the obsolescent XSB predicates abort/1 and abort/2.

12.3.2 Predicates used in Handling Errors

For best results, output for handling errors should be sent to XSB’s standard error stream using the alias user_error or one of the predicates described below.

catch(?Goal,?CatchTerm,+Handler)  ISO

Calls Goal, and sets up information so that future throws will be able to access CatchTerm under the mechanism mentioned above. catch/3 does not attempt to clean up system level resources with the exception of incomplete tables, which are abolished as discussed in Section 12.1.2. However, it is left up to the handler to close any open files, reset current input and output, and so on 8.

default_user_error_handler(?CatchTerm)

Handles any error terms that unify with CatchTerm that are not caught by invocations of catch/3. This predicate closes open tables and release mutexes held by the calling thread, but does not attempt to clean up other system level resources, which is left to the handler.

error_write(?Message)  module: standard

error_writeln(?Message)  module: standard

Utility routines for user-defined error catching. These predicates output Message to XSB’s STDERR stream, rather than to XSB’s STDOUT stream, as does write/1 and writeln/1. In addition, if Message is a comma list, the elements in the comma list are output as if they were concatenated together. Each of these predicates must be implicitly from the module standard.

8cf. the default system error handler, which performs these functions, if needed.
CHAPTER 12. EXCEPTION HANDLING

xsb_error_get_message(Err, Message) module: error_handler
Obtains the message associated with an error in XSB’s standard format. All
errors in standard format have messages.

xsb_error_get_goal(Err, ?Goal) module: error_handler
Obtains the goal (represented as a Prolog term), if any, from an error term that
is in XSB’s standard format. If the error term has no goal, the predicate fails.

xsb_error_get_goalatom(Err, ?GoalAtom) module: error_handler
Obtains the goal (represented as a Prolog atom), if any, from an error term that
is in XSB’s standard format. If the error term has no goal, the predicate fails.
This routine is slightly more efficient than xsb_error_get_goal/2.

xsb_error_get_tid(Err, ?Tid) module: error_handler
Obtains the atom ’th <tid>’ indicating the id of the thread that threw the
error. Thread Id information is only present when using the multi-threaded
version of XSB, and even in that version is not present in all error terms.

xsb_error_get_backtrace(+Err, -Backtrace) module: error_handler
Obtains the backtrace — the stack of the forward continuations in the execution
stack at the time the error was thrown. Backtraces are present by default in all
XSB system error terms, and are described in Section 12.5.

12.4 Convenience Predicates

The following convenience predicates are provided to make a commonly used check
and to throw an ISO error if the check is not satisfied; some are written directly in C
for speed. All these predicates must be imported from the module error_handler,
which also contains provides a few other specialized checks.

check_acyclic(?Term, +Predicate, +Arg) module: error_handler
Checks that Term is acyclic. If so, the predicate succeeds; if not it throws a
miscellaneous error.

check_atom(?Term, +Predicate, +Arg) module: error_handler
Checks that Term is an atom. If so, the predicate succeeds; if not it throws a
type error.

check Callable(?Term, +Predicate, +Arg) module: error_handler
Checks that Term is callable. If so, the predicate succeeds; if not it throws a
type error.
**CHAPTER 12. EXCEPTION HANDLING**

check_ground(?Term,+Predicate,+Arg)  
module: error_handler  
Checks that $Term$ is ground. If so, the predicate succeeds; if not it throws an instantiation error.

check_integer(?Term,+Predicate,+Arg)  
module: error_handler  
Checks that $Term$ is an integer. If so, the predicate succeeds; if not it throws a type error.

check_nonvar(?Term,+Predicate,+Arg)  
module: error_handler  
Checks that $Term$ is not a variable. If not, the predicate succeeds; if $Term$ is a variable, it throws an instantiation error.

check_nonvar_list(?Term,+Predicate,+Arg)  
module: error_handler  
Checks that $Term$ is a list, each of whose elements is ground. If so, the predicate succeeds; if not it throws an instantiation error.

check_one_thread(+Operation,+Object_Type,+Predicate)  
module: error_handler  
In the multi-threaded engine, check_one_thread/3 checks that there is only one active thread: if not, a miscellaneous error is thrown indicating that $Operation$ is not permitted on $Object_Type$ as called by $Predicate$, when more than one thread is active. This check provides a convenient way to allow inclusion of certain operations that are difficult to make thread-safe by other means.

In the single-threaded engine this predicate always succeeds.

check_stream(?Stream,+Predicate,+Arg)  
module: error_handler  
Checks that $Stream$ is a stream. If so, the predicate succeeds; if not it throws an instantiation error.\(^9\)

check_var(?Term,+Predicate,+Arg)  
module: error_handler  
Checks that $Term$ is a variable. If so, the predicate succeeds; if not it throws an instantiation error.

### 12.5 Backtraces

Displaying a backtrace of the calling context of an error in addition to an error message can greatly expedite debugging. For XSB’s default error handler, backtraces are printed out by default, a behavior that can be overridden for a given thread by the

\(^9\)The representation of streams in XSB is subject to change.
command: \texttt{set_prolog_flag(backtrace\_on\_error,off)}. For users who write their own error handlers, the following predicates can be used to manipulate backtraces.

It is important to note that Prolog backtraces differ in a significant manner from backtraces obtained from other languages, such as C backtraces produced by GDB. This is because a Prolog backtrace obtains forward continuations from the local environment stack, and in the WAM, local stack frames are only created when a given clause requires permanent variables – otherwise these stack frames are optimized away. The precise conditions for optimizing away a local stack frame require an understanding of the WAM (and of a specific compiler). However in general, longer clauses with many variables require a local stack frame and their forward continuations will be displayed, while shorter clauses with fewer variables do not and their forward continuations will not be displayed.

\begin{verbatim}
\texttt{xsb\_backtrace(-Backtrace)}
\end{verbatim}

\textit{Upon success \texttt{Backtrace} is bound to a structure indicating the forward continuations for a point of execution. This structure should be treated as opaque, and manipulated by one of the predicates below.}

\begin{verbatim}
\texttt{get\_backtrace\_list(+Backtrace,-PredicateList)}
\end{verbatim}

\textit{Given a backtrace structure, this predicate produces a list of predicate identifiers or the form \texttt{Module:Predicate/Arity}. This list can be manipulated as desired by error handling routines.}

\begin{verbatim}
\texttt{print\_backtrace(+Backtrace)}
\end{verbatim}

\textit{This predicate, which is used by XSB’s default error handler, prints a backtrace structure to XSB’s standard error stream.}

When XSB generates a memory exception \textit{at the OS level} (e.g., a segmentation violation or bus error) it prints out a backtrace and exits. This should be caused only by a bug in XSB or included C code. The first predicate in the backtrace that is printed in these circumstances may be incorrect or redundant. This is because the memory structures used to generate the backtrace are not always completely consistent, and so an interrupt at an unexpected point may result in the use of somewhat inconsistent information.
Chapter 13

Foreign Language Interface

When XSB is used to build real-world systems, a foreign-language interface may be necessary to:

- combine XSB with existing programs and libraries, thereby forming composite systems;
- interface XSB with the operating system, graphical user interfaces or other system level programs;
- speed up certain critical operations.

XSB has both a high-level and the low-level interface to C. The low-level interface is much more flexible, but it requires greater attention to details of how the data is passed between XSB and C. To connect XSB to a C program using the high-level interface requires very little work, but the program must be used “as is” and it must take the input and produce the output supported by this high-level interface. Before describing the interfaces themselves, we first describe aspects common to both the lower- and higher-level foreign language interfaces.

The foreign language interface can also support C++ programs. Since XSB is written in C, the interface functions in the foreign C++ module must have the declaration extern "C", and a separate compiler option (e.g. specifying g++ rather than gcc) may need to be given to ensure proper linkage, inclusion of C++ libraries, etc. In addition, on certain platforms compilation may need to be done externally to XSB – see the xasp 1package for a example of using the foreign language interface with C++ files. For the rest of this chapter, we restrict our attention to foreign predicates written in C.
13.1 Foreign Language Modules

Foreign predicates must always appear in modules, and these modules can contain only foreign predicates. A foreign module differs from a Prolog module in that the foreign module’s source file must appear in a *.c file rather than a *.P file (or .pl file). This *.c file cannot contain a main() function. Furthermore, a *.P file with the same name must not be present or else the *.c file is ignored and the module is compiled as a regular Prolog module. The interface part of a foreign module, which has the same syntax as that of a normal module, is written in Prolog and must appear in a *.H file. If the lower-level interface is used, this *.H file contains explicit export/1 declarations for the the foreign predicates that are to be used by other modules; if the higher-level interface is used, the declarations have the form foreign_pred/1.

The Prolog predicates attached to foreign functions are deterministic, in the sense that they succeed at most once for a given call and are not re-entered on backtracking. Note that this requirement imposes no serious limitation, since it is always possible to divide a foreign predicate into the part to be done on the first call and the part to be redone on backtracking. Backtracking can then take place at the Prolog level where it is more naturally expressed.

A foreign module can be compiled or consulted just like a normal Prolog module. Currently, predicates consult/[1,2] recompile both the *.c and the *.H files of a foreign module when at least one of them has been changed from the time the corresponding object files have been created (see the section Compiling and Consulting in Volume 1) \(^1\). The C compiler used to compile the *.c files can be set as a defaults to that used for the configuration of XSB (refer to the section Getting Started with XSB in Volume 1). This default behavior includes the C compilation options used to compile XSB when it was configured, along with a default set of include files so that header files in XSB directories can be obtained. Alternately, the user can add options to be passed to the C compiler. To give an example, the following command will compile file file.c using the default C Compiler with optimization and by including /usr/local/X11/R6/include to the directories that will be searched for header files.

    :- consult(file, [cc_opts('-O2 -I/usr/local/X11/R6/include')]).

Note in particular, that if XSB were compiled with the -g debugging option, then the C file will be also \(^2\). Any Prolog compiler options are ignored when compiling a foreign module.

\(^1\)In addition, if a C module compiled by the single-threaded XSB engine is loaded by the multi-threaded engine, it will be recompiled, and vice-versa.

\(^2\) In a 64-bit platform, users may override the default compilation of XSB by the configuration
CHAPTER 13. FOREIGN LANGUAGE INTERFACE

Prolog-specific directives such as `index`, `hilog`, `table`, `auto_table` or even `import` make no sense in the case of a foreign module and thus are ignored by the compiler. However, another directive, namely `ldoption`, is recognized in a foreign module and is used to instruct the dynamic loading and linking of the module. The syntax of the `ldoption` directive is simply:

```prolog
:- ldoption(Option).
```

where `Option` should either be an atom or a list of atoms. Multiple `ldoption` directives may appear in the same `.H` file of a foreign module. In Unix-derived systems, the foreign language interface of XSB uses `ld` command that combines object programs to create an executable file or another object program suitable for further `ld` processing. Version 3.6 of XSB assumes that the `ld` command resides in the file `/usr/bin/ld`.

13.2 Lower-Level Foreign Language Interface

Creating a foreign predicate using the lower-level foreign language interface is almost entirely a matter of writing C code. Consider the foreign module `$XSBDIR/examples/XSB_calling_c/simple_foreign.c`. The `.H` file has the form:

```prolog
:- export minus_one/2, my_sqrt/2, change_char/4.
:- ldoption('-lm'). % link together with the math library
```

When the lower level foreign language interface is used, C functions that implement foreign predicates must return values of type `int`. The return value is not used by a Prolog argument; rather if a non-zero is returned, the foreign predicate succeeds; a zero return value means failure.

---

options `-with-bits32` or `-with-bits64`. If either of these options is used, the default compilation options will pass along the appropriate memory options. If XSB is compiled with a memory option that is not the default of the platform, and if an externally compiled C file is to be loaded into XSB, it must be ensured that the C file has been compiled with the appropriate memory options: `-m32` or `-m64` if gcc is used.

Mac OSX users using 10.3 or above should have the environment variable `MACOSX_DEPLOYMENT_TARGET` set to 10.3 so that the compiler generates code that can be dynamically linked by XSB. This should be done automatically by XSB on initialization, but it is useful to check if encountering problems.
At the C level, the function that implements the Prolog predicate must have the same name as the Prolog predicate (that is declared in the *.H file), and must have a special context parameter macro. The context parameter macro allows C functions to be used with both the single-threaded and multi-threaded engines, and are described in detail in Section 13.2.1. The Prolog level arguments are converted to C data structures through several predefined functions rather than through direct parameter passing. The C file simple_foreign.c corresponding to the above .H file is as follows.

```c
/*-----------------------------*/
#include <math.h>
#include <stdio.h>
#include <string.h>
#include <alloca.h>

/*----- Make sure your C compiler finds the following header file. -----*
----- One way to do this is to include the directory XSB/emu on the -----*
----- compiler's command line with the -I (/I in Windows) option -----*/
#include "cinterf.h"

/*-----------------------------*/

int minus_one(CTXTdecl)
{
    int i = ptoc_int(CTXTc 1);
    ctop_int(CTXTc 2, i-1);
    return TRUE;
}

/*-----------------------------*/

int my_sqrt(CTXTdecl)
{
    int i = ptoc_int(CTXTc 1);

4 The inclusion of context parameters changes the lower-level interface for Version 3.0. C files written for previous versions of XSB continue to work properly for the single-threaded engine in, but will not work properly for the multi-threaded engine.
```
ctop_float(CTXTc 2, (float) pow((double)i, 0.5));
return TRUE;
}

/*-----------------------------------------*/

int change_char(CTXTdecl)
{
char *str_in;
int pos;
int c;
char *str_out;

str_in = (char *) ptoc_string(CTXTc 1);
str_out = (char *) alloca(strlen(str_in)+1);
strcpy(str_out, str_in);
pos = ptoc_int(CTXTc (2);
c = ptoc_int(CTXTc (3);
if (c < 0 || c > 255) /* not a character */
    return FALSE; /* this predicate will fail on the Prolog side */

str_out[pos-1] = c;

extern_cstop_string(CTXTc 4, str_out);
return TRUE;
}

/*-----------------------------------------*/

Before describing the C program used, here is a sample session illustrating the behavior of the predicates in simple_foreign.

XSB Version 2.0 (Gouden Carolus) of June 26, 1999
[i686-pc-linux-gnu; mode: optimal; engine: slg-wam; scheduling: batched]
| ?- [simple_foreign].
[Compiling C file ./simple_foreign.c using gcc]
[Compiling Foreign Module ./simple_foreign]
[simple_foreign compiled, cpu time used: 0.0099993 seconds]
[simple_foreign loaded]

yes
| ?- change_char('Kostis', 2, w, TempStr),
Consider the function `minus_one()` above. As discussed, it takes a context parameter (explained below), and returns an integer, and as can be seen the return values can be specified by the macros `TRUE` and `FALSE`. From the Prolog perspective the first argument to `minus_one/2` is an (integer) input argument, while the second is an (integer) output argument. Input arguments for basic C types are translated from their Prolog representation to a C representation by functions of the form `ptoc\_<type>()` – here `ctop_int()`. The single parameter of such a function is the number of the Prolog argument that is to be transformed and the function returns the C representation. Output arguments are converted from C to Prolog by corresponding functions of the form `ctop\_<type>()` – here `ctop_int()`. For converting C back to Prolog, the first parameter of `ctop_int()` is the number of the Prolog argument to be transformed and the second is the C value to be transformed. In the session output above, if an improper argument is given to `minus_one/2` it will emit a warning, and succeed. Also note that the call `my_sqrt(23,X)` succeeds once, but fails on backtracking since it is deterministic, as are all other foreign language functions.
The above example illustrates the exchange of \textit{basic} types through the lower-level interface – e.g. atoms, integers, and floating-point numbers. The lower-level interface also allows a user to pass lists and terms between XSB and C as will be discussed in Section 13.2.3.

13.2.1 Context Parameters

When using the lower-level interface, \textit{context parameters} must be added to many C functions in order for the functions to be used with XSB's multi-threaded engine. In the multi-threaded engine, variables for Prolog's virtual machine, as well as for thread-private data structures are stored in a \textit{context structure}. This context structure must be passed to any functions that need to access elements of a thread's virtual machine – including many of the functions that are used to exchange data between Prolog and C. We note in passing that when using the multi-threaded engine, a user must ensure that foreign-language functions are thread-safe, by using standard multi-threaded programming techniques, including XSB's mutex predicates (see the Section \textit{Predicates for Thread Synchronization} in Volume 1 of this manual). On the other hand, in the single-threaded engine virtual machine elements are kept in static variables, so that context parameters are not required.

The lower-level C interface makes use of a set of macros to address the requirements of the different engines. The data exchange functions discussed in this chapter, \texttt{ptoc\_xxx}, \texttt{ctop\_xxx}, \texttt{c2p\_xxx}, \texttt{p2c\_xxx}, and \texttt{p2p\_xxx} usually, but not always, require information about a thread's virtual machine state. If a C function directly or indirectly calls a data interchange function that requires a context parameter, the function must have a context parameter in its declaration, calls, and prototypes in order to be used by the multi-threaded engine. These context parameters have the following forms:

- In function \textit{declarations}, use the macro \texttt{CTXTdecl} in the code for a function that would otherwise be \texttt{void}, and \texttt{CTXTdeclc} as the first argument in the code for a function with parameters (\texttt{CTXTdeclc} and \texttt{CTXTdecl} are similar, except that macro expansion of \texttt{CTXTdeclc} for the multi-threaded engine includes a comma). The example for \texttt{minus\_one}(\texttt{CTXTdecl}) shows use of this macro.

- In function \textit{calls} use the macro \texttt{CTXT} in the code for a function that would otherwise be \texttt{void}, and \texttt{CTXTc} as the first argument in the code for a function with parameters. As an example, a call to \texttt{minus\_one} would have the form \texttt{minus\_one}(\texttt{CTXT}).
In function prototypes use the macro CTXDecltype in the code for a function that would otherwise be void, and CTXDecltypec as the first argument in the code for a function with parameters. As an example, a prototype for minus_one would have the form minus_one(CTXDecltype).

Fortunately, when compiling with the multi-threaded engine, it is easy to determine at compile time whether context parameters are correct. If compilation of a function foo gives an error along the lines of:

```
foofile.c: In function 'foo':
foofile.c:109: error: 'th' undeclared (first use in this function)
```

Then the declaration of foo omitted a context parameter. If compilation gives an error along the lines of

```
foofile.c: In function 'foo_caller':
foofile.c:149: error: too few arguments to function 'foo'
```

Then the call to foo may have omitted a context parameter.

Note that context parameters are only necessary if the lower-level interface is used. The higher-level interface automatically generates any context parameters it needs.

### 13.2.2 Exchanging Basic Data Types

The basic interface assumes that correct modes (i.e., input or output parameters) and types are being passed between the C and Prolog levels. As a result, output unification should be explicitly performed in the Prolog level. The prototypes for the conversion functions between Prolog and C should be declared before the corresponding functions are used. This is done by including the "cinterf.h" header file. Under Unix, the XSB foreign C interface automatically finds this file in the XSB/emu directory. Under Windows (including Cygwin), the user must compile and create the DLL out of the C file manually, so the compiler option '/I...\XSB\emu' is necessary.

The following C functions are used to convert basic types between Prolog and C.

```c
int ptoc_int(CTXDeclc int N)
CTXDeclc is a context parameter; N is assumed to hold a Prolog integer corresponding to the Nth argument of a Prolog predicate. This function returns the value of that argument in as a C int.
```
double ptoc_float(CTXTdeclc int N)

CTXTdeclc is a context parameter; N is assumed to hold a Prolog integer corresponding to the Nth argument of a Prolog predicate. This function returns the value of that argument as a C double. By default, XSB provides double precision, but if XSB was configured with -enable-fast-floats less than single precision can be provided.

char *ptoc_string(CTXTdeclc int N)

CTXTdeclc is a context parameter; N is assumed to hold a Prolog integer corresponding to the Nth argument of a Prolog predicate. This function returns the value the C string (of type char *) that corresponds to this interned Prolog atom. WARNING: the string should be copied before being manipulated in any way: otherwise unexpected results may arise whenever the interned Prolog atom is unified.

void ctop_int(CTXTdeclc int N, int V)

CTXTdeclc is a context parameter; argument N is assumed to hold a Prolog free variable, and this function binds that variable to an integer of value V.

void ctop_float(CTXTdeclc int N, float V)

CTXTdeclc is a context parameter; argument N is assumed to hold a Prolog free variable, and this function binds that variable to a floating point number of value V.

void extern_ctop_string(CTXTdeclc int N, char * V)

CTXTdeclc is a context parameter; argument N is assumed to hold a Prolog free variable. If needed, this function interns the string to which V points as a Prolog atom and then binds the variable in argument N to that atom.

### 13.2.3 Exchanging Complex Data Types

If the lower-level interface is used, exchanging basic data types is sufficient for most applications. Exchanging complex data types is also possible, although doing so is slightly more involved than exchanging basic types. To exchange complex data types, the lower-level interface uses only one C data type: prolog_term, which can point to any XSB term. On the C side, the type of the term can be checked and then

---

5 The fast float configuration option does represent floating point values as directly tagged single precision values rather than as indirectly tagged double precision values. Speed increases in arithmetic can be gained from this optimization, in exchange for significant precision loss on floating point numbers.
processed accordingly. For instance, if the term turns out to be a structure, then it can be decomposed and the functor can be extracted along with the arguments. If the term happens to be a list, then it can be processed in a loop and each list member can be further decomposed into its atomic components. The advanced interface also provides functions to check the types of these atomic components and for converting them into C types.

We begin by presenting the functions used to exchange complex data types, before presenting a detailed example below. As when exchanging basic C types, the file \texttt{emu/cinterf.h} must be included in the C program in order to make the prototypes of the relevant functions known to the C compiler.

The first set of functions is typically used to check the type of Prolog terms passed into the C program.

\begin{verbatim}
xsbBool is_attv((prolog_term) T)
    is_attv(T) returns \texttt{TRUE} if T represents an XSB attributed variable, and \texttt{FALSE} otherwise.

xsbBool is_float((prolog_term) T)
    is_float(T) returns \texttt{TRUE} if T represents an XSB float value, and \texttt{FALSE} otherwise.

xsbBool is_functor((prolog_term) T)
    is_functor(T) returns \texttt{TRUE} if T represents an XSB structure value (not a list), and \texttt{FALSE} otherwise.

xsbBool is_int((prolog_term) T)
    is_int(T) returns \texttt{TRUE} if T represents an XSB integer value, and \texttt{FALSE} otherwise.

xsbBool is_list((prolog_term) T)
    is_list(T) returns \texttt{TRUE} if T represents an XSB list value (not nil), and \texttt{FALSE} otherwise.

xsbBool is_nil((prolog_term) T)
    is_nil(T) returns \texttt{TRUE} if T represents an XSB [] (nil) value, and \texttt{FALSE} otherwise.

xsbBool is_string((prolog_term) T)
    is_string(T) returns \texttt{TRUE} if T represents an XSB atom value, and \texttt{FALSE} otherwise.
\end{verbatim}
CHAPTER 13. FOREIGN LANGUAGE INTERFACE

xsbBool is_var((prolog_term) T)

   is_var(T) returns TRUE if T represents an XSB variable, and FALSE otherwise.

