$\mathcal{E}$RGO Reasoner User's Manual

Version 1.2 (Solon)

Edited by

Michael Kifer
Coherent Knowledge

May 2017

Portions Copyright © 2013–2017 Coherent Knowledge

Contents

1 Introduction **1**

2 Installation **1**

3 Running $\mathcal{E}$RGO **2**

4 $\mathcal{E}$RGO Shell Commands **3**

5 F-logic and $\mathcal{E}$RGO by Example **9**

6 Main Syntactic Elements of $\mathcal{E}$RGO **9**

7 Basic $\mathcal{E}$RGO Syntax **11**

7.1 F-logic Vocabulary 11
7.2 Symbols, Strings, and Comments 16
7.3 Operators 18
7.4 Logical Expressions 20
7.5 Arithmetic (and related) Expressions 21
7.6 Quasi-constants and Quasi-variables 28
7.7 Synonyms 28
7.8 Reserved Symbols 29

8 Path Expressions **29**

9 Set Notation **31**

10 Typed Variables **32**

11 Truth Values and Object Values **33**

12 Boolean Methods **35**

Portions Copyright © 2013–2017 Coherent Knowledge

12.1 Boolean Signatures 36

13 Skolem Symbols **36**

14 Testing Meta-properties of Symbols and Variables **40**

15 Multifile Knowledge Bases **44**

15.1 $\mathcal{E}$RGO Modules 44
15.2 Calling Methods and Predicates Defined in User Modules 45
15.3 Finding the Current Module Name 47
15.4 Finding the Module That Invoked A Rule 48
15.5 Loading Files into User Modules 48
15.6 Adding Rule Bases to Existing Modules 51
15.7 Calling Prolog from $\mathcal{E}$RGO 52
15.8 Calling $\mathcal{E}$RGO from Prolog 54
15.8.1 Importing $\mathcal{E}$RGO Predicates into Prolog 54
15.8.2 Passing Arbitrary Queries to $\mathcal{E}$RGO 56
15.9 $\mathcal{E}$RGO System Modules 59
15.10 Including Files into $\mathcal{E}$RGO Knowledge Bases 59
15.11 More on Variables as Module Specifications 60
15.12 Module Encapsulation 60
15.13 Importing Modules 63
15.14 Persistent Modules 65
15.15 Fast Loading Vast Amounts of Data 65
15.16 Truth about Relative File Names and Current Directories 66

16 HiLog and Meta-programming **68**

16.1 Meta-programming, Meta-unification 70
16.2 Reification 72
16.3 Meta-decomposition 75

16.4 Passing Parameters between $\mathcal{E}$RGO and Prolog 77

17 Negation **79**

17.1 Default Negation `\naf` vs. Prolog Negation `\+` 79

17.2 Default Negation for Non-ground Subgoals 80

17.3 Non-ground Subgoals Under `\+` 82

17.4 True vs. Undefined Formulas 82

17.5 Explicit Negation 83

18 General Formulas in Rule Bodies **85**

18.1 Quantification of Free and Anonymous Variables 87

18.2 The Difference Between `~~>`, `==>` (or, `<~~`, `<==`), and `:-` 88

19 General Formulas in Rule Heads **89**

19.1 Omniformity of `<==`, `==>`, and `<===>` 90

19.2 Omniformity and Builtins 92

20 Inheritance of Default Properties and Types **93**

20.1 Introduction to Inheritance 94

20.2 Monotonic Behavioral Inheritance 97

20.3 Non-monotonic Behavioral Inheritance 98

20.4 Inheritance of Negative Information 102

20.4.1 Negative Monotonic Behavioral Inheritance 103

20.4.2 Negative Structural Inheritance 103

20.4.3 Negative Non-monotonic Behavioral Non-Inheritance 104

20.5 Code Inheritance 106

21 Custom Module Semantics **108**

21.1 Equality Maintenance 109

21.2 Choosing a Semantics for Inheritance 112

Portions Copyright © 2013–2017 Coherent Knowledge

21.3 Choosing a Semantics for the Subclass Relationship . . . 112
21.4 Choosing a Semantics for Tabling . . . 113
21.5 Class Expressions . . . 113
21.6 Ad Hoc Custom Semantics . . . 115
21.7 Querying the Module Semantics . . . 115

22 $\mathcal{E}$RGO and Tabling **116**
22.1 Tabling in a Nutshell . . . 116
22.2 Transactional Methods . . . 119
22.3 Re-examining Query Results: The peek{...} primitive . . . 120
22.4 Operational Semantics of $\mathcal{E}$RGO . . . 121
22.5 Tabling and Performance . . . 122
22.6 Cuts . . . 123

23 User Defined Functions **124**
23.1 Syntax . . . 124
23.2 Higher-order UDFs . . . 129
23.3 User-defined Functions and Multifile Modules . . . 129
23.4 Semantics . . . 130
23.5 Representing equality with user defined functions . . . 131

24 Controlling Context of Symbols **133**

25 Updating the Knowledge Base **135**
25.1 Non-transactional (Non-logical) Updates . . . 136
25.2 Transactional (Logical) Updates . . . 141
25.3 Updates and Tabling . . . 142
25.4 Updates and Meta-programming . . . 146
25.5 Updates and Negation . . . 147
25.6 Stealth Updates . . . 148

Portions Copyright © 2013–2017 Coherent Knowledge

25.7 Counter Support 149

26 Insertion and Deletion of Rules **150**

26.1 Creation of a New Module and Module Erasure at Run-time 150

26.2 Insertion of Rules 150

26.3 Deletion of Rules 151

27 Querying the Rule Base **153**

27.1 Basic Querying of Rules via `clause{...}` 153

27.2 Querying Rules via the Rule Structure Database 156

28 Aggregate Operations **160**

28.1 Syntax of Aggregate Operators 160

28.2 Evaluation of Aggregates 161

28.3 Notes on Scoping for Variables in Aggregate Operators 163

28.4 Aggregation and Sorted Results 164

28.5 Aggregation and Set-Valued Methods 165

29 Control Flow Statements **166**

29.1 If-Then-Else 166

29.2 Loops 167

29.2.1 The while-do and do-until Loops 167

29.2.2 The while-loop and loop-until Loops 169

30 Constraint Solving **170**

31 Low-level Predicates **171**

32 Sensors: Predicates with Restricted Invocation Patterns **173**

33 Rearranging the Order of Subgoals at Run Time **176**

Portions Copyright © 2013–2017 Coherent Knowledge

34 Rule Ids and Meta-information about Rules **180**

34.1 The Current Rule Id Quasi-constant . . . 182

34.2 Enabling and Disabling Rules . . . 182

34.3 Changing the Defeasibility at Run Time . . . 184

34.4 Querying Rule Descriptors . . . 185

34.5 Reserved Properties in Rule Descriptors . . . 186

35 Latent Queries **186**

36 Integrity Constraints **188**

37 Alerts **193**

38 Defeasible Reasoning **196**

38.1 Specifying an Argumentation Theory to Use . . . 198

38.2 Rule Overriding and Conflicts . . . 199

38.3 Cancellation of Rules . . . 200

38.4 Changing the Default Defeasibility Status . . . 202

38.5 Supported Argumentation Theories . . . 204

38.5.1 The Cautious, Original, and Strong Courteous Argumentation Theories . . . 204

38.5.2 Courteous Logic with Exclusion Constraints . . . 205

38.6 Defeasible Rules Must Be Purely Logical . . . 205

38.7 Some Useful and Common Definitions . . . 207

38.8 Debugging Defeasible Knowledge Bases . . . 209

39 Primitive Data Types **210**

39.1 $\mathcal{E}$RGO Symbols . . . 212

39.2 The \iri Data Type . . . 213

39.3 The Primitive Type \dateTime . . . 218

39.4 The Primitive Type \date . . . 221

Portions Copyright © 2013–2017 Coherent Knowledge

39.5 The Primitive Type \time . . . 223
39.6 The Primitive Type \duration . . . 225
39.7 The Primitive Type \boolean . . . 226
39.8 The Primitive Type \double . . . 227
39.9 The Primitive Type \long . . . 228
39.10 The Primitive Types \decimal, \integer, and \short . . . 229
39.11 The Primitive Type \string . . . 229
39.12 The Primitive Type \list . . . 231
39.13 Character Lists . . . 234
39.14 Special Classes for Callable Literals . . . 235
39.15 User-defined Types . . . 235
39.16 Regular Expression Matching . . . 236

40 Cardinality Constraints **238**

41 Exception Handling **240**

42 The Compile-time Preprocessor **241**

43 Debugging User Knowledge bases **242**

43.1 Checking for Undefined Methods and Predicates . . . 242
43.2 Type Checking . . . 245
43.3 Checking Cardinality of Methods . . . 248
43.4 Logical Assertions that Depend on Transactional and Non-logical Features . . . 251
43.5 Explaining Answers . . . 252
43.6 Examining Tables . . . 252
43.7 Examining Incomplete Tables . . . 254
43.8 Non-termination Analysis . . . 255
43.9 Runtime Inspection of Computation . . . 260
43.10 Continuous Runtime Monitoring . . . 262

Portions Copyright © 2013–2017 Coherent Knowledge

43.11 Tracing Tabled Calls via Forest Logging . . . 263
43.12 Controlling Subgoal and Answer Size, Timeouts, Unification Mode . . . 266
43.12.1 Timeouts . . . 267
43.12.2 Subgoal Size Control . . . 268
43.12.3 Answer Size Control . . . 269
43.12.4 Controlling the Number of Active Goals . . . 270
43.12.5 Memory Usage Limit . . . 270
43.12.6 Unification Mode . . . 270

44 Considerations for Improving Performance of Queries 271

45 Compiler Directives 273
45.1 Executable vs. Compile-time Directives . . . 273
45.2 Miscellaneous Compiler Options . . . 275

46 $\mathcal{E}$RGO System Modules 276
46.1 Input and Output . . . 276
46.2 Storage Control . . . 282
46.3 System Control . . . 283
46.4 Type and Cardinality Checking . . . 284
46.5 Data Types . . . 284
46.6 Reading and Compiling Input Terms . . . 284
46.7 Displaying $\mathcal{E}$RGO Terms and Goals . . . 286

47 ErgoText 287
47.1 ErgoText Knowledge Bases . . . 288
47.2 ErgoText Template Definitions . . . 290
47.3 ErgoText Queries and Command Line . . . 292
47.4 ErgoText and $\mathcal{E}$RGO Modules . . . 293
47.5 Debugging ErgoText Templates . . . 294

47.6 ErgoText and Error Reporting . . . 296
47.7 ErgoText and Text Generation . . . 297

48 Unicode and Character Encodings **298**
48.1 What Is a Character Encoding? . . . 298
48.2 Character Encodings in $\mathcal{E}$RGO . . . 299
48.3 Specifying Encodings in $\mathcal{E}$RGO . . . 299

49 Notes on Style and Common Pitfalls **300**
49.1 Facts are Unordered . . . 300
49.2 Testing for Class Membership . . . 300
49.3 Composite Frames in Rule Heads . . . 301

50 Miscellaneous Features **303**
50.1 Suppression of Banners . . . 303
50.2 Production and Development Compilation Modes . . . 303

51 Useful XSB Predicates Without a Counterpart in $\mathcal{E}$RGO **304**
51.1 Time-related Predicates . . . 304
51.2 Hashing . . . 305
51.3 Input/Output . . . 305
51.4 Meta-programming . . . 305

52 Bugs in Prolog and $\mathcal{E}$RGO: How to Report **305**

53 The Expert Mode **308**

Appendices **310**

A A BNF-style Grammar for $\mathcal{E}$RGO **310**

B The $\mathcal{E}$RGO Debugger **314**

1 Introduction

The Coherent $\mathcal{E}$RGO reasoner is a sophisticated object-based knowledge representation and reasoning platform. It is based on decades of research into logic rules systems and it presents a unified language of F-logic [9], HiLog [4], Transaction Logic [2, 1], and defeasible reasoning [15].

$\mathcal{E}$RGO is based on the open-source $\mathcal{F}$LORA-*2* inference system,[1] also known as $\mathcal{E}$RGO^{Lite}, but is much more scalable and it extends $\mathcal{E}$RGO^{Lite} in numerous ways that are crucial for enterprise use. These extensions and enhancements are proprietary to Coherent Knowledge Systems.

The present document is based on the $\mathcal{F}$LORA-*2* manual[2] but is adapted for Coherent needs and expanded with $\mathcal{E}$RGO-specific information.

2 Installation

The $\mathcal{E}$RGO reasoner is part of the Coherent $\mathcal{E}$RGO Suite and comes bundled with $\mathcal{E}$RGO Studio and various $\mathcal{E}$RGO connectors (e.g., to OWL). To install $\mathcal{E}$RGO Suite on Windows, download the $\mathcal{E}$RGO installer, `ergo_customer.exe`, and then click through the installation process. Note that if you choose to install $\mathcal{E}$RGO in `Program Files` or some other system directory, you must use an account with administrative privileges not just during the installation but also when using $\mathcal{E}$RGO afterwards. It is therefore generally recommended to install $\mathcal{E}$RGO in a user-owned folder.

For Linux and Mac, the installer is `ergosuite.run`; it is a self-extracting archive, which you can simply execute as

```
sh ergosuite.run
```

and $\mathcal{E}$RGO Suite will be installed in a subdirectory called `Coherent`.

To start $\mathcal{E}$RGO via the UI, simply double-click on the Coherent $\mathcal{E}$RGO icon that should appear on your desktop after the installation. If things go wrong on startup, the studio should detect the problem and present an error dialog. If your email client is configured properly, it will also even draft an email bug report and offer you to send it. If the startup was successful, but you encounter a problem later, use the Studio's bug reporter. It is found in the listener window and is accessible via the menu `Tools/Send Bug Report`.

If for some reason you prefer to start the $\mathcal{E}$RGO reasoner as a standalone engine on command line, locate the folder where the reasoner is installed and type `runergo` there. See Section 3 for more details.

[1] http://flora.sourceforge.net
[2] http://flora.sourceforge.net/docs/floraManual.pdf

Portions Copyright © 2013–2017 Coherent Knowledge

3 Running $\mathcal{E}$RGO

$\mathcal{E}$RGO is fully integrated into the underlying Prolog engine, including its module system. In particular, $\mathcal{E}$RGO modules can invoke predicates defined in other Prolog modules, and Prolog modules can query the objects defined in $\mathcal{E}$RGO modules. At present, XSB is the only Prolog platform where $\mathcal{E}$RGO can run, because it heavily relies on tabling and the well-founded semantics for negation, both of which are available only in XSB.

The easiest way to get a feel of the system is to start $\mathcal{E}$RGO shell and begin to enter queries interactively. The simplest way to do this is to use the shell script

```
.../ergo/runergo
```

where "..." is the directory where $\mathcal{E}$RGO is downloaded. For instance,

```
~/ENGINEDIR/ergo/runergo
```

At this point, $\mathcal{E}$RGO takes over and F-logic syntax becomes the norm. To get back to the Prolog command loop, type `Control-D` (Unix) or Control-Z (Windows), or

```
ergo> \end.
```

If you are using the $\mathcal{E}$RGO shell frequently, it pays to define an alias, say (in Bash):

```
alias ergo='~/ENGINEDIR/ergo/runergo'
```

$\mathcal{E}$RGO can then be invoked directly from the shell prompt by typing `ergo`. It is even possible to tell $\mathcal{E}$RGO to execute commands on start-up. For instance, typing

```
ergo -e "\help."
```

in the command window of your operating system will cause the system to execute the help command right after after the initialization. Then the usual $\mathcal{E}$RGO shell prompt is displayed.

$\mathcal{E}$RGO comes with a number of demos that live in

```
.../ergo/demos/
```

The demos can be run with the command `demo{demo-filename}`. at the $\mathcal{E}$RGO prompt, *e.g.*,

```
ergo>  demo{flogic_basics}.
```

There is no need to change to the demo directory, as `demo{...}` knows where to find these examples.

The initialization file. When ℰRGO starts up, it first executes the commands found in the initialization file, if it is specified and exists. The initialization file is specified as a value of the `ERGO_RC_FILE` environment variable of the operating system in use. If this variable is not set or if the value of that variable is not a readable plain file, the initialization file is ignored. The commands in the initialization file can be any kind of ℰRGO queries or commands. They must be specified exactly as they would be written in the interactive ℰRGO shell, i.e., without the query prefix "`?-`" and they must be terminated with the period. For instance, if `ERGO_RC_FILE` is set to `~/test/myrc` (note: the `.ergo` suffix is not required) and that file contains

```
writeln('Welcome!')@\plg.
insert{foo(bar)}.
```

then the message "Welcome!" will be printed and the fact `foo(bar)` will be inserted.

There is one restrictions on the initialization file: a comment cannot be the last statement.

Also note that the command-line option `-e` mentioned earlier is executed *after* the initialization file commands.

4 ℰRGO Shell Commands

Loading knowledge bases from files. The most common shell commands you probably need are the commands for loading and compiling knowledge bases:

```
ergo>   [myfile].        // e.g., ['c:/My Documents/data'].
ergo>   [url(myurl)].    // e.g., [url('http://example.com/data')].
```

or

```
ergo>  load{myfile}.     // e.g., load{'/home/me/proj/kb'}.
ergo>  load{url(myurl)}.
```

Here *myfile* or *myurl* are a file names (respectively, a URL) that are assumed to be pointing to an ℰRGO knowledge base or a Prolog program. Both *myfile* and URL must be Prolog atoms. If they contain non-alphanumeric characters (as in the examples above) then they must be single-quoted (as usual for Prolog atoms). A URL is expected when the argument has the form `url(myurl)`.[3] The file can be relative to the directory in which ℰRGO was started. For instance, if that directory has the file `foo.ergo` then one can simply type `[foo]` instead of `['/home/me/foo.ergo']` or even

[3] For this to work, the XSB package `curl` must be configured as described in the XSB manual, volume 2.

['/home/me/foo']. Note that our first example file, 'c:/My Documents/data', is a Windows file name, except that it uses *forward* slashes (like in URLs), which is preferred to backward slashes. Backward slashes can also be used in Windows, but they must be doubled: 'c:\\My Documents\data'. On Unix-based systems, such as Linux and Mac, only forward slashes can be used.

If *myfile*.ergo exists, it is assumed to be a knowledge base written for ℰRGO. The system will compile the knowledge base, if necessary, and then load it. The compilation process is two-stage: first, the knowledge base is compiled into a Prolog program (one or more files with extensions .pl, .fdb, and others) and then into an executable byte-code, which has the extension .xwam. For instance,

```
ergo>  load{url('http://example.com/test1')}.
ergo>  [url('http://example.com/test2')].
```

will compile (if necessary) and load the ℰRGO files test1.ergo and test2.ergo found at the Web site http://example.com/.

If there is no *myfile*.ergo file, the file is assumed to contain a Prolog program and the system will look for the file named *myfile*.P. This file then is compiled into *myfile*.xwam and loaded. Note that in this case the program is loaded into a *Prolog module* of ℰRGO and, therefore, calls to the predicates defined in that program must use the appropriate module attribution — see Section 15.1 for the details about the module system in ℰRGO.

For backward compatibility with Flora-2, ℰRGO will also try to compile and load files with the .flr extension, such as program.flr. However, if there are both program.ergo and program.flr then program.flr is ignored unless the .flr extension is *explicitly* mentioned. For instance, if the command load{'program.flr'} is executed, then program.flr will be taken even if program.ergo exists. On the other hand, load{program} (without an explicit extension) will try to load program.ergo first. If this file does not exist, program.flr will be tried. If that file is not found, program.P will be attempted followed by program.xwam.

By default, all ℰRGO knowledge bases are loaded into the module called main, but you can also load into other modules using the following command:

```
ergo>   [myfile>>modulename].
ergo>   [url(myurl)>>modulename].
```

Understanding ℰRGO modules is very important in order to be able to take full advantage of the system; we will discuss the module system of ℰRGO in Section 15.1. Once the knowledge base is loaded, you can pose queries and invoke methods for the objects defined in that knowledge base.

All the loading commands that apply to files also apply to URLs, so in the future we will be giving examples for files only.

There is an important special case of the `load{...}` and `[...]` commands when the file name is missing. In that case, $\mathcal{E}$RGO creates a scratchpad file and starts reading user input. At this point, the user can start typing in $\mathcal{E}$RGO clauses, which the system saves in a scratchpad file. When the user is done and types the end of file character `Control-D` (Unix) or `Control-Z` (Windows), the file is compiled and loaded. It is also possible to load the scratchpad file into a designated module, rather than the default one, using one of the following commands:

```
ergo>  [>>module].
ergo>  load{>>module}.
```

Adding rulebases to modules. When the `load{...}` command loads a rule base into a module, it first wipes out all the rules and facts that previously formed the knowledge base of that module. Sometimes it is desirable to *add* the facts and rules contained in a certain file to the already existing knowledge base of a module. This operation, called `add{...}`, does not erase the old knowledge base in the module in question. It is also possible to use the `[...]` syntax by prefixing the file name with a `+`-sign. Here are some examples of adding a rulebase contained in files to existing modules:

```
ergo>  [+foo].
ergo>  [+foo>>bar].
ergo>  add{foo}.
ergo>  add{foo>>bar}.
```

When using the `[...]` syntax, adding and loading can be intermixed. For instance,

```
ergo>  [foo>>bar, +foo2>>bar].
```

This first loads the file `foo.ergo` into the module `bar` and then adds the rule base contained in `foo2.ergo` to the same module.

Reloading and re-adding. Recompilation. $\mathcal{E}$RGO's `load{...}` and `add{...}` commands try to be smart in order to simplify maintenance of knowledge bases and to avoid undesirable side effects. First, reloading and re-adding the same file to the same module will have no effect unless one of the *dependent* files has changed since the previous load/add. So, cyclic add/load commands are harmless, albeit they constitute evidence of bad design.

For the purpose of recompilation, a dependent file is one that is included with the `#include` compiler directive. The dependent property is transitive, so if any of the dependent files downstream from the parent changes, loading or adding the parent file will cause that parent to be recompiled.

Similar relationship exists with respect to the load/add dependency. Normally, as we said, reloading a file will have no effect. But what if the file being loaded (or added) has an explicit load/add command that loads another file (which we call *load-dependent*)? If the dependent file was changed since the last loading, it needs to be reloaded and recompiled. In this case, if the parent file is reloaded then this reloading *will* take place and so all the load/add commands in that file will be re-executed causing the reloading of all the relevant load-dependent files. Such loading will take place if any of the load-dependents changes—at any level down-stream from the parent.

Reporting query answers. When the user types in a query to the shell, the query is evaluated and the results are returned. A result is a tuple of values for each variable mentioned in the query, except for the *anonymous variables* represented as "`?_`" or ?, and named *don't care variables*, which are preceded with the underscore, *e.g.*, `?_abc`.

By default, $\mathcal{E}$RGO prints out all answers (if their number is finite). If only one at a time is desired, type in the following command: `\one`. You can revert back to the all-answers mode by typing `\all`. Note: `\one` and `\all` affect only the *subsequent* queries. That is, in

```
ergo>  \one, goallist1.
ergo>  goallist2.
```

the `\one` directive will affect `goallist2`, but *not* `goallist1`. This is because `goallist1` executes in the same query as `\one` and thus is not affected by this directive.

The $\mathcal{E}$RGO shell includes many more commands beyond those mentioned above. These commands are listed below. However, at this point the purpose of some of these commands might seem a bit cryptic, so it is a good idea to come back here after you become more familiar with the various concepts underlying the system.

Summary of shell commands. In the following command list, the suffixes `.ergo` `.P`, `.xwam` are optional. If the file suffix is specified explicitly, the system uses the file with the given name without any modification. The `.ergo` suffix denotes a $\mathcal{E}$RGO knowledge base, the `.P` suffix indicates that it is a Prolog program, and `.xwam` means that it is a bytecode file, which can be executed by Prolog. If no suffix is given, the system assumes it is dealing with a $\mathcal{E}$RGO knowledge base and adds the suffix `.ergo`. If the file with such a name does not exist, it assumes that the file contains a Prolog program and tries the suffix `.P`. Otherwise, it tries `.xwam` in the hope that an executable Prolog bytecode exists. If none of these tries are successful, an error is reported.

- `\help`: Show the help info.
- `compile{file}`: Compile FILE.ergo for the default module `main`.

- `compile{file>>module}`: Compile FILE.ergo for the module `module`.
- `load{file>>module}`: Load `file.ergo` into the module `module`. If you specify `file.P` or `file.xwam` then will load these files.
- `load{file}`: Load `file.ergo` into the default module `main`. If you specify `file.P` or `file.xwam` then will load these files.
- `compile{file}`: Compile FILE.ergo for *adding* to the default module `main`.
- `compileadd{file>>module}`: Compile FILE.ergo for adding to the module `module`.
- `add{file>>module}`: Add `file.ergo` to the module `module`.
- `add{file}`: Add `file.ergo` to the default module `main`.
- `[file.{P|xwam|flr} > > module,...]`: Load the files in the specified list into the module `module`. The files can optionally be prefixed with a "+", which means that the file should be added to the module rather than loaded into it.
- `setmonitor{Secs,Type}`: Start monitoring query execution. Statistics will be printed to the standard output every *Secs* seconds. *Type* is the type of the monitor. The `heartbeat` monitor just shows the elapsed time. The `performance` monitor shows time, memory, and other key statistics. The `extended` monitor shows additional statistics of interest.
- `demo{demofilename}`: Consult a demo from $\mathcal{E}$RGO demos directory.
- `op{Precedence,Associativity,Operator}`: Define an operator in shell mode.
- `\all`: Show all solutions (default). Affects subsequent queries only.
- `\one`: Show solutions to subsequent queries one by one.
- `\trace/\notrace`: Turn on/off $\mathcal{E}$RGO trace.
- `chatter{on}/chatter{off}`: Turn on/off the display of the number of solutions at the end of each query evaluation.
- `feedback{on}/feedback{off}`: Turn on/off the display of query answers. Mostly used in Java applications.
- `setwarnings{type}`: Control the types of warnings to be shown to the user.
  - `all` — show *all* warnings (default)
  - `off` — do not show *any* warnings

  - `compiler=on`, `compiler=off` — turn compiler warnings on (default) or off; no effect on other types of warnings
  - `dependency=on/off` — turn dependency checker warnings on (default) or off; does not affect other types of warnings
  - `runtime=on/off` — turn runtime warnings on (default) or off; does not affect other types of warnings.

- `warnings{?Type}`: tell which warning control options are in effect. `?Type` can be a variable or a pattern like `compiler=?X`.
- `\end`: Say Ciao to ERGO, but stay in Prolog. You can still re-enter ERGO by executing the `ergo_shell` command at the Prolog prompt.
- `\halt`: Quit both ERGO and Prolog.

Of course, many other executable directives and queries can be executed at the ERGO shell. These are described further in this manual

In general, ERGO built-in predicates whose name is of the form `fl[A-Z]...` are either the ERGO shell commands or predicates that can be used in Prolog to control the execution of ERGO modules. We will discuss the latter in Section 15.8. Some of these commands — mostly dealing with loading and compilation of ERGO modules — can also be useful within ERGO applications.

All commands with a FILE argument passed to them use the Prolog `library_directory` predicate to search for the file, except that the command `demo{FILE}` first looks for `FILE` in the ERGO demo directory. The search path typically includes the standard system's directories used by Prolog followed by the current directory.

All Prolog commands can be executed from ERGO shell, if the corresponding Prolog library has already been loaded.