After checking the types of the arguments passed in from the Prolog side, the next task usually is to convert Prolog data into the types understood by C. This is done with the following functions. The first three convert between the basic types. The last two extract the functor name and the arity. Extraction of the components of a list and the arguments of a structured term is explained later.

int p2c_int((prolog_term) V)

   The prolog_term parameter must represent a Prolog integer, and p2c_int returns the C representation of that integer.

double p2c_float((prolog_term) V)

   The prolog_term parameter must represent a Prolog floating point number, and p2c_float returns the C representation of that floating point number.

char *p2c_string((prolog_term) V)

   The prolog_term parameter must represent a (Prolog) atom, and p2c_string returns that atom as a C string. The pointer returned points to the actual atom name in XSB 's atom table, and thus it must NOT be modified by the calling program.

char *p2c_functor((prolog_term) V)

   The prolog_term parameter must represent a structured term (not a list). p2c_functor returns the name of the main functor symbol of that term as a string. The pointer returned points to the actual functor name in XSB 's space, and thus it must NOT be modified by the calling program.

int p2c_arity((prolog_term) V)

   The prolog_term parameter must represent a structured term (not a list). p2c_arity returns the arity of the main functor symbol of that term as a C int.

The next batch of functions support conversion of data in the opposite direction: from basic C types to the type prolog_term. These c2p_ * functions all return a boolean value TRUE if successful and FALSE if unsuccessful. The XSB term argument must always contain an XSB variable, which will be bound to the indicated value as a side effect of the function call.
xsbBool c2p_int(CTXTdeclc (int) N, (prolog_term) V)

  CTXTdeclc is a context parameter; c2p_int binds the prolog_term V (which must be a variable) to the integer value N, creating a Prolog integer.

xsbBool c2p_float(CTXTdeclc (double) F, (prolog_term) V)

  CTXTdeclc is a context parameter; c2p_float binds the prolog_term V (which must be a variable) to the (double) float value F, creating a double Prolog float.

xsbBool c2p_string(CTXTdeclc (char *) S, (prolog_term) V)

  CTXTdeclc is a context parameter; c2p_string binds the prolog_term V (which must be a variable) to the Prolog atom corresponding to the char *S. During this process the Prolog atom is interned into XSB’s atom table.

The following functions create Prolog data structures within a C program. This is usually done in order to pass these structures back to the Prolog side.

xsbBool c2p_functor(CTXTdeclc (char *) S, (int) N, (prolog_term) V)

  CTXTdeclc is a context parameter; c2p_functor binds the prolog_term V (which must be a variable) to an open term whose main functor symbol is given by S (of type char *) and whose arity is N. An open term is one with all arguments as new distinct variables.

xsbBool c2p_list(CTXTdeclc (prolog_term) V)

  CTXTdeclc is a context parameter; c2p_list binds the prolog_term V (which must be a variable) to an open list term, i.e., a list term with both car and cdr as new distinct variables. Note: to create an empty list use the function c2p_nil described below.

xsbBool c2p_nil(CTXTdeclc (prolog_term) V)

  CTXTdeclc is a context parameter; c2p_nil binds the prolog_term V (which must be a variable) to the atom [] (nil).

prolog_term p2p_new()

  Create a new Prolog variable. This is sometimes needed when you want to create a Prolog term on the C side and pass it to the Prolog side.

To use the above functions, one must be able to get access to the components of the structured Prolog terms. This is done with the help of the following functions:

prolog_term p2p_arg((prolog_term) T, (int) A)

  Parameter T must be a prolog_term that is a structured term (but not a list).
A is a positive integer (no larger than the arity of the term) that specifies an argument position of the term $T$. $\text{p2p\_arg}$ returns the $A^{th}$ subfield of the term $T$.

\text{prolog\_term p2p\_car((prolog\_term) T)}

Parameter $T$ must be a \text{prolog\_term} that is a list (not nil). $\text{p2p\_car}$ returns the car (i.e., head of the list) of the term $T$.

\text{prolog\_term p2p\_cdr((prolog\_term) T)}

Parameter $T$ must be a \text{prolog\_term} that is a list (not nil). $\text{p2p\_cdr}$ returns the cdr (i.e., tail of the list) of the term $T$.

It is important to realize that these functions return the actual Prolog term that is, say, the head of a list or the actual argument of a structured term. Thus, assigning a value to such a Prolog term also modifies the head of the corresponding list or the relevant argument of the structured term. It is precisely this feature that allows passing structured terms and lists from the C side to the Prolog side. For instance,

\begin{verbatim}
prolog\_term plist, /* a Prolog list */
   structure; /* something like f(a,b,c) */
prolog\_term tail, arg;

........

tail = p2p\_cdr(plist); /* get the list tail */
arg = p2p\_arg(structure, 2); /* get the second arg */

/* Assume that the list tail was supposed to be a prolog variable */
if (is\_var(tail))
   c2p\_nil(CTXTC tail); /* terminate the list */
else {
   fprintf(stderr, "Something wrong with the list tail!");
   exit(1);
}
/* Assume that the argument was supposed to be a prolog variable */
c2p\_string(CTXTC "abcdef", arg);
\end{verbatim}

In the above program fragment, we assume that both the tail of the list and the second argument of the term were supposed to be bound to Prolog variables. In case of the tail, we check if this is, indeed, the case. In case of the argument, no checks are done; XSB will issue an error (which might be hard to track down) if the second argument is not currently bound to a variable.
The last batch of functions is useful for passing data in and out of the Prolog side of XSB. The first function is the only way to get a prolog_term out of the Prolog side; the second function is sometimes needed in order to pass complex structures from C into Prolog.

```c
prolog_term reg_term(CTXTdeclc (int) R)
    CTXTdeclc is a context parameter. Parameter R is an argument number of the Prolog predicate implemented by this C function (range 1 to 255). The function reg_term returns the prolog_term in that predicate argument.

xsbBool p2p_unify(CTXTdeclc prolog_term T1, prolog_term T2)
    Unify the two Prolog terms. This is useful when an argument of the Prolog predicate (implemented in C) is a structured term or a list, which acts both as input and output parameter. CTXTdeclc is a context parameter.
```

For instance, consider the Prolog call test(X, f(Z)), which is implemented by a C function with the following fragment:

```c
prolog_term newterm, newvar, z_var, arg2;
    ....
    /* process argument 1 */
    c2p_functor(CTXTc "func",1,reg_term(CTXTc 1));
    c2p_string(CTXTc "str",p2p_arg(reg_term(CTXTc 1),1));
    /* process argument 2 */
    arg2 = reg_term(CTXTc 2);
    z_var = p2p_arg(arg2, 1); /* get the var Z */
    /* bind newterm to abc(V), where V is a new var */
    c2p_functor(CTXTc "abc", 1, newterm);
    newvar = p2p_arg(newterm, 1);
    newvar = p2p_new();
    ....
    /* return TRUE (success), if unify; FALSE (failure) otherwise */
    return p2p_unify(CTXTc z_var, newterm);
```

On exit, the variable X will be bound to the term func(str). Processing argument 2 is more interesting. Here, argument 2 is used both for input and output. If test is called as above, then on exit Z will be bound to abc(_h123), where _h123 is some new Prolog variable. But if the call is test(X,f(1)) or test(X,f(Z,V)) then this call will fail (fail as in Prolog, i.e., it is not an error), because the term passed back, abc(_h123), does not unify with f(1) or f(Z,V). This effect is achieved by the use of p2p_unify above.
We conclude this section with two real examples of functions that pass complex data in and out of the Prolog side of XSB. These functions are part of the POSIX regular expression matching package of XSB. The first function uses argument 2 to accept a list of complex Prolog terms from the Prolog side and does the processing on the C side. The second function does the opposite: it constructs a list of complex Prolog terms on the C side and passes it over to the Prolog side in argument 5.

(We should note that this second function could cause a heap overflow in XSB were it to build a large list of values. Instead of building a large list of values on the XSB heap, one would better design the functions to return smaller values, in which case XSB will be able to automatically expand the heap as necessary.)

```c
/* XSB string substitution entry point: replace substrings specified in Arg2
   with strings in Arg3.
   In:
   Arg1: string
   Arg2: substring specification, a list [s(B1,E1),s(B2,E2),...]
   Arg3: list of replacement string
   Out:
   Arg4: new (output) string
   Always succeeds, unless error.
*/
int do_regsubstitute__(CTXTdecl)
{
    /* Prolog args are first assigned to these, so we could examine the types
       of these objects to determine if we got strings or atoms. */
    prolog_term input_term, output_term;
    prolog_term subst_reg_term, subst_spec_list_term, subst_spec_list_term1;
    prolog_term subst_str_term=(prolog_term)0,
        subst_str_list_term, subst_str_list_term1;
    char *input_string=NULL;    /* string where matches are to be found */
    char *subst_string=NULL;
    prolog_term beg_term, end_term;
    int beg_offset=0, end_offset=0, input_len;
    int last_pos = 0;    /* last scanned pos in input string */
    /* the output buffer is made large enough to include the input string and the
       substitution string. */
    char subst_buf[MAXBUFSIZE];
    char *output_ptr;
    int conversion_required=FALSE;    /* from C string to Prolog char list */

    input_term = reg_term(CTXTC 1);    /* Arg1: string to find matches in */
```
if (is_string(input_term)) /* check it */
    input_string = string_val(input_term);
else if (is_list(input_term)) {
    input_string =
        p_charlist_to_c_string(input_term, input_buffer, sizeof(input_buffer),
            "RE_SUBSTITUTE", "input string");
    conversion_required = TRUE;
} else
    xsb_abort("RE_SUBSTITUTE: Arg 1 (the input string) must be an atom or a character list");

input_len = strlen(input_string);

/* arg 2: substring specification */
subst_spec_list_term = reg_term(CTXTc 2);
if (!is_list(subst_spec_list_term) && !is_nil(subst_spec_list_term))
    xsb_abort("RE_SUBSTITUTE: Arg 2 must be a list [s(B1,E1),s(B2,E2),...]" noctum);

/* handle substitution string */
subst_str_list_term = reg_term(CTXTc 3);
if (!is_list(subst_str_list_term))
    xsb_abort("RE_SUBSTITUTE: Arg 3 must be a list of strings");

output_term = reg_term(CTXTc 4);
if (!is_var(output_term))
    xsb_abort("RE_SUBSTITUTE: Arg 4 (the output) must be an unbound variable");

subst_spec_list_term1 = subst_spec_list_term;
subst_str_list_term1 = subst_str_list_term;

if (is_nil(subst_spec_list_term1)) {
    strncpy(output_buffer, input_string, sizeof(output_buffer));
    goto EXIT;
}
if (is_nil(subst_str_list_term1))
    xsb_abort("RE_SUBSTITUTE: Arg 3 must not be an empty list");

/* initialize output buf */
output_ptr = output_buffer;

do {
    subst_reg_term = p2p_car(subst_spec_list_term1);
    subst_spec_list_term1 = p2p_cdr(subst_spec_list_term1);
if (!is_nil(subst_str_list_term1)) {
    subst_str_term = p2p_car(subst_str_list_term1);
    subst_str_list_term1 = p2p_cdr(subst_str_list_term1);

    if (is_string(subst_str_term)) {
        subst_string = string_val(subst_str_term);
    } else if (is_list(subst_str_term)) {
        subst_string =
        p_charlist_to_c_string(subst_str_term, subst_buf, sizeof(subst_buf),
                               "RE_SUBSTITUTE", "substitution string");
    } else
        xsb_abort("RE_SUBSTITUTE: Arg 3 must be a list of strings");
}

beg_term = p2p_arg(subst_reg_term, 1);
end_term = p2p_arg(subst_reg_term, 2);

if (!is_int(beg_term) || !is_int(end_term))
    xsb_abort("RE_SUBSTITUTE: Non-integer in Arg 2");
else{
    beg_offset = int_val(beg_term);
    end_offset = int_val(end_term);
}

/* -1 means end of string */
if (end_offset < 0)
    end_offset = input_len;
if ((end_offset < beg_offset) || (beg_offset < last_pos))
    xsb_abort("RE_SUBSTITUTE: Substitution regions in Arg 2 not sorted");

/* do the actual replacement */
strncpy(output_ptr, input_string + last_pos, beg_offset - last_pos);
output_ptr = output_ptr + beg_offset - last_pos;
if (sizeof(output_buffer) > (output_ptr - output_buffer + strlen(subst_string)))
    strcpy(output_ptr, subst_string);
else
    xsb_abort("RE_SUBSTITUTE: Substitution result size %d > maximum %d",
              beg_offset + strlen(subst_string),
              sizeof(output_buffer));

last_pos = end_offset;
output_ptr = output_ptr + strlen(subst_string);

} while (!is_nil(subst_spec_list_term1));

if (sizeof(output_buffer) > (output_ptr-output_buffer+input_len-end_offset))
    strcat(output_ptr, input_string+end_offset);

EXIT:
    /* get result out */
    if (conversion_required)
        c_string_to_p_charlist(output_buffer,output_term,"RE_SUBSTITUTE","Arg 4");
    else
        /* DO NOT intern. When atom table garbage collection is in place, then
           replace the instruction with this:
           c2p_string(CTXTC output_buffer, output_term);
           The reason for not interning is that in Web page
           manipulation it is often necessary to process the same string many
           times. This can cause atom table overflow. Not interning allows us to
           circumvent the problem. */
        extern_ctop_string(CTXTC 4, output_buffer);

    return(TRUE);
}

/* XSB regular expression matcher entry point
 In:
    Arg1: regexp
    Arg2: string
    Arg3: offset
    Arg4: ignorecase
 Out:
    Arg5: list of the form [match(bo0,e00), match(bo1,e01),...]
          where bo*,e0* specify the beginning and ending offsets of the
          matched substrings.
          All matched substrings are returned. Parenthesized expressions are
          ignored.
 */
int do_bulkmatch__(CTXTdecl)
{
    prolog_term listHead, listTail;
    /* Prolog args are first assigned to these, so we could examine the types
of these objects to determine if we got strings or atoms. */
prolog_term regexp_term, input_term, offset_term;
prolog_term output_term = p2p_new();
char *regexp_ptr=NULL; /* regular expression ptr */
char *input_string=NULL; /* string where matches are to be found */
int ignorecase=FALSE;
int return_code, paren_number, offset;
regmatch_t *match_array;
int last_pos=0, input_len;
char regexp_buffer[MAXBUFSIZE];

if (first_call)
    initialize_regexp_tbl();

regexp_term = reg_term(CTXTc 1); /* Arg1: regexp */
if (is_string(regexp_term)) /* check it */
    regexp_ptr = string_val(regexp_term);
else if (is_list(regexp_term))
    regexp_ptr =
            p_charlist_to_c_string(regexp_term, regexp_buffer, sizeof(regexp_buffer),
                                "RE_MATCH", "regular expression");
else
    xsb_abort("RE_MATCH: Arg 1 (the regular expression) must be an atom or a character list");

input_term = reg_term(CTXTc 2); /* Arg2: string to find matches in */
if (is_string(input_term)) /* check it */
    input_string = string_val(input_term);
else if (is_list(input_term)) {
    input_string =
            p_charlist_to_c_string(input_term, input_buffer, sizeof(input_buffer),
                                "RE_MATCH", "input string");
} else
    xsb_abort("RE_MATCH: Arg 2 (the input string) must be an atom or a character list");

input_len = strlen(input_string);

offset_term = reg_term(CTXTc 3); /* Arg3: offset within the string */
if (! is_int(offset_term))
    xsb_abort("RE_MATCH: Arg 3 (the offset) must be an integer");
offset = int_val(offset_term);
if (offset < 0 || offset > input_len)
    xsb_abort("RE_MATCH: Arg 3 (=%d) must be between 0 and %d", input_len);
/* If arg 4 is bound to anything, then consider this as ignore case flag */
if (! is_var(reg_term(CTXTC 4)))
  ignorecase = TRUE;

last_pos = offset;
/* returned result */
listTail = output_term;
while (last_pos < input_len) {
  c2p_list(CTXTC listTail); /* make it into a list */
  listHead = p2p_car(listTail); /* get head of the list */

  return_code = xsb_re_match(regexp_ptr, input_string+last_pos, ignorecase,
                           &match_array, &paren_number);

  /* exit on no match */
  if (! return_code) break;

  /* bind i-th match to listHead as match(beg,end) */
  c2p_functor(CTXTC "match", 2, listHead);
  c2p_int(CTXTC match_array[0].rm_so+last_pos, p2p_arg(listHead,1));
  c2p_int(CTXTC match_array[0].rm_eo+last_pos, p2p_arg(listHead,2));

  listTail = p2p_cdr(listTail);
  last_pos = match_array[0].rm_eo+last_pos;
}
c2p_nil(CTXTC listTail); /* bind tail to nil */
return p2p_unify(CTXTC output_term, reg_term(CTXTC 5));

13.3 Foreign Modules That Call XSB Predicates

A C function that has been called from XSB through the lower-level foreign language interface may want to call back into XSB to have XSB evaluate a predicate. This can be done by using the interface described in Chapter 3 (Volume 2) on calling XSB from another language. The interface described there allows a caller to initialize XSB and pass queries to it. However, since XSB has already called a foreign module, XSB does not need to be initialized. However it does need to manage the registers that are in use to support interaction with the foreign module currently executing. So there are some minor differences with the interface described in Chapter 3.

First, XSB should not be initialized. I.e., a foreign module should not call
The foreign module must protect the XSB registers it is currently using when it calls XSB. To do this, after it has retrieved its arguments into local variables and before it calls any XSB predicate, it must call \texttt{xsb\_query\_save(NumRegs)} which saves the current XSB registers and initializes them to be able to accept a new query. \texttt{NumRegs} is the number of registers used to interact with the currently executing foreign routine (i.e., the arity of the predicate that called this foreign code.) When the foreign routine has completed its work, it will set the appropriate registers with the appropriate return values and return to the caller. Before it does this, it must call \texttt{xsb\_query\_restore()} to restore the saved registers and prepare XSB for the return. Note that it must be called before any of the output registers are accessed to set return values. (It must also be called even if no values are returned.)

In summary the extra functions needed to call XSB from a foreign module are:

\begin{verbatim}
int xsb_query_save(CTXTc (byte) NumRegs)
    This function is used in a foreign routine that is called from XSB. It is used to save the current contents of the XSB registers and to initialize them to be prepared to accept a query. It must be called after a foreign routine collects its input arguments from the XSB registers and before it invokes any XSB predicate.

int xsb_query_restore(CTXT)
    This function is used in a foreign routine that is called from XSB and in turn calls an XSB predicate. It is used to restore the previously saved contents of the XSB registers. It must be called after all XSB predicates have been called and returned, and before the current foreign routine sets its output parameters and returns to XSB.
\end{verbatim}

An example where a foreign module and XSB call each other recursively can be found in the directory \texttt{$XSB\_DIR/example/XSB\_calling\_c} and files \texttt{fibr.[ch]} and \texttt{fibp.P}.

\section*{13.4 Foreign Modules That Link Dynamically with Other Libraries}

Sometimes a foreign module might have to link dynamically with other (non-XSB) libraries. Typically, this happens when the foreign module implements an interface
to a large external library of utilities. One example of this is the package \texttt{libwww}
in the XSB distribution, which provides a high-level interface to the W3C’s Libwww
library for accessing the Web. The library is compiled into a set of shared objects
and the \texttt{libwww} module has to link with them as well as with XSB.

The problem here is that the loader must know at run time where to look for the
shared objects to link with. On Unix systems, this is specified using the environment
variable \texttt{LD_LIBRARY_PATH}; on Windows, the variable name is \texttt{LIBPATH}. For instance,
under Bourne shell or its derivatives, the following will do:

\begin{verbatim}
LD_LIBRARY_PATH=dir1:dir2:dir3
export LD_LIBRARY_PATH
\end{verbatim}

One problem with this approach is that this variable must be set before starting XSB.
The other problem is that such a global setting might interact with other foreign
modules.

To alleviate the problem, XSB dynamically sets \texttt{LD_LIBRARY_PATH} (\texttt{LIBPATH} on
Windows) before loading foreign modules by adding the directories specified in the \texttt{-L}
option in \texttt{ldoption}. Unfortunately, this works on some systems (Linux), but not on
others (Solaris). One route around this difficulty is to build a runtime library search
path directly into the object code of the foreign module. This can be specified using
a loader flag in \texttt{ldoption}. The problem here is that different systems use a different
flag! To circumvent this, XSB provides a predicate that tries to guess the right flag
for your system:

\begin{verbatim}
runtime_loader_flag(+Hint,-Flag)
\end{verbatim}

Currently it knows about a handful of the most popular systems, but this will be
expanded. The argument \texttt{Hint} is not currently used. It might be used in the future
to provide \texttt{runtime_loader_flag} with additional information that can improve the
accuracy of finding the right runtime flags for various systems.

The above predicate can be used as follows:

\begin{verbatim}
..., 
runtime_loader_flag(_,Flag),
fmt_write_string(LDoptions, '%sdir1:dir2:dir2 %s', args(Flag,OldLDoption)),
fmt_write(File, ':- ldoption(%s).', LDoptions),
file_nl(File).
\end{verbatim}
13.5 Higher-Level Foreign Language Interface

The high-level foreign predicate interface was designed to release the programmer from the burden of having to write low-level code to transfer data from XSB to C and vice-versa. Instead, all the user needs to do is to describe each C function and its corresponding Prolog predicates in the .H files. The interface then automatically generates wrappers that translate Prolog terms and structures to proper C types, and vice-versa. These wrappers also check for type-correctness of arguments to the C function; in addition, in Unix-derived systems the wrappers are automatically compiled and loaded along with the foreign predicates in the .c file.

As with the lower-level foreign interfaces, when predicates are defined in a foreign module myfile.[cH], the predicates must be explicitly imported from the module to be used.

For an example of using the higher level interface, see $XSBDIR/examples/XSB_calling_c/second_foreign.[cH].

13.5.1 Declaration of high level foreign predicates

The basic formats of a foreign predicate declaration are:

\[\begin{align*}
\text{:- foreign_pred} & \text{ predname}([+-]parg1, [+-]parg2,...) \\
& \text{ from funcname}(carg1: \text{type1}, carg2: \text{type2}, \\
& \quad \ldots):\text{functype}. \\
\text{and} \\
\text{:- private_foreign_pred} & \text{ predname}([+-]parg1, [+-]parg2,...) \\
& \text{ from funcname}(carg1: \text{type1}, carg2: \text{type2}, \\
& \quad \ldots):\text{functype}. \\
\end{align*}\]

where:

- foreign_pred, private_foreign_pred declares a new foreign predicate. For most cases, the declaration foreign_pred can be used in both the multi-threaded and the sequential engine. The declaration private_foreign_pred needs to be used only in the multi-threaded engine when the external foreign function, funcname contains a context parameter as

---

6 For Windows, please see special instructions in Section 13.6.
7 In Version 3.6, a foreign module that uses the higher-level C interface must be explicitly consulted before it can be used.
its first argument because \texttt{funcname} needs to access thread-private data or other information from the context of the XSB thread (see Section 13.2.1). This case is uncommon, and mostly occurs for users who are creating XSB packages (e.g. the XASP interface to Smodels).

\texttt{predname}

is the name of the foreign Prolog predicate.

\texttt{parg1, parg2, ...}

are the predicate arguments. Each argument is preceded by either '+' or '-', indicating its mode as input or output respectively. The names of the arguments must be the same as those used in the declaration of the corresponding C function. If a C argument is used both for input and output, then the corresponding Prolog argument can appear twice: once with "+" and once with "-". In addition, a special argument \texttt{retval} is used to denote the argument that corresponds to the return value of the C function; it must always have the mode '-'.

\texttt{funcname}

is the name of the function in the \texttt{.c} file. At compile-time a C function with name \texttt{predname} will be generated which will translate arguments from Prolog to C, call \texttt{funcname}, and then translate arguments back from C to Prolog.

\texttt{carg1, carg2, ...}

is the list of arguments of the C function. The names used for the arguments must match the names used in the Prolog declaration.

\texttt{type1, type2, ...}

are the types associated to the arguments of the C function. This is not the set of C types, but rather a set of descriptive types, as defined in Table 13.5.1.

\texttt{functype}

is the return type of the C function.

Using the higher-level interface, the same C code can be used for both the sequential and the multi-threaded engines, and no context parameters are required in a user’s C code unless thread context information is explicitly needed. However, a foreign module compiled for the single-threaded engine will need to be recompiled for the multi-threaded engine and vice-versa.

Table 13.5.1 provides the correspondence between the types allowed on the C side of a foreign module declaration and the types allowed on the Prolog side of the declaration.
<table>
<thead>
<tr>
<th>Descriptive Type</th>
<th>Mode Usage</th>
<th>Associated C Type</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>int</td>
<td>+</td>
<td>int</td>
<td>integer numbers</td>
</tr>
<tr>
<td>float</td>
<td>+</td>
<td>double</td>
<td>floating point numbers</td>
</tr>
<tr>
<td>atom</td>
<td>+</td>
<td>unsigned long</td>
<td>atom represented as an unsigned long</td>
</tr>
<tr>
<td>chars</td>
<td>+</td>
<td>char *</td>
<td>the textual representation of an atom is passed to C as a string</td>
</tr>
<tr>
<td>chars(size)</td>
<td>+</td>
<td>char *</td>
<td>the textual representation of an atom is passed to C as a string in a buffer of size size</td>
</tr>
<tr>
<td>string</td>
<td>+</td>
<td>char *</td>
<td>a prolog list of characters is passed to C as a string</td>
</tr>
<tr>
<td>string(size)</td>
<td>+</td>
<td>char *</td>
<td>a prolog list of characters is passed to C as a string</td>
</tr>
<tr>
<td>term</td>
<td>+</td>
<td>prolog_term</td>
<td>the unique representation of a term</td>
</tr>
<tr>
<td>intptr</td>
<td>+</td>
<td>int *</td>
<td>the location of a given integer</td>
</tr>
<tr>
<td>floatptr</td>
<td>+</td>
<td>double *</td>
<td>the location of a given floating point number</td>
</tr>
<tr>
<td>atomptr</td>
<td>+</td>
<td>unsigned long *</td>
<td>the location of the unique representation of a given atom</td>
</tr>
<tr>
<td>charsptr</td>
<td>+</td>
<td>char **</td>
<td>the location of the textual representation of an atom</td>
</tr>
<tr>
<td>stringptr</td>
<td>+</td>
<td>char **</td>
<td>the location of the textual representation of a list of characters</td>
</tr>
<tr>
<td>termptr</td>
<td>+</td>
<td>prolog_term *</td>
<td>the location of the unique representation of a term</td>
</tr>
<tr>
<td>intptr</td>
<td>-</td>
<td>int *</td>
<td>the integer value returned is passed to Prolog</td>
</tr>
<tr>
<td>floatptr</td>
<td>-</td>
<td>double *</td>
<td>the floating point number is passed back to Prolog</td>
</tr>
<tr>
<td>charsptr</td>
<td>-</td>
<td>char **</td>
<td>the string returned is passed to Prolog as an atom</td>
</tr>
<tr>
<td>stringptr</td>
<td>-</td>
<td>char **</td>
<td>the string returned is passed back as a list of characters</td>
</tr>
<tr>
<td>atomptr</td>
<td>-</td>
<td>unsigned long *</td>
<td>the number returned is passed back to Prolog as the unique representation of an atom</td>
</tr>
<tr>
<td>termptr</td>
<td>-</td>
<td>prolog_term *</td>
<td>the number returned is passed to Prolog as the unique representation of a term</td>
</tr>
<tr>
<td>chars(size)</td>
<td>+-</td>
<td>char *</td>
<td>the atom is copied from Prolog to a buffer, passed to C and converted back to Prolog afterwards</td>
</tr>
<tr>
<td>string(size)</td>
<td>+-</td>
<td>char *</td>
<td>the list of characters is copied from Prolog to a buffer, passed to C and back to Prolog afterwards</td>
</tr>
<tr>
<td>intptr</td>
<td>+-</td>
<td>int *</td>
<td>an integer is passed from Prolog to C and from C back to Prolog</td>
</tr>
<tr>
<td>floatptr</td>
<td>+-</td>
<td>double *</td>
<td>a float number is passed from Prolog to C, and back to Prolog</td>
</tr>
<tr>
<td>atomptr</td>
<td>+-</td>
<td>unsigned long *</td>
<td>the unique representation of an atom is passed to C, and back to Prolog as a list of characters</td>
</tr>
<tr>
<td>charsptr</td>
<td>+-</td>
<td>char **</td>
<td>the atom is passed to C as a string, and a string is passed to Prolog as an atom</td>
</tr>
<tr>
<td>stringptr</td>
<td>+-</td>
<td>char **</td>
<td>the list of characters is passed to C, and a string passed to Prolog as a list of characters</td>
</tr>
<tr>
<td>termptr</td>
<td>+-</td>
<td>prolog_term *</td>
<td>the unique representation of a term is passed to C, and back to Prolog</td>
</tr>
</tbody>
</table>

Table 13.1: Allowed combinations of types and modes, and their meanings
In all modes and types, checks are performed to ensure the types of the arguments. Also, all arguments of type `-' are checked to be free variables at call time.

13.6 Compiling Foreign Modules on Windows and under Cygwin

Due to the complexity of creating makefiles for the different compilers under Windows, XSB doesn’t attempt to compile and build DLL’s for the Windows foreign modules automatically. However, for almost all typical cases the user should be able to easily adapt the sample makefile for Microsoft VC++:

```
XSB/examples/XSB_calling_c/MakefileForCreatingDLLs
```

It is important that the C program will have the following lines near the top of the file:

```c
#include "xsb_config.h"
#ifdef WIN_NT
#define XSB_DLL
#endif
#include "cinterf.h"
```

Note that these same DLLs will work under Cygwin — XSB’s C interface under Cygwin is like that under Windows rather than Unix.

If the above makefile cannot be adapted, then the user has to create the DLL herself. The process is, roughly, as follows: first, compile the module from within XSB. This will create the XSB-specific object file, and (if using the higher-level C interface) the wrappers. The wrappers are created in a file named `xsb_wrap_modulename.c`.

Then, create a project, using the compiler of choice, for a dynamically-linked library that exports symbols. In this project, the user must include the source code of the module along with the wrapper created by XSB. This DLL should be linked against the library

```
XSB\config\x86-pc-windows\bin\xsb.lib
```

which is distributed with XSB. In VC++, this library should be added as part of the linkage specification. In addition, the following directories for included header files must be specified as part of the preprocessor setup:
In VC++, make sure you check off the “No precompiled headers” box as part of the “Precompiled headers” specification. All these options are available through the Project»Settings menu item.

### 13.7 Functions for Use in Foreign Code

In addition to functions for passing data between Prolog and C, XSB contains other functions that may be useful in Foreign C code. We mention a few here that pertain to throwing exceptions from C code (cf. Volume 1 Chapter 8: Exception Handling). These functions can be used by code that uses either the lower- or higher-level interface.

**void xsb_domain_error(CTXTdeclc char *valid_domain, Cell culprit, char *pred, int arity, int arg)**

Used to throw an ISO-style domain error from foreign code, indicating that culprit is not in domain valid_domain in argument arg of pred/arity.

**Example:** The code fragment

```c
Cell num;
:xsb_domain_error(CTXTc "not_less_than_zero",num,"atom_length",2,2);
```

in atom_length/2 gives rise to the behavior

| ?- atom_length(abcde,-1). | ++Error[XSB/Runtime/P]: [Domain (-1 not in domain not_less_than_zero)] in arg 2 of predicate atom_length/2 |

**void xsb_existence_error(CTXTdeclc char *objType, Cell culprit, char *pred, int arity, int arg)**

Used to throw an ISO-style existence error from foreign code, indicating that an object culprit of type objType does not exist, in argument arg of pred/arity.

**Example:** The code fragment
Cell tid;
:
  xsb_existence_error(CTXTc "thread",reg[2],"xsb_thread_join",1,1);

in thread_join/1 gives rise to a the behavior

| ?- thread_join(7).
++Error[XSB/Runtime/P]: [Existence (No thread 1 exists)]
in arg 1 of predicate thread_join/1)

if a thread with thread id 7 does not exist.

void xsb_instantiation_error(CTXTdeclc char *pred,int arity,int arg,char *state)

Used to throw an ISO-style instantiation error from foreign code. If state
is a NULL pointer, the message indicates that there is an instantiation error for
argument arg of of pred/arity. If state is non-NULL, the message additionally
indicates that argument arg must be state.

Example: The code fragment

    xsb_instantiation_error(CTXTc "atom_length",2,1,NULL);

in atom_length/2 gives rise to a the behavior

| ?- atom_length(X,Y).
++Error[XSB/Runtime/P]: [Instantiation] in arg 1 of predicate atom_length/2

void xsb_misc_error(CTXTdeclc char *message,char *pred,int arity)

Used to throw a non ISO-error from foreign code, printing message and indi-
cating that the error arose in pred/arity.

void xsb_permission_error(CTXTdeclc char *op,char *obj,Cell culprit,char *pred,int arity)

Used to throw an ISO-style permission error from foreign code, indicating that
an operation of type op on type obj is not permitted on culprit, in argument
arg of pred/arity.

Example: The code fragment

    xsb_permission_error(CTXTc "unlock mutex","mutex not held by thread",
xsb_thread_id,"mutex_unlock",2);
in `mutex_unlock/1` gives rise to a behavior

| ?- mutex_unlock(mymut).
++Error[XSB/Runtime/P]: [Permission (Operation) unlock mutex on mutex not held by thread: 0] in predicate mutex_unlock/1)

if thread 0 does not own mutex `mymut`.

```c
void xsb_resource_error(CTXTdeclc char *resource,char *pred,int arity)
```

Used to indicate that there are not sufficient resources of type `resource` for `pred/arity` to succeed.

**Example**: The code fragment

```c
xsb_resource_error(th,"system threads","thread_create",2);
```

in `thread_create/1` gives rise to a behavior

| ?- thread_create(X).
++Error[XSB/Runtime/P]: [Resource (system threads)] in predicate thread_create/2)

If the number of system threads has been exceeded.

```c
void xsb_type_error(CTXTdeclc char *valid_type,Cell culprit,char *pred,int arity,int arg)
```

Used to throw an ISO-style type error from foreign code, indicating that `culprit` is not in ISO type `valid_type` in argument `arg` of `pred/arity`.

**Example**: The code fragment

```c
Cell num;

if (!isinteger(num)) xsb_type_error(CTXTc "integer",num,"atom_length",2,2);
```

in `atom_length/2` gives rise to the behavior

| ?- atom_length(foo,a).
++Error[XSB/Runtime/P]: [Type (a in place of integer)] in arg 2 of predicate atom_length/2)

```c
void xsb_throw(CTXTdeclc prolog_term Ball)
```

Used to throw a Prolog term from C code, when an ISO-style error is not required. The term can be caught and handled by the Prolog predicate `catch/3` just as any other thrown term; however if it is not caught, XSB’s default error handler will treat it as an unhandled exception.
Chapter 14

Embedding XSB in a Process

There are many situations in which it is desirable to use XSB as a rule- or constraint-processing subcomponent of a larger system that is written in another language. Depending on the intended architecture, it may be appropriate for XSB to reside in its own process, separate from other components of an application, and communicating through sockets, a database, or some other mechanism. However it is often useful for XSB to reside in the same process as other components. To do this, one wants to be able to call XSB from the host language, providing queries for XSB to evaluate, and retrieving back the answers. An interface for calling XSB from C is provided for this purpose and is described in this chapter. Based on this C interface, XSB can also be called from Java either through a JNI or a socket-based interface, as described in the documentation for Interprolog, available through xsb.sourceforge.net. To call XSB from Visual Basic, a DLL is created as described in this chapter, and additional declarations must be made in visual basic as described in the web page “How to use XSB DLL from Visual Basic” http://xsb.sourceforge.net/vbdll.html. In addition, the interface described in this chapter has also been extended to allow XSB to be called from Delphi and Ruby. However, since all of these interfaces – Java, Ruby, Delphi and Visual Basic – depend on XSB’s C API, we refer in this chapter to C programs or threads calling XSB, although each of the examples suitably modified can be extended to other calling languages.

New to Version 3.1 are extensions to the C API to allow multiple XSB threads to be called from multiple C threads 1. In this Chapter, we provide an overview of XSB’s C API, and then elaborate its use through a series of examples, beginning with a single XSB thread called by a single C thread, then showing how a C thread can

---

1XSB’s threading model is based on POSIX threads, which can be called in Windows through a variety of POSIX APIs – see Volume 1 chapter 8 Multi-threaded Programming in XSB.
interact with multiple XSB threads, and finally discuss how multiple XSB threads can interact with multiple POSIX threads. Finally, Section 14.3 describes each C function in the API.

14.1 Calling XSB from C

XSB provides several C functions (declared in $XSBDIR/emu/cinterf.h and defined in $XSBDIR/emu/cinterf.c), which can be called from C to interact with XSB as a subroutine. These functions allow a C program to interact with XSB in a number of ways.

- XSB may be initialized, using most of the parameters available from the command-line.

- XSB may then execute a series of commands or queries. A command is a deterministic query which simply succeeds or fails without performing any unification on the query term. On the other hand, a non-deterministic query can be evaluated so that its answer substitutions are retrieved one at a time, as they are produced, just as if XSB were called on a command line. Alternately a non-deterministic query can be closed in the case where not every answer to the query is needed. Only one query per thread can be active at a time. I.e., an application must completely finish processing one query to a given thread $T$ (either by retrieving all the answers for it, or by issuing a call to `xsb_close_query()` before trying to evaluate another using $T$).

- Finally, XSB can be closed, so that no more queries can be made to any XSB threads.

In general, while any functions in the C API to XSB can be intermixed, the functions can be classified as belonging to three different levels.

- A VarString level which uses an XSB-specific C-type definition for variable-length strings (Section 14.4), to return answers.

- A fixed-string level provides routines that return answers in fixed-length strings.

- A register-oriented level that requires users to set up queries by setting registers for XSB which are made globally available to calling functions. The mechanisms for this resemble the lower-level C interface discussed in Chapter 13. This level
of interface should only be used for the single-threaded applications, as it is
difficult to prevent race-conditions at this level of interface when multiple C
threads are used to call XSB.

The appropriate level to use depends on the nature of the calling program, the
speed desired, and the expertise of the programmer. By and large, functions in
the \texttt{VarString} level are the the easiest and safest to use, but they depend on a C
type definition that may not be available to all calling programs (e.g. it may be
difficult to use if the calling program is not directly based on C, such as Visual Basic
or Delphi). For such applications functions from the fixed-string level would need to
be used instead. In general, most applications should use either functions from the
\texttt{VarString} or the fixed-string level, rather than the register-oriented level. This latter
level should only be used by programmers who are willing to work at a low interface
level, when the utmost speed is needed by an application, and when multiple threads
do not need to interact with XSB.

\section{Examples of Calling XSB}

We introduce a series of examples of how XSB would be called using the string-
level interfaces. Simple examples of the register-level interface are given in the
XSB/examples/c_calling_XSB subdirectory, in files \texttt{cmain.c}, \texttt{cmain2.c}, \texttt{ctest.P},
and \texttt{Makefile}, but are not discussed in this section.

We structure our discussion by first showing how to construct a C program to call
the single-threaded engine alone in Section \ref{subsec:sequential-only}. This example is mostly pedagogic:
with a small amount of extra coding a C program can be constructed to call both
the single- and the multi-threaded engine, and these extensions are discussed in Sec-
tion \ref{subsec:multi-threaded}. Next, we show how to a C program can call and manage multiple XSB
threads in Section \ref{subsec:multi-c_threads}. Finally, we show how multiple XSB threads can interact
with multiple C threads in Section \ref{subsec:multi-threaded}.

\subsection{The XSB API for the Sequential Engine Only}

We start with a simple program shown, in Figure \ref{fig:sequential-only}, that will call the following
XSB predicate

\begin{verbatim}
p(a,b,c).
p(1,2,3).
p([1,2],[3,4],[5,6]).
\end{verbatim}
CHAPTER 14. EMBEDDING XSB IN A PROCESS

438

p(A,B,A).
r(c,b,a).
r(3,2,1).
r([5,6],[3,4],[1,2]).
r(_A,B,B).

and backtrack through unifying answers (cf. $XSBDIR/examples/c_calling_xsb/edb.P$).

This example will only compile properly if the sequential engine is used, and its
style is not recommended: it will be shown in Section 14.2.2 how to extend the style.

We discuss the program in Figure 14.1 in detail. This program, slightly modified so
that it compiles with the multi-threaded engine is in $XSBDIR/examples/c_calling_xsb/cvartest.c.$
An executable for this program can be made most easily by calling $XSBDIR/examples/c_calling_xsb/make$, which makes the executable cvtest.

The program begins by including some standard C headers: note that string.h
is needed for string manipulation routines such as strcpy. In addition, the XSB
library header cinterf.h is necessary for the XSB C API. Since the program in
Figure 14.1 uses functions in the VarString interface, within main() the routine
XSB_StrDefine(return_string) declares and initializes a structure of type VarString,
named return_string.

The next order of business is to initialize XSB. In order to do this, xsb_init_string()
needs to know the installation directory for XSB, which must be passed as part of the
initialization string. In Figure 14.1 this is done by manipulating the path of the exec-
tutable (cvstest) that calls XSB. In fact any other approach would also work as long
as the XSB installation directory were passed. Within the initialization string, other
command line arguments can be passed to XSB if desired with the following exceptions:
the arguments -B (boot module), -D (command loop driver), -i (interpreter)
and -d (disassembler) cannot be used when calling XSB from a foreign language.

As a final point on initialization, note that the function xsb_init() can also be used
to initialize XSB based on an argument vector and count (see Section 14.3).

Note that the calling program checks for any errors returned by xsb_init_string() and
other API commands. In general, xsb_init_string() may throw an error if
the XSB’s installation directory has become corrupted, or for similar reasons. This
mechanism for error handling is different than that used if XSB is called in its usual

2In previous versions of XSB, initialization from the C level required a -n option to be passed. This is no longer required.
```c
#include <stdio.h>
#include <string.h>
/* cinterf.h is necessary for the XSB API, as well as the path manipulation routines*/
#include "cinterf.h"

extern char *xsb_executable_full_path(char *);
extern char *strip_names_from_path(char*, int);

int main(int argc, char *argv[]) {
    char init_string[1024];
    int rc;
    XSB_StrDefine(return_string);

    /* xsb_init_string() relies on the calling program to pass the absolute or relative
     path name of the XSB installation directory. We assume that the current
     program is sitting in the directory ../examples/c_calling_xsb/
     To get the installation directory, we strip 3 file names from the path. */
    strcpy(init_string,strip_names_from_path(xsb_executable_full_path(argv[0]),3));

    if (xsb_init_string(init_string) == XSB_ERROR) {
        fprintf(stderr,"++initializing XSB: %s/%s\n",xsb_get_init_error_type(),
                xsb_get_init_error_message());
        exit(XSB_ERROR);
    }

    /* Create command to consult a file: edb.P, and send it. */
    if (xsb_command_string("consult('edb.P').") == XSB_ERROR)
        fprintf(stderr,"++Error consulting edb.P: %s/%s\n",xsb_get_error_type(),xsb_get_error_message());

    rc = xsb_query_string_string("p(X,Y,Z).",return_string,"|”);
    while (rc == XSB_SUCCESS) {
        printf("Return %s\n",(return_string.string));
        rc = xsb_next_string(&return_string,"|”);
    }

    if (rc == XSB_ERROR)
        fprintf(stderr,"++Query Error: %s/%s\n"xsb_get_error_type(),xsb_get_error_message());

    xsb_close();
}
```

Figure 14.1: Calling the Sequential Engine Using the **VarString** Interface
stand-alone mode, in which case such an error would cause XSB to exit). An error returned by XSB’s API are similar to an error ball described in Volume 1 Exception Handling in that it has both a type and a message. For normal Prolog exceptions, XSB’s API will throw the same kinds of errors as XSB called in a stand-alone (or server) mode, i.e. instantiation errors, type errors, etc. However XSB’s API adds two new error types:

- **init_error** is used as the type of an error discovered upon initialization of XSB, before query and command processing has begun. If an init_error is raised, XSB has not been properly initialized and will not run.

- **unrecoverable_error** is used to indicate that XSB has encountered an error, (such as a memory allocation error), during command or query processing from which it cannot recover. Such an error would cause XSB to immediately exit if it were called in a stand-alone mode. In general the calling program should handle unrecoverable errors as fatal since there is a good chance that the error conditions will affect the calling program as well as XSB.

Errors raised by `xsb_init_string()` usually have type `init_type`.

and a string pointer to the associated message can be found by the function `xsb_get_init_error_message()`.

As can be seen from the example, handling errors from commands is done in manner similar to that of initialization. For non-initialization errors, a string pointer to the type can be obtained by `xsb_get_error_type()`, while a string pointer to the message can be obtained by `xsb_get_error_message()`.

Next in Figure 14.1 the file `edb.P` is consulted (containing the p/3 and r/3 predicates shown above). Note, that the argument to `xsb_command_string` must be a syntactically valid Prolog term ending with a period, otherwise a syntax error will be thrown, which may be displayed through `xsb_get_error_type()` and `xsb_get_error_message()` 3.

Queries to XSB are a little more complicated than commands. Since a query may return multiple solutions, a query should usually be called from inside a loop. In Figure 14.1, the query is opened with `xsb_query_string()`. If the query has at least one answer, `xsb_query_string()` will return XSB_SUCCESS; if the query fails, it will return XSB_FAILURE, and if there is an exception it will return XSB_ERROR as usual. Any answer will be returned as a string in the VarString return_string, and each

---

3Most XSB errors are handled in this manner when XSB is called through its API. A few errors will print directly to stderr and some XSB warnings will print to stdwarn which upon startup is dup-ed to stderr.
argument of the query will be separated by the character \( | \). Thus, in our example, the first answer will write the string

\[ a \mid b \mid c \]

Once a query has been opened, subsequent answers can be obtained via \texttt{xsb\_next\_string()}. These answers are written to \texttt{return\_string} in the same manner as \texttt{xsb\_query\_string\_string()}. 

\[ 1 \mid 2 \mid 3 \]
\[ [1,2] \mid [3,4] \mid [5,6] \]
\[ \_h102 \mid \_h116 \mid \_h102 \]

A query is automatically closed when no more answers can be derived from it. Alternatively, a query that may have answers remaining can be closed using the command \texttt{xsb\_close\_query()}. If the calling application will need to pass more queries or commands to XSB nothing need be done at this point: a new queries or commands can be invoked using one of the functions just discussed. However if the calling process is finished with XSB and will never need it again during the life of the process, it can call \texttt{xsb\_close()}. 

### An Example using Fixed Strings

Figure 14.2 shows a fragment of code indicating how the previous example would be modified if the fixed-string interface were used. Note that \texttt{return\_string} now becomes a pointer to explicitly malloc-ed memory. To open the query \( p(X,Y,Z) \) the function \texttt{xsb\_query\_string\_string\_b()} is called, with the \_b indicating that a fixed buffer is being used rather than a \texttt{VarString}. The call is similar to \texttt{xsb\_query\_string\_string()}, except that the length \texttt{anslen} of the buffer pointed to by \texttt{return\_string} is now also required. If the answer to be returned (including separators) is longer than \texttt{anslen}, \texttt{xsb\_query\_string\_string\_b()} will return \texttt{XSB\_OVERFLOW}. If this happens, a new answer buffer can be used (here the old one is realloc-ed) and the answer retrieved via \texttt{xsb\_get\_last\_answer\_string}. Similarly, further answers are obtained via \texttt{xsb\_next\_string\_b()} whose length must be checked. Thus the only difference between the fixed-string level and the \texttt{VarString} level is that the length of each answer should be checked and \texttt{xsb\_get\_last\_answer\_string()} called if necessary.