After a parsing or compilation error, ERGO shell will discard tokens read from the current input stream until the end of file or a rule delimiter (".") is encountered. If ERGO shell seems to be hanging after the message

```
++Warning[Ergo]: discarding tokens (rule delimiter '.' or EOF expected)
```

hit the `Enter` key once, type ".", and then `Enter` again. This should reset the current input buffer and you should see the ERGO command prompt:

```
ergo>
```

5 F-logic and $\mathcal{E}$RGO by Example

In the future, this section will contain a number of small introductory examples illustrating the use of F-logic and $\mathcal{E}$RGO. In addition, the reader can consult the $\mathcal{E}$RGO tutorial available on the Coherent Web site: https://sites.google.com/a/coherentknowledge.com/ergo-suite-tutorial/. Another source of examples and background material is the $\mathcal{F}$LORA-*2* tutorial: http://flora.sourceforge.net/t

6 Main Syntactic Elements of $\mathcal{E}$RGO

$\mathcal{E}$RGO has rich syntax, so it is useful to first list the various types of statements one may encounter in this manual. First, we should note that $\mathcal{E}$RGO does *not* have any alphanumeric *reserved* keywords, so the user is not restricted in that name space in any way. The only reserved keywords are those that start with a backslash, e.g., `\and`, `\or`, `\if`, etc.

The main types of $\mathcal{E}$RGO statements are compiler and runtime directives, rules, queries, latent queries, and facts:

- **Compiler directives** have the form

  ```
  :- directiveName{arguments}.
  ```

  Some directives do not have arguments. Compiler directives affect the compilation of the file in which they appear—typically the semantic and optimization options.

- **Runtime directives** have the form

  ```
  ?- directiveName{arguments}.
  ```

  Runtime directives are typically used to change the semantics of the runtime environment at run time.

- **Rules** have the form

  ```
  @!{statementDescriptors}  ruleHead  :-  ruleBody.
  ```

  Rules constitute the key part of an $\mathcal{E}$RGO knowledge base as they (along with the class hierarchy) represent the actual *knowledge*. The presence of rules is the main difference between knowledge bases and mere databases.

  The statement descriptor part (`@!{...}`) is optional. The body of a rule is sometimes also called a *premise*.

- **Queries** have the form

  ```
  ?- ruleBody.
  ```

  Syntactically queries have the same form as the rule bodies, but they use the symbol "`?-`" to distinguish the two. Queries are used to request information from the knowledge base.

- **Latent queries** have the form

  ```
  @!{statementDescriptors}  !- ruleBody.
  ```

  Latent queries are similar to regular queries. However, regular queries are *immediate* requests for information from the knowledge base, while latent queries are requests that intended to be posed at a later time. A latent query also has descriptors, which are used to refer to the query and to invoke it.

- **Facts** are statements that are considered to be unconditionally true. They have the form

  ```
  @!{statementDescriptors}  ruleHead.
  ```

  Syntactically and conceptually, a fact is a rule without a premise. The descriptor part of the syntax for facts is optional.

Rule heads, bodies, queries, and their arguments are typically composed out of *base formulas* with the help of connectives, such as conjunction, disjunction, the various negations, and more. The main forms of base formulas are

- F-logic **frames** are used for object-oriented knowledge representation.
- HiLog **predicates** are used for more traditional knowledge representation. However, in $\mathcal{E}$RGO, predicates can be higher-order and variables are allowed to range over them.

The different types of frames and predicates are described in the respective sections. The main components used to construct predicates and frames include:

- **Variables**, which are expressions of the form *?Varname*.
- **Constants**, which includes *symbols*, *strings*, *numbers*, and various other data types. There are certain builtin constants, like `\true`, `\false`, and `\undefined`, which represent the three truth values in $\mathcal{E}$RGO: *true*, *false*, and *undefined*.
- **Operators**, including arithmetic operators.

- **Quasi-constants** and **quasi-variables**. Quasi-constants are symbols that get substituted with real constants at compile time or at the time the knowledge base is loaded.

  Examples of quasi-constants are `\@F` and `\@L`, which get substituted with the file name and the line number in which these constants occur. Quasi-constants let one write statements that refer to objects, such as the file name or line number, which are either unknown at the time of writing or may change later. Quasi-constants never change during runtime—it is just that their values are typically unknown at the time these constants are written into the knowledge base by the knowledge engineer.

  As example of a quasi-variable is `\?C`. This variable can appear in the body of a rule and it gets substituted with the name of the module that invoked that rule. This is a *quasi*-variable because it gets instantiated by the runtime system and is not under the control of the knowledge engineer. It is a *variable* (rather than a constant) because it may get instantiated with different values during runtime (because, different modules may invoke the same rule).

- **Auxiliary symbols**, such as `->`, `!`, `[`, `(`, `[|`, etc., are used to glue together the aforesaid components to form base formulas.

7 Basic $\mathcal{E}$RGO Syntax

In this section we describe the basic syntactic structures used to specify $\mathcal{E}$RGO knowledge base. Subsequent sections describe the various advanced features that are needed to build practical applications. The complete syntax is given in Appendix A. However, it should be noted that BNF cannot describe the syntax of $\mathcal{E}$RGO precisely, because it is based on operator grammar (as in Prolog) mixed with context free grammars in places where operator grammar is inadequate (as, for example, in parsing if-then-else).

7.1 F-logic Vocabulary

- *Symbols*: The F-logic alphabet of *object constructors* consists of the sets $\mathcal{C}$ and $\mathcal{V}$ (constants and variables). Variables are symbols that begin with a questionmark, followed by a letter or an underscore, and then followed by zero or more letters and/or digits and/or underscores (e.g., `?X`, `?name`, `?_`, `?_v_5`). All other symbols, including the constants (which are 0-ary object constructors), are symbols that start with a letter followed by zero or more letters and/or digits and/or underscores (e.g., `a, John, v_10`). They are called *general constant symbols* or *Prolog atoms*. General constant symbols can also be any sequence of symbols enclosed in single quotes (e.g., `'AB@*c'`). Later, in Section 39, we introduce additional constants, called typed literals.

In addition to the usual first-order connectives and symbols, $\mathcal{E}$RGO has a number of special symbols: `]`, `[`, `}`, `{`, "`,`", "`;`", `%`, `#`, `\#`, `->`, `=>`, `:`, `::`, `->->`, `->>`, `:=:`, etc.

- *Numeric constants*: These include *integers*, like 123 or 5063; *decimals* of the form 123.45; or *floating point numbers*, like 12.345e12 ($= 12.345 * 10^{12}$), 0.34e+3 (same as 0.34e3), or 360.1e-2 ($= 360.1 * 10^{-2}$).

- *Anonymous and don't care variables*: Variables of the form `?_` or `?` are called *anonymous* variable. They are used whenever a *unique* new variable name is needed. In particular, two different occurrences of `?_` or `?` in the same clause are treated as *different* variables. Named variables that start with `?_`, e.g., `?_foo`, are called *don't care* variables. Unlike anonymous variables, two different occurrences of such a variable in the same clause refer to the *same* variable. Nevertheless, don't care variables have special status when it comes to error checking and returning answers. The practice of logic programming shows that a singleton occurrence of a variable in a clause is often a mistake due to misspelling. Therefore, $\mathcal{E}$RGO issues a warning when it finds that some variable is mentioned only once in a clause. If such an occurrence is truly intended, it must be replaced by an anonymous variable or a don't care variable to avoid the warning message from $\mathcal{E}$RGO. Also, bindings for anonymous and don't care variables are not returned as answers.

- *Id-Terms/Oids*: Instead of the regular first-order terms used in Prolog, $\mathcal{E}$RGO uses HiLog terms. HiLog terms [4] generalize first-order terms by allowing variables in the position of function symbols and even other terms can serve as functors. For instance, `p(a)(?X(f,b))` is a legal HiLog term. Formally, a HiLog term is a constant, a variable, or an expression of the form $\mathtt{t}(\mathtt{t_1}, ..., \mathtt{t_n})$ where t, t_1, ..., t_n is a HiLog term.

  HiLog terms over $\mathcal{C}$ and $\mathcal{V}$ are called *Id-terms*, and are used to name objects, methods, and classes. Ground Id-terms (i.e., terms with no variables) correspond to *logical object identifiers* (*oid*s), also called object *names*. Numbers (including integers and floats) can also be used as Id-terms, but such use might be confusing and is not recommended.

- *Base formulas*: Let $\mathtt{O}, \mathtt{M}, \mathtt{R_i}, \mathtt{X_i}, \mathtt{C}, \mathtt{D}, \mathtt{T}$ be Id-terms. In addition to the usual first-order predicate formulas, like $p(X_1, \ldots, X_n)$, $\mathcal{E}$RGO allows higher-order *HiLog* base formulas of the form $?\mathtt{X}(\mathtt{s}, ?\mathtt{Y})$, $?\mathtt{X}(\mathtt{f}, ?\mathtt{Y})(?\mathtt{X}, \mathtt{g}(\mathtt{k}))$, etc., where `?X` and `?Y` are variables, while the symbols not prefixed with a `?` are constants. Furthermore, the following *frame* formulas are supported in $\mathcal{E}$RGO:

  1. `O[M->V]`, `C[|M->V|]`
  2. `O[M->{`$V_1, ..., V_n$`}]`, `C[|M->{`$V_1, ..., V_n$`}|]`
  3. `O[M=>T]`, `C[|M=>T|]`, `C[|M{L..H}=>T|]`
  4. `O[V]`, `C[|V|]`

5. `O[=>T]`, `C[|=>T|]`
6. `O[ ]`, `C[| |]`

Here `O`, `C`, `M`, V_i, T_i are *HiLog terms* of the form `a`, `f(?X)`, `?X(s,?Y)`, `?X(f,?Y)(?X,g(k))`, etc., where `?X` and `?Y` are variables and `f`, `s`, etc., are constants.

Expressions (1) and (2) above are *data frames* for *value-returning* methods. They specify that a *method expression* `M` applied to an object `O` returns the result object `V` in case (1), or a set of objects, V_1, ..., V_n, in case (2). In all cases, methods are assumed to be set-valued. However, later we will see that cardinality constraints can be imposed on methods, so it would be possible to state that a particular method is functional or has some other cardinality constraints. The formula (2) says that the result consists of several objects, which *includes* V_1, V_2, ..., V_n. Note that we emphasized "includes" to make it plain that other facts and rules in the knowledge base can specify additional objects that must be included among the method result.

In (1) and (2), when `M` is a constant, *e.g.*, `abc`, then we say that it is an *attribute* or a *property*; for example, `John123[name -> 'John Doe']`. When `M` has the form `f(X,Y,Z)` then we refer to it as a method `f` with arguments `X`, `Y`, and `Z`; for example, `John123[salary(2017) -> 50000]`. However, as we saw earlier, method expressions can be much more general than these two possibilities: they can be arbitrary HiLog terms.

The formulas in (1) and (2) that use `[|...|]` apply to *classes* and specify the *default values* inherited by the objects that belong to those classes. To make it easier to remember, we use the letter `C` in those cases. The letter `O` (for "object") is used with formulas of the form `[...]`, which apply to individual objects and specify *concrete* values of their attributes/methods as opposed to the default values inherited from superclasses. These concepts are explained in greater detail in Section 20.

The expression (3) above is a *signature frame* (or *typing frame*). It specifies a *type constraint* that says that the method expression, `M`, when applied to objects that belong to class `C`, must yield objects that belong to class `T`. The first form of type constraints in (3) applies to individual objects, the second to classes, and this second form is inherited by subclasses and individual objects. Third, while also applying to classes, imposes a cardinality constraint on the possible number of values `T` that can correspond to the same `M`. Here `L>0` is a number that specifies the lower bound on that number of values and `H` specifies the upper limit. See Section 20 for additional details on this type of formulas.

Note: ERGO does not automatically enforce type constraints. However, one can *request* that an integrity constraint is to be maintained automatically, as described in Section 36. Also, run-time type checking is possible—see Section 43.2. □

The form (4) is used for Boolean methods. Unlike the methods in (1) and (2), Boolean methods can be either true or false: they do not return any values. Apart from that, the

previous conventions apply: `O[V]` says that object `O` has a Boolean property `V` and `C[|V|]` says that class `C` has a *default* Boolean property that is inherited by `C`'s subclasses and member objects.

The form (5) specifies signatures, i.e., types for the formulas in (4). Note that, unlike in (3), cardinality constraints do not apply in this case.

Empty frames. Normally, a frame has at least one statement about the frame's object, as in

```
John[spouse -> Mary]
```

However, as seen from case (6) above, a frame can also be *empty* like this: e.g.,

```
Mary[]
Mary[| |]
```

Here we have an empty object specification and an empty class signature. This is to be interpreted as a statement that the corresponding object (`Mary`, in our case) is *known to exist* in the domain of discourse. This means that some statement about that object is derivable from user specifications. For instance, if the knowledge base has statements that imply, say, `Mary:person`, `Mary[age->25]`, or if the "empty" fact `Mary[]` exists in the knowledge base, then `Mary[]` holds true. This rule excludes builtin datatypes (see Section 39), such as `\integer`, `\object`, and `\symbol`. For instance, even though `2:\integer` and `foobar:\object` are true, by definition, both `2[]` and `foobar[]` would be false unless these facts are implied by other parts of the knowledge base given by the user, i.e., unless the user "told" the knowledge base about the existence of these objects.

Objects are grouped into classes using what we call *ISA-literals*, which account both for class membership and subclass relationships:

5. `O:C`
6. `C::D`

The expression (5) states that `O` is an *instance* of class `C`, while (6) states that `C` is a *subclass* of D.

User-defined equality

7. `O1 :=: O2`

enables the user to state that two syntactically different (and typically non-unifiable) terms represent the same object. For instance, one can assert that `a :=: b` and from then on everything that is true about `a` will be true about `b`, and vice versa. Note that this is different and more powerful than the unification-based equality built-in `=`, which exists both

in ERGO and Prolog. For instance, `=`-based formulas can never occur as a fact or in a rule head, and `a = b` is always false. More on user-defined equality in Section 21.1.

8. *Composite frames.*

F-logic frames (or *frame literals*) and ISA-literals, can be combined in various ways, reducing long conjunctions into very compact forms. For instance, the conjunction of `John:person`, `Bill:Student`, `John[age->31]`, `John[children->Bob]`, `John[children->Mary]`, `John[children->Bill]`, and `Mary[age->5]` can be compacted into the following complex frame:

```
John:person[age->31, children->{Bob,Mary[age->5],Bill:Student}]
```

Note that this shows that frames can be nested (e.g., `Mary[age->5]`) and values pertaining to the same method and object can be grouped into sets (e.g., `John[children->{Bob,Mary,Bill}]`). Furthermore, ISA-literals can be attached both to the outermost object as well as the inner ones (as in `John:person` and `Bill:Student`).

- *Atomic formulas.* The base formulas of the types (1)-(7) above are called *atomic formulas*. Atomic formulas are also base formulas, but the latter can also contain non-atomic formulas: complex frames.

- *Rules* are constructs of the form *head*`:-`*body*, where

  – *head* is a frame/HiLog literal or a conjunction of such literals. These literals can also be negated with `\neg`. and

  – *body* is a conjunction and/or disjunction of frame/HiLog literals or negated (with `\+`, `\neg`, or `\naf`) frame/HiLog literals. Each rule must be terminated with a ".".

  Conjunction is specified, as in Prolog, using the "," symbol but `\and` is also accepted. Disjunction is denoted using the semicolon symbol ";" or using `\or`. Negation is specified using `\+`, `\neg`, or `\naf`— the difference will be explained later. For example,

```
p(?X), \neg ?Y[foo->bar(?X)] :- (q(?X,?Y) \or ?X[foo->moo,abc->cde(?Y)]),\naf w(?X).
```

  ERGO supports much richer syntax in the rule heads, including disjunction and quantifiers. See Section 19 for details.

  As usual in logic languages, a single rule with a disjunction in the body

$$\begin{array}{ll} \textit{head} \texttt{ :- } & \texttt{John}[\texttt{age->31}], \\ & (\texttt{John}[\texttt{children->}\{\texttt{Bob},\texttt{Mary}\}] \texttt{ ; } \texttt{John}[\texttt{children->Bill}]). \end{array} \tag{1}$$

  is equivalent to the following pair of rules:

```
head :- John[age->31], John[children->{Bob,Mary}].
head :- John[age->31], John[children->Bill].
```

Disjunction is also allowed inside frame literals. For instance, rule (1) can be equivalently rewritten as:

```
head :- John[age->31, (children->{Bob,Mary} ; children->Bill)].
```

Note that the conjunction "," binds stronger than disjunction ";", so the parentheses in the above example are essential.

- *Knowledge bases and queries*: A *knowledge base* is a set of rules. A *query* is a rule without the head. In $\mathcal{E}$RGO, such headless rules use `?-` instead of `:-`, *e.g.*,

```
?- John[age->?X].
```

The symbol `:-` in headless $\mathcal{E}$RGO expressions is used for various directives, which are plenty and will be introduced in due course.

Example 7.1 (Publications Database) Figure 1 depicts a fragment of a $\mathcal{E}$RGO knowledge base that represents a database of scientific publications.

7.2 Symbols, Strings, and Comments

Symbols. $\mathcal{E}$RGO symbols (that are used for the names of constants, predicates, and object constructors) begin with a letter followed by zero or more letters (`A...Z`, `a...z`), digits (`0...9`), or underscores (_), e.g., `student`, `apple_pie`. Symbols can also be *any* sequence of characters enclosed in a pair of single quotes, e.g., `'JOHN SMITH','default.ergo'`. Internally, $\mathcal{E}$RGO symbols are represented as *Prolog symbols*, which are also called Prolog atoms. They are typically used as names of predicates and function symbols. All $\mathcal{E}$RGO symbols belong to the class `\symbol`.

$\mathcal{E}$RGO also recognizes escape sequences inside single quotes (`'`, symbols). An escape sequence begins with a backslash (\). Table 1 lists the special escape character sequences and their corresponding special symbols. An escape sequence can also represent a Unicode character. Such a character is preceded with a backslash followed by the letters `x`, `u`, `X`, or `U` followed by 1 to 8 hexadecimal digits (0–F) representing the character's Unicode value. The sequence of digits **must** be terminated with a vertical bar, |. For example, `\xd|` is the ASCII character Carriage Return, `\x3A|` represents the semicolon, while `\u05D0|` is the Hebrew letter Alef. In other contexts, a backslash is recognized as itself.

If it is necessary to include a quote inside a quoted symbol, that single quote must be escaped by another single quote, e.g., `'isn''t'` or by a backslash, e.g., `'isn\'t'`.

Schema:
paper[|authors => person, title => string|].
journal_p :: paper[|in_vol => volume|].
conf_p :: paper[|at_conf => conf_proc|].
journal_vol[|of => journal, volume => integer, number => integer, year => integer|].
journal[|name => string, publisher => string, editors => person|].
conf_proc[|of_conf => conf_series, year => integer, editors => person|].
conf_series[|name => string|].
publisher[|name => string|].
person[|name => string, affil(integer) => institution|].
institution[|name => string, address => string|].

Objects:
*o_j*1 : journal_p[title -> 'Records, Relations, Sets, Entities, and Things',
authors -> {*o_mes*}, in_vol -> *o_i*11].
o_di : conf_p[title -> 'DIAM II and Levels of Abstraction',
authors -> {*o_mes*, *o_eba*}, at_conf -> *o_v*76].
*o_i*11 : journal_vol[of -> *o_is*, number -> 1, volume -> 1, year -> 1975].
o_is : journal[name -> 'Information Systems', editors -> {*o_mj*}].
*o_v*76 : conf_proc[of -> vldb, year -> 1976, editors -> {*o_pcl*, *o_ejn*}].
o_vldb : conf_series[name -> 'Very Large Databases'].
o_mes : person[name -> 'Michael E. Senko'].
o_mj : person[name -> 'Matthias Jarke', affil(1976) -> *o_rwt*].
o_rwt : institution[name -> 'RWTH_Aachen'].

Figure 1: A Publications Object Base and its Schema in $\mathcal{E}$RGO

Numbers. Normal $\mathcal{E}$RGO integers are decimals represented by a sequence of digits, e.g., `892`, `12`. $\mathcal{E}$RGO also recognizes integers in other bases (2 through 36). The base is specified by a decimal integer followed by a single quote (`'`). The digit string immediately follows the single quote. The letters `A...Z` or `a...z` are used to represent digits greater than 9. Table 2 lists a few example integers.

Underscore (`_`) can be put inside any sequence of digits as delimiters. It is used to partition some long numbers. For instance, `2'11_1111_1111` is the same as `2'1111111111`. However, "`_`" cannot be the first symbol of an integer, since variables can start with an underscore. For example, 1_2_3 represents the number 123 whereas ?_12_3 represents a variable named ?_12_3.

Floating point numbers normally look like `24.38`. The decimal point must be preceded by an integral part, even if it is 0, e.g., `0.3` must be entered as `0.3`, but not as `.3`. Each floating number may also have an optional exponent. It begins with a lowercase `e` or an uppercase `E` followed by an

Escaped Sequence	ASCII (decimal)	Symbol
\\	92	\
\n or \N	10	NewLine
\t or \T	9	Tab
\r or \R	13	Return
\v or \V	11	Vertical Tab
\b or \B	8	Backspace
\f or \F	12	Form Feed
\e or \E	27	Escape
\d or \D	127	Delete
\s or \S	32	Whitespace

Table 1: Escaped Character Sequences and Their Corresponding Symbols

Integer	Base (decimal)	Value (decimal)
1023	10	1023
2'1111111111	2	1023
8'1777	8	1023
16'3FF	16	1023
32'vv	32	1023

Table 2: Representation of Integers

optional minus sign (`-`) or plus sign (`+`) and an integer. This exponent is recognized as in base 10. For example, `2.43E2 is 243` whereas `2.43e-2 is 0.0243`.

Other data types. ℰRGO supports an array of primitive data types, including string, Boolean, dateTime, iri, and more. Primitive data types are described in Section 39.

Comments. ℰRGO supports two kinds of comments: (1) all characters following `//` until the end of the line; (2) all characters inside a pair of `/*` and `*/`. Note that only (2) can span multiple lines.

Comments are recognized as whitespace by the compiler. Therefore, tokens can also be delimited by comments.

7.3 Operators

As in Prolog, ℰRGO allows the user to define operators, to make the syntax more natural. There are three kinds of operators: infix, prefix, and postfix. An infix operator appears between its two arguments, while a prefix operator appears in front of its single argument. A postfix operator is

written after its single argument. For instance, if `foo` is defined as an infix operator, then `?X foo a` will be parsed as `foo(?X,a)` and if `bar` is a postfix operator then `?X bar` is parsed as `bar(?X)`.

Each operator has a *precedence level*, which is a positive integer. Each operator also has a *type*. The possible types for infix operators are: `xfx`, `xfy`, `yfx`; the possible types for prefix operators are: `fx`, `fy`; and the possible types for postfix operators are: `xf`, `yf`. In each of these expressions, `f` stands for the operator, and `x` and `y` stand for the arguments. The symbol `x` in an operator expression means that the precedence level of the corresponding argument should be *strictly less* than that of the operator, while `y` means that the precedence level of the corresponding argument should be *less than or equal* to that of the operator.

The precedence level and the type together determine the way the operators are parsed. The general rule is that precedence of a constant or a functor symbol that has not been defined as an operator is zero. Precedence of a Prolog term is the same as the precedence of its main functor. An expression that contains several operators is parsed in such a way that the operator with the highest precedence level becomes the main functor of the parsed term, the operator with the next-highest precedence level becomes the main functor of one of the arguments, and so on. If an expression cannot be parsed according to this rule, a parse error is reported.

It is not our goal to cover the use of operators in any detail, since this information can be found in any book on Prolog. Here we just give an example that illustrates the main points. For example, in ℰRGO, `-` has precedence level `800` and type `yfx`, `*` has precedence level `700` and type `yfx`, `->` has precedence level 1100 and type `xfx`. Therefore, `8-2-3*4` is the same as `-(-(8,2),*(3,4))` in prefix notation, and `a -> b -> c` will generate a parsing error.

Any symbol can be defined as an operator. The general syntax is

```
:- op{Precedence,Type,Name}.
```

For instance,

```
:- op{800, xfx, foo}
```

As a notational convenience, the argument `Name` can also be a list of operator names of the same type and precedence level, for instance,

```
:- op{800,yfx,[+,-]}.
```

It is possible to have more than one operator with the same name provided they have different uses (*e.g.*, one infix and the other postfix). However, the ℰRGO built-in operators are not allowed to be redefined. In particular, any symbol that is part of F-logic syntax, such as “,”, “.”, “[“, “:”, etc., as well as any name that begins with `flora` or `fl` followed by a capital letter should be considered as reserved for internal use.

Although this simple rule is sufficient, in most cases, to keep you out of trouble, you should be aware of the fact that symbols such as "`,`", "`;`", "`+`", "`.`", "`->`", "`::`", "`:-`", "`?-`" and many other parts of ℰRGO syntax are operators. Therefore, there is a chance that precedence levels chosen for the user-defined operators may conflict with those of ℰRGO and, as a result, your specification might not parse. If in doubt, check the declarations in the file `flroperator.P` in the ℰRGO source code.

The fact that some symbols are operators can sometimes lead to surprises. For instance,

```
?- (a,b,c).
:- (a,b).
```

will be interpreted as terms `'?-'(a,b,c)` and `':-'(a,b)` rather than a query and a directive, respectively. The reason for this is that, first, such terms are allowed in Prolog and there is no good reason to ban them in ℰRGO; and, second, the above syntax is ambiguous and the parser makes the choice that is consistent with the choice made in Prolog. Typically, users do not put parentheses around subgoals in such cases, and would instead write

```
?- a,b,c.
:- a,b.
```

Note that things like

```
?- (a,b),c.
?- ((a,b,c)).
```

will be interpreted as queries, so there are plenty of ways to satisfy one's fondness for redundant parentheses.

7.4 Logical Expressions

In a ℰRGO, any combination of conjunction, disjunction, and negation of literals can appear wherever a logical formula is allowed, e.g., in a rule body.

Conjunction is made with the infix operator "`,`" and disjunction is made using the infix operator "`;`". Negation is specified using the prefix operators "`\+`" and "`\naf`".[4] When parentheses are

[4] In brief, "`\+`" represents Prolog-style negation, which does not have an acceptable logical semantics. It is useful, however, when applied to non-tabled Prolog predicates, F-logic frames, or HiLog predicates. "`\naf`", on the other hand, is negation that implements the logical well-founded semantics. Refer to Section 17 for more information on the difference between negation operators.

omitted, conjunction binds stronger than disjunction and the negation operators bind their arguments stronger than the other logical operators. For example, in $\mathcal{E}$RGO the following expression: `a, b; c, \naf d`, is equivalent to the logical formula: $(\mathtt{a} \wedge \mathtt{b}) \vee (\mathtt{c} \wedge (\neg \mathtt{d}))$.

Logical formulas can also appear inside the frame specification of an object. For instance, the following frame:

```
o[\naf att1 -> val1, att2 -> val2; meth -> res]
```

is equivalent to the following formula:

```
(\naf o[att1 -> val1], o[att2 -> val2]) ; o[meth -> res]
```

7.5 Arithmetic (and related) Expressions

In $\mathcal{E}$RGO arithmetic expressions are *not* always evaluated. As in Prolog, the arithmetic operators such as `+`, `-`, `/`, and `*`, are defined as normal binary functors. To evaluate an arithmetic expression, $\mathcal{E}$RGO provides another operator, `\is`. For example, `?X \is 3+4` will bind `?X` to the value `7`. In addition, $\mathcal{E}$RGO provides a powerful feature of inline evaluation of such expressions, which allows these expressions to appear as arguments to predicates and frames, and be automatically evaluated at runtime. Ergo also provides a number of additional operators, including list/set append/union/intersect, and difference, and string concatenation, which also automatically converts arbitrary terms to their printable form.

When dealing with arithmetic expressions, the order of literals is *sometimes* important. The comparison and evaluation operators for which the order is unimportant (the *logical* operators) are:

- `>`, `<`, `=<`, `>=`
- `!=`, `!==`, `=:=`, `=\=`
- `\is`
- `~`, `\~`, `!~`

The operators for which the order is important (the *non-logical* operators) are:

- `==`
- `\=`, `\==`, `?=`
- `@<`, `@>`

Logical operators commute among themselves, but non-logical operators generally do not commute with either logical or non-logical operators, and different orders of these operators in an expression may produce different results. For instance, if ?X is not bound then ?X == abc, ?X = abc will fail, while ?X = abc, ?X == abc will succeed with ?X = abc. The reason for this is, of course the non-logical operator ==.

Arithmetic expression must be instantiated at the time of evaluation. Otherwise, a runtime error will occur. However, $\mathcal{E}$RGO tries to delay the evaluation of arithmetic expressions until the variables become bound and it will issue a runtime error only if it determines that some variable will never get bound. For instance,

```
?- ?X > 1, ?X \is 1+1.
```

will *not* produce an error, while the following query will:

```
?- ?X > 1.
```

As in Prolog, the operands of an arithmetic expression can be any variable or a constant. However, in $\mathcal{E}$RGO, an operand can also be a *path expression*. For the purpose of this discussion, a path expression of the form p.q should be understood as a shortcut for p[q->?X], where ?X is a new variable, and p.q.r is a shortcut for p[q->?X], ?X[r->?Y], where ?X and ?Y are new variables. More detailed discussion of path expressions appears in Section 8.

In arithmetic expressions, all variables are considered to be existentially quantified. For example, the following query

```
ergo> John.bonus + Mary.bonus > 1000.
```

should be understood as

```
ergo> John[bonus->?_V1], Mary[bonus->?_V2], ?_V1+?_V2 > 1000.
```

Note that the first query does not have any variables, so after the evaluation the system would print either yes or no. To achieve the same behavior, we use *don't care variables*, ?_V1 and ?_V2. If we used ?V1 and ?V2 instead, the values of these variables would have been printed out.

$\mathcal{E}$RGO recognizes numbers as oids and, thus, it is perfectly normal to have arithmetic expressions inside path expressions such as this: 1.2.(3+4*2).7. When parentheses are omitted, this might lead to ambiguity. For instance, does the expression

```
1.m+2.n.k
```

correspond to the arithmetic expression `(1.m)+(2.n.k)`, to the path expressions `(1.m+2.n).k`, by `(1.m + 2).n.k`, or to `1.(m+2).n.k`? To disambiguate such expressions, we must remember that the operator "." used in path expressions binds stronger than the arithmetic operators +, −, etc.

Precedence	Operator	Use	Associativity	Arity
not applicable	`( )`	parentheses; used to change precedence	not applicable	not applicable
not applicable	`.`	decimal point	not applicable	not applicable
	`.`	object reference	left	binary
400	`: ::`	class membership and subclass relationships	left	binary
600	`-`	minus sign	right	unary
	`+`	plus sign	right	unary
700	`*`	multiplication	left	binary
700	`**`	power	left	binary
	`/`	division	left	binary
800	`- +`	subtraction and addition	left	binary
	`=<`	less than or equals to	not applicable	binary
	`>=`	greater than or equals to	not applicable	binary
1000	`=:=`	numeric equals-to	not applicable	binary
	`=\=`	unequal to	not applicable	binary
	`\is`	arithmetic assignment	not applicable	binary
	`=`	unification	left	binary
	`!= or \=`	disunification	left	binary
	`==`	identity	left	binary
	`=..`	meta decomposition	left	binary
	`!== or \==`	not identical	left	binary
	`@<`	lexicographical less-than	left	binary
	`?=`	identical or not unifiable	left	binary
1200	`\naf`	well-founded negation	not applicable	unary
	`\neg`	explicit negation	not applicable	unary
	`\+`	Prolog-style negation	not applicable	unary
1250	∼	semantic unification	left	binary
	! ∼ or \ ∼	semantic disunification	left	binary
990	||	string/term concatenation	right	binary
290	`++ -- &&`	list append/set union/difference/intersection	right	binary

Table 3: Operators in increasing precedence order

Even more interesting is the following example: `2.3.4`. Does it represent the path expression `(2).(3).(4)`, or `(2.3).4`, or `2.(3.4)` (where in the latter two cases 2.3 and 3.4 are interpreted as decimal numbers)? The answer to this puzzle (according to $\mathcal{E}$RGO conventions) is `(2.3).4`: when tokenizing, $\mathcal{E}$RGO first tries to classify tokens into meaningful categories. Thus, when 2.3 is first found, it is identified as a decimal. Thus, the parser receives the expression (2.3).4, which it identifies as a path expression that consists of two components, the oids 2.3 and 4.

Another ambiguous situation arises when the symbols `-` and `+` are used as minus and plus signs, respectively. $\mathcal{E}$RGO follows the common arithmetic interpretation of such expressions, where the `+/-` signs bind stronger than the infix operators and thus `4-7` and `4-+7` are interpreted as `4-(-7)`

and `4-(+7)`, respectively.

Table 3 lists various operators in decreasing precedence order, their associativity, and arity. When in doubt, use parentheses. Here are some more examples of valid arithmetic expressions:

`o1.m1+o2.m2.m3`	same as `(o1.m1)+(o2.m2)`
`2.(3.4)`	the value of the attribute `3.4` on object `2`
`3 + - - 2`	same as `3+(-(-2))`
`5 * - 6`	same as `5*(-6)`
`5.(-6)`	the value of the attribute `-6` on object `5`

Note that the parentheses in `5.(-6)` are needed, because otherwise "`.-`" would be recognized as a single token. Similarly, the whitespace around "`+`", "`-`", and "`*`" are also needed in these examples to avoid `*-` and `+-` being interpreted as distinct tokens.

In ℰRGO, `\is/2` also understands the string concatenation operator, ||, and the list append, difference, and intersect operators `++`, `--`, and `&&`. For instance,

```
?- ?X \is abc || 'cde:  '  || f(?P) || ' ;', writeln(?X)@\io.
```
abccde: f(?A) ;

Note that this operator converts terms into their printable representation and tries to give meaningful names to variables.

Examples of list append/difference/intersection include:

```
?- ?X \is (([a,c,?L,b]--[c])++[?L,1]) -- [?L,a].
```

Result: `?X = [b, 1]`

```
?- ?X \is [a,c,?L,b]--[c]++[?L,1].
```

Result: `?X = [a, b]`

```
?- ?X \is [a,b]++[c]++[?L,1].
```

Result: `?X = [a, b, c, ?_h9660, 1]`

```
?- ?X \is [a,b]++[c]++[p,1]&&[b,c,p].
```

Result `:?X = [a, b, c, p]`

To understand the results of these queries, the reader needs to keep in mind that `++`, `--`, and `&&` are *right-associative* operators. In the first example, we explicitly placed the parentheses to change the order of the operations, but in the last three examples parenthesizing happens implicitly, through right-association. For example, in the second case we really have `[a,c,?L,b]--([c]++[?L,1])`. The other important thing to note is that variables are not unified, but are compared via lexical identity. This is why, in the second query, `?L` disappears from the result, but if the list expression were `[a,c,?L,b]--([c]++[?K,1])` then the result of the evaluation would be `[a,?L,b]`.

In addition to the operators, the builtin function, and constants listed in Table 4 can occur in arithmetic expressions, i.e., on the right side of `\is/2` and on either side of the inequalities (>, <, =<, and >=).

Note: the expression on the right side of `\is/2` must not contain any variables at the time the `\is/2` predicate is evaluated (except in case of the || operator illustrated above). Otherwise, an error is issued. If you want to solve equations and constraints (e.g., find $?X$ such that $5 = ?X + 2$), `\is/2` is a wrong predicate to do so: see Section 30 for the proper way to do this.

Also, in the arithmetic expressions on the right side of `\is/2`, both `pi`, `e`, `\pi`, and `\e` are recognized. However, outside of the context of `\is/2`, only `\pi` and `\e` are recognized as special constants. The other functions in the above table (like `min`, `max`, trigonometry) are recognized both in `\is/2` expressions and inside constraints (see Section 30). Yet others (`div`, `rem`, `mod`) are recognized only inside `\is/2` expressions.

ℰRGO also supplies additional functions that can be used with `\is`, including `sum`, `min`, `max`, `avg`, `last`, `count`, `delete`, `reverse`, and `nth`. For instance,

```
?-  ?X \is count([a,b,c]).  // ?X = 3
?-  ?X \is avg([1,2,3]).    // ?X = 2
?-  ?X \is sum([1,2,3]).    // ?X = 6
?-  ?X \is nth(2,[a,b,c]).  // ?X = b
```

Inline evaluation of arithmetic expressions. In addition to the various extensions of `\is/2`, ℰRGO provides a more natural and powerful way to evaluate expressions by placing them directly as arguments of the predicates and frames, and prefixing them with a "`=`" or with "`\is`". This can be done both in rule heads and rule bodies. For example,

```
q(2).
```

Function	Arity	Meaning
min, max	2	minimum and maximum of the arguments
abs	1	absolute value
ceiling, floor, round	1	ceiling, floor, and rounding of a real
float	1	convert to float
truncate	1	truncate the decimal part of a real
mod	2	integer division modulo
rem	2	remainder of integer division
div, `%%`	2	integer division
exp, **	2	exponent of argument 1
exp	1	exponent of number E
sqrt	1	square root
sign	1	sign of number (1 or -1)
sin, cos	1	sine and cosine
asin,acos	1	arcsine and arccosine
tan, atan	1	tangent and arctangent
log	1	logarithm base E
log10	1	logarithm base 10
$\wedge$, $\vee$	2	bit-wise AND and bit-wise OR
\e, \pi	0	the E and PI numbers
min,max,sum,avg	1	take a list of numbers; return min, max, sum, and average
count,last	1	take a list; return the number of members and the last member
nth	2	take a number, N, and a list; return the N-th element of the list
delete	2	delete element (arg1) from list (arg2); return the resulting list
reverse	1	reverse list

Table 4: Arithmetic functions that can be used in arithmetic expressions

```
p(=?X+3) :- q(?X).  // alternatively: p(\is ?X+3) :- q(?X).

?- p(?X).


?X = 5
```

Note that if we used the rule `p(?X+3) :- q(?X)` instead (i.e., without the prefix "`=`" or "`\is`") then the answer to the above query would have been `?X = 2+3` because arithmetic expressions are not evaluated by default.

The following example illustrates the inline evaluation feature in a query (it works identically in rule bodies):

```
q(2), p(5).

?- q(?X), p(=?X+3).  //  alternatively: ?- q(?X), p(\is ?X+3).
```

```
Yes
```

Perhaps the most interesting way to use this feature is in combination with the string concatenation operator introduced earlier. This combination provides Java-like printing facility as follows:

```
p(a,b).

?- p(?X,?Y), writeln(= result ||': '|| ?X|| + || ?Y)@\io.

result: a+b
```

Observe that *all* arithmetic functions, as well as list/string concatenation, list operations, etc., are allowed in inline expressions. For instance,

```
p(=sin(5)+atan(3)).

?- p(?X).

 Resut: ?X = 0.2901


q(a,b), r([1,2,3],[1,3],[3,8]).

p(=?X||?Y, \is ?Z++?V--?W) :- q(?X,?Y), r(?Z,?V,?W).

?- p(ab,[1,2,3,1]).

Yes
```

Note: Aggregate operators described in Section 28 and user-defined functions (UDFs) described in Section 23 are *always* evaluated when they are found in an argument position, so they never need to be prefixed with an "=" or an `\is`. For instance,

```
\udf f(?X) := =?X+5 \if p(?X).
p({1,2,3}).
?- insert{q(6)}.
?- insert{r(=f(?X))}.   // UDF f/1 used inline without a =
?- q(sum{?V|p(?V)}).   // Yes. Aggr sum{?V|p(?V)} used inline without a =
?- r(?Z).              // ?X = 6,7,8
```

7.6 Quasi-constants and Quasi-variables

In some cases the developer might require the knowledge base to refer to the information about the source code where the various ℰRGO statements occur. To this end, the compiler provides a number of *quasi-constants*, which get substituted for real constants at compile or a loading time. The supported constants are:

- `\#`, `\#1`, `\#2`, etc. – Skolem constants.
- `\@!` – the Id of the rule where this quasi-constant occurs.
- `\@` – the module into which the file containing this quasi-constant is loaded. In the ℰRGO shell, this quasi-constant is set to be the symbol `main`.
- `\@F` – the file in which this quasi-constant occurs. In the ℰRGO shell, this quasi-constant is set to be the symbol `'(runtime)'`.
- `\@L` – the line in the source code on which this quasi-constant occurs.

Some of these constants are further illustrated in later in this manual. A typical use of these constants is to put them somewhere in place of a constant or a variable. For instance,

```
p(\@F,?X,?Y) :- ?X = \@L, ?Y = \@!.
```

In addition, ℰRGO provides a number of *quasi-variables*. Quasi-variables are similar to quasi-constants in that ℰRGO replaces them by constants. However, this happens at *run time*, during the evaluation. The supported quasi-variables are:

- `\?C` – the module from which the given rule was called. Used in the bodies of rules and queries.
- `\?F` – the file name from which a sensor was called. Used only in the bodies of the rules that are part of a sensor definition.
- `\?L` – the line number from which a sensor was called. Used only in the bodies of the rules that are part of a sensor definition.

7.7 Synonyms

For better readability by non-experts and for documentation purposes, ℰRGO provides the following useful mnemonic keywords, which can be used in place of the most commonly used symbols:

Symbol	Synonym
:	\isa, \memberof
::	\sub, \subclassof
=>	\hastype
->	\hasvalue
+>>	\contains

7.8 Reserved Symbols

$\mathcal{E}$RGO reserves all symbols that begin with the backslash, such as `\or`, `\hastype`, `\io`, `\neg`, etc. In addition, the symbols `->`, `=>`, `==>`, `<==`, `<==>`, `~~>`, `<~~`, `<~~>`, `:-`, `-:`, `!-`, `?-`, `?`, `|`, `#`, “,”, “;”, “:”, “::”, `->->`, `-»`, `+»`, arithmetic operators, `\/`, and `/\` are also reserved and cannot be used as names of methods, function symbols, predicates, and the like.

8 Path Expressions

In addition to the basic F-logic syntax, the $\mathcal{E}$RGO system also supports *path expressions* to simplify object navigation along value-returning method applications, and to avoid explicit join conditions [6]. The basic idea is to allow the following *path expressions* wherever Id-terms are allowed:

7. `O.M`

The path expression in (7) refers to an object $\mathtt{R_0}$ for which $\mathtt{O[M\text{->}R_0]}$ holds. The symbols `O` and `M` stand for an Id-term or a path expression. As a special case, `M` can be a method that takes arguments. For instance, $\mathtt{O.M(P_1,\ldots,P_k)}$ is a valid path expression.

Path expressions associate to the left, so `a.b.c` is equivalent to `(a.b).c`, which specifies the object `o` such that $\mathtt{a[b\text{->}x]} \wedge \mathtt{x[c\text{->}o]}$ holds (note that $\mathtt{x} = \mathtt{a.b}$). To change that, parentheses can be used. For instance, `a.(b.c)` is that object `o1` for which $\mathtt{b[c\text{->}x1]} \wedge \mathtt{a[x1\text{->}o1]}$ holds (note that in this case, $\mathtt{x1} = \mathtt{b.c}$). In general, `o` and `o1` can be different objects. Note also that in `(a.b).c`, `b` is a method name, whereas in `a.(b.c)` it is used as an object name and `b.c` as a method. Observe that function symbols can also be applied to path expressions, since path expressions, like Id-terms, represent objects. Thus, `f(a.b)` is a valid expression.

Note: Since a path expression represents an object Id, it can appear wherever an oid can, and it *cannot* appear in place of a truth-valued expression (e.g., a subquery). Thus,

```
?- ?P.authors.
```

is illegal. To use a path expression as a query, square brackets must be attached. For instance, the following are legal queries:

```
?- ?P.authors[].
?- ?P.authors[name->?N].
```

As path expressions and frames can be arbitrarily nested, this leads to a concise and flexible specification language for object properties, as illustrated in the following example. □

Example 8.1 (Path Expressions) Consider again the schema given in Figure 1. If $?n$ represents the name of a person, the following path expression is a query that returns all editors of conferences in which $?n$ had a paper:

```
?- ?P[authors->?[name->?n]].at_conf.editors[].
```

Likewise, the answer to the query

```
?- ?P[authors->?[name->?n]].at_conf[editors->?E].
```

is the set of all pairs (P,E) such that P is (the logical oid of) a paper written by $?n$, and E is the corresponding proceedings editor. If we also want to see the affiliations of the above editors, we only need to modify our query slightly:

```
?- ?P[authors->?[name->?n]].at_conf[year->?Y].editors[affil(?Y)->?A].
```

Thus, $\mathcal{E}$RGO path expressions support navigation along the method application dimension using the operator ".". In addition, intermediate objects through which such navigation takes place can be selected by specifying the properties of such objects inside square brackets.[5]

To access intermediate objects that arise implicitly in the middle of a path expression, one can define the method self as

```
?X[self->?X].
```

and then simply write ...`[self->?O]`... anywhere in a complex path expression. This would bind the Id of the current object to the variable `?O`.

[5] A similar feature is used in other languages, e.g., XSQL [8].

Example 8.2 (Path Expressions with self) To illustrate the convenience afforded by the use of the `self` attribute in path expressions, consider the second query in Example 8.1. If, in addition, we want to obtain the names of the conferences where the respective papers were published, that query can be reformulated as follows:

```
?X[self->?X].
?- ?P[authors->?[name->?n]].at_conf[self->?C,year->?Y].editors[affil(?Y)->?A].
```

9 Set Notation

The original F-logic [9] permitted convenient set notation as return values of set-valued methods. For instance,

```
John[children->{Mary,Bob}]
```

is a shortcut for the conjunction

```
John[children->Mary], John[children->Bob]
```

whether this expression occurs in the head or the body of a rule. $\mathcal{E}$RGO makes a leap forward in this direction and permits set notation anywhere a path expression is allowed. Here are some examples of what is possible:

```
{Mary,Joe}:{Student,Worker}.
happily_married(?X,?Y) :- person({?X,?Y}), ?X[{spouse,loves}->?Y].
child(John,{Mary,Kate,Bob}).
```

The above statements are, respectively, shortcuts for

```
Mary:Student, Mary:Student, Joe:Student, Joe:Worker.
happily_married(?X,?Y) :- person(?X), person(?Y),
                          ?X[spouse->?Y], ?X[loves->?Y].
child(John,Mary), child(John,Kate), child(John,Bob).
```

For some more extreme examples, consider these:

```
{p,q}(a,g(f({b,c}))).
r({?X,?Y}) :- {p,q}(?,?(f({?X,?Y}))).
{a,m}[{prop1,prop2} -> {1,{2,3},4}].
```

Portions Copyright © 2013–2017 Coherent Knowledge

These are shortcuts for the following statements, respectively:

```
p(a,g(f(b))), p(a,g(f(c))), q(a,g(f(b))), q(a,g(f(c))).
r(?X), r(?Y) :- p(?,?(f(?X))), p(?,?(f(?Y))), q(?,?(f(?X))), q(?,?(f(?Y))).
a[prop1->{1,2,3,4}, prop2->{1,2,3,4}], m[prop1->{1,2,3,4}, prop2->{1,2,3,4}].
```

The set term `{1,{2,3},4}` above is just a less readable way of writing `{1,2,3,4}`, by the way—there are no true set-objects in $\mathcal{E}$RGO or F-logic. Other extremely useful cases of the use of the set notation involve equality and related operators. For instance,

```
?X={?Y,?Z,?W}
?X!={1,2,3}
```

instead of the much more tedious

```
?X=?Y, ?Y=?Z, ?Z=?W
?X!=1, ?X!=2, ?X!=3
```

10 Typed Variables

Apart from the regular variables, $\mathcal{E}$RGO supports *typed variables*. Typed variables can be bound to classes, which will restrict them so that they will be unifiable only with members of the specified classes. This includes the classes associated with all the primitive data types such as `\integer`, `\real`, `\time`, `\list`, `\duration`, etc., which are discussed in Section 39. The syntax is `?Var^^Class`. The class can be a class name (which may have variables) or a class expression. For instance,

```
father(?X^^\integer,?Y^^Person) :- parent(?X,?Y), male(?X).
grandfather(?X^^\integer,?Y^^Person) :- father(?X,?Z^^Person),parent(Z,Y).
?- foo(?Y^^(A,(B-C) ; \dateTyme).
```

As seen from the example, the type declaration can appear both in the head and the body of a rule. The semantics is that once the variable is bound then the binding is checked for belonging to the specified class. Theoretically this is equivalent to making a test like `?X:Class` but in practice typed variables can be much more efficient. Testing `?X:Class` means enumerating the entire class `Class`, which can be expensive when the class is large and may not terminate, if the class is infinite (e.g., `?X:\integer`).

The `?Var^^Class` should be understood as a declaration, so it should be done only once per variable in the same rule. Multiple declarations are treated as intersections. For instance,

```
?- ?X^^foo=?, ?X^^moo=?.

?X = ?_h6272 { \$typed variable : type = (moo  ','  foo) }

?- ?X^^foo=?Y^^moo.

?X = ?_h5593 { \$typed variable : type = (moo  ','  foo) }
?Y = ?_h5593 { \$typed variable : type = (moo  ','  foo) }
```

Here are some additional examples:

```
{1,2}:foo, {2,3}:moo.                        // facts

?- ?^^foo = 4.                               // false
?- ?^^foo = 1                                // true
?- ?X^^foo = ?^^moo, ?X=3                    // false
?- ?X^^foo = ?^^moo, ?X=2                    // true
?- ?^^(foo-moo) = 2  // false
?- ?^^(foo-moo) = 1  // true
?- ?^^((foo-moo);(moo-foo)) = 2              // false
?- ?^^((foo-moo);(moo-foo)) = 3              // true
?- ?X^^\real=?, ?X^^\short=1                 // false
?- {1,2,-3}:{\integer,\short,\long},         // true
      {1.2,3.4}:{\real,\float,\double},
      [a,3] = ?X^^\list,
      ?X:\list,
      "abc"^^\charlist=?Y,
      ?Y:\charlist,
      writeln(test6=ok)@\plg.
```

11 Truth Values and Object Values

Id-terms, terms and path expressions can be also understood as objects. This is clear for Id-terms. The object interpretation for path expressions of the form (7) was given on page 29. On the other hand, frame formulas, class membership, and subclassing are typically understood as truth-valued formulas. However, there also is a natural way to interpret them as objects. For example, `o:c[m->r]` has object value `o` and some truth value. However, unlike the object value, the truth value depends on the database (on whether `o` belongs to class `c` in the database and whether the value of the attribute `m` is, indeed, `r`).

Although previously we discussed only the object interpretation for path expressions, it is easy to see that they have truth values as well, because a path expression corresponds to a conjunction of F-logic frames. Consequently, all frame literals of the form (1) through (7) have dual readings: As logical formulas (*the deductive perspective*), and as expressions that represent one or more objects (*the object-oriented perspective*). Given an intended model, $\mathcal{I}$, of an F-logic knowledge base, an expression has:

- An *object value*, which yields the Id(s) of the object(s) that are reachable in $\mathcal{I}$ by the corresponding expression, and
- A *truth value*, like any other literal of the language.

An important property that relates the above interpretations is: a frame, r, evaluates to *false* if $\mathcal{I}$ has no object corresponding to r.

Consider the following path expression and an equivalent, decomposed expression:

$$\texttt{a.b[c -> d.e]} \quad \Leftrightarrow \quad \texttt{a[b -> ?X}_{\texttt{ab}}\texttt{]} \wedge \texttt{d[e -> ?X}_{\texttt{de}}\texttt{]} \wedge \texttt{?X}_{\texttt{ab}}\texttt{[c -> ?X}_{\texttt{de}}\texttt{]}. \tag{2}$$

Such decomposition is used to determine the truth value of arbitrarily complex path expressions in the *body* of a rule. Let *obj*(`path`) denote the Ids of all objects represented by the path expression. Then, for (2) above, we have:

$$obj(\texttt{d.e}) = \{\texttt{x}_{\texttt{de}} \mid \mathcal{I} \models \texttt{d[e -> x}_{\texttt{de}}\texttt{]}\}$$

where $\mathcal{I} \models \varphi$ means that φ holds in $\mathcal{I}$. Observe two formulas can be equivalent, but their object values might be different. For instance, `d[e -> f]` is equivalent to `d.e` as a formula. However, *obj*(`d.e`) is f, while *obj*(`d[e -> f]`) is `d`.

In general, for an F-logic database $\mathcal{I}$, the object values of ground path expressions are given by the following mapping, *obj*, from ground frame literals to sets of ground oids (t, o, c, d, m can be oids or path expressions):

$$\begin{array}{rcl}
obj(t) & := & \{t \mid \mathcal{I} \models t[]\}, \text{ for a ground Id-term } t \\
obj(o[\ldots]) & := & \{o1 \mid o1 \in obj(o), \mathcal{I} \models o1[\ldots]\} \\
obj(o:c) & := & \{o1 \mid o1 \in obj(o), \mathcal{I} \models o1:c\} \\
obj(c::d) & := & \{c1 \mid c1 \in obj(c), \mathcal{I} \models c1::d\} \\
obj(o.m) & := & \{r1 \mid r1 \in obj(r), \mathcal{I} \models o[m \texttt{->} r]\}
\end{array}$$

Observe that if `t[]` does not occur in $\mathcal{I}$, then *obj*(t) is $\emptyset$. Conversely, a ground frame r is called *active* if *obj*(r) is not empty.

Portions Copyright © 2013–2017 Coherent Knowledge

Meta-predicates. Since path expressions can appear wherever Id-terms are allowed, the question arises whether a path expression is intended to indicate a truth value or an object value. For instance, we may want to call a predicate `foobar/1`, which expects as an argument a formula because the predicate calls this formula as part of its definition. For instance, the predicate may take a formula and a variable that occurs in that formula and joins this formula with some predicate using that variable:

```
foobar(?Form,?Var) :- ?Form, mypred(?Var).
?- foobar($a[b->?X], ?X).
```

Here `$a[b->?X]` is a `reification` of the formula `a[b->?X]` (see Section 16.2), i.e., an object that represents the formula. $\mathcal{E}$RGO does not allow one to write `foobar(a[b->?X],?X)` because this notation proved to be error-prone and confusing to the user in the previous versions of the system.

12 Boolean Methods

As a syntactic sugar, $\mathcal{E}$RGO provides boolean methods, which can be considered as value-returning methods that return some fixed value, e.g., `void`. For example, the following facts:

```
John[is_tall -> void].
John[loves(tennis) -> void].
```

can be simplified as boolean methods as follows:

```
John[is_tall].
John[loves(tennis)].
```

Conceptually, boolean methods are statements about objects whose truth value is the only concern. Boolean methods do not return any value (not even the value `void`). Therefore, boolean methods `cannot` appear in path expressions. For instance, `John.is_vegetarian`, where `is_vegetarian` is a binary method, is illegal.

Like other methods, boolean methods can be inherited, if specified as part of the class information:

```
Buddhist[|is_vegetarian|].
John:Buddhist.
```

The above says that all Buddhists are vegetarian by default and John (the object with oid `John`) is a Buddhist. Since `is_vegetarian` is specified as a property of the entire class `Buddhist`, it follows that John is also a vegetarian, i.e., `John[is_vegetarian]`.