### 14.2.2 The General XSB API

The previous section showed how to use the XSB API with both the \texttt{VarString} type and without, but did not consider the multi-threaded engine. In fact, there
int retsize = 15;
char *return_string;
int anslen;

return_string = malloc(retsize);

rc = xsb_query_string_string_b(CTXTc "p(X,Y,Z).", return_string, retsize, &anslen, "|");

while (rc == XSB_SUCCESS || rc == XSB_OVERFLOW) {
    if (rc == XSB_OVERFLOW) {
        return_string = (char *) realloc(return_string, anslen);
        return_size = anslen;
        rc = xsb_get_last_answer_string(CTXTc return_string, retsize, &anslen);
    }

    printf("Return %s %d\n", return_string, anslen);
    rc = xsb_next_string_b(CTXTc return_string, 15, &anslen, "|");
}

Figure 14.2: Calling XSB using the Fixed String Interface
are different ways to use XSB’s multi-threading that can have advantages for various situations. In the first mode, threads are managed from Prolog, with a single XSB thread called from the API; that XSB thread can then create another XSB thread that does work, and the first thread can return almost immediately to handle more requests from the API’s caller. A second model allows the caller to manipulate a pool of several XSB threads, so that different XSB threads may be called from different threads over the API. In this model each C, Java, Ruby, or other thread could a number of different Prolog threads. In this section we sketch how to use the API to illustrate the first model, and sketch the second model in the next section.

Figure 14.3 shows how relevant portions of the previous VarString example can be adapted to use the multi-threaded engine. The main change is that a new variable is introduced on the C side that points to the context of the main thread. As pointed out in Chapter 13, each thread in the multi-threaded engine has a context in which is kept much of its thread-specific data (excluding tables and dynamic code). Of the threads running in the multi-threaded engine the thread created upon the call to xsb_init() is designated as the main thread, and is closed only upon calling xsb_close().

Within the multi-threaded engine, a call to an API function such as xsb_query_string_string() is actually a call to a specific thread to do some work (using a thread context pointer). Accordingly, since any errors produced will be specific to a given thread, all calls to error reporting functions are also thread-specific. If no specific thread is needed, it may be best just to use the main thread, which is what is done in Figure 14.3. The thread context pointer th is initialized to the main thread using the API macro xsb_get_main_thread(). Afterwards, this pointer is passed into the various interface functions by making use of XSB macros defined in context.h In the multi-threaded engine, these macros are defined as

#define CTXT th
#define CTXTc th,

while in the single-threaded engine they are defined as empty strings, as is xsb_get_main_thread(). As a result the code in Figure 14.3 will compile and run properly both for the single-threaded and the multi-threaded engines.

At this stage, suppose one wanted a new thread to execute a specific command, say do_foo. In this case, a C call such as

xsb_query_string_string(CTXTc "thread_create(do_foo,Id).", &return_string,"|")

creates a thread to execute the command, and returns the thread id of the newly created thread in return_string. The behavior of this newly created thread is
/* context.h is necessary for the type of a thread context. */
#include "context.h"

int main(int argc, char *argv[])
{
    char init_string[MAXPATHLEN];
    int rc;
    XSB_StrDefine(return_string);
    strcpy(init_string,strip_names_from_path(xsb_executable_full_path(argv[0]),3));
    if (xsb_init_string(init_string) == XSB_ERROR) {
        fprintf(stderr,"++initializing XSB: %s/%s
",xsb_get_init_error_type(),
                xsb_get_init_error_message());
        exit(XSB_ERROR);
    }
    #ifdef MULTI_THREAD
    th_context *th = xsb_get_main_thread();
    #endif
    /* Create command to consult a file: edb.P, and send it. */
    if (xsb_command_string(CTXTc "consult('edb.P').") == XSB_ERROR)
        fprintf(stderr,"++Error consulting edb.P: %s/%s
",xsb_get_error_type(CTXT),
                xsb_get_error_message(CTXT));
    rc = xsb_query_string_string(CTXTc "p(X,Y,Z).",&return_string,"|" );
    while (rc == XSB_SUCCESS) {
        printf("Return %s\n",(return_string.string));
        rc = xsb_next_string(CTXTc &return_string,"|" );
    }
    if (rc == XSB_ERROR)
        fprintf(stderr,"++Query Error: %s/%s
",xsb_get_error_type(CTXT),xsb_get_error_message(CTXT));
    xsb_close();
}

Figure 14.3: Calling the Single- or Multi-Threaded Engine Using the VarString Interface
exactly the same as if it were created from the XSB command line: in particular the newly created thread will automatically exit upon completion of its command. As a somewhat technical point, there are two different ways of referring to XSB threads. The foreign language interfaces described in Chapter 13 and here use pointers to thread contexts so that the interfaces use much of the same code as the XSB engine. However Prolog refers to threads using thread identifiers. The two different forms can be converted into each other by the functions \texttt{xsb\_thread\_id\_to\_context()} and \texttt{xsb\_thread\_context\_to\_id()}.

### 14.2.3 Managing Multiple XSB Threads through the API

The ability to pass thread contexts into query and command functions allows a great deal of flexibility\(^4\). Once XSB is initialized, XSB threads can be created from C and can execute independently of each other, effectively giving the ability for different calling threads to query XSB in a mechanism reminiscent of database cursors.

Figure 14.4 illustrates a very simple example of this. XSB is initialized and the file \texttt{edb.P} consulted exactly as in Figure 14.4. However, the function \texttt{xsb\_ccall\_thread\_create()} causes the XSB thread \texttt{p\_th} to create a new thread, causes the new thread to call the same command loop as the main thread, and sets \texttt{r\_th} to point to the context of the new thread. The new thread \texttt{r\_th} can be used for commands or queries just as \texttt{p\_th}. Figure 14.4 shows that queries to the two threads can be interleaved, and errors for both threads can be checked and reported independently.

It is important to note that since each thread created by \texttt{xsb\_ccall\_thread\_create()} goes into a command-loop similar to the command loop, it will stay around until it is explicitly killed or until XSB is closed. The call

\[
xsb\_kill\_thread(r\_th);
\]

is needed to make \texttt{r\_th} to exit. Once a thread is exited, all of its data structures will be freed, including those that support \texttt{xsb\_get\_error\_type()} and \texttt{xsb\_get\_error\_message()}\(^5\).

---

\(^4\)For the sake of brevity, we sometimes abuse notation and do not always distinguish between thread-contexts and their pointers.

\(^5\)Note that causing XSB’s main thread to exit will cause the entire process to exit – not just XSB.
/* context.h is necessary for the type of a thread context. */
#include "context.h"

int main(int argc, char *argv[])
{
    static th_context *p_th, *r_th;
    char init_string[MAXPATHLEN];
    int rcp, rcr;
    XSB_StrDefine(p_return_string);
    XSB_StrDefine(r_return_string);

    strcpy(init_string,strip_names_from_path(xsb_executable_full_path(argv[0]),3));

    if (xsb_init_string(init_string)) {
        fprintf(stderr,"%s initializing XSB: %s/%s\n",xsb_get_init_error_type(),
            xsb_get_init_error_message());cin
        exit(XSB_ERROR);
    }

    p_th = xsb_get_main_thread();

    /* Create command to consult a file: edb.P, and send it. */
    if (xsb_command_string(p_th, "consult('edb.P').") == XSB_ERROR)
        fprintf(stderr,"++Error consulting edb.P: %s/%s\n",xsb_get_error_type(p_th),
            xsb_get_error_message(p_th));

    xsb_ccall_thread_create(p_th,&r_th);

    rcp = xsb_query_string_string(p_th,"p(X,Y,Z).",&p_return_string,"|");
    rcr = xsb_query_string_string(r_th,"r(X,Y,Z).",&r_return_string,"|");

    while (rcp == XSB_SUCCESS && rcr == XSB_SUCCESS) {
        printf("Return p %s\n",(p_return_string.string));
        rcp = xsb_next_string(p_th, &p_return_string,"|");

        printf("Return r %s\n",(r_return_string.string));
        rcr = xsb_next_string(r_th, &r_return_string,"|");
    }

    if (rcp == XSB_ERROR)
        fprintf(stderr,"++Query Error p: %s/%s\n",xsb_get_error_type(p_th),xsb_get_error_message(p_th));
    if (rcr == XSB_ERROR)
        fprintf(stderr,"++Query Error r: %s/%s\n",xsb_get_error_type(r_th),xsb_get_error_message(r_th));

    xsb_close();
}

Figure 14.4: Manipulating Multiple Threads Using the VarString Interface
14.2.4 Calling Multiple XSB Threads using Multiple C Threads

Figure 14.4 shows how two XSB threads can be created, can receive different queries and can interleave their backtracking and answer return. Although Figure 14.4 demonstrated only backtracking through simple predicates, the mechanism employed works for complicated examples using tabling, dynamic code, and other features. All this provides a sophisticated interface, but it is not “fully” multi-threaded in the following sense. When a C thread \( T \) causes XSB to execute a command or query the thread must wait until the calling function returns before proceeding. In certain applications it may be useful, for example, for \( T \) to create a C thread \( T_{\text{new}} \) which runs asynchronously from \( T \), executing the XSB command or query and then exiting. Alternately, an application may want to have a pool of C threads that can interact with a pool of XSB threads.

XSB’s C API has been designed to support these features. Figure 14.5 shows fragments of Figure 14.4 rewritten so that the routines to print out the answers to the queries \( p(X,Y,Z) \) and \( r(X,Y,Z) \) can be called from C threads specially designed for this purpose. More specifically, the routine \( \text{query_ps()} \) calls \( p_{\text{th}} \) to query \( p(X,Y,Z) \) and backtrack through its answers – its use of a single \( \text{void *} \) argument and a \( \text{void *} \) return reflect the requirements of functions that are to be called using \( 
\text{pthread_create()} \).

We note several points about this example. First the XSB API is a low-level API that can be used to build application specific interfaces, and some experience with pthread programming is useful if multiple XSB threads are called from multiple C threads. For instance, one issue is fairness. When called from the C API each XSB thread \( X_T \) makes use of mutexes to ensure that it answers only one query or command at a time. If multiple C threads are are waiting for \( X_T \) to respond to requests or queries, there is no guarantee that the requests will be processed in any sort of order, or even that a request will eventually be handled (In order to ensure this, the calling program would have to use a queue or some other scheduling mechanism to send requests to the XSB thread). In addition, it is important to note that, the main XSB thread should only be called from the C thread that initialized XSB.. This restriction is due to the current design of synchronizing an XSB thread with calling threads, and may be lifted in the future.

Protected and Non-Protected API Functions

Example 14.5 shows that, when the \text{Varstring} functions are used, if a single calling thread opens a query to an XSB thread \( X_T \), \( X_T \) will be protected from queries
void *query_ps(void * arg) {
    int rc;
    th_context *p_th;
    XSB_StrDefine(p_return_string);
    p_th = (th_context *)arg;

    rc = xsb_query_string_string(p_th,"p(X,Y,Z).",&p_return_string,"|")
        while (rc == XSB_SUCCESS) {
            printf("Return p %s\n",(p_return_string.string));
            rc = xsb_next_string(p_th, &p_return_string,"|")
        }
    
    if (rc == XSB_ERROR)
        fprintf(stderr,"++Query Error p: %s/%s\n",xsb_get_error_type(p_th),
                  xsb_get_error_message(p_th));
    return NULL;
}

int main(int argc, char *argv[]) {

    char init_string[MAXPATHLEN];
    static th_context *p_th, *r_th;
    int pstatus, rstatus;
    pthread_t pthread_id,rthread_id;
    XSB_StrDefine(p_return_string);
    XSB_StrDefine(r_return_string);

    main_th = xsb_get_main_thread();
    /* Create command to consult a file: edb.P, and send it. */
    if (xsb_command_string(xsb_get_main_thread(), "consult('edb.P').") == XSB_ERROR)
        fprintf(stderr,"++Error consulting edb.P: %s/%s\n",xsb_get_error_type(main_th),
                xsb_get_error_message(main_th));

    xsb_ccall_thread_create(main_th,&r_th);
    xsb_ccall_thread_create(main_th,&p_th);

    pthread_create(&rthread_id,NULL,command_rs,r_th);
    pthread_create(&pthread_id,NULL,command_ps,p_th);
    pthread_create(&rthread_id,NULL,command_rs,r_th);
    pthread_create(&pthread_id,NULL,command_ps,p_th);

    rstatus = pthread_join(rthread_id,&rreturn);
    if (rstatus != 0) fprintf(stderr,"R join returns status %d\n",rstatus);
    pstatus = pthread_join(pthread_id,&preturn);
    if (pstatus != 0) fprintf(stderr,"P join returns status %d\n",pstatus);

    xsb_kill_thread(r_th);
    xsb_close();
}
and commands posed by other C threads until the query is closed, failed out of, or exits via an error. In fact, queries (and commands) are protected when the \texttt{Varstring} or fixed string interfaces are used. However, consider what may happen when the register level interface is used. In this case, a calling thread may call one or more API functions to set up the registers, execute a command or query, call several more API functions to obtain the output, and so on. For this reason, if an application uses API commands that depend on user manipulation of registers (\texttt{xsb\_command()}, \texttt{xsb\_query()}, \texttt{xsb\_query\_string()}, and \texttt{xsb\_next()}) the user must ensure that only one calling thread interacts with an XSB thread when that thread in the course of executing a command or query. See $XSB\_DIR/examples/c\_calling\_xsb/cregs\_thread2.c for an example of how mutexes can be used to protect XSB threads.

When writing multi-threaded applications in XSB, be sure to be aware of how multiple threads share (and do not share) dynamic data and tables. By default dynamic predicates (and tables) are unique to a given thread. For data to be shared by multiple threads, a predicate must be declared to be \texttt{shared}. See section 7.2 for details.

14.3 A C API for XSB

14.3.1 Initializing and Closing XSB

\begin{verbatim}
int xsb_init_string(char *options)
This function is used to initialize XSB via an initialization string \*options, and must be called before any other calls can be made. The initialization string must include the path to the XSB directory installation directory $XSB\_DIR, which is expanded to an absolute path by XSB. Any other command line options may be included just as in a command line except \texttt{-D}, \texttt{-d}, \texttt{-B} and \texttt{-i}. For example, a call from an executable in a sibling directory of XSB might have the form

\texttt{xsb\_init\_string("..\!/XSB -e startup.");}

which initializes XSB with the goal ?- startup.

Return Codes
\end{verbatim}

- \texttt{XSB\_SUCCESS} indicates that initialization returned successfully.
- \texttt{XSB\_ERROR}
  - \texttt{init\_error} if any error occurred during initialization.
– permission_error if xsb_init_string() is called after XSB has already been correctly initialized.

```c
int xsb_init(int argc, char *argv[])

This function is a variant of xsb_init_string() which passes initialization arguments as an argument vector: argc is the count of the number of arguments in the argv vector. The argv vector is exactly as would be passed from the command line to XSB.

• argv[0] must be an absolute or relative path name of the XSB installation directory (i.e., $XSB_DIR). Here is an example, which assumes that we invoke the C program from the XSB installation directory.

```c
int main(int argc, char *argv[])
{
    int myargc = 1;
    char *myargv[1];

    /* XSB_init relies on the calling program to pass the addr of the XSB installation directory. From here, it will find all the libraries */
    myargv[0] = ".";

    /* Initialize xsb */
    xsb_init(myargc,myargv);
}
```

The return codes for xsb_init() are the same as those for xsb_init_string().

```c
int xsb_close()

This routine closes the entire connection to XSB. After this, no more calls can be made (not even calls to xsb_init_string() or xsb_init()). In Version 3.6, no guarantee is made that all space used by XSB will be restored to the process (even when the process has dynamically linked to XSB), but space for any XSB tables is freed.

Return Codes

• XSB_SUCCESS indicates that XSB was closed successfully.
• XSB_ERROR
  – permission_error if xsb_closed() when XSB has not been (correctly) initialized.
14.3.2 Passing Commands to XSB

```c
int xsb_command_string(th_context *th, char *cmd)
```

This function passes a command to the XSB thread designated by `th` (the first argument is not used in the single-threaded engine). No query can be active in `th` when the command is called. The command is a string consisting of a Prolog (or HiLog) term terminated by a period (\`).

When used in the multi-threaded engine, `xsb_command_string` protects the called thread from API calls from other pthreads until the command is finished.

**Return Codes**

- `XSB_SUCCESS` indicates that the command succeeded.
- `XSB_FAILURE` indicates that the command failed.
- `XSB_ERROR`
  - `permission_error` if `xsb_command_string()` is called while a query is open in `th`.
  - Otherwise, any queries thrown during execution of the command are accessible through `xsb_get_error_type(th)` and `xsb_get_error_message(th)`.

```c
int xsb_command(th_context *th)
```

This function passes a command to the XSB thread designated by `th` (the first argument is not used in the single-threaded engine). Any previous query must have already been closed. Before calling `xsb_command()`, the calling program must construct the term representing the command in register 1 in the XSB thread’s space. This can be done by using the `c2p_*` (and `p2p_*`) routines, which are described in Section 13.2.3 below. Register 2 may also be set before the call to `xsb_query()` (using `xsb_make_vars(int)` and `xsb_set_var_*()`) in which case any variables set to values in the `ret/n` term will be so bound in the call to the command goal. `xsb_command` invokes the command represented in register 1 and returns `XSB_SUCCESS` if the command succeeds, `XSB_FAILURE` if it fails, and `XSB_ERROR` if an error is thrown while executing the command.

When used in the multi-threaded engine, `xsb_command_string` does not protect the called thread from API calls from other pthreads until the command is finished. It is the user’s responsibility to protect the XSB thread, using a mutex or other concurrency control, from the time the goal begins to be constructed in the register 1 until the command has completed.

Apart from the steps necessary to formulate the query and the lack of protection of the XSB thread, the behavior of `xsb_command()` is similar to that of `xsb_command_string()`, including its return codes.
14.3.3 Querying XSB

int xsb_query_string_string(th_context *th, char *query, VarString *buff, char *sep)

This function opens a query to the XSB thread designated by th (the first argument is not used in the single-threaded engine); it returns the first answer (if there is one) as a VarString. Any previous query to th must have already been closed. Any query may return multiple data answers. The first is found and made available to the caller as a result of this call. To get any subsequent answers, xsb_next_string() must be called. An example call is:

```c
rc = xsb_query_string_string(th, "append(X,Y,[a,b,c]).",buff,";");
```

The second argument is the period-terminated query string. The third argument is a pointer to a variable string buffer in which the subroutine returns the answer (if any.) The variable string data type VarString is explained in Section 14.4. (Use xsb_query_string_string_b() if you cannot declare a parameter of this type in your programming language.) The last argument is a string provided by the caller, which is used to separate arguments in the returned answer. For the example query, buff would be set to the string:

```c
[];[a,b,c]
```

which is the first answer to the append query. There are two fields of this answer, corresponding to the two variables in the query, X and Y. The bindings of those variables make up the answer and the individual fields are separated by the sep string, here the semicolon (;). In the answer string, XSB atoms are printed without quotes. Complex terms are printed in a canonical form, with atoms quoted if necessary, and lists produced in the normal list notation.

When used in the multi-threaded engine, xsb_query_string_string protects the called thread from API calls from other pthreads until the entire query is finished.

Return Codes

- **XSB_SUCCESS** indicates that the query succeeded.
- **XSB_FAILURE** indicates that the query failed.
- **XSB_ERROR**
  - permission_error if xsb_query_string_string() is called while a query to th is open.
– Otherwise, any errors thrown during execution of the query are accessible through `xsb_get_error_type()` and `xsb_get_error_message()`.

```c
int xsb_query_string_string_b(th_context *th, char *query, char *buff, int bufflen, int *anslen, char *sep)
```

This function provides a lower-level alternative to `xsb_query_string_string` (not using the `VarString` type), which makes it easier for non-C callers (such as Visual Basic or Delphi) to access XSB functionality. Any previous query to `th` must have already been closed. Any query may return possibly multiple data answers. The first is found and made available to the caller as a result of this call. To get any subsequent answers, `xsb_next_string_b()` or a similar function must be called. The first and last arguments are the same as in `xsb_query_string_string()`. The `buff`, `bufflen`, and `anslen` parameters are used to pass the answer (if any) back to the caller. `buff` is a character array provided by the caller in which the answer is returned. `bufflen` is the length of the buffer (`buff`) and is provided by the caller. `anslen` is returned by this routine and is the length of the computed answer. If that length is less than `bufflen`, then the answer is put in `buff` (and null-terminated). If the answer is longer than will fit in the buffer (including the null terminator), then the answer is not copied to the buffer and `XSB_OVERFLOW` is returned. In this case the caller can retrieve the answer by providing a bigger buffer (of size greater than the returned `anslen`) in a call to `xsb_get_last_answer_string()`.

When used in the multi-threaded engine, `xsb_query_string_string_string_b` protects the called thread from API calls from other pthreads until the entire query is finished.

**Return Codes**

- **XSB_SUCCESS** indicates that the query succeeded.
- **XSB_FAILURE** indicates that the query failed.
- **XSB_ERROR**
  - `permission_error` if `xsb_query_string_string_b()` is called while a query to `th` is open.
  - Otherwise, any queries thrown during execution of the command are accessible through `xsb_get_error_type()` and `xsb_get_error_message()`.
- **XSB_OVERFLOW** indicates that the query succeeded, but the answer was too long for the buffer.

```c
int xsb_query(th_context *th)
```

This function passes a query to the XSB thread `th`. Any previous query to `th` must have already been closed. Any query may return possibly multiple
data answers. The first is found and made available to the caller as a result of this call. To get any subsequent answers, \texttt{xsb\_next()} or a similar function must be called. Before calling \texttt{xsb\_query()} the caller must construct the term representing the query in the XSB thread's register 1 (using routines described in Section 13.2.3 below.) If the query has no answers (i.e., just fails), register 1 is set back to a free variable and \texttt{xsb\_query()} returns \texttt{XSB\_FAILURE}. If the query has at least one answer, the variables in the query term in register 1 are bound to those answers and \texttt{xsb\_query()} returns \texttt{XSB\_SUCCESS}. In addition, register 2 is bound to a term whose main functor symbol is \texttt{ret/n}, where \(n\) is the number of variables in the query. The main subfields of this term are set to the variable values for the first answer. (These fields can be accessed by the functions \texttt{p2c\_*}, or the functions \texttt{xsb\_var\_*}, described in Section 13.2.3 below.) Thus there are two places the answers are returned. Register 2 is used to make it easier to access them. Register 2 may also be set before the call to \texttt{xsb\_query()} (using \texttt{xsb\_make\_vars(int)} and \texttt{xsb\_set\_var\_*()}) in which case any variables set to values in the \texttt{ret/n} term will be so bound in the call to the goal.

When used in the multi-threaded engine, \texttt{xsb\_query} \textit{does not protect} the called thread from API calls from other pthreads until the query is finished, or even when the registers are being accessed. It is the user's responsibility to protect the XSB thread, using a mutex or other concurrency control, from the time the goal begins to be constructed in the register 1 until the query is closed, failed, or exited upon error.

\begin{verbatim}
int xsb_get_last_answer_string(th_context *th, char *buff, int bufflen, int *anslen)

This function is used only when a call \texttt{xsb\_query\_string\_string\_b()} or \texttt{xsb\_next\_string\_b()} to \texttt{th} returns \texttt{XSB\_OVERFLOW}, indicating that the buffer provided was not big enough to contain the computed answer. In that case the user may allocate a larger buffer and then call this routine to retrieve the answer (that had been saved.) Only one answer is saved per thread, so this routine must called immediately after the failing call in order to get the right answer. The parameters are the same as the 2nd through 4th parameters of \texttt{xsb\_query\_string\_string\_b()}.