Portions Copyright © 2013–2017 Coherent Knowledge

12.1 Boolean Signatures

Boolean methods can have signatures like the value-returning methods. These signatures can be specified as part of the object-level information (in which case they apply to specific objects) or as part of the class information (in which case they apply to all objects in the class and to all subclasses):

```
Obj[=>Meth]
Class[|=>Meth|]
```

The first statement refers to `Class` as an individual object, while the second is a statement about the type of the object `Class` as a class. It thus is inherited by every member of `Class` and all its subclasses. For instance,

```
Person[|=>loves(game)|].
```

13 Skolem Symbols

For applications where the particular names for oids are not important, $\mathcal{E}$RGO provides the quasi-constants `\#` and `\#1`, `\#2`, etc., to automatically generate a new *Skolem constant* or a *Skolem function symbol* to test for such automatically generated constants. We call such symbols *Skolem symbols*. Skolem symbols are interpreted *differently* in the rule heads and bodies.

Skolems in rule heads and descriptors outside of the reification operator. Outside of reification (see Section 16.2 for the details of reification), the head-occurrences of Skolem symbols are interpreted, as described below.

Each occurrence of `\#` in the rule head or a rule descriptor (see Section 34 to learn about rule descriptors) represents a *new* Skolem constant, which is unique throughout the source code.

Uniqueness is achieved through the use of a special "weird" naming schema for such oids, which internally prefixes them with several `"_$"`s. However, as long as the user does not use a similar naming convention (who, on earth, would give names that begin with lots of `"_$"`s?), uniqueness is guaranteed.

For example, in the following example

```
\#[ssn->123, father->\#[name->John, spouse->\#[name->Mary]]].
foo[\#(?X)->?Y] :- bar[?Y->?X].
```

the compiler will generate unique oids for each occurrence of `\#`. Note that, in the second clause, only one oid is generated and it serves as a method name.

Co-reference of Skolem symbols. In some situations, it is necessary to be able to create a new oid and use it within the same rule head or a fact multiple times. Sometimes the *same* Skolem might need to be used across *several* different rules and facts. Since such an oid needs to be referenced inside the same clause, it is no longer possible to use `\#`, because each occurrence of `\#` causes the compiler to generate a new Skolem symbol. To solve this problem, $\mathcal{E}$RGO allows *numbered Skolem symbols*, which are of the form `\#132`, *i.e.*, `\#` with a number attached to it. For instance,

```
\#1[ssn->123, father->f(\#1)[name->John, spouse->\#[name->Mary]]].
\#1[self->\#1].
```

The first time the compiler finds `\#1` in the first clause above, it will generate a new oid. However, the second occurrence of `\#1` in the *same* clause (i.e., `f(\#1)`) will use the oid previously generated for the first occurrence. On the other hand, occurrences of `\#1` in *different* clauses are substituted with different oids. Thus, the occurrences of `\#1` in the first and second clauses above refer to different objects.

The same numbered Skolem can be co-referenced in *different* facts and rules. The general rule is that the scope of a numbered Skolem is a $\mathcal{E}$RGO sentence terminated with a period. For instance, in the following

```
(h1(?X,\#1) :- b1(?X),c1(?X)),
p(a,\#1),
q(b(\#1)),
(h2(?X,\#1,?Y) :- b2(?X,?Y)).
```

the four occurrences of `\#1` represent the same Skolem constant. Note that the first and the last statement above are rules, while the two middle statements are facts. The reason why `\#1` is because all four statements are withing the same scope: all four are terminated by the same period. Note also that each rule must be enclosed in parentheses or else ambiguity arises. For instance,

```
h1(?X,\#1) :- b1(?X),c1(?X),
p(a,\#1),
q(b(\#1)).
```

would be interpreted as a single rule whose body has four literals, *not* as a statement that contains a rule `h1(?X,\#1) :- b1(?X),c1(?X)` and two additional, separate facts. Also note that it is not necessary to put each of the four sub-statement above on a separate line—this was done for readability.

Portions Copyright © 2013–2017 Coherent Knowledge

Generating Skolems at run time. Normally, Skolem constants are generated at compile time without regard for the oids that might exist at run time. Sometimes it is necessary to generate a Skolem constant at *run time*. This can be accomplished with the `skolem{...}` built-in. For instance,

```
ergo>  skolem{?X}.

?X = _$_$_ergo'dyn_skolem308   (#)

1 solution(s) in 0.0000 seconds

Yes

ergo>  skolem{?X}.

?X = _$_$_ergo'dyn_skolem309  (#)
```

etc.

Skolems in subgoals of rule bodies outside of the reification operator. In a body subgoal outside of the reification operator, a Skolem symbol (numbered or not) is interpreted differently that in the head: not as a new Skolem constant, but as a *test* of whether or not the corresponding argument is bound to a Skolem constant. One can think of it as a variable that is tested for being bound to a Skolem constant. Numbered Skolems within the same rule are interpreted as the same variable, so the occurrences of the same numbered Skolem are expected to be bound to the same constant. For instance, in

```
?- insert{abc[prop1->\#, prop2->\#3, prop3->\#3],
          cde[prop1->\#, prop2->\#3, prop3->\#]}.
test1 :- abc[prop1->\#, prop2->\#5, prop3->\#5].
test2 :- cde[prop1->\#, prop2->\#5, prop3->\#5].
```

Here the query `test1` will return `true` because in the object `abc` the properties `prop1` and `prop2` are bound to the same Skolem constant. In contrast, in `cde`, these properties are *not* bound to the same Skolem (since `\#3` and `\#` are different Skolems), so the query `test2` fails.

Note that, as a test, both `\#` and `\#`N, where N is an positive integer, match not only Skolem constants, but also Skolem functions. For instance, the query

```
?- insert{p(\#(a,b))}, p(\#).
```

succeeds because `\#(a,b)` is a term obtained from an application of a Skolem function, `\#`, to its arguments.

Note that, since Skolems have a different semantics in the rule head and in the body, the following, somewhat counter-intuitive situation might occur:

```
p(\#1) :- q(\#1).
```

Here `\#1` inside `p(...)` is a constant, while in inside `q(...)` it is a check that the argument of `q/1` is bound to a Skolem term.

Skolems in reified rules. If a reified rule, R, occurs somewhere in the body or a head of another, normal rule (or fact), the Skolem symbols that occur in R are interpreted the same way as if the rule were outside of reification. That is, the head occurrences in R are interpreted as Skolem constants and the body occurrences are interpreted as tests for Skolems. For instance, in

```
head1(${head11(?X,\#) :- body11(\#1,\#1)}).
head2 :- body2(${head22(?X,\#) :- body22(\#1,\#1)}).
```

the symbol `\#` in the fact `head11(...)` is a Skolem constant, while the two occurrences of the symbol `\#1` in `body11` test if `body11(?X,?Y)` can succeed with `?X` and `?Y` bound to the same Skolem constant. The interpretation of `\#` and `\#1` in the second rule above is similar.

Skolems in reified subgoals. If a reified subgoal (not rule), G, contains a Skolem symbol, that symbol is interpreted as a Skolem constant. This is true both if G occurs in the head of a rule or in the body. Note that this implies that if one later uses G as a *query* then this query is likely to fail. For instance, in

```
?- insert{p(a), p(\#)}.
?- ?X = ${p(\#)}, ?X.
```

the second query will fail because the occurrence of `\#` there is a Skolem constant that differs from the Skolem constant in the first rule. However,

```
?- ?X = ${p(\#)}, insert{?X}, p(\#).
```

succeeds because when the reified subgoal `${p(\#)}` is inserted into the database, the Skolem symbol is interpreted as a constant and the subsequent check `p(\#)` tests that the argument of `p` is bound to a Skolem constant.

Portions Copyright © 2013–2017 Coherent Knowledge

Numbered Skolems are interpreted the same way as non-numbered ones except that different occurrences of the same numbered Skolem symbol are treated as identical within the same rule or query. For instance, below,

```
?- ?X = ${p(\#2,\#2)}, insert{?X}, p(\#7,\#7).
?- erase{p(?,?)}, ?X = ${p(\#2,\#)}, insert{?X}, p(\#7,\#7).
```

the first query succeeds, while the second fails.

More on Skolems. A Skolem constant can also be used as a function symbol and even a predicate symbol. For instance, `\#(a,b)`. Since $\mathcal{E}$RGO terms are actually HiLog terms, we can also have higher-order Skolem functions, as in these terms:

```
\#1(foo,bar)(123)
\#(a,c,d)(o,w)(1,2)
```

Sometimes it is useful to know whether a particular term is a *Skolem term*, i.e., whether it is a Skolem constant or is constructed using a Skolem function or a higher-order Skolem function. To this end, $\mathcal{E}$RGO provides a special built-in class, `\skolem`. For instance, in the following case

```
p(\#1(foo,bar)(123)).
?- p(?X), ?X:\skolem.
```

the variable `?X` gets bound to a higher-order Skolem term, so the query succeeds. But the query `?- f(a):\skolem` fails, since `f` is not a Skolem symbol. In addition to this, $\mathcal{E}$RGO provides a builtin for checking if a particular symbol (symbol, not just any term) is Skolem. For instance,

```
?- skolem{?X}, isskolem{?X}.
```

will succeed. Of course, `?- skolem?X, ?X:\skolem` will also succeed, but the `\skolem` class contains all Skolem terms, while the `isskolem{...}` primitive is true of Skolem constants only.

Finally, we note that both the `isskolem{...}` primitive and the `\skolem` class require that the argument (the class instance being tested) must be bound.

14 Testing Meta-properties of Symbols and Variables

Sometimes it useful to be able to be able to find out to what kind symbol a particular variable is bound or even whether that variable is bound to a term or not. In Section 13, we already saw the

`isskolem{...}` primitive, which can tell whether a particular symbol is a Skolem constant. Here is a summary of all such meta-predicates. For the meaning of the IRI and string constants, please refer to Section 39.

- `isnumber{Arg}` — tests whether the argument is (or is bound to) a number.
- `isinteger{Arg}` — tests whether the argument is (or is bound to) an integer number.
- `isfloat{Arg}` — tests whether the argument is (or is bound to) a floating point number.
- `isdecimal{Arg}` — tests if the argument is (or is bound to) a decimal number. At present, this is the same as `isnumber{Arg}`.
- `isatom{Arg}` — tests whether the argument is (or is bound to) a Prolog atom.
- `iscompound{Arg}` — tests whether the argument is (or is bound to) a compound term, i.e., a term that has a non-zero-ary function symbol.
- `isatomic{Arg}` — tests if the argument is (or is bound to) a Prolog atom or a number.
- `islist{Arg}` — tests if the argument is (or is bound to) a list term.
- `ischarlist{Arg}` — tests whether the argument is (or is bound to) a list or ASCII characters.

- `isiri{Arg}` — tests whether the argument is (or is bound to) an IRI data type.
- `isstring{Arg}` — tests whether the argument is (or is bound to) a string data type.
- `issymbol{Arg}` — tests whether the argument is (or is bound to) an abstract symbol. An abstract symbol is any atom that is not an internal representation of a string or an IRI. (These internal representations involve special unprintable characters and thus are unlikely to be used by a normal user directly.)
- `isvar{Arg}` — tests if the argument is an unbound variable.
- `isnonvar{Arg}` — tests if the argument is not an unbound variable.
- `isground{Arg}` — tests if the argument is a ground term (or is bound to one).
- `isnonground{Arg}` — tests if the argument is not a ground term.
- `variables{Term,List}` — binds `List` to the list of all the variables that occur in `Term`.
- `cloneterm{Term,ClonedTerm}` — creates a copy of `Term` with the same constants and function symbols, but variables are consistently renamed to become new variables.

Portions Copyright © 2013–2017 Coherent Knowledge

In addition, some of the above primitives have *delayable* 2-argument versions. A delayable version differs in that if the first argument is a variable then evaluation of such a builtin is delayed until the argument gets bound. If the argument does not get bound at the end of the query computation, then the outcome depends on the second *mode*-argument: if the mode is `must` then an error is issued; if the mode is `wish` then the builtin is quietly evaluated to *false*. For instance,

```
ergo>  isinteger{?X}.

No

ergo>  isinteger{?X}, ?X=1.

No

ergo>  isinteger{?X,wish}, ?X=1.

?X = 1

Yes

ergo>  isinteger{?X,must}, ?X=1.

?X = 1

Yes

ergo>  isinteger{?X,must}.

++Abort[Ergo]> in file (runtime) on line 1: instantiation error in builtin:
                isinteger{_h3464}; unbound argument

ergo>  isinteger{?X,wish}.

No
```

Thus, the delayable versions of the above primitives are insensitive to the order in which they appear in the rule body, which makes them sometimes easier to use and renders their behavior more logical. On the other hand, the delayable versions cannot serve as *guards* for subsequent evaluations. For example, if `foo(?X)` expects `?X` to be an integer then

```
... :- ..., isinteger{?X,must}, foo(?X), ...
```

will not prevent calling `foo/1` with `?X` an unbound variable, since `isinteger{Arg,Mode}` will be delayed past the moment `foo(?X)` is evaluated. Using `isinteger{?X}` instead will do the job.

The delayable versions of the aforesaid builtins are listed below. The delayable versions of `isvar{Arg,Mode}`, `isnonvar{Arg,Mode}`, `isground{Arg,Mode}`, and `isnonground{Arg,Mode}` have slightly different semantics.

- `isnumber{Arg,Mode}` — the delayable version of `isnumber{Arg}`.
- `isinteger{Arg,Mode}` — the delayable version of `isinteger{Arg}`.
- `isfloat{Arg,Mode}` — the delayable version of `isfloat{Arg}`.
- `isdecimal{Arg,Mode}` — the delayable version of `isdecimal{Arg}`.
- `isatom{Arg,Mode}` — the delayable version of `isatom{Arg}`.
- `iscompound{Arg,Mode}` — the delayable version of `iscompound{Arg}`.
- `isatomic{Arg,Mode}` — the delayable version of `isatomic{Arg}`.
- `islist{Arg,Mode}` — the delayable version of `islist{Arg}`.
- `ischarlist{Arg,Mode}` — the delayable version of `ischarlist{Arg}`.
- `isiri{Arg,Mode}` — the delayable version of `isiri{Arg}`.
- `isstring{Arg,Mode}` — the delayable version of `isstring{Arg}`.
- `issymbol{Arg,Mode}` — the delayable version of `issymbol{Arg}`.
- `isvar{Arg,Mode}` — the delayable version of `isvar{Arg}`. If the argument `?X` in `isvar{?X,must}` *at the end of the query computation* is *not* an unbound variable, an error is issued. Otherwise, the result is true. For `isvar{?X,wish}`, the evaluation is delayed and then `?X` is tested. If it is unbound, the result is true; otherwise, it is false. No errors are issued.
- `isnonvar{Arg,Mode}` — the delayable version of `isnonvar{Arg}`. If `?X` in `isnonvar{?X,must}` *at the end of the query computation* is a variable, an error is issued. Otherwise, the result is true. For `isnonvar{?X,wish}`, the evaluation is delayed and then `?X` is tested. If it is bound, the result is true; otherwise, it is false. No errors are issued.
- `isground{Arg,Mode}` — the delayable version of `isground{Arg}`. If `?X` in `isground{?X,must}` *at the end of the query computation* is non-ground, an error is issued. Otherwise, the result is true. For `isground{?X,wish}`, the evaluation is delayed and then `?X` is tested. If it is bound to a ground term, the result is true; otherwise, it is false. No errors are issued.

- `isnonground{Arg,Mode}` — the delayable version of `isnonground{Arg}`. If `?X` in `isnonground{?X,must}` *at the end of the query computation* is ground, an error is issued. Otherwise, the result is true. For `isnonground{?X,wish}`, the evaluation is delayed and then `?X` is tested. If it is bound to a ground term, the result is false; otherwise, it is true. No errors are issued.

15 Multifile Knowledge Bases

$\mathcal{E}$RGO supports many ways in which a knowledge base can be modularized. First, it can be split into many files with separate namespaces. Each such file can be considered an independent library, and the different libraries can call each other. In particular, the same method name (or a predicate) can be used in different files and the definitions will not clash. Second, a file can be composed of several files, and these files can be included by the preprocessor prior to the compilation. In this case, all files share the same namespace in the sense that the different rules that define the same method name (or a predicate) in different files are assumed to be part of one definition. Third, $\mathcal{E}$RGO knowledge bases can invoke Prolog modules and vice versa. In this way, a large system can be built partly in Prolog and partly in $\mathcal{E}$RGO.

We discuss each of these modularization methods in turn.

15.1 $\mathcal{E}$RGO Modules

A $\mathcal{E}$RGO *module* is an abstraction that allows a large knowledge base to be split into separate libraries that can be reused in multiple ways in the same system. Formally, a module is a pair that consists of a *name* and a *contents*. The name must be an alphanumeric symbol (the underscore character, `_`, is also allowed), and the contents consists of part of the knowledge base that is typically loaded from some file (but can also be constructed runtime by inserting facts into another module).

The basic idea behind $\mathcal{E}$RGO modularization is that reusable code libraries are to be placed in separate files. To use a library, it must be *loaded into a module*. Other parts of the knowledge base can then invoke this library's methods by providing the name of the module (and the method/predicate names, of course). The exported methods and predicates can be called by other parts of the knowledge base. (A module can have non-public methods, if the module is encapsulated — see Section 15.12.) In this way, the library loaded into a module becomes that module's content.

Note that there is no a priori association between files and modules. Any file can be loaded into any module and the same file can be loaded even into two different modules at the same time. The same module can be reused during the same $\mathcal{E}$RGO session by loading another file into that module. In this case, the old contents is erased and the module gets new contents from the second file.

Portions Copyright © 2013–2017 Coherent Knowledge

In $\mathcal{E}$RGO, modules are completely decoupled from file names. A knowledge base knows only the module names it needs to call, but not the file names. Specific files can be loaded into modules by some other, unrelated bootstrapping code. Moreover, a knowledge base can be written in such a way that it invokes a method of some module without knowing that module's name. The name of the module can be passed as a parameter or in some other way and the concrete binding of the method to the module will be done at runtime.

This dynamic nature of $\mathcal{E}$RGO modules stands in sharp contrast to the module system of Prolog, which is static and associates modules with files at compile time. Moreover, to call a predicate from another module, that predicate must be imported explicitly and referred to by the same name.

As a pragmatic measure, $\mathcal{E}$RGO defines *three kinds of modules* rather than just one. The kind described above is actually just one of the three: the *user module*. As explained, these modules are decoupled from the actual code, and so they can contain different code at different times. The next kind is a *Prolog module*. This is an abstraction in $\mathcal{E}$RGO, which is used to call Prolog predicates. Prolog modules are static and are assumed to be closely associated with their code. We describe these modules in Section 15.7. (Do not confuse $\mathcal{E}$RGO Prolog modules — an abstraction used in the language of $\mathcal{E}$RGO— with Prolog modules, which is an abstraction used in Prolog.) The third type of modules are the $\mathcal{E}$RGO `system modules`. These modules are preloaded with $\mathcal{E}$RGO and provide useful methods and predicates (*e.g.*, I/O) and, thus, are also static. These modules are described in Section 15.9 and 46. The abstraction of system modules is a convenience provided by $\mathcal{E}$RGO, which enables users to perform common actions using standard names of predicates and methods implemented in those modules. The syntactic conventions for calling each of these types of modules are similar, but distinct.

15.2 Calling Methods and Predicates Defined in User Modules

If *literal* is a frame or a predicate defined in another user module, it can be called using the following syntax:

literal @ *module*

The name of the module can be any alphanumeric symbol.[6] For instance, `foo(a) @ foomod` tests whether `foo(a)` is true in the user module named `foomod`, and `Mary[children -> ?X]@genealogy` queries the information on `Mary`'s children available in the module `genealogy`. More interestingly, the module specifier can be a variable that gets bound to a module name at run time. For instance,

```
..., ?Agent=Zagat, ..., NewYork[dinner(Italian) -> ?X]@?Agent.
```

[6] In fact, any symbol is allowed. However, it cannot contain the quote symbol, "'".

A call to a literal with an unbound module specification or one that is not bound to a symbol will result in a runtime error.

When calling the literals defined in the same module, the `@`*module* notation is not needed, of course. (In fact, since knowledge bases do not know where they will be loaded, using the `@`*module* idiom to call a literal in the same module is difficult. However, it is easy to do with the help of the quasi-constant `\@`, which is described later, and is left as an exercise.)

The following rules apply when calling a literal defined in another module:

1. Literal reference cannot appear in a rule head or be specified as a fact. For example, the following will generate a parsing error

   ```
   John[father->Smith] @ foomod.
   foo(?X) @ foomod :- goo(?X).
   ```

   because defining a literal that belongs to another module does not make sense.

2. Module specification is distributive over logical connectives, including the conjunction operator, "`,`", the disjunction, "`;`", and the negation operators, "`\+`" and "`\naf`". For example, the formula below:

   ```
   (John[father->Smith], \naf Smith[spouse->Mary]) @ foomod
   ```

   is equivalent to the following formula:

   ```
   John[father->Smith] @ foomod, \naf (Smith[spouse->Mary] @ foomod)
   ```

3. Module specifications can be nested. The one closest to a literal takes effect. For example,

   ```
   (foo(a), goo(b) @ goomod, hoo(c)) @ foomod
   ```

   is equivalent to

   ```
   foo(a) @ foomod, goo(b) @ goomod, hoo(c) @ foomod
   ```

4. The module specification propagates to any frame appearing in the argument of a predicate for which the module is specified. For example,

   ```
   foo(a.b[c->d]) @ foomod
   ```

   is equivalent to

```
a[b->?X] @ foomod, ?X[c->d] @ foomod, foo(?X) @ foomod
```

5. Module specifications do not affect function terms that are not predicates or method names, unless such a specification is explicitly attached to such a term. For instance, in

```
?- foo(goo(a)) @ foomod.
```

`goo/1` refers to the same functor both in module `foomod` and in the calling module. However, if the argument is *reified* (i.e., is an object that represents a formula — see Section 16.2), as in

```
?- foo(${goo(a) @ goomod}) @ foomod.
```

then `foo/1` is assumed to be a meta-predicate that receives the query `goo(a)` in module `goomod` as a parameter. Moreover, module specification propagates to any reified formula appearing in the argument of a predicate for which the module is specified. For example,

```
?- foo(${goo(a)}) @ foomod.
```

is equivalent to

```
?- foo(${goo(a) @ foomod}) @ foomod.
```

15.3 Finding the Current Module Name

Since a $\mathcal{E}$RGO knowledge base can be loaded into any module, it does not have a priori knowledge of the module it will be executing in. However, the knowledge base can find its module at runtime using the quasi-constant `\@`, which is replaced with the current module name when the module is loaded. More precisely, if `\@` occurs anywhere as an oid, method name, value, etc., in file `foo.ergo` then when `foo.ergo` is loaded into a module, say, `bar`, then all such occurrences of `\@` are replaced with `bar`. For instance,

```
a[b->\@].
?- a[b->?X].

?X=main

Yes
```

15.4 Finding the Module That Invoked A Rule

Sometimes it is useful to find out which module called any particular rule at run time. This can be used, for example, when the rule performs different services for different modules. The name of the caller-module can be obtained by calling the primitive `caller{?X}` in the body of a rule. Alternatively, the `\?C` quasi-variable can be used. For instance,

```
p(?X) :- caller{?X}, (write('I was called by module: '), writeln(?X))@\prolog.
p(?X) :- ?X=\?C, (write('I was called by module: '), writeln(?X))@\prolog.
```

When a call to predicate `p(?X)` is made from any module, say `foobar`, and the above rule is invoked as a result, then the message "I was called by module: foobar" will be printed.

15.5 Loading Files into User Modules

$\mathcal{E}$RGO provides several commands for compiling and loading files into specified user modules.

Compilation. The command

?- compile{*myfile*>>*mymodule*}.

generates the byte code for the program to be loaded into the user module named `module`. In practice this means that the compiler generates files named *myfile_mymodule.P* and *myfile_mymodule.xwam* with symbols appropriately renamed to avoid clashes.

If no module is specified, the command

?- compile{*myfile*}.

compiles *myfile*.ergo for the default module `main`.

Loading. The above commands compile files without actually loading their contents into the in-memory knowledge base. To load a file, the following commands can be used:

?- [*myfile*].
or
?- load{*myfile*}.

This command loads the byte code of the program in the file *myfile*.ergo into the default user module main. If a compiled byte code of the program for the module main already exists and is newer than the source file, the byte code is used. If *myfile*.ergo is newer than the compiled byte code (or if the byte code does not exist), then the source file is recompiled and then loaded.

An optional module name can be given to tell ℰRGO to load the program into a specified module:

```
?- [myfile >> foomod].
?- load{myfile >> foomod}.
```

This loads the byte code of the ℰRGO program *myfile*.ergo into the user module named foomod. As with the previous form of that command, if a compiled byte code of the program for the module foomod already exists and is newer than the source file, the byte code is used. If *myfile*.ergo is newer than that compiled byte code (or if the byte code does not exist), then the source file is recompiled and then loaded.

The user can compile and load several program files at the same time: If the file was not compiled before (or if the program file is newer), the program is compiled before being loaded. For instance, the following command:

```
?- [mykb1, mykb2]
```

will load both mykb1 and mykb2 into the default module main. However, simply loading several knowledge bases into the same module is not very useful: the content of the last file will wipe out the code of the previous ones. This is a general rule in ℰRGO. Thus, loading multiple files is normally used in conjunction with the module targets:

```
?- ['mykb1.ergo', mykb2 >> foomod].
```

which loads mykb1.ergo into the module main and mykb2.ergo into the module foomod.

Adding to already loaded modules. Files can also be *added* to an existing module, as explained in the following subsection.

Note that the [...] command can also load and compile Prolog programs. The overall algorithm is as follows. If the file suffix is specified explicitly, the corresponding file is assumed to be a ℰRGO file, a Prolog file, or a byte code depending on the suffix: .ergo, .P, or .xwam. If the suffix is not given explicitly, the compiler first checks if mykb.ergo exists. If so, the file assumed to contain ℰRGO code and is compiled as such. If mykb.ergo is not found, but *mykb*.P or *mykb*.xwam is, the file is passed to Prolog for compilation.

Sometimes it is useful to know which user modules are loaded or if a particular user module is loaded (say, because you might want to load it, if not). To find out which modules are loaded at the present time, use the primitive `isloaded{...}`. For instance, the first query, below, succeeds if the module `foo` is loaded. The second query succeeds and binds `L` to the list of all user modules that are loaded at the present time.

```
?- isloaded{foo}.
?- ?L= setof{?X|isloaded{?X}}.
```

`setof` and other aggregate operators are discussed in Section 28.

One can also check which files are loaded in what modules and in what mode. The mode is `load` or `add`, and the file names are absolute path names. For instance,

```
?- isloaded{?F,?Module,?Mode}.
?F = /a/b/c/foo.flr
?Module = bar
?Mode = load

?F = /a/b/foo2.flr
?Module = bar
?Mode = add

?F = /a/b/d/bar.flr
?Module = main
?Mode = add
```

There is also a four-place version of `isloaded`: `isloaded{FileAbsName,Module,FileLocalName,Mode}`. The difference is that argument 3 is now the local version of the file name (without the directory part). This version is useful in many ways. For instance, it lets one find the full name of the current file. To do so, recall that $\mathcal{E}$RGO has a quasi-variable `\@F`, which represents the *local* file name. So, the query

```
?- isloaded{?FullFile,\@,\@F,?}.
```

will bind `?FullFile` to the absolute name of the file where the above query occurs.