Return Codes

- \texttt{XSB\_OVERFLOW} indicates that the answer was still too long for the buffer.
\end{verbatim}

\begin{verbatim}
int xsb_query_string(th_context *th, char *query)

This function passes a query to the XSB thread \texttt{th}. The query is a string
consisting of a term that can be read by the XSB reader. The string must be terminated with a period (.). Any previous query must have already been closed. In all other respects, `xsb_query_string()` is similar to `xsb_query()`, except the only way to retrieve answers is through Register 2. The ability to create the return structure and bind variables in it is particularly useful in this function.

When used in the multi-threaded engine, `xsb_query_string` *does not protect* the called thread from API calls from other pthreads until the query is finished, or even when the registers are being accessed. It is the user’s responsibility to protect the XSB thread, using a mutex or other concurrency control, from the time the goal begins to be constructed in the register 1 until the query is closed, failed, or exited upon error.

**Return Codes**

- `XSB_SUCCESS` indicates that the query succeeded.
- `XSB_FAILURE` indicates that the query failed.
- `XSB_ERROR` indicates that an error occurred while executing the query.

```c
int xsb_next_string(th_context *th, VarString *buff, char *sep)
```

This routine is called after `xsb_query_string()` to retrieve a subsequent answer in `buff`. If a query is not open in `th`, an error is returned. This function treats answers just as `xsb_query_string_string()`. For example after the example call

```
rc = xsb_query_string_string(th,"append(X,Y,[a,b,c]).",buff,";");
```

which returns with `buff` set to

```
[;][a,b,c]
```

Then a call:

```
rc = xsb_next_string(th,buff,";");
```

returns with `buff` set to

```
[a];[b,c]
```
the second answer to the indicated query.

In the multi-threaded engine,\texttt{xsb\_next\_string()} protects the XSB thread from concurrent access by other threads as long as the query was invoked by \texttt{xsb\_query\_string\_string(_b)}.

**Return Codes**

- \texttt{XSB\_SUCCESS} indicates that the query succeeded.
- \texttt{XSB\_FAILURE} indicates that the query failed.
- \texttt{XSB\_ERROR} indicates that an error occurred while executing the query.

```c
int \texttt{xsb\_next\_string\_b(th\_context *th, char *buff, int bufflen, int *anslen, char *sep)}
```

This function is a variant of \texttt{xsb\_next\_string()} that does not use the \texttt{VarString} type. Its parameters are the same as the 3rd through 6th parameters of \texttt{xsb\_query\_string\_string\_b(_b)}. The next answer to the current query is returned in \texttt{buff}, if there is enough space. If the buffer would overflow, this routine returns \texttt{XSB\_OVERFLOW}, and the answer can be retrieved by providing a larger buffer in a call to \texttt{xsb\_get\_last\_answer\_string\_b(_b)}. In any case, the length of the answer is returned in \texttt{anslen}.

In the multi-threaded engine, \texttt{xsb\_next\_string()} protects the XSB thread from concurrent access by other threads as long as the query was invoked by \texttt{xsb\_query\_string\_string(_b)}.

**Return Codes**

- \texttt{XSB\_SUCCESS} indicates that backtracking into the query succeeded.
- \texttt{XSB\_FAILURE} indicates that backtracking into the query failed.
- \texttt{XSB\_ERROR} indicates that an error occurred while further executing the query.
- \texttt{XSB\_OVERFLOW} indicates that backtracking into the query succeeded, but the new answer was too long for the buffer.

```c
int \texttt{xsb\_next(th\_context *)}
```

This function is called after \texttt{xsb\_query()} (which must have returned \texttt{XSB\_SUCCESS}) to retrieve more answers. It rebinds the query variables in the term in register 1 and rebinds the argument fields of the \texttt{ret/n} answer term in register 2 to reflect the next answer to the query. Its return codes are as with \texttt{xsb\_next\_string(_)}. When used in the multi-threaded engine, \texttt{xsb\_next does not protect} the called thread from API calls from other pthreads until the query is finished, or even when the registers are being accessed. It is the user's responsibility to protect
the XSB thread, using a mutex or other concurrency control, through the time that registers are accessed by the calling program.

```c
int xsb_close_query(th_context *th)
```

This function allows a user to close a query to `th` before all its answers have been retrieved. Since XSB is (usually) a tuple-at-a-time system, answers that are not retrieved are not computed so that closing a query may save time. If a given query \( Q \) is open, it is an error to open a new query without closing \( Q \) either by retrieving all its answers or explicitly calling `xsb_close_query()` to close \( Q \). Calling `xsb_close_query()` when no query is open gives an error message, but otherwise has no effect.

**Return Codes**

- `XSB_SUCCESS` indicates that the current query was closed.
- `XSB_ERROR` – `permission_error` if `xsb_close_query()` is called while no query is open.

### 14.3.4 Obtaining Information about Errors

```c
char * xsb_get_init_error_message()
```

Used to find error messages if `xsb_init_string()` or `xsb_init()` returns `XSB_ERROR`. Any errors returned by these functions have type `init_error`. Because initialization errors occur before XSB or any of its threads have been initialized, initialization errors do not require a thread context for input.

```c
char * xsb_get_error_type(th_context *th)
```

If a function called for `th` returned `XSB_ERROR` this function provides a pointer to a string representing the type of the error. Types are as in Volume 1 *Exception Handling* with the addition of `init_error` for errors that occur during initialization of XSB, and `unrecoverable_error` for errors from which no recovery is possible for XSB (e.g. inability to allocate new memory).

```c
char * xsb_get_error_message(th_context *th)
```

If a function called for `th` returned `XSB_ERROR` this function provides a pointer to a string representing a message associated with the error. For errors raised within the Prolog portion of execution, messages are as in Volume 1 *Exception Handling*. 
14.3.5 Thread Management from Calling Programs

```c
int xsb_ccall_thread_create(th_context *callingThread, th_context **newThread)
```

Causes `callingThread` to create a thread pointed to by `newThread`. `newThread` runs exactly the same interpreter loop as `callingThread` and all API functions will work on `newThread` just as on the main thread, or any other thread. `newThread` will be non-detached, and will inherit any private parameters from `callingThread`. To create a thread to do a specific task or a detached thread, rather than one that executes a command loop, simply call the query `thread_create/[2,3]` from one of the query functions.

```c
th_context *xsb_get_main_thread()
```

Returns a pointer to the thread context of XSB’s main thread. If XSB has not been initialized or has been closed this function returns 0.

```c
xsb_tid xsb_thread_id_to_context(th_context *th)
```

```c
th_context *xsb_thread_context_to_id(xsb_tid id)
```

14.4 The Variable-length String Data Type

XSB uses variable-length strings to communicate with certain C subroutines when the size of the output that needs to be passed from the Prolog side to the C side is not known. Variable-length strings adjust themselves depending on the size of the data they must hold and are ideal for this situation. For instance, as we have seem the two subroutines `xsb_query_string_string(query,buff,sep)` and `xsb_next_string(buff,sep)` use the variable string data type, `VarString`, for their second argument. To use this data type, make sure that

```c
#include "cinterf.h"
```

appears at the top of the program file. Variables of the `VarString` type are declared using a macro that must appear in the declaration section of the program:

```c
XSB_StrDefine(buf);
```
There is one important consideration concerning VarString with the automatic storage class: they must be destroyed on exit (see XSB_StrDestroy, below) from the procedure that defines them, or else there will be a memory leak. It is not necessary to destroy static VarString’s.

The public attributes of the type are int length and char *string. Thus, buf.string represents the actual contents of the buffer and buf.length is the length of that data. Although the length and the contents of a VarString string is readily accessible, the user must not modify these items directly. Instead, he should use the macros provided for that purpose:

- XSB_StrSet(VarString *vstr, char *str): Assign the value of the regular null-terminated C string to the VarString vstr. The size of vstr is adjusted automatically.

- XSB_StrSetV(VarString *vstr1, VarString *vstr2): Like XSB_StrSet, but the second argument is a variable-length string, not a regular C string.

- XSB_StrAppend(VarString *vstr, char *str): Append the null-terminated string str to the VarString vstr. The size of vstr is adjusted.

- XSB_StrPrepend(VarString *vstr, char *str): Like XSB_StrAppend, except that str is prepended.

- XSB_StrAppendV(VarString *vstr1, VarString *vstr2): Like XSB_StrAppend, except that the second string is also a VarString.

- XSB_StrPrependV(VarString *vstr1, VarString *vstr2): Like XSB_StrAppendV, except that the second string is prepended.

- XSB_StrCompare(VarString *vstr1, VarString *vstr2): Compares two VarString. If the first one is lexicographically larger, then the result is positive; if the first string is smaller, than the result is negative; if the two strings have the same content (i.e., vstr1->string equals vstr2->string) then the result is zero.

- XSB_StrCmp(VarString *vstr, char *str): Like XSB_StrCompare but the second argument is a regular, null-terminated string.

- XSB_StrAppendBlk(VarString *vstr, char *blk, int size): This is like XSB_StrAppend, but the second argument is not assumed to be null-terminated. Instead, size characters pointed to by blk are appended to vstr. The size of vstr is adjusted, but the content is not null terminated.
• XSB_StrPrependBlk(VarString *vstr, char *blk, int size): Like XSB_StrPrepend, but blk is not assumed to point to a null-terminated string. Instead, size characters from the region pointed to by blk are prepended to vstr.

• XSB_StrNullTerminate(VarString *vstr): Null-terminates the VarString string vstr. This is used in conjunction with XSB_StrAppendBlk, because the latter does not null-terminate variable-length strings.

• XSB_StrEnsureSize(VarString *vstr, int minsize): Ensure that the string has room for at least minsize bytes. This is a low-level routine, which is used to interface to procedures that do not use VarString internally. If the string is larger than minsize, the size might actually shrink to the nearest increment that is larger minsize.

• XSB_StrShrink(VarString *vstr, int increment): Shrink the size of vstr to the minimum necessary to hold the data. increment becomes the new increment by which vstr is adjusted. Since VarString is automatically shrunk by XSB_StrSet, it is rarely necessary to shrink a VarString explicitly. However, one might want to change the adjustment increment using this macro (the default increment is 128).

• XSB_StrDestroy(VarString *vstr): Destroys a VarString. Explicit destruction is necessary for VarString’s with the automatic storage class. Otherwise, memory leak is possible.

14.5 Passing Data into an XSB Module

The previous chapter described the low-level XSB/C interface that supports passing the data of arbitrary complexity between XSB and C. However, in cases when data needs to be passed into an executable XSB module by the main C program, the following higher-level interface should suffice. (This interface is actually implemented using macros that call the lower level functions.) These routines can be used to construct commands and queries into XSB ‘s register 1, which is necessary before calling xsb_query() or xsb_command().

```c
void xsb_make_vars((int) N)
    xsb_make_vars creates a return structure of arity N in Register 2. So this routine may called before calling any of xsb_query, xsb_query_string, xsb_command, or xsb_command_string if parameters are to be set to be sent to the goal. It
```
must be called before calling one of the \texttt{xsb\_set\_var\_*} routines can be called. \texttt{N} must be the number of variables in the query that is to be evaluated.

\begin{verbatim}
void xsb_set_var_int((int) Val, (int) N)
    set_and_int sets the \texttt{N}^{th} field in the return structure to the integer value \texttt{Val}. It is used to set the value of the \texttt{N}^{th} variable in a query before calling \texttt{xsb\_query} or \texttt{xsb\_query\_string}. When called in XSB, the query will have the \texttt{N}^{th} variable set to this value.

void xsb_set_var_string((char *) Val, (int) N)
    set_and_string sets the \texttt{N}^{th} field in the return structure to the atom with name \texttt{Val}. It is used to set the value of the \texttt{N}^{th} variable in a query before calling \texttt{xsb\_query} or \texttt{xsb\_query\_string}. When called in XSB, the query will have the \texttt{N}^{th} variable set to this value.

void xsb_set_var_float((float) Val, (int) N)
    set_and_float sets the \texttt{N}^{th} field in the return structure to the floating point number with value \texttt{Val}. It is used to set the value of the \texttt{N}^{th} variable in a query before calling \texttt{xsb\_query} or \texttt{xsb\_query\_string}. When called in XSB, the query will have the \texttt{N}^{th} variable set to this value.

prolog_int xsb_var_int((int) N)
    xsb_var_int is called after \texttt{xsb\_query} or \texttt{xsb\_query\_string} returns an answer. It returns the value of the \texttt{N}^{th} variable in the query as set in the returned answer. This variable must have an integer value (which is cast to \texttt{long} in a 64-bit architecture).

char* xsb_var_string((int) N)
    xsb_var_string is called after \texttt{xsb\_query} or \texttt{xsb\_query\_string} returns an answer. It returns the value of the \texttt{N}^{th} variable in the query as set in the returned answer. This variable must have an atom value.

prolog_float xsb_var_float((int) N)
    xsb_var_float is called after \texttt{xsb\_query} or \texttt{xsb\_query\_string} returns an answer. It returns the value of the \texttt{N}^{th} variable in the query as set in the returned answer. This variable must have a floating point value (which is cast to \texttt{double} in a 64-bit architecture).
\end{verbatim}
14.6 Creating an XSB Module that Can be Called from C

To create an executable that includes calls to the above C functions, these routines, and the XSB routines that they call, must be included in the link (ld) step.

Unix instructions: You must link your C program, which should include the main procedure, with the XSB object file located in

\$XSBDIR/config/<your-system-architecture>/saved.o/xsb.o

Your program should include the file cinterf.h located in the XSB/emu subdirectory, which defines the routines described earlier, which you will need to use in order to talk to XSB. It is therefore recommended to compile your program with the option -I\$XSBDIR/XSB/emu.

The file \$XSBDIR/config/your-system-architecture/modMakefile is a makefile you can use to build your programs and link them with XSB. It is generated automatically and contains all the right settings for your architecture, but you will have to fill in the name of your program, etc.

It is also possible to compile and link your program with XSB using XSB itself as follows:

```
:- xsb_configuration(compiler_flags,CFLAGS),
    xsb_configuration(loader_flags,LDFLAGS),
    xsb_configuration(config_dir,CONFDIR),
    xsb_configuration(emudir,EMUDIR),
    xsb_configuration(compiler,Compiler),
    str_cat(CONFDIR, '/saved.o/', ObjDir),
    write('Compiling myprog.c ... '),
    shell([Compiler, '-I', EMUDIR, '-c ', CFLAGS, ' myprog.c ]),
    shell([Compiler, CFLAGE, '-o ', './myprog ',
               ObjDir, 'xsb.o ' , ' myprog.o ', LDFLAGS]),
    writeln(done).
```

This works for every architecture and is often more convenient than using the make files. There are simple examples of C programs calling XSB in the \$XSBDIR/examples/c_calling_XSB

---

6The variable CFLAGS is needed in the linking stage in order to ensure that the appropriate memory option is passed if XSB is configured --with-bits32 or --with-bits64 to override the default on a 64-bit platform.
directory, in files cmain.c, ctest.P, cmain2.c.

**Windows instructions:** To call XSB from C, you must build it as a DLL, which is done as follows:

```
cd $XSB_DIR\XSB\build
makeXsb_wind DLL="yes"
```

The DLL, which you can call dynamically from your program is then found in

```
$XSB_DIR\config\x86-32-bit-windows\bin\xsb.dll
```

Since your program must include the file cinterf.h, it is recommended to compile it with the option `/I$XSB_DIR\XSB\emu`. 
Chapter 15

Restrictions and Current Known Bugs

In this chapter we indicate some features and bugs of XSB that may affect the users at some point in their interaction with the system.

If at some point in your interaction with the system you suspect that you have run across a bug not mentioned below, please report it to (xsb-contact@cs.sunysb.edu). Please try to find the smallest program that illustrates the bug and mail it to this address together with a script that shows the problem. We will do our best to fix it or to assist you to bypass it.

15.1 Current Restrictions

- The maximum arity for predicate and function symbols is 255.

- The maximum length of atoms that can be stored in an XSB object code file is in principle $2^{32} - 1$.

- Not all of XSB’s tabling and builtins currently take account of cyclic terms, so using them may lead to XSB hanging or crashing (cf. Section 6.8). Cyclic terms can be checked using the predicate is_cyclic/1.

- In the current version, you should never try to rename a byte code file generated for a module, though you can move it around in your file system. Since the module name is stored in the file, renaming it causes the system to load it into wrong places. However, byte code files for non-modules can be renamed at will.
• XSB allows up to 1 Gigabyte of address space for 32-bit chips. There are various tagging schemes, which depend on the operating system and where in the 32-bit virtual address space it allocates user memory. The most general tagging scheme (named GENERAL_TAGGING) adjusts itself to the address space in use. Other more specific tagging schemes are available for specific architectures. Floating point numbers are by default double precision when computed at runtime. Floating point numbers in the compiler are only single precision (due to the way they are represented in object byte-code files.) If \texttt{-enable-fast-floats} is specified, then 28-bit floats are used. For 64-bit platforms, addresses are stored in 60 bits. However, as the object code file format is the same as for the 32-bit versions, compiled constants are subject to 32-bit limitations.

• Indexing on floating-point numbers is suspect, since, as implemented in XSB, the semantics of floating-point unification is murky in the best case. Therefore, it is advisable that if you use floating point numbers in the first argument of a procedure, that you explicitly index the predicate in some other argument.

• The XSB compiler cannot distinguish the occurrences of a 0-ary predicate and a name of a module (of an import declaration) as two different entities. For that reason it fails to characterise the same symbol table entry as both a predicate and a module at the same time. As a result of this fact, a compiler error is issued and the file is not compiled. For that reason we suggest the use of mutually exclusive names for modules and 0-ary predicates, though we will try to amend this restriction in future versions of XSB.

• Tabled predicates that use call-subsumption do not handle calls that use attributed variables, and may not use answer subsumption or incremental tabling.

15.2 Known Bugs

• The reader cannot read an infix operator immediately followed by a left parenthesis. In such a case you get a syntax error. To avoid the syntax error just leave a blank between the infix operator and the left parenthesis. For example, instead of writing:

\begin{verbatim}
| ?- X=(a,b).
\end{verbatim}

write:

\begin{verbatim}
| ?- X= (a,b).
\end{verbatim}
• The reader cannot properly read an operator defined as both a prefix and an infix operator. For instance the declaration

\begin{verbatim}
:- op(1200,xf,'<=').
:- op(1200,xfx,'<=').
\end{verbatim}

will lead to a syntax error.

• When the code of a predicate is reloaded many times, if the old code is still in use at the time of loading, unexpected errors may occur, due to the fact that the space of the old code is reclaimed and may be used for other purposes.

• Currently, term comparisons (==, @=<, @<, @>, and @=>) do not work for terms that overflow the C-recursion stack (terms that contain more than 10,000 variables and/or function symbols).
Appendix A

GPP - Generic Preprocessor

Version 2.0 - (c) Denis Auroux 1996-99
http://www.math.polytechnique.fr/cmat/auroux/prog/gpp.html

As of version 2.1, XSB uses gpp as a source code preprocessor for Prolog programs. This helps maintain consistency between the C and the Prolog parts of XSB through the use of the same .h files. In addition, the use of macros improves the readability of many Prolog programs, especially those that deal with low-level aspects of XSB. Chapter 3.10 explains how gpp is invoked in XSB.

A.1 Description

gpp is a general-purpose preprocessor with customizable syntax, suitable for a wide range of preprocessing tasks. Its independence on any programming language makes it much more versatile than cpp, while its syntax is lighter and more flexible than that of m4.

gpp is targeted at all common preprocessing tasks where cpp is not suitable and where no very sophisticated features are needed. In order to be able to process equally efficiently text files or source code in a variety of languages, the syntax used by gpp is fully customizable. The handling of comments and strings is especially advanced.

Initially, gpp only understands a minimal set of built-in macros, called metac- macros. These meta-macros allow the definition of user macros as well as some basic operations forming the core of the preprocessing system, including conditional tests, arithmetic evaluation, and syntax specification. All user macro definitions
are global, i.e. they remain valid until explicitly removed; meta-macros cannot be redefined. With each user macro definition gpp keeps track of the corresponding syntax specification so that a macro can be safely invoked regardless of any subsequent change in operating mode.

In addition to macros, gpp understands comments and strings, whose syntax and behavior can be widely customized to fit any particular purpose. Internally comments and strings are the same construction, so everything that applies to comments applies to strings as well.

A.2 Syntax

```sh
gpp [-o outfile] [-I/include/path] [-Dname=val ...]
   [-z|+z] [-x] [-m] [-n] [-C|-T|-H|-P|-U ... [-M ...]]
   [+c<n> str1 str2] [-c str1]
   [+s<n> str1 str2 c] [infile]
```

A.3 Options

`gpp` recognizes the following command-line switches and options:

- **-h**
  Print a short help message.

- **-o outfile**
  Specify a file to which all output should be sent (by default, everything is sent to standard output).

- **-I /include/path**
  Specify a path where the `#include` meta-macro will look for include files if they are not present in the current directory. The default is `/usr/include` if no `-I` option is specified. Multiple `-I` options may be specified to look in several directories.

- **-D name=val**
  Define the user macro `name` as equal to `val`. This is strictly equivalent to using the `#define` meta-macro, but makes it possible to define macros from the command-line. If `val` makes references to arguments or other macros, it should
conform to the syntax of the mode specified on the command-line. Note that macro argument naming is not allowed on the command-line.

• +z
Set text mode to Unix mode (LF terminator). Any CR character in the input is systematically discarded. This is the default under Unix systems.

• -z
Set text mode to DOS mode (CR-LF terminator). In this mode all CR characters are removed from the input, and all output LF characters are converted to CR-LF. This is the default if gpp is compiled with the WIN_NT option.

• -x
Enable the use of the \#exec meta-macro. Since \#exec includes the output of an arbitrary shell command line, it may cause a potential security threat, and is thus disabled unless this option is specified.

• -m
Enable automatic mode switching to the cpp compatibility mode if the name of an included file ends in '.h' or '.c'. This makes it possible to include C header files with only minor modifications.

• -n
Prevent newline or whitespace characters from being removed from the input when they occur as the end of a macro call or of a comment. By default, when a newline or whitespace character forms the end of a macro or a comment it is parsed as part of the macro call or comment and therefore removed from output. Use the -n option to keep the last character in the input stream if it was whitespace or a newline.

• -U arg1 ... arg9
User-defined mode. The nine following command-line arguments are taken to be respectively the macro start sequence, the macro end sequence for a call without arguments, the argument start sequence, the argument separator, the argument end sequence, the list of characters to stack for argument balancing, the list of characters to unstack, the string to be used for referring to an argument by number, and finally the quote character (if there is none an empty string should be provided). These settings apply both to user macros and to meta-macros, unless the -M option is used to define other settings for meta-macros. See the section on syntax specification for more details.
-M arg1 ... arg7
User-defined mode specifications for meta-macros. This option can only be used together with -M. The seven following command-line arguments are taken to be respectively the macro start sequence, the macro end sequence for a call without arguments, the argument start sequence, the argument separator, the argument end sequence, the list of characters to stack for argument balancing, and the list of characters to unstack. See below for more details.