Scratchpad code. In some cases—primarily for testing—it is convenient to be able to type up and load small excerpts of code into a running $\mathcal{E}$RGO session. To this end, the system provides special idioms, `[]`, `[>>module]`, `[+]`, and `[+>>module]`. This causes $\mathcal{E}$RGO to start reading input

clauses from the standard input and load them into the default module or the specified module. To indicate the end of the input, the user can type `Control-D` in Unix-like systems or `Control-Z` in Windows. For instance,

```
ergo>  [>>foo].
aaa[bbb->ccc].
?X[foo->?Y] :- ?Y[?X->bar].
Control-D
```

A word of caution. It is dangerous to place the `load{...}` command in the body of a rule if `load{...}` loads a file into the same module where the rule belongs. For instance, if the following rule is in module `bar`

```
p(X) :- ..., [foo>>bar], ...
```

then execution of such a rule is likely to crash Prolog. This is because this very rule will be wiped out before it finishes execution — something that XSB is not ready for. $\mathcal{E}$RGO tries to forewarn the user about such dangerous occurrences of `load{...}`, but it cannot intercept all such cases reliably.

15.6 Adding Rule Bases to Existing Modules

Loading a file into a module causes the knowledge base contained in that module to be erased before the new information is loaded. Sometimes, however, it is desirable to *add* knowledge (rules and facts) contained in a file to an existing module. This operation does not erase the old contents of the module. For instance each of the following commands

```
?- [+mykb >> foomod].
or
?- add{mykb >> foomod}.
```

will *add* the rules and facts contained in the file `mykb.ergo` into the module `foomod` without erasing the old contents. The following commands

```
?- [+mykb].
?- add{mykb}.
```

will do the same for module `main`. Note that, in the `[...]` form, loading and adding can be freely mixed. For instance,

```
?- [foo1, +foo2]
```

will first load the file `foo1.ergo` into the default module `main` and then add the contents of `foo2.ergo` to that same module.

Like the loading commands, the addition statements first compile the files they load if necessary. It is also possible to compile files for *later* addition without actually adding them. Since files are compiled for addition a little differently from files compiled for loading, we use a different command:

```
?- compileadd{foo}.
?- compileadd{foo >> bar}.
```

15.7 Calling Prolog from $\mathcal{E}$RGO

Prolog predicates can be called from $\mathcal{E}$RGO through the $\mathcal{E}$RGO module system $\mathcal{E}$RGO models Prolog programs as collections of static *Prolog modules*, i.e., from $\mathcal{E}$RGO's point of view, Prolog modules are always available and do not need to be loaded explicitly because the association between Prolog programs and modules is fixed.

@\prolog and @\plg. The syntax to call Prolog predicates is one of the following:

?- *predicate*`@\prolog(`*module*`)`

For instance, since the predicate `member/2` is defined in the Prolog module `basics`, we can call it as follows:

```
?- member(abc,[cde,abc,pqr])@\prolog(basics).
```

`\plg` instead of `\prolog` also works.

To use this mechanism, you must know which Prolog module the particular predicate is defined in. Some predicates are defined by programs that do not belong to any module. When such a Prolog program is loaded, the corresponding predicates become available in the default Prolog module. In XSB, the default module is called `usermod` and $\mathcal{E}$RGO can call such predicates as follows:

```
?- foo(?X)@\prolog(usermod).
```

Note that variables are not allowed in the module specifications of Prolog predicates, i.e.,

```
?- ?M=usermod, foo(?X)@\prolog(?M).
```

will cause a compilation error.

Some Prolog predicates are considered "well-known" and, even though they are defined in various Prolog modules, the user can just use those predicates without remembering the corresponding Prolog module names. These predicates (that are listed in the XSB manual) can be called from $\mathcal{E}$RGO with particular ease:

```
?- writeln('Hello')@\prolog
```

i.e., we can simply omit the Prolog module name (but parentheses must be preserved).

@\prologall and @\plgall. The Prolog module specification `@\prolog` has one subtlety: it does not affect the arguments of a call. For instance,

```
?- foo(f(?X,b))@\prolog.
```

will call the Prolog predicate foo/1. Recall that $\mathcal{E}$RGO uses HiLog terms to represent objects, while Prolog uses Prolog terms. Thus, the argument `f(?X,b)` above will be treated as a HiLog term. Although it looks like a Prolog term and, in fact, HiLog terms generalize Prolog terms, the internal representation of HiLog and Prolog terms is different. Therefore, if the fact `foo(f(a,b))` is defined somewhere in the Prolog program then the above query will fail, since a Prolog term `f(?X,b)` and a HiLog term `f(?X,b)` are *different* even though their textual representation in $\mathcal{E}$RGO is the same.

A correct call to `foo/1` in this case would be as follows:

```
?- foo(f(?X,b)@\prolog)@\prolog.
```

Here we explicitly tell the system to treat `f(?X,b)` as a Prolog term. Clearly, this might be too much writing in some cases, and it is also error prone. Moreover, bindings returned by Prolog predicates are Prolog terms and they somehow need to be converted into HiLog.

To simplify calls to Prolog, $\mathcal{E}$RGO provides another, more powerful primitive: `@\prologall`. In the above case, one can call

```
?- foo(f(?X,b))@\prologall.
```

without having to worry about the differences between the HiLog representation of terms in $\mathcal{E}$RGO and the representation used in Prolog.

One might wonder why is there the `@\prolog` module call in the first place. The reason is efficiency. The `@\prologall` call does automatic conversion between Prolog and HiLog, which

is not always necessary. For instance, to check whether a term, `f(a)`, is a member of a list, `[f(b),f(a)]`, one does not need to do any conversion, because the answer is the same whether these terms are HiLog terms or Prolog terms. Thus,

```
?-  member(f(a), [f(b),f(a)])@\prolog(basics).
```

is perfectly acceptable and is more efficient than

```
?-  member(f(a), [f(b),f(a)])@\prologall(basics).
```

$\mathcal{E}$RGO provides a special primitive, `p2h{...,...}`, which converts terms to and from the HiLog representation, and the knowledge engineer can use it in conjunction with `@\prolog` to achieve a greater degree of control over argument conversion. This issue is further discussed in Section 16.4.

Builtin Prolog predicates. There is a large number of low-level builtin predicates in the underlying Prolog system. Most of these builtins have been lifted to the $\mathcal{E}$RGO level in various ways and made easier to use. Other predicates were not lifted because they were subsumed by various $\mathcal{E}$RGO's constructs. Nevertheless, there still are XSB predicates whose lifting is not being planned because of their rare use, but which can still be useful in some cases. A partial list of such predicates appears in Section 51.

15.8 Calling $\mathcal{E}$RGO from Prolog

Since Prolog does not understand $\mathcal{E}$RGO syntax, it can call only predicates (not frames) defined in $\mathcal{E}$RGO knowledge bases. To expose such predicates to Prolog, they must be imported by the Prolog program.

15.8.1 Importing $\mathcal{E}$RGO Predicates into Prolog

To import a $\mathcal{E}$RGO predicate into a Prolog shell, the following must be done:

- The query

  ```
  ?- [flora2], bootstrap_ergo.
  ```

  must be executed first. If you are importing $\mathcal{E}$RGO predicates into a Prolog program, say `test.P`, and not just into a Prolog shell then the above must be executed *before compiling or loading* `test.P` and the following additional directive must appear near the top of `test.P`, *prior to any call* to $\mathcal{E}$RGO predicates:

```
:- import ('\\flimport')/1 from flora2.
```

- One of the following `'\\flimport'` queries must be executed in the shell:

  ?- '\\flimport' *flora-predicate/arity* as *xsb-name*(_,_,...,_)
  from *filename* >> *flora-module-name*
  ?- '\\flimport' *flora-predicate/arity* as *xsb-name*(_,_,...,_)
  from *flora-module-name*

  We will explain shortly which `'\\flimport'` query should be used in what situation. Note that a double-backslash is required in front of `flimport` because in Prolog the backslash is an escape character. This is also the reason why the backslash appears twice in the various commands in the rest of this section.

Note: You must let Prolog know the location of your installation of $\mathcal{E}$RGO. This is done by executing the prolog instruction `asserta(library_directory(path-to-flora))`. For instance

```
?- add_lib_dir(a('/home/me/ergo')).
```

before calling any of the $\mathcal{E}$RGO modules. Observe that `asserta` and *not* `assert` must be used.

The first form for `'\\flimport'` above is used to both import a predicate and also to load the file containing it into a given $\mathcal{E}$RGO user module. The second syntax is used when the $\mathcal{E}$RGO knowledge base is already loaded into a module and we only need to import the corresponding predicate.

In `'\\flimport'`, *flora-predicate* is the name of the imported predicate as it is known in the $\mathcal{E}$RGO module. For non-tabled predicates, whose names start with `%` in $\mathcal{E}$RGO, *flora-predicate* should have the following syntax: `%(predicate-name)`. For instance, to import a $\mathcal{E}$RGO non-tabled predicate `%foobar` of arity 3 one can use the following statement:

```
?- '\\flimport' '%'(foobar)/3 as foobar(_,_,_) from mymodule.
```

The imported predicate must be given a name by which this predicate will be known in Prolog. (This name can be the same as the name used in $\mathcal{E}$RGO.) It is important, however, that the Prolog name be specified as shown, i.e., as a predicate skeleton with the same number of arguments as in the $\mathcal{E}$RGO predicate. For instance, `foo(_,_,_)` will do, but `foo/3` will not. Once the predicate is imported, it can be used under its Prolog name as a regular predicate.

Prolog programs can also load and compile $\mathcal{E}$RGO knowledge bases using the following queries (again, `bootstrap_ergo` must be executed in advance):

```
:- import '\\load'/1, '\\compile'/1 from flora2.
?- '\\load'(flora-file >> flora-module).
?- '\\load'(flora-file).
?- '\\compile'(flora-file >> flora-module).
?- '\\compile'(flora-file).
```

The first query loads the file *flora-file* into the given user module and compiles it, if necessary. The second query loads the knowledge base into the default module `main`. The last two queries compile the file for loading into the module *flora-module* and `main`, respectively, but do not load it.

Finally, a Prolog program can check if a certain $\mathcal{E}$RGO user module has been loaded using the following call:

```
:- import '\\isloaded'/1 from flora2.
?- '\\isloaded'(flora-module-name).
```

Note that in Prolog the `'\\isloaded'` predicate must be quoted and the backslash doubled.

Note: You must make sure that Prolog will find this installation and use it. One way of doing this was described earlier (by executing an appropriate `asserta/1`). This method works best if your application consists of both $\mathcal{E}$RGO and Prolog modules, but the initial module of your application (i.e., the one that bootstraps everything) is a Prolog program. If the initial module is a $\mathcal{E}$RGO knowledge base, then the best way is to start XSB and $\mathcal{E}$RGO using the `runergo` script (page 2) located in the distribution of $\mathcal{E}$RGO. □

15.8.2 Passing Arbitrary Queries to $\mathcal{E}$RGO

The method of calling $\mathcal{E}$RGO from Prolog, which we just described, assumes that the user knows which predicates and methods to call in the $\mathcal{E}$RGO module. Sometimes, it is useful to be able to pass arbitrary queries to $\mathcal{E}$RGO. This is particularly useful when $\mathcal{E}$RGO runs under the control of a Java or C program.

To enable such unrestricted queries, $\mathcal{E}$RGO provides a special predicate, `flora_query/4`, which is called from Prolog and takes the following arguments:

- *String*: A string that contains a $\mathcal{E}$RGO query. It can be an atomic frame (e.g., `'foo[bar->?X].'`) or a list of character codes (e.g., `"foo[bar->?X]."`).
- *Vars*: A list of the form `['?Name1'=Var1, '?Name2'=Var2,...]` or `["?Name1"=Var1, "?Name2"=Var2,...]`. `?Name` is a name of a variable mentioned in *String*, for instance, `'?X'` (note: the name must be quoted, since it is a Prolog atom). Var is a *Prolog* (not $\mathcal{E}$RGO!)

variable where you want the binding for the variable `Name` in *String* to be returned. For instance, if *String* is `'p(?X,?Y).'` then *Vars* can be `['?X' = Xyz, "?Y" = Qpr]`. In this case, `Xyz` will be bound to the value of `?X` in `p(?X,?Y)` after the execution, and `Qpr` will be bound to the value of `?Y` in `p(?X,?Y)`.

- *Status*: Indicates the status of compilation of the command in `String`. It is a list, which contains various indicators. The most important ones are `success` and `failure`.
- *Exception*: If the execution of the query is successful, this variable is bound to `normal`. Otherwise, this variable will contain an exception term returned by XSB (see the XSB manual, if you need to process exceptions in sophisticated ways).

In order to use the `flora_query/4` predicate from within Prolog, the following steps are necessary:

1. The $\mathcal{E}$RGO installation directory must be added to the XSB search path:

   ```
   ?- add_lib_dir(a('/home/myHomeDir/ergo')).
   ```

2. The query

   ```
   ?- [flora2], bootstrap_ergo
   ```

   must be executed *before compiling or loading* the Prolog file.

3. `flora_query/4` must be imported from `flora2`.

Here is an example of a Prolog file, `test.P`, which loads and then queries a $\mathcal{E}$RGO file, flrtest.ergo:

```
:- import bootstrap_ergo/0 from flora2.
?- add_lib_dir(a('/home/myHomeDir/ergo')),
   [flora2],
   bootstrap_ergo.
:- import ergo_query/5 from flora2.
:- import '\\load'/1 from flora2.

?- '\\load'(flrtest).

?- Str="?X[b->?Y].",
   ergo_query(Str,["?X"=YYY,"?Y"=PPP], _Status,_XWamState,_Exception).
```

After the query to `flrtest.ergo` is successfully executed, the bindings for the variable `?X` in the $\mathcal{E}$RGO query will be returned in the Prolog variable `YYY`. The binding for `?Y` in the query will be returned in the Prolog variable `PPP`. If there are several answers, you can get them all using a fail-loop, as usual in Prolog. For instance,

```
?- Str='?X[b->?Y].',
   ergo_query(Str,['?X'=YYY,'?Y'=PPP], _Status,_XWamState,_Exception),
   writeln('?X' = YYY),
   writeln('?Y' = PPP),
   \false.
```

Note that the Prolog variables in the variable list (like `YYY` and `PPP` above) can be bound and in this way input to the $\mathcal{E}$RGO query can be provided. For instance,

```
?- YYY=abc,
   ergo_query('?X[b->?Y].',['?X'=YYY,'?Y'=PPP],_Status,_XWamState,_Exception).
```

yields the same result as

```
?- ergo_query('abc[b->?Y].',['?Y'=PPP],_Status,_XWamState,_Exception).
```

However, the user should be aware of the fact that if a query is going to be used many times with different parameters then the first form is much faster. That is,

```
?- YYY=abc1,
   ergo_query('?X[b->?Y].',['?X'=YYY,'?Y'=PPP], _Status,_XWamState,_Exception).
?- YYY=abc2,
   ergo_query('?X[b->?Y].',['?X'=YYY,'?Y'=PPP], _Status,_XWamState,_Exception).
......
```

is noticeably faster than

```
?- ergo_query('abc1[b->?Y].',['?Y'=PPP],_Status,_XWamState,_Exception).
?- ergo_query('abc2[b->?Y].',['?Y'=PPP],_Status,_XWamState,_Exception).
......
```

if the above queries are executed thousands of times with different parameters `abc1`, `abc2`, etc.

The rest of the arguments for `flora_query/5` are as follows:

- `Status`: If the query was parsed successfully, this variable is bound to `normal`. Otherwise, it will be bound to an appropriate compiler error message.
- `XWamState`: If the query succeeds, the answer may be true or undefined in the well-founded semantics. If `XWamState =:= 0`, it means the answer is true. Otherwise, it is undefined.
- `Exception`: If a runtime exception is thrown during the execution of the query, this variable will be bound to the term representing that exception.

15.9 $\mathcal{E}$RGO System Modules

$\mathcal{E}$RGO provides a special set of modules that are *preloaded* with useful utilities, such as data type manipulation or I/O. These modules have special syntax, `\modname`, and cannot be loaded by the user. For this reason, these modules are called $\mathcal{E}$RGO *system modules*. For instance, to write to the standard output one can use

```
?- write(Something)@\io.
```

For more details on the currently existing $\mathcal{E}$RGO system modules see Section 46.

15.10 Including Files into $\mathcal{E}$RGO Knowledge Bases

The last and the simplest way to construct multi-file $\mathcal{E}$RGO knowledge bases is via the `#include` preprocessing directive. For instance if file `foo.ergo` contains the following instructions:

```
#include "file1"
#include "file2"
#include "file3"
```

the effect is the same as if the above three files were concatenated together and stored in `foo.ergo`. Note that the file names must be enclosed in *double* quotes.

Unix-style path names (i.e., using forward slashes) are understood universally and irrespective of the actual OS under which $\mathcal{E}$RGO runs. So, for portability, it is recommended to use only Unix-style path names relative to the directory of the host file that contains the include-directive (e.g., `"../abc/cde.foo"`).

If one does need to use Windows-style path names for some reason, keep in mind that backslashes in path names must be doubled. For instance,

```
#include "..\\foo\\bar.ergo "
```

15.11 More on Variables as Module Specifications

Earlier we mentioned that a user module specification can be a variable, *e.g.*, `a[m->b]@?X`, which ranges over module names. This variable does not need to be bound to a concrete module name before the call is made. If it is a variable, then `?X` will get successively bound to the user modules where `a[m->b]` is true. However, these bindings will not include `\prolog`, `\prolog(module)`, or `\module`.

Dynamic module bindings can be used to implement *adaptive methods*, which are used in many types of applications, *e.g.*, agent programming. Consider the following example:

```
Module foo                                  Module moo
something :- ...                            ......
something_else :-                           ......
        a[someservice(\@,?Arg)->?Res]@moo   ......
......                                      a[someservice(?Module,?Arg)-> ?Res] :-
......                                                  something@?Module, ...
......                                      ......
```

Here the method `someservice` in user module `moo` performs different operation depending on who is calling it, because `something` can be defined differently for different callers. When `something_else` is called in module `foo`, it invokes the method `someservice` on object `a` in module `moo`. The current module name (`foo`) is passed as a parameter (with the quasi-constant `\@`). When `someservice` is executed in module `moo` it therefore calls the predicate `something` in module `foo`. If `someservice` is called from a different module, say `bar`, it will invoke `something` defined in *that* module and the result might be different, since `something` in module `bar` may have a different definition than in module `foo`.

An example of the use of the above idea is the pretty printing module of $\mathcal{E}$RGO. A pretty-printing method is called on an object in some user module, and to do its job the pretty-printing method needs to query the object *in the context of the calling module* to find the methods that the object has.

It is also possible to view adaptive methods as a declarative counterpart of callback functions in C/C++, which allows the callee to behave differently for different clients.

15.12 Module Encapsulation

So far, in multi-module knowledge bases, any module could invoke any method or predicate in any other module. That is, modules were not encapsulated. However, $\mathcal{E}$RGO lets the user encapsulate

any module and export the methods and predicates that other modules are allowed to call. Calling an unexported method or predicate will result in a runtime error.

A module is encapsulated by placing an `export` directive in it or by executing an `export` directive at run time. Modules that do not have `export` directives in them are not encapsulated, which means that any method or predicate defined inside such a module can be called from the outside.

Syntax. The `export` directive has the form:

`:- export`{*MethodOrPredExportSpec1*, *MethodOrPredExportSpec2*, ...}.

There can be one or more export specifications (*MethodOrPredExportSpec*) in each `export` statement, and there can be any number of different `export` statements in a module. The effect of all these statements is cumulative.

Each *MethodOrPredExportSpec* specifies three things, two of which are optional:

- The list of methods or predicates to export.
- The list of modules to which to export. This list is optional. If it is not given then the predicates and modules are exported to *all* modules.
- Whether the above are exported as *updatable* or not. If a method or a predicate is exported as updatable, then the external modules can add or delete the corresponding facts. Otherwise, these modules can only query these methods and predicates. If `updatable` is not specified, the calls are exported for querying only.

The exact syntax of a *MethodOrPredExportSpec* is as follows:

[`updatable`] *ExportList* [>> *ModuleList*]

The square brackets here denote optional parts. The module list is simply a comma-separated list of modules and *ExportList* is a comma-separated list of *predicate/method/ISA templates*. Method templates have the form

`?[` *termTemplate* `-> ?]` or
`?[` *termTemplate* `]`

and predicate templates are the same as term templates. A *term template* is a HiLog term that has no constants or function symbols in it. For instance, `p(?,?)(?)` and `q(?,?,?)` are term templates, while `p(a,?)(?)` and `q(?,?,f(?))` are not.

ISA templates have the form `?:?` or `?::?`. Of course, `?_` can also be used instead of `?`.

Examples. Here are some examples of export directives:

```
:- export{?[a(?) -> ?]}.
:- export{?[b ->?], ?[c(?,?)], ?[d(?)(?,?) -> ?]}.
:- export{(?[e -> ?], ?[f(?,?)]) >> (foo, bar)}.
:- export{updatable (?[g -> ?], ?[h(?,?)]) >> (foo, bar)}.
:- export{updatable (?[g -> ?], ?[h(?,?)]) >> (foo, bar),
            (?[k -> ?], m(?,?)(?)) >> abc}.
```

Observe that the method `g` and the boolean method `h` have been exported in the updatable mode. This means that the modules `foo` and `bar` can insert and delete facts of the form `a[g->b]` and `a[h(b,c)]` using the statements like (assuming that `moo` is the name of the module that includes the above directives):

```
?- insert{a[g->b]@moo}.
?- delete{a[h(b,c)]@moo}.
```

Parenthesizing rules. Note that the last three export statements above use parentheses to disambiguate the syntax. Without the parentheses, these statements would be understood differently:

```
:- export{?[e -> ?], (?[f(?,?)] >> foo), bar}.
:- export{updatable ?[g -> ?], (?[h(?,?)]) >> foo), bar}.
:- export{updatable ?[g -> ?], (?[h(?,?)] >> foo), bar,
            ?[k -> ?], (m(?,?)(?) >> abc)}.
```

We should also note that `updatable` binds stronger than the comma or `>>`, which means that an `export` statement such as the one bellow

```
:- export{updatable ?[g -> ?], ?[h(?,?)] >> foo}.
```

is actually interpreted as

```
:- export{updatable(?[g -> ?]), (?[h(?,?)] >> foo)}.
```

Exporting frames other than `->`. In order to export any kind of call to a non-Boolean method, one should use only `->`. This will allow other modules to make calls, such as `a[|d(c)(e,f) -> ?X|]`, `a[b ->-> ?Z]`, and `c[e=>t]` to the exported methods. The `export` directive does not allow the user to separately control calls to the F-logic frames that involve method specifiers such as `+>>`, `=>`, `->->`, etc.

The `export` directive has an executable counterpart. For instance, at run time a module can execute an export instruction such as

```
?- export{?[e -> ?], (?[f(?,?)] >> foo), bar}.
```

and export the corresponding methods. If the module was not encapsulated before, it will become now. Likewise, it is possible to execute export directives in another module. For instance executing

```
?- export{?[e -> ?], (?[f(?,?)] >> foo), bar}@foo.
```

will cause the module `foo` to export the specified methods and to encapsulate it, if it was not encapsulated before.

15.13 Importing Modules

Referring to methods and predicates defined in other modules is one way to invoke knowledge defined separately elsewhere. Sometimes, however, it is convenient to *import* the entire module into another module. This practice is particularly common when it comes to reusing ontologies.

$\mathcal{E}$RGO supports import of entire modules through the `importmodule` compile-time directive. Its syntax is as follows:

```
:- importmodule{module1, module2, ..., module-k}.
```

Once a module is imported, its methods and predicates can be referenced without the need to use the *@module* idiom.

Importing a module is *not* the same as including another module as a file with the `#include` statement. First, only *exported* methods and predicates can be referenced by the importing module. The non-exported elements of an imported module are encapsulated. Second, even when everything is exported (as in the case when no explicit `export` directive is provided), import is still different from inclusion. To see why, consider one module, `main`, that looks like this:

```
?- [mykb>>foo].
```

```
:- importmodule{foo}.

p(abc).
?- q(?X).
```

This module loads the contents of the file `mykb.ergo` into a module `foo` and then imports that module. The importing module itself contains a fact and a query.

Suppose `mykb.ergo` is as follows:

```
q(?X) :- p(?X).
p(123).
```

It is easy to see that the query `q(?X)` in the importing module `main` will return the answer `?X = 123`. In contrast, if the module `main` *included* `mykb.ergo` instead of importing it, i.e., if it looked like this:

```
#include "mykb.ergo "
p(abc).
?- q(?X).
```

then the same query would have returned two answers: `?X = 123` and `?X = abc`. This is because the latter is simply

```
q(?X) :- p(?X).
p(123).
p(abc).
?- q(?X).
```

In other words, in the first case, the query `q(?X)` still queries module `foo` even though the query does not use the *@foo* idiom. The module `foo` has only one answer to the query, so only one answer is returned. In contrast, when `mykb.ergo` is included then the resulting knowledge base has two `p`-facts and two answers are returned.

Note: If module `modA` imports module `modB` then `modB` must be loaded before any query is issued against `modA`. Otherwise, `modA` might attempt to query `modB` and an error will be issued telling the user that `modB` is not loaded. A good practice to avoid this sort of errors is to load `modB` before `modA`.

15.14 Persistent Modules

Normally, the data in a $\mathcal{E}$RGO module is *transient* — it is lost as soon as the system terminates. The $\mathcal{E}$RGO package `persistentmodules` allows one to make $\mathcal{E}$RGO modules *persistent*. This package is described in the document *A Guide to $\mathcal{E}$RGO Packages*.

15.15 Fast Loading Vast Amounts of Data

When one needs to load vast amounts of data (hundreds of thousands to millions of facts), running them through the regular $\mathcal{E}$RGO compiler can take too long because that compiler is built to recognize the rich syntax of $\mathcal{E}$RGO. In contrast, large data sets invariably have very simple structure and it would be a waste to run a complex compiler to import all that data. To this end, $\mathcal{E}$RGO provides special primitives to load, query, and destroy such vast amounts of data provided the input has very regular structure: Prolog (not HiLog) terms with no infix, prefix, or postfix operators. It can contain $\mathcal{E}$RGO preprocessor commands (Section 42) and comments, but the comments must be either enclosed between `/* ... */` or be one-line comments beginning with a `%` (*i.e.*, Prolog style, not $\mathcal{E}$RGO style).

For instance, the following input would work

```
foo(1,bar(abc,cde),q2).
foo(2,bar(ab,ce),q3).
bar(23,foo(123,[a,b,c]),"q33").
```

but the following will cause errors and loading will be aborted:

```
foo(2,bar(ab,ce),q3).  %% ok
foo(1+2,abc).          %% nope: + is an infix operator
bar(2,foo(123,"q33").  %% nope: parentheses missing
pqr(1,?f(2,g(3))).     %% nope: ?f(2,...) is a HiLog term
pqr(1,ff(2,${g(3)})).  %% nope: reification (${g(3)}) isn't allowed
```

The fast-loaded facts are placed in a special storage container that is not accessible to regular queries: these facts are to be queried via a special primitive. The primitives for fast-loading and working with these storage containers are as follows:

- `fastload{`*FileName*`,`*Storage*`}` — fast load *Filename* into storage container *Storage*. *Filename* must have the extension `.P` and *Storage* must be an alphanumeric string. Executing multiple `fastload` commands on the same storage will add all facts up from the files involved. Duplicates are eliminated.
  Example: `?- fastload{myfile,mystorage}`.

- `fastquery{`*Storage*`,`*Query*`}` — query the information stored in the storage container *Storage*. *Query* must be a HiLog term. Other forms are also allowed, but they will return nothing, as they will not match the input of the form described above. Variables in the query term are allowed. Note also that even though the input consists of Prolog terms, they are converted to HiLog so HiLog terms can be used to query the information stored in these containers.
  Examples:
  ```
  ?- fastquery{mystorage,?X}.
  ?- fastquery{mystorage,?Y(p,q)}.
  ?- fastquery{mystorage,r(p,q)}.
  ```

- `fasterase{`*Storage*`}` — erase all the information in the given storage container.
  Example: `?- fasterase{mystorage}`.

Notes:

- The `fastload` command should not occur in the body of any rule that has a tabled head-literal—directly or indirectly (via other rules). The most straightforward way to call `fastload` is via a top-level query, as in the above examples.

- The intended use of `fastload` is that the data will be loaded early in the process, before any kind of significant queries were computed. It is not an error to call `fastload` *after* such queries, but then the tables associated with those queries will be abolished. The result will be that, if the same or related queries are asked again, they will be recomputed from scratch, resulting in a delay.

15.16 Truth about Relative File Names and Current Directories

Absolute file names (the ones starting with `'/'` in Linux and Mac and those starting with \ in Windows) do not depend on the location of the $\mathcal{E}$RGO file in which they appear and so they can be used in load/add commands and in I/O operations without the fear of ambiguity. Not so with *relative* file names like `foo/bar.flr` or `../moo.flr`. These depend on the *current directory* in which $\mathcal{E}$RGO executes at the moment when these files are being used. So, what is this "current directory?" This can be summarized as follows:

- When $\mathcal{E}$RGO starts, the current directory is the directory in which $\mathcal{E}$RGO was started. Namely,
  - If $\mathcal{E}$RGO was started from a command window then the current directory of $\mathcal{E}$RGO is the current directory of that window unless that directory is non-writable or is $\mathcal{E}$RGO's installation directory. In the latter cases, the current directory is user's home directory.

  - If it was started by clicking on an icon (via $\mathcal{E}$RGO Studio) then the current directory is the user's home directory.

- During loading or adding of an $\mathcal{E}$RGO's file, the current directory is that file's directory. If, during loading, that file loads another file, the current directory temporarily is switched to the directory of that latter file.

- When file loading is finished, the current directory returns to what it was before the loading.

- The current directory of $\mathcal{E}$RGO can be changed at run time by executing the subgoal `File[chdir(myNewDir)]@\io` — see Section 46.1.

Now, regarding the relative file names that occur in the `load{...}` and `add{...}` commands, the following natural rule holds:

- Any relative file name used in the `load{...}` or `add{...}` command is always relative to the directory of the file in which that load/add command occurs.

This behavior is desirable because it greatly facilitates creation of location-independent multi-file knowledge bases.

The other common place where file names occur are I/O operations, like `tell(myfile)`, which opens `myfile` for writing (see Section 46.1 for more details on I/O). Here the situation is more subtle because I/O commands can execute in many different contexts, not necessarily in the context of the file in which they are physically located. For that reason, the rule is that

- any relative file name that occurs in an I/O command is always relative to the directory that is current at the time that command is executed.

For example, we know from the above that during loading of a file the current directory is the directory of that file. Therefore, if a query like

```
?- tell(abc)@\io, writeln(abcdefg)@\io, told@\io.
```

occurs in a file `/foo/bar/moo/myfile.flr` then, when that file will be loaded, the file `abc` will be created in the directory `/foo/bar/moo/` and its contents will be a single line `abcdefg`. This is because, as we discussed, the current directory during loading of a file is that file's directory. In contrast, if our file has a rule

```
makefile(?X) :- tell(?X)@\io, writeln(12345678)@\io, told@\io.
```

and `makefile(cde)` is called at some point, this will create a file, called `cde` and containing the single line `12345678`, in the directory that was current at the time `tell(?X)@\io` (and `makefile(cde)`) was called.

Thus, the context in which files are interpreted in I/O commands is not uniform and depends on how these commands are called—a somewhat unpleasant situation. Fortunately, to rectify this problem, $\mathcal{E}$RGO provides a primitive, `here{...}`, which takes any subgoal and calls it in the directory of the file in which `here{...}` occurs. For instance, with a slightly modified rule

```
makefile(?X) :- here{tell(?X)@\io}, writeln(12345678)@\io, told@\io.
```

a call `makefile(cde)` will create `cde` in the directory of our example file, i.e. `/foo/bar/moo/`, and not in the directory where `makefile(cde)` was called.

Finally, we note that `here{...}` can take any goal, even a complex one, provided it is enclosed in parentheses. For instance, `here{(p(?X,?Y),q(?Y))}`. In fact, this primitive can be used to temporarily localize the context of execution for *any* subgoal, not only for I/O commands.

16 HiLog and Meta-programming

HiLog [4] is the default syntax that $\mathcal{E}$RGO uses to represent functor terms (including object Ids) and predicates. In HiLog, complex terms can appear wherever a function symbol is allowed. For example, `group(?X)(?Y,?Z)` is a HiLog term where the functor is no longer a symbol but rather a complex term `group(?X)`. Variables in HiLog can range over terms, predicate and function symbols, and even over base formulas. For instance,

$$?-\mathtt{p}(?\mathtt{X}), ?\mathtt{X}(\mathtt{p}).$$

and

$$?-\mathtt{p}(?\mathtt{X}),\ ?\mathtt{X}(\mathtt{p}),\ ?\mathtt{X}. \tag{3}$$

are perfectly legal queries. If `p(a(b))`, `a(b)(p)`, and `a(b)` are all true in the database, then $?\mathtt{X} = \mathtt{a}(\mathtt{b})$ is one of the answers to the query in HiLog.

Although HiLog has a higher order syntax, its semantics is first order [4]. Any HiLog term can be consistently translated into a Prolog term. For instance, `group(?X)(?Y,?Z)` can be represented by the Prolog term `apply(apply(group,?X),?Y,?Z)`. The translation scheme is pretty straightforward and is described in [4].

Any Id-term in $\mathcal{E}$RGO, including function symbols and predicate symbols, are considered to be HiLog terms and therefore are subject to translation. That is, even a normal Prolog term will by default be represented using the HiLog translation, e.g., `foo(a)` will be represented as

`apply(foo,a)`. This guarantees that HiLog unification will work correctly at runtime. For instance, `foo(a)` will unify with `?F(a)` and bind the variable `?F` to `foo`.

There is one important difference between HiLog, as described in [4], and its implementation in $\mathcal{E}$RGO. In HiLog, functor terms that appear as arguments to predicates and the base formulas (i.e., predicates that are applied to some arguments) belong to the same domain. In contrast, in $\mathcal{E}$RGO they are in different domains.[7] For instance, suppose `p(a(b))` is true, and consider the following query:

```
?- ?X ~ a(b), p(?X).
```

Here `~` is a meta-unification operator to be discussed shortly, in Section 16.1; it binds `?X` to the base formula `a(b)` in the current module. The answer to this query is 'No' because `?X` is bound to the base formula `a(b)`, while `a(b)` in `p(a(b))` is a HiLog term.

Our earlier query, (3), will also not work (unlike the original HiLog) because `?X` is bound to a term and not a formula: if we execute the query (3), we will get an error stating that `?X` is bound to a HiLog term, not a predicate, and therefore the query `?X` is meaningless. To correct the problem, `?X` must be promoted to a predicate and relativized to a concrete module—in our case to the current module. So, the following query *will* work and produce a binding `a(b)` for `?X`.

```
?- p(?X), ?X(p), ?X@ \@.
```

Like in classical logic, `foo` and `foo()` are different terms. However, it is convenient to make these terms synonymous when they are treated as predicates. Prologs often disallow the use of the `foo()` syntax altogether. The same distinction holds in HiLog: `foo`, `foo()` and `foo()()` are all different. In terms of the HiLog to Prolog translation, this means that `foo` is different from `flapply(foo)` is different from `flapply(apply(foo))`. However, just like in Prolog, we treat `p` as syntactic sugar for `p()` when both occur as predicates. Thus, the following queries are the same:

```
?- p.
?- p().
```

In the following,

```
p.
q().
?- p(), ?X().
?- q, ?X().
?- r = r().
```

[7] This is allowed in *sorted HiLog* [3].

the first two queries will succeed (with `?X` bound to `p` or `q`), but the last one will fail. Making `p` and `p()` synonymous does not make them synonymous with `p()()`—the latter is distinct from both `p` and `p()` not only as a term but also as a formula. Thus, in the following, all queries fail:

```
p.
q().
?- p()().
?- q()().
?- p = p()().
?- q() = q()().
```

16.1 Meta-programming, Meta-unification

F-logic together with HiLog is powerful stuff. In particular, it lends itself naturally to meta-programming. For instance, it is easy to examine the methods and types defined for the various classes. Here are some simple examples:

```
// all unary methods defined for John
?- John[?M(?) -> ?].

// all unary methods that apply to John,
// for which a signature was declared
?- John[?M(?) => ?].

// all method signatures that apply to John,
// which are either declared explicitly or inherited
?- John[?M => ?].

// all method invocations defined for John
?- John[?M -> ?].
```

However, a number of meta-programming primitives are still needed since they cannot be directly expressed in F-logic. Many such features are provided by the underlying Prolog system and $\mathcal{E}$RGO simply takes advantage of them:

```
?- functor(?X,f,3)@\prolog.
?X = f(_h455,_h456,_h457)@\prolog
Yes
```

```
?- arg(2,f(?X,3)@\prolog,?Y)@\prolog.
?Y = 3
Yes
```

Note that these primitives are used for Prolog terms only and are described in the XSB manual. These primitives have not been ported to work with HiLog terms yet.

Meta-unification. In $\mathcal{E}$RGO, a variable can be bound to either a formula or a term. For instance, in `?X = p(a)`, `p(a)` is viewed as a term and `?X` is bound to it. Likewise, in `?X = a[m->v]`, the frame is evaluated to its object value (which is `a`) and then unified with `?X`. To bind variables to formulas instead, $\mathcal{E}$RGO provides a `meta-unification` operator, `~`. This operator treats its arguments as formulas and unifies them as such. For instance, `?X ~ a[m->v,k->?V]` binds `?X` to the frame `a[m->v,k->?V]` and `a[m->v,k->?V] ~ ?X[?M->v,k->p]` unifies the two frames by binding `?X` to `a`, `?M` to `m`, and `?V` to `p`.

Meta-unification is very useful when it is necessary to determine the module in which a particular formula lives. For instance,

```
?- ?X@?M ~ a[b->c]@foo.
```

would bind `?X` to the formula `a[b->c]`, `?M` to the module of `?X`. Note that in meta-unification the variable `?X` in the idiom `?X@?M` or `?X@foo` is viewed as a meta-variable that is bound to a formula. More subtle examples are

```
?- ?X ~ f(a), ?X ~ ?Y@?M.
?- f(a)@foo ~ ?Y@?M.
```

`?M` is bound to the current module in the first query and `foo` in the second one. `?Y` is bound to the (internal representation of the) HiLog formula `f(a)@ \@` in the first query and `f(a)@foo` in the second — *not* to the HiLog term `f(a)`!

Another subtlety has to do with the scope of the module specification. In $\mathcal{E}$RGO, module specifications have scope and inner specifications override the outer ones. For instance, in

```
..., (abc@foo, cde)@bar, ...
```

the term `abc` is in module `foo`, while `cde` is in module `bar`. This is because the inner module specification, `@foo`, overrides the outer specification `@bar` for the literal in which it occurs (*i.e.*, `abc`). These scoping rules have subtle impact on literals that are computed dynamically at run time. For instance, consider

```
?- ?X@?M ~ a[b->c]@foo, ?X@bar.
```

Because `?X` gets bound to `a[b->c]@foo`, the literal `?X@bar` becomes the same as `(a[b->c]@foo)@bar`, i.e., `a[b->c]@foo`. Thus, both of the following queries succeed:

```
?- ?X@?M ~ a[b->c]@foo, ?X@bar ~ a[b->c]@foo.
?- ?X@?M ~ a[b->c]@foo, ?X@?N ~ a[b->c]@foo.
```

Moreover, in the second query, the variable `?N` is *not* bound to anything because, as noted before, the literal `?X@?N` becomes `(a[b->c]@foo)@?N)` at run time and, due to the scoping rules, is the same as `a[b->c]@foo`.

Meta-disunification. The negation of meta-unification is *meta-disunification*, `!~`. For instance,

```
?- abc !~ neg cde.

Yes
```

Note that `neg`, `\naf`, and `\+` tie stronger than `~` and `!~`, so the above is parsed as `abc!~neg(cde)` and `neg cde!~abc` is parsed as `neg(cde)!~abc` — *not* as `neg(cde!~abc)`.[8]

16.2 Reification

It is sometimes useful to be able to treat $\mathcal{E}$RGO frames and predicates as objects. For instance, consider the following statement:

```
Tom[believes-> Alice[thinks->Flora2:coolThing]].
```

The intended meaning here is that one of Tom's beliefs is that Alice thinks that Flora2 is a cool thing. However, minute's reflection shows that the above has a different meaning:

```
Tom[believes-> Alice].
Alice[thinks->Flora2:coolThing].
```

[8] This is in contrast to `=`, `==`, `<`, etc., which bind stronger than `neg`, `\naf`, and `\+`. For instance, `neg cde=abc` is parsed as `neg(cde=abc)`. This difference in parsing makes sense because `~` and `!~` expect formulas as its operands, while `=`, `==`, and other comparison operators expect terms.

Portions Copyright © 2013–2017 Coherent Knowledge

That is, Tom believes in Alice and Alice thinks that Flora2 is cool. This is different from what we originally intended. For instance, we did not want to say that Alice likes $\mathcal{E}$RGO (she probably does, but she did not tell us). All we said was what Tom has certain beliefs about what Alice thinks. In other words, to achieve the desired effect we must turn the formula `Alice[thinks->Flora2:coolThing]` into an object, i.e., *reify* it.

Reification is done using the operator `${...}`. For instance, to say that Tom believes that Alice thinks that Flora2 is a cool thing one should write:

```
Tom[believes-> ${Alice[thinks->Flora2:coolThing]}].
```

When reification appears in facts or rule heads, then the module specification and the predicate part of the reified formula must be bound. For instance, the following statements are illegal:

```
p(${?X@foo}) :- q(?X).
p(${q(a)@?M}).
?- insert{p(${?X@?M})}.
```

The semantics of reification in $\mathcal{E}$RGO is described in [19].

Reification of complex formulas. In $\mathcal{E}$RGO, one can reify not only simple facts, but also anything that can occur in a rule body. Even a set of rules can be reified! The corresponding objects can then be manipulated in ways that are semantically permissible for them. For instance, reified conjunctions of facts can be inserted into the database using the `insert{...}` primitive. Reified conjunctions of rules can be inserted into the rulebase using the `insertrule{...}` primitive. Reified rule bodies, which can include disjunctions, negation, and even things like aggregate functions and update operators(!), can be called as queries.

```
request[
  input       -> ${?Ticket[from->?From, to->?To, \naf international],
  inputAxioms -> ${(?Ticket[international] :-
                       ?Ticket[from->?From:?Country1, to->?To:?Country2],
                       ?Country1 \= ?Country2)
                  }
].

?- ?Request[input->?Input, inputAxioms->?Rules],
   insertrule{?Rules},
   ?Input.
```

In the above example, the object `request` has two attributes, which return reified formulas. The `input` attribute returns a Boolean combination of frames, while `inputAxioms` returns a reified rule. In general, conjunctions of rules are allowed inside the reification operator (*e.g.*, `${(rule1), (rule2)}`), where each rule is enclosed in a pair of parentheses. Such a conjunction can then be inserted (or deleted) into the rulebase using the `insertrule{...}` primitive.[9]

Note that rule Ids and other meta-data (see Section 34) can be supplied with reified rules just like it can be with regular rules. For instance,

```
... :- ?X=${@!{abc[tag->foo]} head(?X):-body(?X,?Y)}.
```

Reification and meta-unification. Reification should not be confused with meta-unification, although they are closely related concepts. A reified formula reflects the exact structure that is used to encode it, so structurally similar, but syntactically different, formulas might meta-unify, but their internal representations could be very different. For instance,

```
?- a[b->?X]@?M ~ ?Y[b->d]@foo.
```

will return *true*, because the two frames are structurally similar and thus meta-unify. On the other hand,

```
?- ${a[b->?X]@?M} = ${?Y[b->d]@foo}.
```

will be false, because `a[b->?Y]@?X` and `?Z[b->d]@foo` have different internal representations (even though their conceptual structures are similar), so they do not unify (using "=", i.e., in the usual first-order sense). Note, however, that the queries

```
?- ${a[b->?Y]@foo} = ${?Z[b->d]@foo}.
?- ?M=foo, ${a[b->?Y]@?M} = ${?Z[b->d]@?M}.
?- a[b->?Y]@foo ~ ?Z[b->d]@foo.
?- ?M=foo, a[b->?Y]@?M ~ ?Z[b->d]@?M.
```

will all return *true*, because `a[b->?Y]@foo` and `?Z[b->d]@foo` are structurally similar — both conceptually and as far as their internal encoding is concerned (and likewise are `a[b->?Y]@foo` and `?Z[b->d]@foo`).

[9] In fact, Boolean combinations of rules are also allowed inside the reification operator. However, such combinations cannot be inserted into the rulebase. $\mathcal{E}$RGO does not impose limitations here, since is impossible to rule out that a knowledge base designer might use such a feature in creative ways.

16.3 Meta-decomposition

$\mathcal{E}$RGO supports an extended version of the Prolog meta-decomposition operator "`=..`". On Prolog terms, it behaves the same way as one would expect in Prolog. For instance,

```
?- ?X=p(a,?Z)@\prolog, ?X=..?Y.

?X = p(a,?_h4094)@\prolog
?Z = ?_h4094
?Y = [p, a, ?_h4094]
```

The main use of the `=..` operator in $\mathcal{E}$RGO is, however, for decomposing HiLog terms or reifications of HiLog predicates and F-logic frame literals. The meta-decomposition operator uses special conventions for these new cases.

For HiLog terms, the head of the list on the right-hand side of `=..` has the form `hilog(HiLogPredicateName)`. For instance,

```
?- p(a,b) =.. ?L.

?L = [hilog(p), a, b]
```

For HiLog predicates the head of the list has the form `hilog(HiLogPredicateName,Module)`. For instance,

```
?- ${p(a,b)@foo} =.. ?L.

?L = [hilog(p,foo), a, b]
```

For non-tabled HiLog predicates, which represent actions with side-effects, the head of the list is similar except that `'%hilog'` (quoted!) is used instead of `hilog`. For instance,

```
?- ${%p(a,b)@foo} =.. ?L.

?L = ['%hilog'(p,foo), a, b]
```

For frame literals, the head of the list has the form `flogic(FrameSymbol,Module)`. The `FrameSymbol` argument represents the type of the frame and can be one of the following: `->`, `*->`, `=>`, `*=>`, `+>>`, `*+>>`, `->->`, `*->->`, `:`, `::`, `boolean` (tabled Boolean methods), `*boolean` (class-level tabled Boolean methods), `%boolean` (transactional, nontabled Boolean methods), `:=:`, `[]` (empty frames, such as `a[]`). Here are a number of examples that illustrate the use of `=..` for decomposition of frames:

```
${a[b->c]@foo} =.. [flogic(->,foo), a, b, c]
${a[|b->c|]@foo} =.. [flogic('*->',foo), a, b, c]
${a[b=>c]@foo} =.. [flogic(=>,foo), a, b, c]
${a[|b=>c|]@foo} =.. [flogic('*=>',foo), a, b, c]
${a[b+>>c]@foo} =.. [flogic(+>>,foo), a, b, c]
${a[|b+>>c|]@foo} =.. [flogic('*+>>',foo), a, b, c]
${a[b->->c]@foo} =.. [flogic(->->,foo), a, b, c]
${a[|b->->c|]@foo} =.. [flogic('*->->',foo), a, b, c]
${a:b@foo} =.. [flogic(:,foo), a, b]
${a::b@foo} =.. [flogic(::,foo), a, b]
${a:=:b@foo} =.. [flogic(:=:,foo), a, b]
${a[]@foo} =.. [flogic([],foo), a]
${a[p]@foo} =.. [flogic(boolean,foo), p]
${a[|p|]@foo} =.. [flogic('*boolean',foo), a, p]
${a[%p]@foo} =.. [flogic('%boolean',foo), a, p]
```

The `=..` operator supports explicit negation (Section 17.5). The corresponding type designators are `neg_hilog` and `neg_flogic`. For instance,

```
${\neg a[b->c]@foo} =.. [negation(neg), $a[b->c]@main]
${\neg p(a,b)@foo} =.. [negation(neg), $p(a,b)@main]
${\naf a[b->c]@foo} =.. [negation(naf), $a[b->c]@main]
${\naf p(a,b)@foo} =.. [negation(naf), $p(a,b)@main]
```

Additional examples:

```
${foo;bar} =.. [logic(or),${foo},${bar}]
${\if foo\then bar\else moo} =.. [control(ifthenelse),main,${foo},${bar},${moo}]
${\while foo \do bar} =..[control(whiledo),main,${foo},${bar}]
${insert{?V|p(?V)}} =.. [update(insert),[?V],${p(?V)}]
${deleteall{?V|p(?V)}} =.. [update(deleteall),[?A],${p(?A)}]
${?X=min{?V|p(?V)}} =.. [logic(and),${?T=min{?V|p(?V)}},?X=?T]
avg{?V[?T]|p(?V,?T)} =.. [aggregate(avg),?V,[?T],${p(?V,?T)},?Result]
count{?V[?G]|p(?V,?G)} =.. [aggregate(count),?V,[?G],${p(?C,?G)},?R]
${insert{p,(q:-r)}} =.. [update(insert),main,[${p},${q :- r}]].
${delete{p,(q:-r)}} =.. [update(delete),main,[${p},${q :- r}]].
${wish(nonvar(?X))^p(?X)} =.. [quantifier(delay),wish,${nonvar(?X)@\plg},${p(?X)}].
```

The `=..` operator is bi-directional, which means that either one or both of its arguments can be bound. For instance,

```
?- ?X =.. [flogic('*boolean',foo),a ,p].

?X = ${a[|p|]@foo}
```

However, that last feature, while it works for most statements, is *not* fully implemented in the sense that the terms produced (if produced at all) on the left-hand side might not be a valid $\mathcal{E}$RGO term or formula or the internal representations may be slightly different.

The statements for which the reverse mode of `?Variable =.. ?List` has not yet been implemented include `clause{...}`, `newmodule{...}`, and some others. The statements for which `=..` is not fully reversible are the update operators, statements, and a few others. For instance, while

```
?- ${insert{p,(q:-r)}} =.. [update(insert),main,[${p}, ${q :- r}]].
```

succeeds, the following will fail:

```
?- ${delete{p,(q:-r)}} =.. [update(delete),main,[${p}, ${q :- r}]].
```

because the internal representations of the two sides of `=..` happen to be slightly different. Nevertheless, `=..` does produce valid insert- and delete-statements from appropriate list on the right side. For instance,

```
?- ?I =.. [update(insert),main,[${p}, ${r}, ${q :- r}]], ?I.
?- p,q,r.
```

will succeed, i.e., a valid insert statement was constructed in `?I` and the call `?I` did perform the insertion. Likewise,

```
?- ?D =.. [update(delete),main,[${p}, ${r}, ${q :- r}]], ?D.
```

constructs a proper delete statement for `?D` and a call to `?D` then performs the deletion.

The low-level Prolog `=..` is also available using the idiom `(?Term =.. ?List)@\prolog`. This is rarely needed, however. One might use this when the term to be decomposed is known to be a Prolog term (in this case the Prolog's operator will run slightly faster) or if one wants to process the Prolog terms into which $\mathcal{E}$RGO literals are encoded internally (which is probably hardly ever necessary).

16.4 Passing Parameters between $\mathcal{E}$RGO and Prolog

The native HiLog support in $\mathcal{E}$RGO causes some tension when crossing the boundaries from one system to another. The reason is that $\mathcal{E}$RGO terms and Prolog terms have different internal

representations. Even though XSB supports HiLog (according to the manual, anyway), this support is incomplete and is not integrated well into the system — most notably into the XSB module system. As a result, XSB does not recognize terms passed to it from $\mathcal{E}$RGO as HiLog terms and, thus, many useful primitives will not work correctly. (Try `?- writeln(foo(abc))@\prolog` and see what happens.)

To cope with the problem, $\mathcal{E}$RGO provides a primitive, `p2h{?Plg,?Hlg}`, which does the translation. If the first argument, `?Plg`, is bound, the primitive binds the second argument to the HiLog representation of the term. If `?Plg` is already bound to a HiLog term, then `?Hlg` is bound to the same term without conversion. Similarly, if `?Hlg` is bound to a HiLog term, then `?Plg` gets bound to the Prolog representation of that term. If `?Hlg` is bound to a non-HiLog term, then `?Plg` gets bound to the same term without conversion. In all these cases, the call to `p2h{...}` succeeds. If `both` arguments are bound, then the call succeeds if and only if

- `?Plg` is a Prolog term and `?Hlg` is its HiLog representation.
- Both `?Plg` and `?Hlg` are identical Prolog terms.

Note that if both `?Plg` and `?Hlg` are bound to the same *HiLog term* then the predicate *fails*. Thus, if you type the following queries into the $\mathcal{E}$RGO shell, they both succeed:

```
?- p2h{?X,f(a)}, p2h{?X,?X}.
```

but the following will fail:

```
?- p2h{f(a),?X}, p2h{?X,?X}.
?- p2h{f(a),f(a)}.
```

The first query succeeds because `?X` is bound to a Prolog term, and by the above rules `p2h{?X,?X}` is supposed to succeed. The second query fails because `?X` is bound to a HiLog term and, again by the above rules, `p2h{?X,?X}` is supposed to fail. The reason why the last query fails is less obvious. In that query, both occurrences of `f(a)` are HiLog terms, as are all the terms that appear in a $\mathcal{E}$RGO knowledge base (unless they are marked with `@\prolog` or `@\prologall` module designations). Therefore, again by the rules above the query should fail.

One should not try to convert certain Prolog terms to HiLog and expect them to be the same as similarly looking $\mathcal{E}$RGO terms. In particular, this applies to reified statements. For instance, if `?X = $a[b->c]` then `?- p2h{?X,?Y}, ?Y = $a[b->c]` is not expected to succeed. This is because `p2h{...}` does not attempt to mimic the $\mathcal{E}$RGO compiler in cases where conversion to HiLog (such as in the case of reified statements) makes no sense. Doing so would have substantially increased the run-time overhead.

Not all arguments passed back and forth to Prolog need conversion. For instance, `sort/2`, `ground/1`, `compound/1`, and many others do not need conversion because they work the same for Prolog and HiLog representations. On the other hand, most I/O predicates require conversion. $\mathcal{E}$RGO provides the `io` library, described in Section 46, which provides the needed conversions for the I/O predicates.

Another mechanism for calling Prolog modules, described in Section 15.7, is use of the `@\prologall` and `@\prologall(module)` specifiers (`@\plgall` also works). These specifiers cause the compiler to include code for automatic conversion of arguments to and from Prolog representations. However, as mentioned above, such conversion is sometimes not necessary and the use of `@\prologall` might incur unnecessary overhead.

17 Negation

$\mathcal{E}$RGO supports three kinds of negation: a *Prolog-style negation* \+ [5]; *default negation* based on *well-founded semantics*, denoted `\naf`, [13, 14]; and *explicit negation* `\neg`, which is analogous to what is called "classical" negation in [7].

These three types of negation are quite different and should not be confused. Prolog-style negation does not have a model-theoretic semantics and it is unsatisfactory in many other respects. It is included in $\mathcal{E}$RGO for completeness and is primarily used when one needs to negate a Prolog predicate (in which case it is much faster than default negation). Explicit negation is mostly syntactic sugar that enables one to represent negative information explicitly. Default negation is a logically sound version of Prolog-style negation.