(default mode)
The default mode is a vaguely cpp-like mode, but it does not handle comments, and presents various incompatibilities with cpp. Typical meta-macros and user macros look like this:

```cpp
#define x y
macro(arg,...)
```

This mode is equivalent to

```cpp
-U " " " "(" " "")" "(" "")" "#" "\"
-M "#" "\n" " " " "\n" "(" "")"
```

-C
cpp compatibility mode. This is the mode where gpp’s behavior is the closest to that of cpp. Unlike in the default mode, meta-macro expansion occurs only at the beginning of lines, and C comments and strings are understood. This mode is equivalent to

```cpp
-n -U " " " "(" " "")" "(" "")" "#" "
-M "\n#\w" "\n" " " "\n" " " "
+c "/*" "*/" +c "//" "\n" +c "\\\n" "
+s "\" "\" "\" "\" "\"
```

-T
TeX-like mode. In this mode, typical meta-macros and user macros look like this:

```latex
\define{x}{y}
\macro{arg}\{\ldots\}
```

No comments are understood. This mode is equivalent to
-U "\\" " " \" \" \" \" \" \" \{" \} \{ \} \{" \} \"\" \"\"

- H
HTML-like mode. In this mode, typical meta-macros and user macros look like this:

```c
#define x|y>
#define macro arg|...
```

No comments are understood. This mode is equivalent to

-U "<# " " " " " " " " " " " " " " " " " \\"

- P
Prolog-compatible cpp-like mode. This mode differs from the cpp compatibility mode by its handling of comments, and is equivalent to

```c
-n -U " " " " ( " , " ) " ( " ) " " " " "
-M \n#w " \n" " " " " \n" " " " 
+ccss \\o/* " */ +ccss \\# " \n +cii \n\n " " 
+s "\" "\" " " +s "\" "\" " " "
```

- +c <n> str1 str2
Specify comments. Any unquoted occurrence of str1 will be interpreted as the beginning of a comment. All input up to the first following occurrence of str2 will be discarded. This option may be used multiple times to specify different types of comment delimiters. The optional parameter <n> can be specified to alter the behavior of the comment and e.g. turn it into a string or make it ignored under certain circumstances, see below.

- -c str1
Un-specify comments or strings. The comment/string specification whose start sequence is str1 is removed. This is useful to alter the built-in comment specifications of a standard mode, e.g. the cpp compatibility mode.

- +s <n> str1 str2 c
Specify strings. Any unquoted occurrence of str1 will be interpreted as the beginning of a string. All input up to the first following occurrence of str2 will be output as is without any evaluation. The delimiters themselves are output. If c is non-empty, its first character is used as a string-quote character, i.e. a character whose presence immediately before an occurrence of str2 prevents it
from terminating the string. The optional parameter \(<n>\) can be specified to alter the behavior of the string and e.g. turn it into a comment, enable macro evaluation inside the string, or make the string specification ignored under certain circumstances, see below.

- \(-s \text{ str1}\)
  Un-specify comments or strings. Identical to \(-c\).

- \(\text{infile}\)
  Specify an input file from which gpp reads its input. If no input file is specified, input is read from standard input.

### A.4 Syntax Specification

The syntax of a macro call is the following: it must start with a sequence of characters matching the **macro start sequence** as specified in the current mode, followed immediately by the name of the macro, which must be a valid **identifier**, i.e. a sequence of letters, digits, or underscores ("_"). The macro name must be followed by a **short macro end sequence** if the macro has no arguments, or by a sequence of arguments initiated by an **argument start sequence**. The various arguments are then separated by an **argument separator**, and the macro ends with a **long macro end sequence**.

In all cases, the parameters of the current context, i.e. the arguments passed to the body being evaluated, can be referred to by using an **argument reference sequence** followed by a digit between 1 and 9. Macro parameters may alternately be named (see below). Furthermore, to avoid interference between the gpp syntax and the contents of the input file a **quote character** is provided. The quote character can be used to prevent the interpretation of a macro call, comment, or string as anything but plain text. The quote character "protects" the following character, and always gets removed during evaluation. Two consecutive quote characters evaluate as a single quote character.

Finally, to facilitate proper argument delimitation, certain characters can be 'stacked' when they occur in a macro argument, so that the argument separator or macro end sequence are not parsed if the argument body is not balanced. This allows nesting macro calls without using quotes. If an improperly balanced argument is needed, quote characters should be added in front of some stacked characters to make it balanced.

The macro construction sequences described above can be different for metamacros and for user macros: this is e.g. the case in cpp mode. Note that, since meta-
macros can only have up to two arguments, the delimitation rules for the second argument are somewhat sloppier, and unquoted argument separator sequences are allowed in the second argument of a meta-macro.

Unless one of the standard operating modes is selected, the above syntax sequences can be specified either on the command-line, using the -M and -U options respectively for meta-macros and user macros, or inside an input file via the \#mode meta and \#mode user meta-macro calls. In both cases the mode description consists of 9 parameters for user macro specifications, namely the macro start sequence, the short macro end sequence, the argument start sequence, the argument separator, the long macro end sequence, the string listing characters to stack, the string listing characters to unstack, the argument reference sequence, and finally the quote character. As explained below these sequences should be supplied using the syntax of C strings; they must start with a non-alphanumeric character, and in the first five strings special matching sequences can be used (see below). If the argument corresponding to the quote character is the empty string that functionality is disabled. For meta-macro specifications there are only 7 parameters, as the argument reference sequence and quote character are shared with the user macro syntax.

The structure of a comment/string is the following: it must start with a sequence of characters matching the given comment/string start sequence, and always ends at the first occurrence of the comment/string end sequence, unless it is preceded by an odd number of occurrences of the string-quote character (if such a character has been specified). In certain cases comment/strings can be specified to enable macro evaluation inside the comment/string: in that case, if a quote character has been defined for macros it can be used as well to prevent the comment/string from ending, with the difference that the macro quote character is always removed from output whereas the string-quote character is always output. Also note that under certain circumstances a comment/string specification can be disabled, in which case the comment/string start sequence is simply ignored. Finally, it is possible to specify a string warning character whose presence inside a comment/string will cause gpp to output a warning (this is useful e.g. to locate unterminated strings in cpp mode). Note that input files are not allowed to contain unterminated comments/strings.

A comment/string specification can be declared from within the input file using the \#mode comment meta-macro call (or equivalently \#mode string), in which case the number of C strings to be given as arguments to describe the comment/string can be anywhere between 2 and 4: the first two arguments (mandatory) are the start sequence and the end sequence, and can make use of the special matching sequences (see below). They may not start with alphanumeric characters. The first character of the third argument, if there is one, is used as string-quote character (use an empty
string to disable the functionality), and the first character of the fourth argument, if there is one, is used as string-warning character. A specification may also be given from the command-line, in which case there must be two arguments if using the +c option and three if using the +s option.

The behavior of a comment/string is specified by a three-character modifier string, which may be passed as an optional argument either to the +c/+s command-line options or to the \#mode comment/\#mode string meta-macros. If no modifier string is specified, the default value is "ccc" for comments and "sss" for strings. The first character corresponds to the behavior inside meta-macro calls (including user-macro definitions since these come inside a \#define meta-macro call), the second character corresponds to the behavior inside user-macro parameters, and the third character corresponds to the behavior outside of any macro call. Each of these characters can take the following values:

- i: disable the comment/string specification.
- c: comment (neither evaluated nor output).
- s: string (the string and its delimiter sequences are output as is).
- q: quoted string (the string is output as is, without the delimiter sequences).
- C: evaluated comment (macros are evaluated, but output is discarded).
- S: evaluated string (macros are evaluated, delimiters are output).
- Q: evaluated quoted string (macros are evaluated, delimiters are not output).

Important note: any occurrence of a comment/string start sequence inside another comment/string is always ignored, even if macro evaluation is enabled. In other words, comments/strings cannot be nested. In particular, the 'Q' modifier can be a convenient way of defining a syntax for temporarily disabling all comment and string specifications.

Syntax specification strings should always be provided as C strings, whether they are given as arguments to a \#mode meta-macro call or on the command-line of a Unix shell. If command-line arguments are given via another method than a standard Unix shell, then the shell behavior must be emulated, i.e. the surrounding '"' quotes should be removed, all occurrences of '\\' should be replaced by a single backslash, and similarly '"' should be replaced by '"'. Sequences like '\n' are recognized by gpp and should be left as is.
Special sequences matching certain subsets of the character set can be used. They are of the form \( \backslash x \), where \( x \) is one of:

- \( b \): matches any sequence of one or more spaces or TAB characters (\( \backslash b \) is identical to ‘ ’).
- \( w \): matches any sequence of zero or more spaces or TAB characters.
- \( B \): matches any sequence of one or more spaces, tabs or newline characters.
- \( W \): matches any sequence of zero or more spaces, tabs or newline characters.
- \( a \): an alphabetic character (‘a’ to ‘z’ and ‘A’ to ‘Z’).
- \( A \): an alphabetic character, or a space, tab or newline.
- \( # \): a digit (‘0’ to ‘9’).
- \( i \): an identifier character. The set of matched characters is customizable using the \#mode charset id command. The default setting matches alphanumeric characters and underscores (‘a’ to ‘z’, ‘A’ to ‘Z’, ‘0’ to ‘9’ and ‘_’).
- \( t \): a TAB character.
- \( n \): a newline character.
- \( o \): an operator character. The set of matched characters is customizable using the \#mode charset op command. The default setting matches all characters in "+-*\/^\\ <>=~:.?@#&%!\", except in Prolog mode where ‘!’, ‘%’ and ‘|’ are not matched.
- \( O \): an operator character or a parenthesis character. The set of additional matched characters in comparison with \( \backslash o \) is customizable using the \#mode charset par command. The default setting is to have the characters in "(()[]{}" as parentheses.

Moreover, all of these matching subsets except \( \backslash w \) and \( \backslash W \) can be negated by inserting a ‘!’, i.e. by writing \( \backslash !x \) instead of \( \backslash x \).

Note an important distinctive feature of start sequences: when the first character of a macro or comment/string start sequence is ‘ ’ or one of the above special sequences, it is not taken to be part of the sequence itself but is used instead as a context check: for example a start sequence beginning with ‘\n’ matches only at the beginning of a line, but the matching newline character is not taken to be part of the sequence. Similarly
a start sequence beginning with ' ' matches only if some whitespace is present, but the matching whitespace is not considered to be part of the start sequence and is therefore sent to output. If a context check is performed at the very beginning of a file (or more generally of any body to be evaluated), the result is the same as matching with a newline character (this makes it possible for a cpp-mode file to start with a meta-macro call).

A.5 Evaluation Rules

Input is read sequentially and interpreted according to the rules of the current mode. All input text is first matched against the specified comment/string start sequences of the current mode (except those which are disabled by the 'i' modifier), unless the body being evaluated is the contents of a comment/string whose modifier enables macro evaluation. The most recently defined comment/string specifications are checked for first. Important note: comments may not appear between the name of a macro and its arguments (doing so results in undefined behavior).

Anything that is not a comment/string is then matched against a possible meta-macro call, and if that fails too, against a possible user-macro call. All remaining text undergoes substitution of argument reference sequences by the relevant argument text (empty unless the body being evaluated is the definition of a user macro) and removal of the quote character if there is one.

Note that meta-macro arguments are passed to the meta-macro prior to any evaluation (although the meta-macro may choose to evaluate them, see meta-macro descriptions below). In the case of the #mode meta-macro, gpp temporarily adds a comment/string specification to enable recognition of C strings ("...") and prevent any evaluation inside them, so no interference of the characters being put in the C string arguments to #mode with the current syntax is to be feared.

On the other hand, the arguments to a user macro are systematically evaluated, and then passed as context parameters to the macro definition body, which gets evaluated with that environment. The only exception is when the macro definition is empty, in which case its arguments are not evaluated. Note that gpp temporarily switches back to the mode in which the macro was defined in order to evaluate it: so it is perfectly safe to change the operating mode between the time when a macro is defined and the time when it is called. Conversely, if a user macro wishes to work with the current mode instead of the one that was used to define it it needs to start with a #mode restore call and end with a #mode save call.

A user macro may be defined with named arguments (see #define description
below). In that case, when the macro definition is being evaluated, each named parameter causes a temporary virtual user-macro definition to be created; such a macro may only be called without arguments and simply returns the text of the corresponding argument.

Note that, since macros are evaluated when they are called rather than when they are defined, any attempt to call a recursive macro causes undefined behavior except in the very specific case when the macro uses \texttt{#undef} to erase itself after finitely many loop iterations.

Finally, a special case occurs when a user macro whose definition does not involve any arguments (neither named arguments nor the argument reference sequence) is called in a mode where the short user-macro end sequence is empty (e.g. \texttt{cpp} or TeX mode). In that case it is assumed to be an \textit{alias macro}: its arguments are first evaluated in the current mode as usual, but instead of being passed to the macro definition as parameters (which would cause them to be discarded) they are actually appended to the macro definition, using the syntax rules of the mode in which the macro was defined, and the resulting text is evaluated again. It is therefore important to note that, in the case of a macro alias, the arguments actually get evaluated twice in two potentially different modes.

### A.6 Meta-macros

These macros are always pre-defined. Their actual calling sequence depends on the current mode; here we use \texttt{cpp}-like notation.

- \texttt{#define} \texttt{x y}
  This defines the user macro \texttt{x} as \texttt{y}. \texttt{y} can be any valid gpp input, and may for example refer to other macros. \texttt{x} must be an identifier (i.e. a sequence of alphanumeric characters and '_'), unless named arguments are specified. If \texttt{x} is already defined, the previous definition is overwritten. If no second argument is given, \texttt{x} will be defined as a macro that outputs nothing. Neither \texttt{x} nor \texttt{y} are evaluated; the macro definition is only evaluated when it is called, not when it is declared.

It is also possible to name the arguments in a macro definition: in that case, the argument \texttt{x} should be a user-macro call whose arguments are all identifiers. These identifiers become available as user-macros inside the macro definition; these virtual macros must be called without arguments, and evaluate to the corresponding macro parameter.
• **#defeval** \ x \ y  
  This acts in a similar way to **#define**, but the second argument \ y \ is evaluated immediately. Since user macro definitions are also evaluated each time they are called, this means that the macro \ y \ will undergo two successive evaluations. The usefulness of **#defeval** is considerable, as it is the only way to evaluate something more than once, which can be needed e.g. to force evaluation of the arguments of a meta-macro that normally doesn’t perform any evaluation. However since all argument references evaluated at define-time are understood as the arguments of the body in which the macro is being defined and not as the arguments of the macro itself, usually one has to use the quote character to prevent immediate evaluation of argument references.

• **#undef** \ x  
  This removes any existing definition of the user macro \ x \.

• **#ifdef** \ x  
  This begins a conditional block. Everything that follows is evaluated only if the identifier \ x \ is defined, until either a **#else** or a **#endif** statement is reached. Note however that the commented text is still scanned thoroughly, so its syntax must be valid. It is in particular legal to have the **#else** or **#endif** statement ending the conditional block appear as only the result of a user-macro expansion and not explicitly in the input.

• **#ifndef** \ x  
  This begins a conditional block. Everything that follows is evaluated only if the identifier \ x \ is not defined.

• **#ifeq** \ x \ y  
  This begins a conditional block. Everything that follows is evaluated only if the results of the evaluations of \ x \ and \ y \ are identical as character strings. Any leading or trailing whitespace is ignored for the comparison. Note that in cpp-mode any unquoted whitespace character is understood as the end of the first argument, so it is necessary to be careful.

• **#ifneq** \ x \ y  
  This begins a conditional block. Everything that follows is evaluated only if the results of the evaluations of \ x \ and \ y \ are not identical (even up to leading or trailing whitespace).

• **#else**  
  This toggles the logical value of the current conditional block. What follows is evaluated if and only if the preceding input was commented out.
APPENDIX A. GPP - GENERIC PREPROCESSOR

- `#endif`
  This ends a conditional block started by a `#if...` meta-macro.

- `#include` file
  This causes gpp to open the specified file and evaluate its contents, inserting the resulting text in the current output. All defined user macros are still available in the included file, and reciprocally all macros defined in the included file will be available in everything that follows. The include file is looked for first in the current directory, and then, if not found, in one of the directories specified by the `-I` command-line option (or `/usr/include` if no directory was specified). Note that, for compatibility reasons, it is possible to put the file name between "" or <>.

  Upon including a file, gpp immediately saves a copy of the current operating mode onto the mode stack, and restores the operating mode at the end of the included file. The included file may override this behavior by starting with a `#mode restore` call and ending with a `#mode push` call. Additionally, when the `-m` command line option is specified, gpp will automatically switch to the cpp compatibility mode upon including a file whose name ends with either `.c` or `.h`.

- `#exec` command
  This causes gpp to execute the specified command line and include its standard output in the current output. Note that this meta-macro is disabled unless the `-x` command line flag was specified, for security reasons. If use of `#exec` is not allowed, a warning message is printed and the output is left blank. Note that the specified command line is evaluated before being executed, thus allowing the use of macros in the command-line. However, the output of the command is included verbatim and not evaluated. If you need the output to be evaluated, you must use `#defeval` (see above) to cause a double evaluation.

- `#eval` expr
  The `#eval` meta-macro attempts to evaluate `expr` first by expanding macros (normal gpp evaluation) and then by performing arithmetic evaluation. The syntax and operator precedence for arithmetic expressions are the same as in C; the only missing operators are `<<, >>, ?: and assignment operators. If unable to assign a numerical value to the result, the returned text is simply the result of macro expansion without any arithmetic evaluation. The only exceptions to this rule are the `==` and `!=` operators which, if one of the sides does not evaluate to a number, perform string comparison instead (ignoring trailing and leading spaces).
Inside arithmetic expressions, the `defined(...)` special user macro is also available: it takes only one argument, which is not evaluated, and returns 1 if it is the name of a user macro and 0 otherwise.

- **#if** `expr`
  This meta-macro invokes the arithmetic evaluator in the same manner as `#eval`, and compares the result of evaluation with the string '0' in order to begin a conditional block. In particular note that the logical value of `expr` is always true when it cannot be evaluated to a number.

- **#mode** `keyword` ...
  This meta-macro controls gpp’s operating mode. See below for a list of `#mode` commands.

The key to gpp’s flexibility is the `#mode` meta-macro. Its first argument is always one of a list of available keywords (see below); its second argument is always a sequence of words separated by whitespace. Apart from possibly the first of them, each of these words is always a delimiter or syntax specifier, and should be provided as a C string delimited by double quotes ('"'). The various special matching sequences listed in the section on syntax specification are available. Any `#mode` command is parsed in a mode where "..." is understood to be a C-style string, so it is safe to put any character inside these strings. Also note that the first argument of `#mode` (the keyword) is never evaluated, while the second argument is evaluated (except of course for the contents of C strings), so that the syntax specification may be obtained as the result of a macro evaluation.

The available `#mode` commands are:

- **#mode save** / **#mode push**
  Push the current mode specification onto the mode stack.

- **#mode restore** / **#mode pop**
  Pop mode specification from the mode stack.

- **#mode standard** `name`
  Select one of the standard modes. The only argument must be one of: default (default mode); cpp, C (cpp mode); tex, TeX (tex mode); html, HTML (html mode); prolog, Prolog (prolog mode). The mode name must be given directly, not as a C string.

- **#mode user** "s1" ... "s9"
  Specify user macro syntax. The 9 arguments, all of them C strings, are the
mode specification for user macros (see the -U command-line option and the section on syntax specification). The meta-macro specification is not affected.

- **#mode meta** \{user | 's1' ... 's7'}
  Specify meta-macro syntax. Either the only argument is user (not as a string), and the user-macro mode specifications are copied into the meta-macro mode specifications, or there must be 7 string arguments, whose significance is the same as for the -M command-line option (see section on syntax specification).

- **#mode quote** ['c']
  With no argument or "" as argument, removes the quote character specification and disables the quoting functionality. With one string argument, the first character of the string is taken to be the new quote character. The quote character cannot be alphanumeric nor '_', and cannot be one of the special matching sequences either.

- **#mode comment** [xxx] 'start' 'end' ['c' ['c']] 
  Add a comment specification. Optionally a first argument consisting of three characters not enclosed in ' ' can be used to specify a comment/string modifier (see the section on syntax specification). The default modifier is ccc. The first two string arguments are used as comment start and end sequences respectively. The third string argument is optional and can be used to specify a string-quote character (if it is "" the functionality is disabled). The fourth string argument is optional and can be used to specify a string delimitation warning character (if it is "" the functionality is disabled).

- **#mode string** [xxx] 'start' 'end' ['c' ['c']] 
  Add a string specification. Identical to #mode comment except that the default modifier is sss.

- **#mode nocomment / #mode nostring** ['start']
  With no argument, remove all comment/string specifications. With one string argument, delete the comment/string specification whose start sequence is the argument.

- **#mode preservelf** \{ on | off | 1 | 0 \}
  Equivalent to the -n command-line switch. If the argument is on or 1, any newline or whitespace character terminating a macro call or a comment/string is left in the input stream for further processing. If the argument is off or 0 this feature is disabled.
• **#mode charset**  \{ id \mid op \mid par \} "string"

Specify the character sets to be used for matching the \(\backslash o\), \(\backslash O\) and \(\backslash i\) special sequences. The first argument must be one of *id* (the set matched by \(\backslash i\)), *op* (the set matched by \(\backslash o\)) or *par* (the set matched by \(\backslash O\) in addition to the one matched by \(\backslash o\)). *"string"* is a C string which lists all characters to put in the set. It may contain only the special matching sequences \(\backslash a\), \(\backslash A\), \(\backslash b\), \(\backslash B\), and \(\backslash #\) (the other sequences and the negated sequences are not allowed). When a '-' is found in-between two non-special characters this adds all characters in-between (e.g. "A-Z" corresponds to all uppercase characters). To have '-' in the matched set, either put it in first or last position or place it next to a \(\backslash x\) sequence.

### A.7 Examples

Here is a basic self-explanatory example in standard or cpp mode:

```cpp
#define FOO This is
#define BAR a message.
#define concat #1 #2
concat(FOO,BAR)
#ifeq (concat(foo,bar)) (foo bar)
This is output.
#else
This is not output.
#endif
```

Using argument naming, the *concat* macro could alternately be defined as

```cpp
#define concat(x,y) x y
```

In TeX mode and using argument naming, the same example becomes:

```latex
\define{FOO}{This is}
\define{BAR}{a message.}
\define{\concat{x}{y}}{\x \y}
\concat{\FOO}{\BAR}
\ifeq{\concat{foo}{bar}}{foo bar}
This is output.
\else
```

This is not output.
\endif

In HTML mode and without argument naming, one gets similarly:

```<#define FOO|This is>
<#define BAR|a message.>
<#define concat|#1 #2>
<#concat <#FOO>|<#BAR>>
<#ifeq <#concat foo|bar>|foo bar>
This is output.
<#else>
This is not output.
<#endif>
```

The following example (in standard mode) illustrates the use of the quote character:

```
#define FOO This is \
  a multiline definition.
#define BLAH(x) My argument is x
BLAH(urf)
\BLAH(urf)
```

Note that the multiline definition is also valid in cpp and Prolog modes despite the absence of quote character, because `\` followed by a newline is then interpreted as a comment and discarded.

In cpp mode, C strings and comments are understood as such, as illustrated by the following example:

```
#define BLAH foo
BLAH "BLAH" /* BLAH */
'It\'s a /*string*/ !'
```

The main difference between Prolog mode and cpp mode is the handling of strings and comments: in Prolog, a `...` string may not begin immediately after a digit, and a `/*...*/` comment may not begin immediately after an operator character. Furthermore, comments are not removed from the output unless they occur in a `#command`.

The differences between cpp mode and default mode are deeper: in default mode `#commands` may start anywhere, while in cpp mode they must be at the beginning
of a line; the default mode has no knowledge of comments and strings, but has
a quote character (\"), while cpp mode has extensive comment/string specifications
but no quote character. Moreover, the arguments to meta-macros need to be correctly
parenthesized in default mode, while no such checking is performed in cpp mode.