17.1 Default Negation `\naf` vs. Prolog Negation \+

$\mathcal{E}$RGO has *three* operations for negation: `\naf`, \+, and `\neg`. In this subsection we discuss the first two and Section 17.5 describes the third.

Prolog negation is specified using the operator \+. Negation based on the well-founded semantics is specified using the operator `\naf`. The well-founded negation, `\naf`, applies to predicates that are tabled (i.e., predicates that do not have the `%` prefix to be discussed in detail in Section 22) or to frames that do not contain transactional methods (i.e., methods prefixed with a `%`).

The semantics for Prolog negation is simple. To find out whether `\+ Goal` is true, the system first asks the query `?- Goal`. If the query fails then `\+ Goal` is said to be satisfied. Unfortunately, this semantics is problematic. It cannot be characterized model-theoretically and in certain simple cases the procedure for testing whether `\+ Goal` holds may send the system into an infinite loop. For instance, in the presence of the rule `%p :- \+ %p`, the query `?- %p` will not terminate. Prolog

negation is the recommended type of negation for non-tabled predicates (but caution is advised).

The well-founded negation, `\naf`, has a model-theoretic semantics and is much more satisfactory from the logical point of view. Formally, this semantics uses three-valued models where formulas can be true, false, or undefined. For instance, if we have the rule `p :- \naf p` then the truth value of `p` is *undefined*. Although the details of this semantics are somewhat involved [14], it is usually not necessary to know them, because this type of negation yields the results that the user normally expects. The implementation of the well-founded negation in XSB requires that it be applied to goals that consist entirely of tabled predicates or frames. Although $\mathcal{E}$RGO allows `\naf` to be applied to non-tabled goals, this may lead to unexpected results. For instance, Section 25 discusses what might happen if the negated formula is defined in terms of an update primitive.

For more information on the implementation of the negation operators in XSB, we refer the reader to the XSB manual.

Both `\+` and `\naf` can be used as operators inside and outside of the frames. For instance,

```
?- \naf %p(a).
?- \+ %p(a).
?- \naf X[foo->bar, bar->foo].
?- X[\naf foo->bar, bar->foo, \+ %p(?Y)].
```

are all legal queries. Note that `\+` applies only to non-tabled constructs, such as non-tabled $\mathcal{E}$RGO predicates and transactional methods.

To apply negation to multiple formulas, simply enclose them in parentheses. (Parentheses are not needed for singleton formulas used in earlier examples.)

```
?- \+ (%p(a),%q(?X)).
?- \naf (p(a),q(?X)).
?- \naf (?X[foo->bar], ?X[bar->foo]).
```

17.2 Default Negation for Non-ground Subgoals

One major difference with other implementations of the well-founded default negation is that $\mathcal{E}$RGO lets one apply it to formulas that contain variables. Normally, systems either require that the formula under `\naf` is ground or they interpret something like `\naf p(?X)` as meaning "not exists `?X` such that `p(?X)` is true" — the so-called *Not-Exists* semantics. However, this is not the right semantics in many cases. The right semantics is usually "there exists `?X` such that `p(?X)` is *not* true." This semantics is known as the *Exists-Not* semantics. Indeed, the standard convention for variables that occur in the rule body but not in the head is existential. For instance, if `?X` does not occur in the head of some rule, `p(?X)` in the body of that rule is interpreted as ∃`?X p(?X)`. Negation should

not be treated differently, i.e., `\naf p(?X)` should be interpreted as ∃`?X \naf p(?X)`. Worse yet, if `?X` *does* occur in the rule head then it is confusing and error-prone to interpret `h(?X) :- \naf p(?X)` as `h(?X) :- \naf ∃?X p(?X)` using the Naf-Exists semantics. And without the Naf-Exists semantics, `\naf p(?X)` has no meaning, if `?X` happens to be non-ground.

$\mathcal{E}$RGO takes a different approach. For body-only variables that appear under `\naf` the semantics is Exists-Not. In addition, `\neg p` implies `\naf p`. So,

```
\neg p({1,2}).
?- \naf p(?X).
```

returns the bindings 1 and 2 for `?X`. For variables that occur both under `\naf` *and* in the rule head, the semantics is also standard: universal quantification that applies to the entire rule. What happens if `?X` is not ground at the time of the call to `p(?X)` (whether `?X` does or does not occur in the rule head)? In that case, $\mathcal{E}$RGO *defers* the call in the hope that `?X` might become ground later. For instance, in

```
p(2).
?- \naf p(?X), ?X=1.
```

the query succeeds because the call `\naf p(?X)` will be delayed past the moment when `?X` becomes ground. Finally, what happens if `?X` does not become ground even after the delay? Still the Exists-Not semantics is used. If `p(?X)` is false for all `?X` (i.e., if `\naf ∃?X p(?X)` holds) then, in particular, ∃ `\naf?X p(?X)` is also true (assuming an infinite number of constants) and the query succeeds with the truth value *true*. If ∀`?X p(?X)` is true then `\naf ∃?X p(?X)` is false. However, if none of the above cases apply then we are in a gray area and there is not enough information to tell whether `\naf ∃?X p(?X)` is true or false, so this subgoal succeeds with the truth value *undefined*. For instance, suppose that the only facts in the KB are the ones below:

```
p({1,2}).
?- \naf p(?X).
?- \naf q(?X).
```

Then the first query succeeds with the truth value *undefined*, since `\naf ∃?X p(?X)` is not true. On the other hand, the second query succeeds with the truth value *true*. In both cases, $\mathcal{E}$RGO makes an open domain assumption by refusing to commit to true or false based only on the explicitly known elements of the domain of discourse.

Finally, what if the user does want the Not-Exists semantics after all? In $\mathcal{E}$RGO one must then say this explicitly and in a natural way through the use of existential quantifier:

```
?- \naf exists(?X)^p(?X).
```

More on the logical quantifiers is given in Section 18.

17.3 Non-ground Subgoals Under \+

When \+ is applied to a non-ground goal, the semantic is the standard Prolog's one: If for *some* values of the variables in `Goal` the query succeeds, then `\+ Goal` is false; it is true only if for *all* possible substitutions for the variables in `Goal` the query is false (fails). Therefore `\+ Goal` intuitively means ∀`?Vars` ¬ `Goal`, where `?Vars` represents all the nonbound variables in `Goal`. However, here ¬ should be understood not as classical negation but rather as a statement that `Goal` cannot be proved to be true.

17.4 True vs. Undefined Formulas

The fact that the well-founded semantics for negation is three-valued brings up the question of what exactly does the success or failure of a call mean. Is undefinedness covered by success or by failure? The way this is implemented in XSB is such that a call to a literal, P, succeeds if and only if P is true *or* undefined. Therefore, it is sometimes necessary to be able to separate true from undefined facts. In $\mathcal{E}$RGO, this separation is accomplished with the $\mathcal{E}$RGO primitives `true{`*Goal*`}` and `undefined{`*Goal*`}`. For good measure, the primitive `false{`*Goal*`}` is also thrown in. For instance,

```
a[b->c].
e[f->g] :- \naf e[f->g].

?- true{a[b->c]}.

Yes

?- undefined{e[f->g]}.

Yes

?- false{k[l->m]}.

Yes
```

It should be noted that the primitives `true{...}` and `undefined{...}` can be used only in top-level queries or in the rules whose heads are not mutually recursive with any of the query literals. Otherwise, the result is undefined. The expression `false{Goal}` is equivalent to `\naf Goal`, and can be used anywhere.

In addition, sometimes it may be necessary to to check if a succeeding query succeeds with the truth value *true* or *undefined*. (A failing query never succeeds, so there is nothing to test in that case.) To this end, $\mathcal{E}$RGO provides the primitive `truthvalue` which takes a variable that gets bound to the truth value of the query. For instance, suppose that the query `q(?X)` has two answers: `?X=a` is a true answer and `?X=b` is undefined (not false!). Then

```
q(a).
q(b) :- \undefined.
?- q(?X), truthvalue{?_P},
   \if ?_P == \true \then writeln(?X=true)@\io
   \else writeln(?X=undefined)@\io.
b = undefined
a = true

?X = a
?X = b  - undefined
```

As with `true{...}` and `undefined{...}`, the `truthvalue{...}` primitive can be used only in top level queries or in rules that are not mutually recursive with any of the query literals.

17.5 Explicit Negation

Explicit negation is denoted using the connective `\neg`, for instance, `\neg p`. It is a weaker form of classical negation and is quite unlike the default negation. First, explicit negation can appear both in rule heads and rule bodies. Second, unlike the default negation or Prolog negation, in order to conclude `\neg p` one must actually *prove* `\neg p`—*not* simply fail to prove `p`. In other words, establishing `\neg p` is a harder requirement than establishing `\naf p`, and `\neg p` *always* implies `\naf p`.

Also, unlike classical negation, the law of excluded middle does not hold for `\neg`, so both `p` and `\neg p`. In contrast, `p` and `\naf p` cannot be both true and cannot be both false—it is always the case that one is true and the other is false. Can both `p` and `\neg p` be true? The answer may surprise: *yes*. $\mathcal{E}$RGO does not check that by default because this is somewhat expensive. To tell it to check this inconsistency, two things must be done:

1. The module where these facts occur must be declared as defeasible using the

`:- use_argumentation_theory.`
directive—see Section 38 and

2. The facts and the rules where the above literals occur in the rule heads must be made defeasible (also see Section 38). This can be done either by giving these rules/facts the defeasible property or by assigning them explicit defeasibility tags:

For example,

```
:- use_defeasibility_theory.
@@{defeasible} p.    // fact has defeasibility property
@{abc} \neg p.       // explicit defeasibility tag
?- p.
No
?- \neg p.
No
```

As you can see, in this case, $\mathcal{E}$RGO will detect an inconsistency and will "defeat" the offending inferences by making both of them false. One can also find out why the various inferences were defeated—see Section 38.

The explicit negation connective, `\neg`, can be applied to conjunctions or disjunctions of literals. In a rule body, the idioms `\neg\naf p` and `\neg \+ p` are illegal, but `\naf\neg p` and `\+\neg p` are legal and should be informally understood as statements that `p` is not known to be false.

As with the default negation `\naf`, `\neg` is allowed inside the frames in front of `->` and `=>` and in front of Boolean (non-transactional) methods. For instance,

```
a[\neg b->c], a[|\neg b => c|], a[\neg b], a[|\neg b|]
```

are allowed, but `a[\neg %b]`, `a[\neg b ->-> c]`, `a[\neg b +> c]` are illegal.

At present, explicit negation works with defeasible reasoning, but it is not treated in any special way in modules that do not use defeasible reasoning. For instance, no consistency check is done to ensure that `p` and `\neg p` are not true at the same time. This is future work.

Explicit negation is fully integrated with meta-programming. For instance, the following is valid syntax:

```
?- ?X = ${a[b->c]}, \neg ?X.
```

This is equivalent to `?- \neg a[b->c]`. The meta-unification and meta-decomposition operators, $\sim$ and `=..`, introduced in Sections 16.1 and 16.3, are also aware of explicit negation. For instance,

```
?- ${\neg a[b->c]@foo} =.. [neg_flogic(->,foo), a, b, c].
?- \neg a[b->?C]@foo ~ \neg ?X@?Z, \neg ?X ~ ${\neg ?A[?B->cc]@foo}.
```

The second query produces the following answer:

```
?C = cc
?X = ${\neg \neg a[b -> cc]@foo}
?Z = foo
?A = a
?B = b
```

The answer for `?X` might look a little strange, but double negation works as expected:

```
?- ${\neg \neg a[b -> cc]@foo} ~ ${a[b -> cc]@foo}.
```

Transactional HiLog literals and methods cannot be negated using `\neg`. For instance, the following literals are syntax errors:

```
\neg %p
\neg a[%p]
```

Similarly, $\mathcal{E}$RGO update or aggregate operations cannot be negated.

18 General Formulas in Rule Bodies

Unlike most other rule engines, $\mathcal{E}$RGO supports a much larger variety of formulas in the rule body. In due time we will see various if-then-else clauses, loops, etc. But the free use of logical quantifiers `forall` (`all`, `each` are also accepted) and `exists` (`exist`, `some` are also accepted) is, perhaps the most unique. In fact, the rule bodies in $\mathcal{E}$RGO can have arbitrary formulas involving quantifiers, `\and`, `\or`, `\naf`, and the logical implications `~~>`, `<~~`, `<~~>`, `==>`, `<==`, and `<==>`. Here are some examples:

```
h1(?Z) :- \naf exists(?X)^(\naf ((p(?X) ~~> q(?X,Z)))).
h2(?Y) :- forall(?X)^(p(?X,?Y) ~~> q(?X)).
h3(?V,?W) :- forall(?Y,?Z)^(\naf exists(?X)^(pp(?X,?W,?Y,?Z) , \naf qq(?X,?Y,?V,?Z))).
?- forall(?X)^(p(?X)<~~>p(?X)).
```

Other connectives, such as `==>`, `<==`, `<==>`, `if-then-else`, etc., can also be used in conjunction with the quantifiers. The definitions of all these implications are as follows:

- ϕ`~~>`ψ (and `\if` ϕ `\then` ψ) is a shorthand for `\naf` ϕ `\or` ψ.
  ψ`<~~`ϕ is the same as ϕ`~~>`ψ, while ϕ`<~~>`ψ is (ϕ`~~>`ψ `\and` ϕ`<~~`ψ).
- ϕ`==>`ψ is a shorthand for `\neg` ϕ `\or` ψ.
  ψ`<==`ϕ is the same as ϕ`==>`ψ, while ϕ`<==>`ψ is (ϕ`==>`ψ `\and` ϕ`<==`ψ).
- `if` ϕ `then` ψ `else` η is (ϕ`~~>`ψ `\and` (`\naf` ϕ)`~~>`η).

It should be noted that it makes no sense to place non-logical operators in the scope of the quantifiers or `\naf`. For instance,

```
?- \naf load{abc}.
```

does not make a whole lot of sense (try it!). But some uses of quantification in conjunction with actions do make sense and are quite natural: For instance,

```
?- forall(?X)^(data(?X)~~>writeln(?X)@\io).
```

will print out the contents of the predicate `data/1`.

Quantifiers in $\mathcal{E}$RGO are implemented through the well-known Lloyd-Topor transformation [12] except that the role of negation here is played by `\naf`, not `\+`. However, unlike Prolog for which the Lloyd-Topor transform was originally defined, $\mathcal{E}$RGO gives a logical treatment to negation of non-ground formulas. In particular, as discussed in Section 17.2, when negating non-ground subgoals the truth value may be *undefined*. Quantifiers and `~~>` have implicit negation in them, so, in order to know when to anticipate undefinedness, it is useful to understand what it means to have a non-ground subgoal under `\naf`.

We say that a variable in the body of a rule occurs *positively* if it occurs in the scope of an even number of `\naf`s. A variable occurs *negatively* if it occurs in the scope of an odd number of `\naf`s. So, undefinedness can happen if some variable in the body of a rule occurs negatively and does not occur positively.

To make the above more precise, we explain how to count the `\naf`s. This is not immediate because of the implicit negations mentioned earlier. So, for the purpose of counting the `\naf`s, one must break each `forall(Varlist)^`ϕ into (`\naf exists(Varlist)^( \naf` ϕ)). Each ϕ`~~>`ψ must be broken into (`\naf` ϕ `or` ψ), and similarly with `<~~`; `if-then` is synonymous with `~~>`; and `<~~>` is a conjunction of `~~>`, and `<~~`. The connective `\if` ϕ `\then` ψ `\else` η is treated for the purpose of counting `\naf`s as ϕ `\and` `\naf` ψ `\or` `\naf` η. Note that `==>` and related connectives do not involve `\naf`s.

18.1 Quantification of Free and Anonymous Variables

In logic, free variables (i.e., variables that do not occur in any quantifier) do not normally make sense because it is unclear how to determine truth of a formula with such a variable. Nevertheless, non-quantified variables are widely used. In fact, languages like Prolog and SQL have no syntax for quantifiers, so one might be misled into thinking that all variables there are non-quantified. This is incorrect, however: both languages have a set of conventions by which all variables are quantified *implicitly*. In Prolog, for example, the convention is that *all* variables are implicitly quantified outside of the rule, i.e., $\forall X, Y, Z(head(...) : -body(...))$. This rule is often also stated equivalently as follows: variables that occur in the body *only* are quantified existentially in the body and all the rest are universal in the scope of the entire rule, i.e., $\forall X, Y, Z(head(X, Y, Z) : -\exists V, W(body(X, Y, Z, V, W)))$. The situation gets muddies when negation as failure is taken into account and we will not go there right now.

In $\mathcal{E}$RGO, the situation is a bit more interesting because here we have *explicit* quantifiers, but it is not mandatory to quantify all variables. According to the previous discussion, variables that are not quantified explicitly must be quantified implicitly according to some rules. What are these rules? The implicit quantification rules for free variables make a distinction between *named* variables (which includes *don't-care* variables) and *anonymous* variables. In addition, it matters whether a free variable occurs under the scope of a `\naf` or not:

- *Anonymous variables*:
  First, note that an anonymous variable is never explicitly quantified simply because it has no name (at least, no name the user knows about) and so one does not have anything to refer to that variable in the quantifier (something like `forall(?)` is not allowed because it can be highly ambiguous).
  So, the rule for such variables is that each anonymous variable that occurs in some literal, `p(...,?,...)` or `\neg p(...,?,...)`, is implicitly quantified with `exists` immediately before that literal, i.e., `exists(?var123)^p(...,?var123,...)` or `exists(?var123)^\neg p(...,?var123,...)`, where `var123` is some internal compiler-generated name for that anonymous variable. This existential quantifier takes precedence over `\naf` (but not over `\neg`, as we just saw).

- *Named variables that do <u>not</u> occur under the* `\naf`:
  The rules for these variables are the same as in Prolog: if they occur in the head (and possibly in the body also) then they are universal with respect to the entire rule; if they occur in the body only then they are existential with respect to the body.

- *Named variables under the* `\naf`:
  This implicit quantification rule has a bit of computational flavor and is not completely declarative, which has to do with the fact that the declarative definition for `\naf` exists for

top-down computation essentially only for the variable-free case.
If a named variable gets grounded during the evaluation *before* the evaluation of `\naf` starts then it is not a variable any more and the question is moot. If it remains unbound by the time `\naf` gets computed then the variable is treated existentially *outside* of the scope of `\naf`. In many cases, this situation leads to answers whose truth value is *undefined*. This because $\mathcal{E}$RGO quantifiers do not specify the domain for the quantified variables and for some domains such queries may be true and for some false.

18.2 The Difference Between `~~>`, `==>` (or, `<~~`, `<==`), and `:-`

Note that the connectives `==>`, `<==`, and `<==>` do not involve `\naf`: they are translated using `\neg`. For instance, ϕ`==>`ψ is a shorthand for `(\neg` ϕ`) \or` ψ, while ϕ`~~>`ψ is a shorthand for `(\naf` ϕ`) \or` ψ (and `\if-\then` is synonymous to `~~>`). This means that there is a significant semantic difference between these types of formulas:

- ϕ`==>`ψ is true iff: when ϕ is *not known* to be false (i.e., when `\neg` ϕ *cannot* be proven true) then ψ is also true.
- ϕ`~~>`ψ is true iff: when ϕ is true then ψ is also true.

The second statement is true in *more* cases. To see this, recall that ϕ`==>`ψ is `(\neg` ϕ`) \or` ψ while ϕ`~~>`ψ is `(\naf` ϕ`) \or` ψ. Note that `(\neg` ϕ`)` implies `(\naf` ϕ`)` i.e., the latter is true more often than the former. Therefore, `(\naf` ϕ`) \or` ψ is true more often than `(\neg` ϕ`) \or` ψ.

Experience shows that in most cases ϕ`~~>`ψ is the correct intended usage in rule bodies. If the user feels that `==>` is needed, a careful analysis of the intended meaning based on above cases is strongly suggested.

The other important differences are syntactic: ϕ`==>`ψ can appear both in the body of a rule (i.e., on the right of $:-$ and in the head (on the left of $:-$) provided that neither ϕ nor ψ use the `\naf`. In contrast, ϕ`~~>`ψ can be used only in the body of a rule because its very definition involves `\naf`.

As to `:-`, this connective is very different from the aforementioned ones, as it is the only one that forms what we call *rules*. What's on the left of `:-` is the head of the rule and what's on the right is the body. Neither `~~>` nor `==>` (nor their leftward brethren `<~~` and `<==`) form rules: they can appear only in rule bodies in $\mathcal{F}$LORA-*2* (but in $\mathcal{E}$RGO `==>` can also appear in rule heads). But the main difference is that these connectives *test* if, say, ϕ`~~>`ψ is true (false, or undefined) for some variable bindings in ϕ and ψ. These tests are part of the process of determining if the body of a rule is true (false or undefined). In contrast, *head*`:-`*body* tests if *body* is true (undefined) and, if so, *derives* that *head* is also true (respectively, undefined). For instance, consider the following example:

```
p(1), q(1), r(2,1).
w(?X) :- r(?X,?Y), (p(?Y) ~~> q(?Y)).
```

Here one can verify that `p(?Y) ~~>q(?Y)` is true when $?Y$ is bound to 1, and it so happens that `r(?X,?Y)` is also true for this binding if `?X` is bound to 2. Then the rule on the second line of the example *derives* the fact `w(2)`. It is important to understand that, in contrast, neither `p(?Y) ~~>q(?Y)` nor `p(?Y) ==>q(?Y)` derive anything. In $\mathcal{E}$RGO, `p(?Y) ==>q(?Y)` may also occur in rule heads, in which case it will take part in derivation of new facts. This is discussed in Section 19.

19 General Formulas in Rule Heads

$\mathcal{E}$RGO greatly extends the syntactic form of the rules in $\mathcal{F}$LORA-*2* by allowing general first-order formulas in the rule heads. This feature is particularly important for translating English sentences into logic, but also in other kinds of reasoning about knowledge, especially when contrapositive forms of the rules are required during the reasoning.

In $\mathcal{E}$RGO, the head formulas can be statements composed out of the following elements in a logically correct syntactic forms. That is, a *head-formula* is a logical formula recursively defined as follows:

- an atomic formula (frame or HiLog);
- `\neg` *head*, where *head* is a head-formula.
- $head_1, head_2$ and $head_1; head_2$, where $head_1$ and $head_2$ are head formulas. As before, `\and` and `\or` can be used in place of "," and ";", respectively.
- (*head*), where *head* is a head-formula.
- $head_1$ *op* $head_2$, where $head_1$ and $head_2$ are head formulas and *op* is `==>`, `<==`, or `<==>`.
- `forall`(*VarList*)^(*head*) and `exist`(*VarList*)^(*head*), where *head* is a head-formula and `VarList` is a comma-separated list of variables. The keyword `exists` can be used in place of `exist` and parentheses can be omitted, if *head* is an atomic formula.

Note that `\naf`, the implications `~~>`, `<~~`, `<~~>`, and the rule connective `:-` are *not permitted* in rule heads. Likewise, builtins (e.g., =, ! =, `\is`, etc.), calls to Prolog, and non-logical operators (e.g., `load`, `compile`, updates, etc.) are prohibited. This type of formulas must always go into rule bodies. Also note that formulas like

```
ϕ <== ϕ.
ϕ ==> ϕ.
```

are syntactically classified in $\mathcal{E}$RGO as *facts*, *not* rules, even though it is tempting to call these constructs rules because they are implications. Thus, both ϕ and ψ appear in the *head* there (since facts can be viewed as body-less rules). In $\mathcal{E}$RGO, a rule always has the form ϕ`:-`ψ, and only this type of constructs have rule bodies (ψ in this case). We will explain the meaning of `==>` and `<==` in rule heads shortly.

For clarity, consider some examples first. Suppose we know that `Agent666` is either a terrorist or works for an enemy state and that either case is dangerous:

```
enemystate(Agent666) \or terrorist(Agent666).
dangerous(?X) :- terrorist(?X) ; enemystate(?X).
```

In classical logic, it would be possible to derive `dangerous(Agent666)`, but $\mathcal{E}$RGO cannot reason by cases and it cannot reach such a conclusion just yet. However, if we also know that `\neg enemystate(Agent666)` is true then $\mathcal{E}$RGO would be able to derive `terrorist(Agent666)` and thus also `dangerous(Agent666)`.

The inability of $\mathcal{E}$RGO to reason by cases is a fundamental limitation of this class of logic languages, and working around it is difficult.[10] For instance, even if `foo \or foo` is given, $\mathcal{E}$RGO still cannot conclude `foo`.

For another example, let us represent the statement that U.S. Senate has exactly two senators from each state. We give two alternative representations:

```
// version 1
forall(?St)^exist(?Sn1,?Sn2)^(senator(?St,{?Sn1,?Sn2}) <== state(?St)).
// version 2
exist(?Sn1,?Sn2)^senator(?St,{?Sn1,?Sn2}) :- state(?St).
state(NY).
state(AL).
state(TX).
...
```

The two statements have somewhat different properties with respect omniformity—a concept discussed next.

19.1 Omniformity of `<==`, `==>`, and `<===>`

We have already seen that the implication `<==` and its related forms are different from `<~~` and friends in that the latter cannot occur in the rule heads. They are also different from the rule

[10] But this has very important advantages from the point of view of computational complexity the reasoning.

connective `:-` in that the body of a `:-` can be more general and `<==` – statements are not even called rules. There are also other differences. The most interesting one is the *omniformity* property. Consider the following statements:

```
mother1(?X,?Y) :- female(?X), parent(?X,?Y).       // (*)
mother2(?X,?Y) <== female(?X), parent(?X,?Y).      // (**)
female(Mary), parent(Mary,Bob).
\neg mother1(Bob,Bill), \neg mother2(Bob,Bill), parent(Bob,Bill).
\neg mother1(Sally,Peter), \neg mother2(Sally,Peter), female(Sally).
```

Statement (*) lets us derive `mother1(Mary,Bob)`, while statement (**) allows us to conclude `mother2(Mary,Bob)`. So far the conclusions are similar. However, the facts given in the example should also let us conclude more, if we use the contrapositive statements corresponding to the implications (*) and (**). For instance, since we know that Bob is not the mother of Bill, we should be able to conclude that Bob is not a female. The catch is that the semantics of `:-` does *not* sanction any contrapositive inferences, while the semantics of `<==` *does*. Thus, the contrapositive forms of (**)

```
\neg female(?X) <== parent(?X,?Y), \neg mother2(?X,?Y).
\neg parent(?X,?Y) <== female(?X), \neg mother2(?X,?Y).
```

let us conclude `\neg female(Bob)` and `\neg parent(Sally,Peter)`. This sanctioning of the contrapositive implications that are not explicitly written is called *omniformity*, and it is the property of `<==`, `==>`, and `<==>` *when they occur in the rule heads* (or facts). (Recall from Section 18.2 that in rule bodies these connectives act as tests only, no omniformity.)

For a more involved example, let us come back to the statement that every U.S. state has two senators:

```
// version 1
forall(?St)^exist(?Sn1,?Sn2)^(senator(?St,{?Sn1,?Sn2}) <== state(?St)).
// version 2
exist(?Sn1,?Sn2)^senator(?St,{?Sn1,?Sn2}) :- state(?St).
state(NY).
...
```

Since Puerto Rico is not a state, it would not be listed among the states, but this is still not enough to conclude `\neg state(PuertoRico)`.[11] However, if we also assert that Puerto Rico has no senators

[11] But is enough to evaluate `\naf state(PuertoRico)` to true. However, remember that `\naf` cannot occur in rule heads, so it is not possible to *derive* `\naf state(PuertoRico)`.