This makes it easier to nest meta-macro calls in default mode than in cpp mode.
For example, consider the following HTML mode input, which tests for the availability
of the \#exec command:

\begin{verbatim}
<#ifeq <#exec echo blah>|blah
 > #exec allowed <#else> #exec not allowed <#endif>
\end{verbatim}

There is no cpp mode equivalent, while in default mode it can be easily translated as

\begin{verbatim}
#ifeq (#exec echo blah
 ) (blah
 )
\#exec allowed
#else
\#exec not allowed
#endif
\end{verbatim}

In order to nest meta-macro calls in cpp mode it is necessary to modify the mode
description, either by changing the meta-macro call syntax, or more elegantly by
defining a silent string and using the fact that the context at the beginning of an
evaluated string is a newline character:

\begin{verbatim}
#mode string QQQ "$" "$"
#ifeq $#exec echo blah
 $ $blah
 $\n\#exec allowed
#else
\#exec not allowed
#endif
\end{verbatim}

Note however that comments/strings cannot be nested (\"...
\" inside $...$ would go
undetected), so one needs to be careful about what to include inside such a silent
evaluated string.

Remember that macros without arguments are actually understood to be aliases
when they are called with arguments, as illustrated by the following example (default
or cpp mode):

\begin{verbatim}
#mode string QQQ "$" "$"
#ifeq $#exec echo blah
 $ $blah
 $\n\#exec allowed
#else
\#exec not allowed
#endif
\end{verbatim}
#define DUP(x) x x
#define FOO and I said: DUP
FOO(blah)

The usefulness of the `#defeval` meta-macro is shown by the following example in HTML mode:

```
<#define APPLY|<#defeval TEMP|<\##1 \#1><#TEMP #2>>
<#define <#foo x>|<#x> and <#x>>
<#APPLY foo|BLAH>
```

The reason why `#defeval` is needed is that, since everything is evaluated in a single pass, the input that will result in the desired macro call needs to be generated by a first evaluation of the arguments passed to APPLY before being evaluated a second time.

To translate this example in default mode, one needs to resort to parenthesizing in order to nest the `#defeval` call inside the definition of APPLY, but need to do so without outputting the parentheses. The easiest solution is

```
#define BALANCE(x) x
#define APPLY(f,v) BALANCE(#defeval TEMP f
TEMP(v))
#define foo(x) x and x
APPLY(foo,BLAH)
```

As explained above the simplest version in cpp mode relies on defining a silent evaluated string to play the role of the BALANCE macro.

The following example (default or cpp mode) demonstrates arithmetic evaluation:

```
#define x 4
The answer is:
#eval x*x + 2*(16-x) + 1998%x

#if defined(x) && !(3*x+5>17)
This should be output.
#endif
```

To finish, here are some examples involving mode switching. The following example is self-explanatory (starting in default mode):

```
#define x 4
The answer is:
#eval x*x + 2*(16-x) + 1998%x

#if defined(x) && !(3*x+5>17)
This should be output.
#endif
```
A good example where a user-defined mode becomes useful is the gpp source of this document (available with gpp’s source code distribution).

Another interesting application is selectively forcing evaluation of macros in C strings when in cpp mode. For example, consider the following input:

```
#define blah(x) "and he said: x"
blah(foo)
```

Obviously one would want the parameter $x$ to be expanded inside the string. There are several ways around this problem:

```
#define push
#define nostring "\\"
#define blah(x) "and he said: x"
#define pop

#define quote "'
#define blah(x) "'and he said: x"

#define string QQQ "$" "$"
#define blah(x) $$"and he said: x"$$
```

The first method is very natural, but has the inconvenient of being lengthy and neutralizing string semantics, so that having an unevaluated instance of ‘$x$’ in the string, or an occurrence of ‘/*’, would be impossible without resorting to further contortions.

The second method is slightly more efficient, because the local presence of a quote character makes it easier to control what is evaluated and what isn’t, but has the
drawback that it is sometimes impossible to find a reasonable quote character without having to either significantly alter the source file or enclose it inside a \texttt{#mode push/pop} construct. For example any occurrence of ‘/*’ in the string would have to be quoted.

The last method demonstrates the efficiency of evaluated strings in the context of selective evaluation: since comments/strings cannot be nested, any occurrence of ‘’ or ‘/*’ inside the ‘$$’ gets output as plain text, as expected inside a string, and only macro evaluation is enabled. Also note that there is much more freedom in the choice of a string evaluation delimiter than in the choice of a quote character.

### A.8 Advanced Examples

Here are some examples of advanced constructions using gpp. They tend to be pretty awkward and should be considered as evidence of gpp’s limitations.

The first example is a recursive macro. The main problem is that, since gpp evaluates everything, a recursive macro must be very careful about the way in which recursion is terminated, in order to avoid undefined behavior (most of the time gpp will simply crash). In particular, relying on a \texttt{#if/#else/#endif} construct to end recursion is not possible and results in an infinite loop, because gpp scans user macro calls even in the unevaluated branch of the conditional block. A safe way to proceed is for example as follows (we give the example in TeX mode):

```latex
\define{countdown}{
  \if{#1}
    #1...\define{loop}{\countdown}
  \else
    Done.
    \define{loop}{}
  \endif
  \loop{\eval{#1-1}}
}
\countdown{10}
```

The following is an (unfortunately very weak) attempt at implementing functional abstraction in gpp (in standard mode). Understanding this example and why it can’t be made much simpler is an exercise left to the curious reader.

```
#mode string "" "" "" \"
```
#define ASIS(x) x
#define SILENT(x) ASIS()
#define EVAL(x,f,v) SILENT(
    #mode string QQQ "'" "'" \"
    #defeval TEMP0 x
    #defeval TEMP1 ( 
        #define \TEMP2(TEMP0) f
    )
    TEMP1
)TEMP2(v)
#define LAMBDA(x,f,v) SILENT(
    #ifeq (v) ()
    #define TEMP3(a,b,c) EVAL(a,b,c)
    #else
    #define TEMP3(a,b,c) \LAMBDA(a,b)
    #endif
    TEMP3(x,f,v)
#define EVALAMBDA(x,y) SILENT(
    #defeval TEMP4 x
    #defeval TEMP5 y
)
#define APPLY(f,v) SILENT(
    #defeval TEMP6 ASIS(\EVA)f
    TEMP6
)EVAL(TEMP4,TEMP5,v)

This yields the following results:

LAMBDA(z,z+z)
=> LAMBDA(z,z+z)

LAMBDA(z,z+z,2)
=> 2+2

#define f LAMBDA(y,y*y)
f
=> LAMBDA(y,y*y)

APPLY(f,blah)
=> blah*blah
APPLY(LAMBDA(t,t t),(t t))
   => (t t) (t t)

LAMBDA(x,APPLY(f,(x+x)),urf)
   => (urf+urf)*(urf+urf)

APPLY(APPLY(LAMBDA(x,LAMBDA(y,x*y)),foo),bar)
   => foo*bar

#define test LAMBDA(y,'#ifeq y urf
 y is urf#else
 y is not urf#endif

APPLY(test,urf)
   => urf is urf

APPLY(test,foo)
   => foo is not urf

A.9 Author

Denis Auroux, e-mail: auroux@math.polytechnique.fr.

Please send me e-mail for any comments, questions or suggestions.

Many thanks to Michael Kifer for valuable feedback and for prompting me to go beyond version 1.0.
Bibliography


Index

\+/1, 181
\=/2, 187
\==/2, 187
^ 223
^ /2, 179
!/0, 181, 228, 379, 380, 387
\+/1, 98
\^'/2, 178
\^'/2, 178
!'/2, 177
'//'/2, 177
'<!'/2, 178
'!'>'/2, 178
'!'>'/2, 178
**/2, 180
*/2, 177
+/2, 177
-/-1, 177
-/-2, 177
->/2, 184
=../2, 203
=/2, 186
==/2, 187
?=/2, 187
@<!'/2, 187
@= /2, 188
@=< /2, 188
@>/2, 188
@>= /2, 188
ISO, 194, 236, 253
[]/1 (consult), 41

$trace/0, 363
^ /2, 227
^=.../2, 205
xsb_error_get_message/2, 403
‘C’/3, 387
abolish/1, 264
abolish_all_private_tables/0, 308
abolish_all_shared_tables/0, 308
abolish_all_tables/0, 307
abolish_module_tables/1, 309
abolish_nonincremental_tables/0, 307
abolish_table_pred/1, 304
abolish_table_pred/2, 305
abolish_table_subgoal/1, 306
abolish_table_subgoal/2, 306
abolish_table_subgoals/1, 305
abolish_table_subgoals/2, 306
abstract_modes/2, 378
acos/1, 180
acyclic_term/1, 210
add_xsb_hook/1, 355
analyze_an_scc/2, 378
analyze_an_scc/3, 378
arg/3, 201
arg0/3, 202
asin/1, 180
assert/1, 262
assert/3, 262
asserta/1, 261
assertz/1, 262
at_end_of_stream/0, 145
INDEX

at_end_of_stream/1, 145
atan/1, 180
atan/2, 180
atan2/2, 180
atom/1, 193
atom_chars/2, 214
atom_codes/2, 210
atom_concat/3, 216
atom_length/2, 215
atomic/1, 194
bagof/3, 224
between/3, 185
bounded_call/3, 231
bounded_call/4, 40, 231
break/0, 252
c2p_float, 417
c2p_functor, 417
c2p_int, 417
c2p_list, 417
c2p_nil, 417
c2p_string, 417
call/1, 228
call/[2,10], 229
call_cleanup/2, 232
call_tv/2, 229
callable/1, 197
catch/3, 402
ceiling/1, 178
cchar_code/2, 215
check_acyclic/3, 403
check_atom/3, 403
check Callable/3, 403
check_ground/3, 404
check_integern/3, 404
check_nonvar/3, 404
check_nonvar_list/3, 404
check_one_thread/3, 404
check_stream/3, 404
check_var/3, 404
check_variant/1, 189
check_variant/2, 189
clause/2, 265
close/1, 143
close/2, 142
compare/3, 188
compile/1, 41, 44
compile/2, 41, 44
compound/1, 195
concat_atom/2, 216
concat_atom/3, 216
console_write/1, 164
console_writeln/1, 164
copy_term/2, 206
copy_term_nat/2, 206
cos/1, 180
current_atom/1, 246
current_functor/1, 245
current_index/2, 245
current_input/1, 236
current_module/1, 244
current_module/2, 245
current_op/3, 251
current_predicate/1, 243
current_prolog_flag/2, 236
cvt_canonical/2, 275
datetime/1, 173
dcg/2, 388
debug/0, 361
debug_ctl/2, 361
debugging/0, 361
default_user_error_handler/1, 402
delete_returns/2, 309
delete_trie/1, 353
div/2, 177
domain_error/4, 400
dynamic/1, 130, 269
e/0, 180
else/0, 58
elseif/1, 58
endif/0, 58
ensure_loaded/1, 42
ensure_loaded/2, 275
epsilon/0, 180
error_write/1, 164, 402
error_writeln/1, 164, 402
evaluation_error/3, 400
excess_vars/4, 227
existence_error/4, 401
expand_term/2, 385
explain_u_val/3, 120, 241, 299
explain_u_val/6, 241, 299
fail/0, 181
fail_if/1, 181
false/0, 181
file_clone/3, 146
file_exists/1, 149
file_file_getbuf_atom/4, 170
file_getbuf_atom/3, 170
file_getbuf_list/3, 170
file_getbuf_list/4, 169
file_putbuf/4, 171
file_putbuf/5, 170
file_read_line_atom/1, 168
file_read_line_atom/2, 169
file_read_line_list/1, 168
file_read_line_list/2, 168
file_reopen/3, 146
file_truncate/3, 146
file_write_line/2, 169
file_write_line/3, 169
find_n/4, 228
findall/3, 225
findall/4, 225
float/1, 178
floor/1, 178
flush_all_output_streams/0, 147
flush_output/0, 144
flush_output/1, 144
fmt_read/3, 165
fmt_read/4, 165
fmt_write/2, 166
fmt_write/3, 166
fmt_write_string/3, 168
forall/2, 232
forest_log_overview/0, 377
functor/3, 197
gc_atoms/0, 222
gc_dynamic/1, 266
gc_heap/0, 254
gc_tables/1, 309
get_backtrace_list/2, 405
get_byte/1, 152
get_byte/2, 151
get_call/3, 284
get_calls/3, 285
get_calls_for_table/2, 286
get_char/1, 150
get_char/2, 150
get_code/1, 151
get_code/2, 151
get_incr_sccs/1, 133
get_incr_sccs/2, 133
get_incr_sccs_with_deps/2, 133
get_incr_sccs_with_deps/3, 133
get_residual/2, 289, 297, 302
get_residual_sccs/3, 120, 241, 297
get_residual_sccs/5, 120, 241, 299
get_returns/2, 287
get_returns/3, 288
get_returns_and_tvs/3, 288
get_returns_for_call/2, 288
get_scc_dumpfile/0, 241
get_scc_dumpfile/1, 296
get_scc_size/3, 378
ground/1, 188
ground_and_acyclic/1, 188
ground_or_cyclic/1, 188
hilog_arg/3, 202
hilog_functor/3, 199
hilog_op/3, 252
INDEX

hilog_symbol/1, 251
if/1, 58
include/1, 57
incr_assert/1, 131
incr_asserta/1, 131
incr_assertz/1, 131
incr_directly_depends/2, 133
incr_invalid_subgoals/1, 135
incr_invalidate_calls/1, 131
incr_is_invalid/1, 135
incr_retract/1, 131
incr_retractall/1, 131
incr_trans_depends/2, 133
incr_trie_intern/2, 132
incr_trie_uninternall/2, 132
index/2, 266
instantiation_error/4, 401
integer/1, 194
intern_term/2, 207
is/2, 176
is_acyclic/1, 210
is_attv/1, 196
is_attv, 415
is_charlist/1, 196
is_charlist/2, 196
is_cyclic/1, 210
is_float, 415
is_functor, 415
is_incremental_subgoal/1, 133
is_int, 415
is_list/1, 195
is_list, 415
is_most_general_term/1, 196
is_nil, 415
is_number_atom/1, 196
is_string, 415
is_var, 416
keysort/2, 191
library_directory/1, 31
listing/0, 248
listing/1, 249
load_dyn/1, 272
load_dyn/2, 273
load_dync/1, 273
load_dync/2, 274
load_forest_log/1, 369
local_datetime/1, 173
log/1, 180
log10/1, 180
log_forest/2, 369
max/2, 178
message/1, 165
message_queue_create/2, 333
message_queue_destroy/1, 333
message_in/1, 165
min/2, 178
misc_error/1, 402
mod/2, 179
module_property/2, 248
mutex_create/1, 336
mutex_destroy/1, 337
mutex_lock/1, 337
mutex_property/2, 339
mutex_trylock/1, 338
mutex_unlock/1, 338
mutex_unlock_all/0, 339
name/2, 212
new_trie/1, 351
nl/0, 155
nl/1, 155
nodebug/0, 361
nonvar/1, 193
nospy/1, 361
not/1, 181
not_exists/1, 182
notrace/0, 358
number/1, 194
number_chars/2, 214
number_codes/2, 211
number_digits/2, 215
numbervars/1, 164  
numbervars/3, 163  
numbervars/4, 163  
one/1, 232  
op/3, 78  
open/3, 140  
open/4, 141  
otherwise/0, 181  
p2c_arity, 416  
p2c_float, 416  
p2c_functor, 416  
p2c_int, 416  
p2c_string, 416  
p2p_arg, 417  
p2p_car, 418  
p2p_cdr, 418  
p2p_new, 417  
p2p_unify, 419  
parsort/4, 192  
path_sysop/2, 174, 175  
path_sysop/3, 174, 175  
peek_byte/1, 153, 154  
peek_char/1, 152  
peek_char/2, 152  
peek_code/1, 153  
peek_code/2, 153  
permission_error/4, 401  
phrase/2, 385  
phrase/3, 385  
pi/0, 180  
predicate_property/2, 246  
print_backtrace/1, 405  
print_incomplete_tables/0, 242, 296, 397  
prompt/2, 253  
proper_hilog/1, 197  
put_char/1, 154  
put_char/2, 154  
put_code/1, 155  
put_code/2, 154  
read/1, 156  
read/2, 156  
read_canonical/1, 156  
read_canonical/2, 157  
read_term/2, 157  
read_term/3, 157  
real/1, 194  
reclaim_uninterned_rn/1, 353  
reg_term, 419  
rem/2, 179  
remove_xsb_hook/1, 356  
repeat/0, 184  
representation_error/3, 401  
resource_error/3, 401  
retract/1, 263  
retractall/1, 264  
round/1, 179  
runtime_loader_flag/2, 427  
see/1, 147  
seeing/1, 148  
seen/0, 148  
set_dcg_style/1, 388  
set_dcg_supported_table/1, 388  
set_forest_logging_for_pred/2, 378  
set_global_compiler_options/1, 46  
set_input/1, 143  
set_output/1, 143  
set_prolog_flag/2, 243  
set_stream_position/2, 145  
setof/3, 223  
shell/1, 171  
shell/2, 172  
shell_to_list/3, 172  
shell_to_list/4, 172  
sign/1, 180  
sin/1, 180  
sort/2, 191  
spy/1, 361  
sqrt/1, 180  
statistics/0, 241, 254
statistics/1, 241, 257
statistics/2, 258
storage_commit/1, 278
storage_delete_fact/3, 277
storage_delete_fact_bt/2, 278
storage_delete_keypair/3, 276
storage_delete_keypair_bt/3, 278
storage_find_fact/2, 277
storage_find_keypair/3, 277
storage_insert_fact/3, 277
storage_insert_fact_bt/2, 278
storage_insert_keypair/4, 276
storage_insert_keypair_bt/4, 277
storage_reclaim_space/1, 278
stream_property/2, 143
string_substitute/4, 219
structure/1, 195
sub_atom/5, 217
subsumes/2, 189
subsumes_chk/2, 189
subsumes_term/2, 189
tab/1, 155
table/1, 82, 130, 280
table_dump/2, 292, 371
table_dump/3, 292, 371
table_once/1, 232
table_state/1, 291
table_state/4, 291
tan/1, 180
tell/1, 148
telling/1, 149
term_depth/2, 206
term_expansion/2, 381, 386
term_size/2, 207
term_to_atom/2, 222
term_to_atom/3, 221
term_to_codes/2, 222
term_to_codes/3, 222
term_variables/2, 206
tfindall/3, 226
thread_cancel/1, 329
thread_create/1, 327
thread_create/2, 327
thread_create/3, 325
thread_detach/1, 328
thread_disable_cancel/0, 331
thread_enable_cancel/0, 331
thread_exit/1, 328
thread_get_message/1, 335
thread_get_message/2, 334
thread_join/2, 327
thread_peek_message/1, 335
thread_peek_message/2, 335
thread_property/2, 331
thread_self/1, 328
thread_send_message/2, 334
thread_signal/2, 330
thread_sleep/1, 332
thread_yield/0, 331
three_valued_scc/1, 378
throw/1, 400
time/1, 260
timed_call/3, 230
timed_call/4, 229
tmpfile_open/1, 147
tnot/1, 98, 182
told/0, 149
tphrase/1, 385
tphrase_set_string/1, 387
tphrase_set_string_auto_abolish/1, 387
tphrase_set_string_keeping_tables/1, 387
trace/0, 358
trace/2, 360
trie_bulk_delete/2, 349
trie_bulk_insert/2, 348
trie_bulk_unify/3, 350
trie_create/2, 344
trie_delete/2, 347
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>trie_drop/1</td>
<td>348</td>
</tr>
<tr>
<td>trie_insert/2</td>
<td>345</td>
</tr>
<tr>
<td>trie_intern/2</td>
<td>351</td>
</tr>
<tr>
<td>trie_intern/5</td>
<td>351</td>
</tr>
<tr>
<td>trie_interned/2</td>
<td>352</td>
</tr>
<tr>
<td>trie_interned/4</td>
<td>352</td>
</tr>
<tr>
<td>trie_property/2</td>
<td>350</td>
</tr>
<tr>
<td>trie_truncate/1</td>
<td>347</td>
</tr>
<tr>
<td>trie_unify/2</td>
<td>346</td>
</tr>
<tr>
<td>trie_unintern2</td>
<td>352</td>
</tr>
<tr>
<td>trimcore/0</td>
<td>254</td>
</tr>
<tr>
<td>true/0</td>
<td>181</td>
</tr>
<tr>
<td>truncate/1</td>
<td>180</td>
</tr>
<tr>
<td>type_error/4</td>
<td>402</td>
</tr>
<tr>
<td>u_not/1</td>
<td>183, 298, 299</td>
</tr>
<tr>
<td>unifiable/3</td>
<td>187</td>
</tr>
<tr>
<td>unify_with_occurs_check/2</td>
<td>186</td>
</tr>
<tr>
<td>unmark_uninterned_nr/2</td>
<td>353</td>
</tr>
<tr>
<td>unnumbervars/3</td>
<td>164</td>
</tr>
<tr>
<td>url_decode/2</td>
<td>150</td>
</tr>
<tr>
<td>url_encode/2</td>
<td>149</td>
</tr>
<tr>
<td>var/1</td>
<td>193</td>
</tr>
<tr>
<td>variant/2</td>
<td>189</td>
</tr>
<tr>
<td>variant_get_residual/2</td>
<td>289, 297</td>
</tr>
<tr>
<td>warning/1</td>
<td>164</td>
</tr>
<tr>
<td>with_mutex/2</td>
<td>336</td>
</tr>
<tr>
<td>word/3</td>
<td>384</td>
</tr>
<tr>
<td>write/1</td>
<td>160</td>
</tr>
<tr>
<td>write/2</td>
<td>160</td>
</tr>
<tr>
<td>write_canonical/1</td>
<td>161</td>
</tr>
<tr>
<td>write_canonical/2</td>
<td>162</td>
</tr>
<tr>
<td>write_prolog/1</td>
<td>162</td>
</tr>
<tr>
<td>write_term/2</td>
<td>158</td>
</tr>
<tr>
<td>write_term/3</td>
<td>159</td>
</tr>
<tr>
<td>writeln/1</td>
<td>162</td>
</tr>
<tr>
<td>writeln/2</td>
<td>162</td>
</tr>
<tr>
<td>writeq/1</td>
<td>161</td>
</tr>
<tr>
<td>writeq/2</td>
<td>161</td>
</tr>
<tr>
<td>xor/2</td>
<td>178</td>
</tr>
<tr>
<td>xsb_assert_hook/1</td>
<td>356</td>
</tr>
<tr>
<td>xsb_backtrace/1</td>
<td>405</td>
</tr>
<tr>
<td>xsb_ccall_thread_create</td>
<td>458</td>
</tr>
<tr>
<td>xsb_close_query</td>
<td>457</td>
</tr>
<tr>
<td>xsb_close</td>
<td>450</td>
</tr>
<tr>
<td>xsb_command_string</td>
<td>451</td>
</tr>
<tr>
<td>xsb_command</td>
<td>451</td>
</tr>
<tr>
<td>xsb_configuration/2</td>
<td>250</td>
</tr>
<tr>
<td>xsb_domain_error()</td>
<td>432</td>
</tr>
<tr>
<td>xsb_error_get_backtrace/2</td>
<td>403</td>
</tr>
<tr>
<td>xsb_error_get_goal/2</td>
<td>403</td>
</tr>
<tr>
<td>xsb_error_get_goalatom/2</td>
<td>403</td>
</tr>
<tr>
<td>xsb_error_get_tid/2</td>
<td>403</td>
</tr>
<tr>
<td>xsb_existence_error()</td>
<td>432</td>
</tr>
<tr>
<td>xsb_exit_hook/1</td>
<td>356</td>
</tr>
<tr>
<td>xsb_get_error_message</td>
<td>457</td>
</tr>
<tr>
<td>xsb_get_error_type</td>
<td>457</td>
</tr>
<tr>
<td>xsb_get_last_answer_string_b</td>
<td>454</td>
</tr>
<tr>
<td>xsb_get_main_thread</td>
<td>458</td>
</tr>
<tr>
<td>xsb_init_string</td>
<td>449</td>
</tr>
<tr>
<td>xsb_init</td>
<td>450</td>
</tr>
<tr>
<td>xsb_instantiation_error()</td>
<td>433</td>
</tr>
<tr>
<td>xsb_make_vars</td>
<td>460</td>
</tr>
<tr>
<td>xsb_misc_error()</td>
<td>433</td>
</tr>
<tr>
<td>xsb_next_string</td>
<td>455, 456</td>
</tr>
<tr>
<td>xsb_next</td>
<td>456</td>
</tr>
<tr>
<td>xsb_permission_error()</td>
<td>433</td>
</tr>
<tr>
<td>xsb_query_restore</td>
<td>426</td>
</tr>
<tr>
<td>xsb_query_save</td>
<td>426</td>
</tr>
<tr>
<td>xsb_query_string_b</td>
<td>453</td>
</tr>
<tr>
<td>xsb_query_string</td>
<td>452</td>
</tr>
<tr>
<td>xsb_query_string</td>
<td>454</td>
</tr>
<tr>
<td>xsb_query</td>
<td>453</td>
</tr>
<tr>
<td>xsb_resource_error()</td>
<td>434</td>
</tr>
<tr>
<td>xsb_retract_hook/1</td>
<td>357</td>
</tr>
<tr>
<td>xsb_set_var_float</td>
<td>461</td>
</tr>
<tr>
<td>xsb_set_var_int</td>
<td>461</td>
</tr>
<tr>
<td>xsb_set_var_string</td>
<td>461</td>
</tr>
<tr>
<td>xsb_thread_context_to_id</td>
<td>458</td>
</tr>
<tr>
<td>xsb_thread_id_to_context</td>
<td>458</td>
</tr>
</tbody>
</table>
xsb_throw(), 434
xsb_type_error(), 434
xsb_var_float, 461
xsb_var_int, 461
xsb_var_string, 461
64-bit architectures, 13, 407, 462

abort
  trace facility, 359
abstraction
  answer, 116
  subgoal, 116
acc, 10
aggregate predicates
  prolog, 223
aliases
  message queues, 332
  mutexes, 336
  streams, 138
    user_error, 138
    user_input, 138
    user_message, 138
    user_output, 138
    user_warning, 138
threads, 325
tries, 344
answer abstraction, 110
answer substitution, 279
attributed variables, 4, 86, 135, 160, 162, 163, 196, 206, 207, 238, 369
backtrackable updates, 276–278
base file name, 22
bounded rationality, 116
byte code
  files
    compiler, 43
canonical format, 42, 161, 273
cc, 10
canonical format, 42, 161, 273
canonical format, 42, 161, 273
character code constants, 68
character sets, 140, 241
compilation
  conditional
    gpp, 47
    Prolog directives, 57
compiler, 43
  cmplib, 43
  directives, 56
  inlines, 64
  invoking, 44
  options, 46
  specialization, 54
compiler options
  mi_warn, 53
  modeinfer, 53
  optimize, 46
  spec_dump, 53
  spec_off, 52
  spec_repr, 52
  ti_dump, 53
  ti_long_names, 53
  unfold_off, 53
  xpp_on, 47
configuration, 8
Constraint Handling Rules, 4
cut, 181, 228, 379, 380, 387
debugger, 358
  ports, 358
declarations
  auto_table, 52, 60, 85
  document_export/1, 29
  document_import/1, 29
export/1, 24
import/1, 33
import/2, 24
index/2, 62

INDEX

local/1, 24
module/2, 24
multifile/2, 43
suppl_table, 52, 61, 85
table as, 88
use_module/2, 24
definite clause grammars, 379
datalog mode, 384
list mode, 383
style, 389
directives
  Compiler, 56
  indexing, 62
  modes, 58
  tabling, 60
dynamic loading of files, 42
dynamic clauses, 266
heap, 254
tables, 309
gcc, 10
GPP, 43, 47
  gpp_include_dir, 47
  gpp_options, 48
  quit_on_error, 52
  xpp_dump/N, 48
  xpp_dump, 48
  xpp_on/N, 48
garbage collection, 38, 39, 237
  atoms, 222
  dynamic clauses, 266
  heap, 254
  tables, 309
garbage collection, 38, 39, 237
  atoms, 222
directives
  Compiler, 56
  indexing, 62
  modes, 58
dynamic loading of files, 42
directory
  command line options, 34
  errors with position, 241
dynamic clauses, 266
heap, 254
tables, 309
garbage collection, 38, 39, 237
  atoms, 222
dynamic clauses, 266
heap, 254
tables, 309
Flora-2, 6, 208
floundering, 182, 183
flash collection, 38, 39, 237
  atoms, 222
  dynamic clauses, 266
  heap, 254
  tables, 309
garbage collection, 38, 39, 237
  atoms, 222
garbage collection, 38, 39, 237
  atoms, 222
directives
  Compiler, 56
  indexing, 62
  modes, 58
dynamic loading of files, 42
modes
  directives, 58
high-level tracing, 358
Incremental Dependency Graph (IDG),
  124, 127, 132, 259, 297, 306
indexing, 260–262, 266, 369
  composite, 268
directives, 62
dynamic predicates, 266
hash-based, 266
multiple-argument, 268
star, 4, 268
transformational, 63
trie-based, 268, 340
inlines
  Compiler, 64
installation into shared directories, 8
Interface
  InterProlog, 11
  ODBC, 11
  Oracle, 11
  SModels, 11
interned terms, 207
invoking the Compiler, 44
ISO
  errors, 397
ISO Compatibility, 65
LATIN-1, 140
LD_LIBRARY_PATH, 427
LIBPATH, 427
load search path, 30
low-level tracing, 363
memory management, 39
message queues, 317
mode analysis
  compiler options, 53
modes
  directives, 58
modules
  compatibility syntax, 24
  name, 24
  XSB syntax, 24
multi-threading, 4, 310–339
mutexes
  user defined, 335
negation
  stable models, 106
  stratified, 96
  unstratified, 100
notational conventions, 7
occurs check, 185, 208, 240
ODBC Interface, 4, 11
options
  command line arguments, 34
  compiler, 46
Oracle Interface, 11
packages, 32
  bootstrap_userpackage/3, 32
  package_configuration/2, 33
  unload_package/1, 33
permanent variables, 404
predicate indicator, 235
preprocessing, 43
Prolog flags, 236
  atom_garbage_collection, 222, 237
  backtrace_on_error, 237, 404
  bounded, 236
  character_set, 241
  clause_garbage_collection, 237
  dcg_style, 237
  debug, 236
  dialect, 236
  double_quotes, 236
  errors_with_position, 241
  exception_action, 176, 241
  exception_pre_action, 241, 242,
        296, 397
  goal, 237
  heap_garbage_collection, 237, 254
  heap_margin, 237
  integer_rounding_function, 236
  max_answer_list_action, 208, 239
  max_answer_list_depth, 208, 239
  max_answer_term_action, 208, 239
  max_answer_term_depth, 208, 239
  max_integer, 236
  max_memory, 40, 240
  max_queue_size, 242
  max_tab_usage, 241
  max_table_answer_action, 121
  max_table_answer_depth, 121
  max_table_subgoal_action, 118,
        208, 239
  max_table_subgoal_depth, 118, 208, 239
  max_threads, 242, 312
  min_integer, 236
  shared_predicates, 242
  table_gc_action, 237, 302, 304–
        306
  thread_complsize, 242, 326
  thread_detached, 242, 326
  thread_glsize, 242, 326
  thread_pdlsize, 242, 326
  thread_tcpsize, 242, 326
  tracing, 238
  unify_with_occurs_check, 185, 208,
        240
  unknown, 29, 236
  version_data, 236
  warning_action, 164, 238
  write_attributes, 160, 238
  write_depth, 238
Prolog-commons, 236
Prologs
SWI, 260
YAP, 260
radial restraint, 120, 299
residual dependency graph, 120, 132, 297
residual program, 105, 290, 297
scheduling strategy, 12
sets, bags, 223
shared_predicates, 38, 243, 313
Silk, 208
skeleton, 280
SModels Interface, 11
source file designator, 22
specialization
  Compiler, 54
compiler options, 52
stable models, 104
stacks
default sizes, 34
expanding, 34
standard predicates, 33, 47, 53
state of the system, 235
streams, 137
  STDDBG, 139
  STDERR, 139, 164, 237, 402
  STDFDBK, 139, 164
  STDIN, 139
  STDERR, 139, 165
  OUT, 139, 164
  STDERR, 139, 164
  system, 139, 147
strongly connected components (SCCs), 90, 296, 395, 397
strongly connected components (SCCs), 368
substitution factor, 279
syntax
  atoms, 69
  escaped characters, 69
  floats, 69
  integers, 67
    binary, 67
    hexadecimal, 67
    octal, 67
  system, state of, 235
tabled subgoals
  complete, 98
  incomplete, 98, 259
tabling
  abolishes, 304, 306–309
    incremental, 306, 307
    multi-threading, 304–306, 308
    transitive vs. single, 304, 305
  and exceptions, 242, 296, 395
  answer completion, 103
  answer subsumption, 85, 107–109, 111, 135, 246, 280
  automatic, 52
  call subsumption, 34, 37, 85, 86, 88, 135, 246, 254, 280, 283, 284
    interaction with meta-logical predicates, 94
  call variance, 85, 86, 135, 246, 254, 280, 283, 284
  compiler options, 52
  complete evaluation, 97, 372
  conditional answers, 100, 302, 304
  consumer, 82
cuts, 92
declarations, 130
directives, 60, 280
dynamic predicates, 280
tabled subgoals, 98, 368, 372
  early completion of subgoals, 98, 368, 372
  incremental, 121, 130, 135, 246, 254, 280, 304, 345
  invalid subgoals, 124
interned terms, 88
multi-threaded, 303, 304
negation, 96
opaque, 130, 280
private, 280
producer, generator, 82
scheduling strategies, 89
shared, 37, 280
similarity measures, 85
strategy selection, 280
supplemental, 52
table deletion, 300
table inspection, 283
Tck/Tk, 11
term depth, 239
definition, 206
term indicator, 235
term size
definition, 207
termination, 61, 84, 85
answer subsumption, 108
radial restraint, 116, 239, 298, 299
subgoal abstraction, 116, 239
terms
comparison of, 185
cyclic, 185, 188, 206–208, 240, 345,
349, 464
unification of, 185
thread
thread status, 316
valid, 315
trace
logging, 360
options, 359
tracing, 358–369
low-level, 363
Prolog Programs, 358
transaction logic, 275
tries
and incremental tabling, 126
asserted, 266
indexing, 268
depth limit, 208
interned, 340–354
Unicode
UTF-8, 69, 140, 150, 155, 210, 214,
222
unification factoring
compiler options, 53
VarString, 458
view consistency, 124, 302
well-founded semantics, 103
XASP, 6, 11
xsbdoc, 6, 29
xsbrc.P initialization file, 31
Index of Predicates Standard in XSB

\+/1, 181
\=/2, 187
\==/2, 187
^/2, 179
!/0, 181
'\vee'/2, 178
'\wedge'/2, 178
'\lhd'/2, 177
'\rhd'/2, 177
'\llhd'/2, 177
'\rrhd'/2, 178
*'2, 177
**/2, 180
*/2, 177
+/2, 177
-/1, 177
-/2, 177
=.../2, 203
=../2, 186
==/2, 187
?=2, 187
\theta<2, 187
\theta/=2, 188
\theta=<2, 188
\theta>/2, 188
\theta>=2, 188
ISO, 194, 236, 253
[]/1 (consult), 41
$trace/0, 363
^/2, 227
^=../2, 205
'C'/3, 387
abolish/1, 264
abolish_all_private_tables/0, 308
abolish_all_shared_tables/0, 308
abolish_all_tables/0, 307
abolish_module_tables/1, 309
abolish_nonincremental_tables/0, 307
abolish_table_pred/1, 304
abolish_table_pred/2, 305
abolish_table_subgoal/1, 306
abolish_table_subgoal/2, 306
abolish_table_subgoals/1, 305
abolish_table_subgoals/2, 306
acos/1, 180
acyclic_term/1, 210
arg/3, 201
arg0/3, 202
asin/1, 180
assert/1, 262
assert/3, 262
asserta/1, 261
assertz/1, 262
at_end_of_stream/0, 145
at_end_of_stream/1, 145
atan/1, 180
atan/2, 180
atan2/2, 180
<table>
<thead>
<tr>
<th>Predicate</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>atom/1, 193</td>
<td></td>
</tr>
<tr>
<td>atom_chars/2, 214</td>
<td></td>
</tr>
<tr>
<td>atom_codes/2, 210</td>
<td></td>
</tr>
<tr>
<td>atom_concat/3, 216</td>
<td></td>
</tr>
<tr>
<td>atom_length/2, 215</td>
<td></td>
</tr>
<tr>
<td>atomic/1, 194</td>
<td></td>
</tr>
<tr>
<td>bagof/3, 224</td>
<td></td>
</tr>
<tr>
<td>break/0, 252</td>
<td></td>
</tr>
<tr>
<td>call/1, 228</td>
<td></td>
</tr>
<tr>
<td>call/[2,10], 229</td>
<td></td>
</tr>
<tr>
<td>call_cleanup/2, 232</td>
<td></td>
</tr>
<tr>
<td>call_tv/2, 229</td>
<td></td>
</tr>
<tr>
<td>callable/1, 197</td>
<td></td>
</tr>
<tr>
<td>catch/3, 402</td>
<td></td>
</tr>
<tr>
<td>ceiling/1, 178</td>
<td></td>
</tr>
<tr>
<td>char_code/2, 215</td>
<td></td>
</tr>
<tr>
<td>clause/2, 265</td>
<td></td>
</tr>
<tr>
<td>close/1, 143</td>
<td></td>
</tr>
<tr>
<td>close/2, 142</td>
<td></td>
</tr>
<tr>
<td>compare/3, 188</td>
<td></td>
</tr>
<tr>
<td>compile/1, 41, 44</td>
<td></td>
</tr>
<tr>
<td>compile/2, 41, 44</td>
<td></td>
</tr>
<tr>
<td>compound/1, 195</td>
<td></td>
</tr>
<tr>
<td>copy_term/2, 206</td>
<td></td>
</tr>
<tr>
<td>cos/1, 180</td>
<td></td>
</tr>
<tr>
<td>current_atom/1, 246</td>
<td></td>
</tr>
<tr>
<td>current_functor/1, 245</td>
<td></td>
</tr>
<tr>
<td>current_index/2, 245</td>
<td></td>
</tr>
<tr>
<td>current_input/1, 236</td>
<td></td>
</tr>
<tr>
<td>current_module/1, 244</td>
<td></td>
</tr>
<tr>
<td>current_module/2, 245</td>
<td></td>
</tr>
<tr>
<td>current_op/3, 251</td>
<td></td>
</tr>
<tr>
<td>current_predicate/1, 243</td>
<td></td>
</tr>
<tr>
<td>current_prolog_flag/2, 236</td>
<td></td>
</tr>
<tr>
<td>debug/0, 361</td>
<td></td>
</tr>
<tr>
<td>debug_ctl/2, 361</td>
<td></td>
</tr>
<tr>
<td>debugging/0, 361</td>
<td></td>
</tr>
<tr>
<td>default_user_error_handler/1, 402</td>
<td></td>
</tr>
<tr>
<td>delete_returns/2, 309</td>
<td></td>
</tr>
<tr>
<td>div/2, 177</td>
<td></td>
</tr>
<tr>
<td>dynamic/1, 130, 269</td>
<td></td>
</tr>
<tr>
<td>e/0, 180</td>
<td></td>
</tr>
<tr>
<td>else/0, 58</td>
<td></td>
</tr>
<tr>
<td>elseif/1, 58</td>
<td></td>
</tr>
<tr>
<td>endif/0, 58</td>
<td></td>
</tr>
<tr>
<td>ensure_loaded/1, 42</td>
<td></td>
</tr>
<tr>
<td>ensure_loaded/2, 275</td>
<td></td>
</tr>
<tr>
<td>epsilon/0, 180</td>
<td></td>
</tr>
<tr>
<td>expand_term/2, 385</td>
<td></td>
</tr>
<tr>
<td>fail/0, 181</td>
<td></td>
</tr>
<tr>
<td>fail_if/1, 181</td>
<td></td>
</tr>
<tr>
<td>false/0, 181</td>
<td></td>
</tr>
<tr>
<td>file_clone/3, 146</td>
<td></td>
</tr>
<tr>
<td>file_exists/1, 149</td>
<td></td>
</tr>
<tr>
<td>file_read_line_atom/1, 168</td>
<td></td>
</tr>
<tr>
<td>file_read_line_atom/2, 169</td>
<td></td>
</tr>
<tr>
<td>file_read_line_list/1, 168</td>
<td></td>
</tr>
<tr>
<td>file_read_line_list/2, 168</td>
<td></td>
</tr>
<tr>
<td>file_reopen/3, 146</td>
<td></td>
</tr>
<tr>
<td>findall/3, 225</td>
<td></td>
</tr>
<tr>
<td>findall/4, 225</td>
<td></td>
</tr>
<tr>
<td>float/1, 178</td>
<td></td>
</tr>
<tr>
<td>floor/1, 178</td>
<td></td>
</tr>
<tr>
<td>flush_output/0, 144</td>
<td></td>
</tr>
<tr>
<td>flush_output/1, 144</td>
<td></td>
</tr>
<tr>
<td>fmt_read/3, 165</td>
<td></td>
</tr>
<tr>
<td>fmt_read/4, 165</td>
<td></td>
</tr>
<tr>
<td>fmt_write/2, 166</td>
<td></td>
</tr>
<tr>
<td>fmt_write/3, 166</td>
<td></td>
</tr>
<tr>
<td>fmt_write_string/3, 168</td>
<td></td>
</tr>
<tr>
<td>forall/2, 232</td>
<td></td>
</tr>
<tr>
<td>functor/3, 197</td>
<td></td>
</tr>
<tr>
<td>gc_atoms/0, 222</td>
<td></td>
</tr>
<tr>
<td>gc_dynamic/1, 266</td>
<td></td>
</tr>
<tr>
<td>gc_heap/0, 254</td>
<td></td>
</tr>
<tr>
<td>gc_tables/1, 309</td>
<td></td>
</tr>
<tr>
<td>get_byte/1, 152</td>
<td></td>
</tr>
<tr>
<td>get_byte/2, 151</td>
<td></td>
</tr>
<tr>
<td>get_call/3, 284</td>
<td></td>
</tr>
<tr>
<td>get_calls/3, 285</td>
<td></td>
</tr>
</tbody>
</table>
get_calls_for_table/2, 286
get_char/1, 150
get_char/2, 150
get_code/1, 151
get_code/2, 151
get_residual/2, 289
get_returns/2, 287
get_returns/3, 288
get_returns_and_tvs/3, 288
get_returns_for_call/2, 288
ground/1, 188
ground_and_acyclic/1, 188
ground_or_cyclic/1, 188
hilog_arg/3, 202
hilog_functor/3, 199
hilog_op/3, 252
hilog_symbol/1, 251
if/1, 58
include/1, 57
index/2, 266
integer/1, 194
is/2, 176
is_acyclic/1, 210
is_attv/1, 196
is_charlist/1, 196
is_charlist/2, 196
is_cyclic/1, 210
is_list/1, 195
is_most_general_term/1, 196
is_number_atom/1, 196
keysort/2, 191
library_directory/1, 31
listing/0, 248
listing/1, 249
load_dyn/1, 272
load_dyn/2, 273
load_dync/1, 273
load_dync/2, 274
log/1, 180
log10/1, 180
max/2, 178
message_queue_create/2, 333
message_queue_destroy/1, 333
min/2, 178
mod/2, 179
module_property/2, 248
mutex_create/1, 336
mutex_destroy/1, 337
mutex_lock/1, 337
mutex_property/2, 339
mutex_trylock/1, 338
mutex_unlock/1, 338
mutex_unlock_all/0, 339
name/2, 212
nl/0, 155
nl/1, 155
nodebug/0, 361
nonvar/1, 193
nospy/1, 361
not/1, 181
not_exists/1, 182
notrace/0, 358
number/1, 194
number_chars/2, 214
number_codes/2, 211
number_digits/2, 215
once/1, 232
op/3, 78
open/3, 140
open/4, 141
otherwise/0, 181
path_sysop/2, 174, 175
path_sysop/3, 174, 175
peek_byte/1, 153, 154
peek_char/1, 152
peek_char/2, 152
peek_code/1, 153
peek_code/2, 153
phrase/2, 385
phrase/3, 385
pi/0, 180
predicate_property/2, 246
prompt/2, 253
proper_hilog/1, 197
put_char/1, 154
put_char/2, 154
put_code/1, 155
put_code/2, 154
read/1, 156
read/2, 156
read_canonical/1, 156
read_canonical/2, 157
read_term/2, 157
read_term/3, 157
real/1, 194
rem/2, 179
repeat/0, 184
retract/1, 263
retractall/1, 264
round/1, 179
see/1, 147
seeing/1, 148
seen/0, 148
set_dcg_style/1, 388
set_global_compiler_options/1, 46
set_input/1, 143
set_output/1, 143
set_prolog_flag/2, 243
set_stream_position/2, 145
setof/3, 223
shell/1, 171
shell/2, 172
shell_to_list/3, 172
shell_to_list/4, 172
sign/1, 180
sin/1, 180
sort/2, 191
spy/1, 361
sqrt/1, 180
statistics/0, 254
statistics/1, 257
statistics/2, 258
storage_commit/1, 278
storage_delete_fact/3, 277
storage_delete_fact_bt/2, 278
storage_delete_keypair/3, 276
storage_delete_keypair_bt/3, 278
storage_find_fact/2, 277
storage_find_keypair/3, 277
storage_insert_fact/3, 277
storage_insert_fact_bt/2, 278
storage_insert_keypair/4, 276
storage_insert_keypair_bt/4, 277
storage_reclaim_space/1, 278
stream_property/2, 143
structure/1, 195
sub_atom/5 , 217
subsumes_term/2, 189
tab/1, 155
table/1, 82, 130, 280
table/1, 82, 130, 280
table_once/1, 232
table_state/1, 291
table_state/4, 291
tan/1, 180
tell/1, 148
telling/1, 149
term_depth/2, 206
term_expansion/2, 381, 386
term_size/2, 207
term_variables/2, 206
tfindall/3, 226
thread_cancel/1, 329
thread_create/1, 327
thread_create/2, 327
thread_create/3, 325
thread_detach/1, 328
thread_exit/1, 328
thread_get_message/1, 335
thread_get_message/2, 334
thread_join/2, 327
INDEX OF PREDICATES STANDARD IN XSB

thread_peek_message/1, 335  thread_peek_message/2, 335  xsb_configuration/2, 250
thread_property/2, 331
thread_self/1, 328
thread_send_message/2, 334
thread_signal/2, 330
thread_sleep/1, 332
thread_yield/0, 331
throw/1, 400
time/1, 260
timed_call/3, 230
timed_call/4, 229
tmpfile_open/1, 147
tnot/1, 182
told/0, 149
tphrase/1, 385
tphrase_set_string/1, 387
trace/0, 358
trace/2, 360
true/0, 181
truncate/1, 180
unify_with_occurs_check/2, 186
url_decode/2, 150
url_encode/2, 149
var/1, 193
variant_get_residual/2, 289
with_mutex/2, 336
word/3, 384
write/1, 160
write/2, 160
write_canonical/1, 161
write_canonical/2, 162
write_prolog/1, 162
write_term/2, 158
write_term/3, 159
writeln/1, 162
writeln/2, 162
writeq/1, 161
writeq/2, 161
xor/2, 178