Portions Copyright © 2013–2017 Coherent Knowledge

```
\neg senator(PuertoRico,?).              // this
\neg exist(?X)^senator(PuertoRico,?X).   // or that
```

then the version 1 implication above but not (version 2!) will sanction the derivation of `\neg state(PuertoRico)`.

Note that one has to be careful in the way the absence of senators is specified. The following might seem to be a correct way to say this and, indeed, it *is* a correct way to express the absence of senators in *classical* logic:

```
\neg exist(?X,?Y)^senator(PuertoRico,{?X,?Y})    //  (***)
```

The trouble is that this is equivalent to

```
forall(?X,?Y)^(\neg senator(PuertoRico,?X) \or \neg senator(PuertoRico,?Y))
```

and, since $\mathcal{E}$RGO cannot reason by cases, it cannot conclude

```
forall(?X)^(\neg senator(PuertoRico,?X))
```

from the above. For that reason, we cannot use a contrapositive instance of

```
forall(?St)^exist(?Sn1,?Sn2)^(senator(?St,{?Sn1,?Sn2}) <== state(?St))
```

namely,

```
\neg state(PuertoRico):- forall(?Sn1,?Sn2)^(\neg senator(PuertoRico,{?Sn1,?Sn2}).
```

since we lack the necessary premise. Therefore, we would not be able to derive that Puerto Rico is not a state, if we the absence of senators were given using (***).

19.2 Omniformity and Builtins

As follows from the above discussion, the head-formulas that use `<==` and related connectives behave very much like rules (since they are translated into sets of rules) and can be used in lieu of them in many contexts. One thing to remember, however, is that these constructs *cannot* contain builtins. Suppose we want to represent the statement "Customers over 65 years old are eligible to receive discounts." We could try

```
eligible(?X) <== customer(?X), birthYear(?X,?YoB),
                 \date[now->?[year->?Yr]]@\basetype,
                 ?Yr-?YoB > 64.
```

but this implication contains a builtin, which cannot occur in this kind of statements. The solution is to make a rule out of the above and put the forbidden predicates in the body:

```
eligible(?X) <== customer(?X) :-                          //  (****)
                 birthYear(?X,?YoB),
                 \date[now->?[year->?Yr]]@\basetype,
                 ?Yr-?YoB>64.
```

Note that the above still enjoys the omniformity feature with respect to the head-implication. That is, if we know

```
customer(Bob), birthYear(Bob,1940).
\neg eligible(Bill), birthYear(Bill,1945).
```

then we can use (****) to conclude `eligible(Bob)` and use the contrapositive form of (****) to conclude `\neg customer(Bill)`.

20 Inheritance of Default Properties and Types

In general, inheritance means that attribute and method specifications for a class are propagated to the subclasses of that class and to the objects that are instances of that class. This section describes the extensive support for inheritance both for default values as well as for typing information.

$\mathcal{E}$RGO supports two types of inheritance: *structural* and *behavioral*. Structural inheritance applies to signatures only, i.e., to *class* frame formulas that use the `=>`-style arrows. These formulas specify the *type information* for classes as a whole. For instance, if `student::person` holds and we have the signature `person[|name=>string|]` then the query `?- student[|name=>?X|]` succeeds with `?X=string`.

Behavioral inheritance is much more involved. $\mathcal{E}$RGO supports two versions of behavioral inheritance—*monotonic* and *non-monotonic*—and the choice can be specified on the per-module basis. In both cases, behavioral inheritance concerns class frame formulas that use the `->`-style arrows or to Boolean class frames, and these formulas are inherited to subclasses and class members. The key difference is that *monotonic* inheritance (both structural and behavioral) is cumulative and resembles the way types are inherited in structural inheritance. In contrast, behavioral inheritance is *non-monotonic* in the sense that the formulas being inherited are understood as default

specifications that can be overridden by the information explicitly specified for subclasses. This also implies that adding new information to subclasses may invalidate previously true facts, i.e., true (inferred) information does not necessarily grow monotonically as we add more data.

20.1 Introduction to Inheritance

$\mathcal{E}$RGO distinguishes between information defined for a class as a whole and information defined for an individual object only. The former, class-wide information is inherited to the members of the class and to its subclasses, and is specified using the frame formulas of the form (note the vertical bars):

```
obj[|Meth->Val|]
obj[|Meth=>Val|]
obj[|BoolProp|]
obj[|=>BoolProp|]
```

These formulas normally occur as part of the specifications for classes. Object-specific information is given using the formulas of the form (no vertical bars!):

```
obj[Meth->Val]
obj[Meth=>Val]
obj[BoolProp]
obj[=>BoolProp]
```

This information is always attached to individual objects. Even if an object represents a class (and in $\mathcal{E}$RGO classes are also objects), this information does *not* apply to the members of that class or its subclasses. For instance,

```
person[avg_age -> 40].
```

does not propagate to an object such as `John` even if `John:person` is true. The property `avg_age->40` is likewise not inherited by the various subclasses of `person`, such as `student` or `employee`: it would not make sense to propagate such information because an average age of all persons is likely to be different from the average age of students and employees, and this is why we wrote `person[avg_age->40]` and not `person[|avg_age->40|]`. Similarly, attributes that typically refer to individuals are better specified as object-level information, because normally there is nothing to inherit these attributes to. For instance,

```
John[age -> 30].
```

makes sense but neither `person[avg->30]` nor `person[|avg->30|]` does.

Class formulas typically define *default* properties for the objects in a class, which are inherited unless there is information to the contrary. For instance, suppose we have

```
British[|nativeLanguage -> English|].
```

If `John:British` is true, then, without evidence to the contrary, we can derive `John[nativeLanguage -> English]`. If we are also told that `Scottish::British`, i.e., Scottish people are also British, then we can derive (again, in the absence of evidence to the contrary) that `Scottish[|nativeLanguage -> English|]`.

Note that a class formula becomes an object-level formula when its information is inherited to the members of the class. For instance, if we are also told that `John:British` then `John` inherits the native language as follows: `John[nativeLanguage->English]` (no bars! because there is nothing to inherit to from `John`). However, the inherited property remains a class formula when it is inherited to a subclass (e.g., `Scottish[|nativeLanguage->English|]`).

Suppose now that we are told that `John` is actually a native speaker of `Gaelic` via the fact `John[nativeLanguage->Gaelic]`. In that case, the explicitly given property `nativeLanguage -> Gaelic` *overrides* the default property `nativeLanguage->English` and the latter is *not* inherited.

Overriding can happen also at the level of classes. Suppose $\mathcal{E}$RGO has this pair of facts: `American[|nativeLanguage->English|]` and `Manuel:American`. In the absence of any other information, we would have derived by inheritance that `Manuel[nativeLanguage->English]`. But if our knowledge base also has these facts:

```
PuertoRican::American.
PuertoRican[|nativeLanguage->Spanish|].
Manuel:PuertoRican.
```

then $\mathcal{E}$RGO would discover that, since `PuertoRican` is a subclass of `American`, the property `nativeLanguage->Spanish` is more specific to `Manuel` than `nativeLanguage->English`, and so, by inheritance, it would derive `Manuel[nativeLanguage->Spanish]`. Had $\mathcal{E}$RGO been told that `Manuel[nativeLanguage->Portuguese]`, this most specific information would override that inheritance of `nativeLanguage->Spanish`.

The signature frame formulas are inherited similarly, but there is no overriding. For instance, suppose that we have a class `language`, which contains objects such as `English`, `Spanish`, `Gaelic`, `French`, etc. The knowledge base might have a fact like

```
American[|nativeLanguage => language|].
```

Portions Copyright © 2013–2017 Coherent Knowledge

which is read as a *type constraint* stating that, for every American, the property `nativeLanguage` can take only the values that are members of the class `language`. That is, in the above example, `Manuel[nativeLanguage->Spanish]` (or even `Gaelic`) would be ok, but `Manuel[nativeLanguage -> abracadabra]` is not. This is because, as we said earlier, `Spanish` and `Gaelic` are known to be in class `language`, but it is impossible to derive `abracadabra:language` given our knowledge base. Section 43.2 explains how type constraints can be checked and Section 36 also explains the mechanism for automatic enforcement of such constraints.

Like defaults, type information is inheritable. For instance, suppose the knowledge base has this information:

```
American[|nativeLanguage => language|].
PuertoRican::American.
Manuel:PuertoRican.
```

$\mathcal{E}$RGO would then derive the following by inheritance:

```
PuertoRican[|nativeLanguage => language|].
Manuel[nativeLanguage => language].
```

Note that since `Manuel` is a member of `PuertoRican` and not a subclass, it has lost the vertical bars during the inheritance, but the class `PuertoRican` still has them. From here on, nothing can be inheritable from `Manuel` even if this object can somehow be considered as a class. For instance, suppose `Manuel` has two alter egos: `ManuelJekyll:Manuel` and `ManuelHyde:Manuel`. Despite the fact that now `Manuel` has its own class members, since `Manuel[nativeLanguage->Spanish]` is a property of `Manuel` *as individual* and not as a class (the missing bars tell us that!), the property `nativeLanguage->Spanish` is not inherited to either `ManuelJekyll` or `ManuelHyde`. Similarly, `Manuel[nativeLanguage => language]` specifies the type of `Manuel` as an individual object and so it is inherited to neither of the two alter egos.

To illustrate the no-overriding property of signature inheritance, suppose the knowledge base also has `PuertoRican[|nativeLanguage => latinLanguage|]`. Unlike the inheritance of default values, this will not override the inheritance of the type `PuertoRican[|nativeLanguage => language|]`: it will still be inherited and this simply means that the actual values must lie in the intersection of the classes `language` and `latinLanguage`, which in this case is simply `latinLanguage`. However, this also means that one must be careful here. If the knowledge base had `American[|nativeLanguage => germanicLanguage|]` then `PuertoRican[|nativeLanguage => germanicLanguage|]` would also be true, but `Spanish` would not be in the intersection of the types `germanicLanguage` and `latinLanguage` (assuming our knowledge base is a good model of the real world). Such a mistake can be detected via type-checking (Section 43.2), which should prompt a revision of the way the types are specified. We should note, however, that an effect similar

to inheritance overriding *can* be achieved via the mechanism of *defeasible reasoning* described in Section 38.

20.2 Monotonic Behavioral Inheritance

The default for behavioral inheritance is *non*-monotonic, so to request *monotonic* inheritance one must use the following compiler directive:

```
:- setsemantics{inheritance=monotonic}.
```

This semantic can also be requested at runtime by executing the command

```
?- setsemantics{inheritance=monotonic}.
```

(Of course, `?-` is to be used only when such a command appears in a file; it should be omitted on the $\mathcal{E}$RGO shell command line.) Here is an example of how monotonic behavioral inheritance works:

```
:- setsemantics{inheritance=monotonic}.
d(?_x):a(?_x).
g(?_x)::a(?_x).
a(r)[|b(y)->c|].
g(r)[|b(y)->e|].
a(u)[|d(1)|].
g(u)[|d(2)|].

?- d(?I)[b(?X)->?Y].

?I = r           // inherited information
?X = y
?Y = c

?- g(?I)[|b(?X)->?Y|].

?I = r           // inherited information
?X = y
?Y = c

?I = r           // explicitly specified information
```

```
?X = y
?Y = e

?- d(?I)[d(?X)].

?I = u           // inherited information
?X = 1

?- g(?I)[|d(?X)|].

?I = u           // inherited information
?X = 1

?I = u           // explicitly specified information
?X = 2
```

20.3 Non-monotonic Behavioral Inheritance

Non-monotonic behavioral inheritance is the default, but sometimes one might also need or want to specify it explicitly—either for documentation or to override a differently specified inheritance at runtime. The compiler directive (that would appear in a file) is

```
:- setsemantics{inheritance=flogic}.
```

and the runtime command is

```
?- setsemantics{inheritance=flogic}.
```

The following is a $\mathcal{E}$RGO specification for the classical `Royal Elephant` example:

```
elephant[|color=>color, color->gray|].
royal_elephant::elephant.
clyde:royal_elephant.
```

The first statement says that the `color` property of an elephant must be of type `color` and that the default value is gray. The rest of the statements say that royal elephants are elephants and that `clyde` is an individual elephant. The question is what is the color of `clyde`? The color of that elephant is not given explicitly, but since `clyde` is an elephant and the default color for elephants is gray, `clyde` must be gray. Thus, we can derive:

```
clyde[color->gray].
```

Observe once again that when class information is inherited by class' individual members, the resulting formula becomes object-level rather than class-level (i.e., `...[...]` instead of `...[|...|]`). On the other hand, when this information is inherited by a subclass from its superclass, then the resulting formula is a class-level formula because it should be still applicable to the members of the subclass and to its subclasses. For instance, if we have

```
circus_elephant::elephant.
```

then we can derive

```
circus_elephant[|color->gray|].
```

Non-monotonicity of behavioral inheritance becomes apparent when new information gets added to the knowledge base. For instance, suppose we learn that

```
royal_elephant[|color->white|].
```

Although we have previously established that `clyde` is gray, this new information renders our earlier conclusion invalid. Indeed, since `clyde` is a royal elephant, it must be white, while being an elephant it must be gray. The conventional wisdom in knowledge representation is that inheritance from more specific classes must take precedence. Thus, we must withdraw our earlier conclusion that `clyde` is gray and infer that he is white:

```
clyde[color->white].
```

Nonmonotonicity also arises due to multiple inheritance. The following example, known as the Nixon Diamond, illustrates the problem. Let us assume the following knowledge base:

```
republican[|policy -> security|].
quaker[|policy -> pacifist|].
Nixon:quaker.
```

Since Nixon is a Quaker, we can derive `Nixon[policy -> pacifist]` by inheritance from the second clause. Let us now assume that the following information is added:

```
Nixon:republican.
```

Now we have a conflict. There are two conflicting inheritance candidates: `quaker[|policy -> pacifist|]` and `republican[|policy -> security|]`. In $\mathcal{E}$RGO, such conflicts cause previously established inheritance to be withdrawn and both `policy->pacifist` and `policy->security` become false for object `Nixon`. This behavior can be altered by adding additional rules and facts. For instance, adding `Nixon[policy->security]` would take precedence and override the inherited information. More generally, one could introduce priority over superclasses, say with a predicate `hasPriority`, and then add the rule

```
?Obj[policy->?P] :-
    ?Obj:?Class, ?Class[|policy->?P|], \naf hasPriority(?AnotherClass,?Class).
```

If we also had `hasPriority(republican,pacifist)` then inheritance from the class `republican` would take precedence.

Behavioral inheritance in F-logic is discussed at length in [16, 18]. The above non-monotonic behavior is just the tip of an iceberg. Much more difficult problems arise when inheritance interacts with regular deduction. To illustrate, consider the following:

```
b[|m->c|].
a:b.
a[m->d] :- a[m->c].
```

In the beginning, it seems that `a[m->c]` should be derived by inheritance, and so we can derive `a[m->d]`. Now, however, we can reason in two different ways:

1. `a[m->c]` was derived based on the belief that attribute `m` is not defined for the object `a`. However, once inherited, we must necessarily have `a[m->{c,d}]`. So, the value of attribute `m` is not really the one produced by inheritance. In other words, inheritance of `a[m->c]` negates the very premise on which the original inheritance was based, so we must give up the earlier conclusion made by inheritance as well as the subsequent inference made by the rule.

2. We did derive `a[m->d]` as a result of inheritance, but that's OK — we should not really be looking back and undoing previously made inheritance inferences. Thus, the result must be `a[m->{c,d}]`.

A similar situation (with similarly conflicting conclusions) arises when the class hierarchy is not static. For instance,

```
d[|m->e|]
d::b.
b[|m->c|].
```

```
a:b.
a:d :- a[m->c].
```

If we inherit `a[m -> c]` from `b` (which seems to be OK in the beginning, because nothing overrides this inheritance), then we derive `a:d`, i.e., we get the following: `a:d`, `d::b`. This means that *now* `d` seems to be negating the reason why `a[m -> c]` was inherited in the first place. Again, we can either undo the inheritance or adopt the principle that inheritance is never undone.

A semantics that favors the second interpretation was proposed in [9]. This approach is based on a fixpoint computation of non-monotonic behavioral inheritance. However, this semantics is unsatisfactory in many respects, including because it is procedural in nature. $\mathcal{E}$RGO uses a different, more cautious semantics for inheritance, which favors the first interpretation above.

Details of this semantics are formally described in [16]. Under this semantics, `clyde` will still inherit color white, but in the other two examples `a[m->c]` is *not* concluded. The basic intuition can be summarized as follows:

1. Method definitions in subclasses override the definitions that appear in the superclasses.
2. In case of a multiple inheritance conflict, the result of inheritance is false.
3. Inheritance from the same source through different paths is *not* considered a multiple inheritance conflict. For instance, in

   ```
   a:c.    c::e.      e[|m->f|].
   a:d.    d::e.
   ```

   Even though we derive `c[|m->f|]` and `d[|m->f|]` by inheritance, these two facts can be further inherited to the object `a`, since they came from a single source `e`.

   On the other hand, in a similar case

   ```
   a:c. c[|m->f|].
   a:d. d[|m->f|].
   ```

   inheritance does not take place (`a[m->f]` is false), because the two inheritance candidates, `c[|m->f|]` and `d[|m->f|]`, are considered to be in conflict.

   Note that in the last example one might argue that even if we did inherit both facts to `a`, there would be no discrepancy because, in both cases, the values of the attribute `m` agree with each other. However, $\mathcal{E}$RGO views this agreement as accidental: had one of the values changed to, say `d[m->g]`, there would be a conflict.

Portions Copyright © 2013–2017 Coherent Knowledge

4. At the level of methods of arity > 1, a conflict is considered to have taken place if there are two non-overridden definitions of the same method attached to two different superclasses. When deciding whether a conflict has taken place we disregard the arguments of the method. For instance, assuming that `c` and `d` are classes that are incomparable with respect to `::`, the following has a multiple inheritance conflict

```
a:c. c[|m(k)->f|].
a:d. d[|m(u)->f|].
```

even though in one case the method `m` is applied to object `k`, while in the other it is applied to object `u`. On the other hand,

```
a:c. c[|m(k)->f|].
a:d. d[|m(k,k)->f|].
```

do not conflict, because `m/1` in the first case is a different method than `m/2` in the second. Similarly,

```
a:c. c[|m(k)()->f|].
a:d. d[|m(u)()->f|].
```

are not considered to be in conflict because here it is assumed that the method names are `m(k)` and `m(u)`, which are distinct names. Finally,

```
a:c. c[|m(k)()->f|].
     c[|m(u)()->f|].
```

is likewise *not* a conflict because inheritance here comes from the same class `c`.

In the examples that we have seen so far, path expressions queried only object-level information. To query class-level information using path expressions, $\mathcal{E}$RGO uses the symbol "!". For instance,

```
royal_elephant!color
```

means: some `?X` such that `royal_elephant[|color->?X|]` is true. In our earlier example, `?X` would be bound to `white`.

20.4 Inheritance of Negative Information

Certain kind of negative information can also be inherited, but inheritance goes in the opposite direction: from class members and subclasses to superclasses. The following subsections discuss this issue.

20.4.1 Negative Monotonic Behavioral Inheritance

If monotonic behavioral inheritance is requested using the `setsemantics` primitive, negative data frames propagate from objects to their classes as class-level frames, *i.e.*, as frames of the form `[|...|]`. Similarly, class-level data frames are propagated from subclasses to superclasses. For example, given the following data

```
:- setsemantics{inheritance=monotonic}.
obj:c1.
c1::c2.
obj[\neg prop->val].        // equivalently:  \neg obj[prop->val]
c1[|\neg prop2->val2|].     // equivalently:  \neg c1[|prop2->val2|]
obj[\neg boolprop].         // equivalently:  \neg obj[boolprop]
c1[|\neg boolprop2|].       // equivalently:  \neg c1[|boolprop2|]
```

the following queries will return the answers as shown:

```
?- c2[|\neg ?prop->?val|].  // equivalently:  \neg c2[|?prop->?val|]
?prop = prop
?val = val

?prop = prop2
?val = val2

?- c2[|\neg ?boolprop|].    // equivalently:  \neg c2[|?boolprop|]

?boolprop = boolprop

?boolprop = boolprop2
```

20.4.2 Negative Structural Inheritance

Since structural inheritance is monotonic, negative structural inheritance works similarly to negative *monotonic* behavioral inheritance. This means that negative signatures from class members propagate to become negative class-level (i.e., `[|...|]`) signatures for superclasses. Similarly, class-level signatures propagate from subclasses to superclasses. For instance, given

```
obj:c1.
c1::c2.
```

```
obj[\neg prop => type].       // equivalently:  \neg obj[|prop=>type|]
obj[\neg =>boolprop].         // equivalently:  \neg obj[|=>boolprop|]
c1[|\neg prop2 => type2|].
c1[|\neg =>boolprop2|].
```

the following queries show how the negative properties propagate towards the class `c2`:

```
?- c2[|\neg ?prop=>?type|].   // equivalently:  \neg c2[|?prop=>?type|]

?prop = prop
?type = type

?prop = prop2
?type = type2

?- c2[|\neg =>?prop|].       // equivalently:  \neg c2[|=>?prop|]

?prop = boolprop

?prop = boolprop2
```

20.4.3 Negative Non-monotonic Behavioral Non-Inheritance

For non-monotonic inheritance, propagation of negative information does not occur. First, it is a logical fallacy to expect to inherit negative information like this:

```
c[|\neg attr->1|].
obj:c.
cc::c.
```

inferring

```
?- obj[\neg attr->1].
?- cc[|\neg attr->1|].
```

This does not make logical sense because the fact `c[|\neg att->1|]` says that *it is known that* `att->1` *is not a default for class* `c`. One cannot logically conclude from this that it is also known that `att->1` is false for `obj` or that `att->1` is not a default for the subclass `cc` of `c`.

Neither does negative information propagate upwards the class hierarchy as was the case with monotonic inheritance of signatures and behavior. For instance, given

```
obj::c.
obj[\neg attr->1].
```

it does not follow that the default for `attr` in class `c` is not `attr->1` (i.e., `c[|\neg attr->1|]` has no logical justification; in fact, `c[|attr->1|]` could well be true).

The overall intuition for the desire to inherit negative information can nevertheless be achieved through a more powerful feature of *defeasible* reasoning described in Section 38. For instance, one could write a rule like

```
@{default(cl)} ?Obj[\neg att->1] :- ?Obj:cl.
```

which says that, by default, `obj[att->1]` is known to be false if `obj` happens to be a member of class `cl`. Other statements for the members of class `cl` may have higher priority and override the above. For instance,

```
abc:cl.
@{highpriority} abc[att->1].
\overrides(highpriority,default(cl)).
```

This says that, for the specific member `abc` of class `cl`, `att->1` is actually true because this information is specified with higher priority (`highpriority`) than the above default rule (`default(cl)`), as indicated by the priority fact `\overrides(highpriority,default(cl))`.

Despite all of the above, negative information does play a role in *blocking* inheritance. More specifically, explicit negative information specified for subclasses and class members is treated as explicit local statements that override (block) the inherited information. For instance, in

```
c[|attr1->{1,2}, attr2->{3,4}, attr3->{5,6}|].
cc::c.
obj:cc.
cc[|\neg attr1->1|].   // blocks inheritance of attr1->1 only
obj[\neg attr2->{}].   // blocks inheritance of attr2->anything
?- obj[attr1->?X].     // ?X = 2. Inheritance of atr->1 is blocked
?- cc[|attr1->?X|].    // same
?- obj[attr2->?X].     // no answers: all inheritance is blocked
?- cc[|attr2->?X|].    // ?X = 3,4: inheritance is blocked at obj, below cc
?- obj[attr3->?X].     // ?X = 5,6: inheritance occurs, nothing is blocked
?- cc[|attr3->?Y|].    // same
```

the explicit negative statement `cc[|\neg attr1->1|]` blocks the inheritance of `c[|attr1->1|]` down to both `cc` and `obj`. The explicit negative information `obj[\neg attr2->{}]` blocks the

inheritance of `c[|attr2->val|]` down to `obj`, for *any* `val`. For `attr3` (and for `attr2` at the level of `cc`), however, no explicit blocking negative information exists, so the last three queries will report that data is inherited in full.

20.5 Code Inheritance

The type of behavioral inheritance defined in the previous subsection is called *value inheritance*. It originates in Artificial Intelligence, but is also found in modern main stream object-oriented languages. For instance, it is related to inheritance of static methods in Java. With this inheritance, one would define a method for a class, e.g.,

```
cl[|foo(?Y) -> ?Z|] :- ?Z \is ?Y**2.
```

and every member of this class will then inherit exactly the same definition of `foo`. Since the method definition has no way to refer to the instances on which it is invoked, this method yields the same result for all class instances. One way to look at this is that class instances do not really inherit the definition of the method. Instead, the method is invoked in the context of the class where it is defined and then the computed value is inherited down to all instances (provided that they do not override the inheritance). So, if `a:cl` and `b:cl` then `a.foo(4)` and `b.foo(4)` will return exactly the same value, 16.

A more common kind of methods is called *instance methods* in Java. In this case, the method definition refers to instances of the class in whose context the method is supposed to be invoked. The invocation takes place as follows. First, a class member inherits the *code* of the method. Then the code is executed in the context of that class member.

In F-logic this kind of inheritance is called *code inheritance* and it was studied in [17, 18]. Code inheritance is not yet supported by $\mathcal{E}$RGO. However, with some loss of elegance and extra work, code inheritance can often be simulated using value inheritance. The idea consists of three steps.

1. Define the desired methods for all appropriate objects irrespective of classes. Definitions of these methods are the ones to be inherited using simulated code inheritance. These are auxiliary methods used in the process.

2. Define the attributes whose values are the names of the methods defined in step 1. These attributes will be subject to value inheritance.

3. Specify how the "real" methods in step 1 represented by the "fake" methods in step 2 are to be invoked on class instances.

We illustrate this process with the following example. First, assume the following information:

```
aa:c1.
bb:c2.
c1::c2.
aa[attr1->7, attr2->2].
bb[attr1->5, attr2->4].
```

We are going to show how code is inherited from `c2` to `bb`. In an attempt to inherit the same code from `c2` to `aa`, it will be overridden by the code from `c1` and the latter (more specific code at `aa`) will be inherited by `aa`.

Step 1: define auxiliary methods.

```
// auxiliary method foo/1 defined for every instance
?X[foo(?Y) -> ?Z] :- ?X[attr1->?V], ?Z \is ?V+?Y.
// auxiliary method bar/1 defined for every instance
?X[bar(?Y) -> ?Z] :- ?X[attr2->?V], ?Z \is ?V*?Y.
```

Unlike Java, the above code is not really local to any class, and this is one aspect in which simulation of code inheritance by value inheritance is inelegant.

Step 2: define the "fake" method `meth`.
Next we define `meth` — the method whose value inheritance will simulate the inheritance of the code of the methods `foo` and `bar`.

```
c1[|dispatch(meth) -> bar|].
c2[|dispatch(meth) -> foo|].
```

Clearly, the object `bb` will inherit `dispatch(meth)->foo` from `c2`, while the object `aa` will inherit `dispatch(meth)->bar` from `c1`; inheritance from `c2` is overridden by the more specific class `c1`.

Step 3: finishing up.
Next, we define how methods are to be invoked in a way that resembles code inheritance:

```
?X[?M(?Y) -> ?Z] :- ?X[dispatch(?M)->?RealMeth], ?X[?RealMeth(?Y) -> ?Z].
```

When `?M` is bound to a particular method, say `meth`, and this method is invoked in the context of a class instance, `?X`, the invocation `?X[meth(?Y)->?Z]` first computes the value of the attribute `dispatch(meth)`, which gives the name of the actual auxiliary method to be invoked. The value of the `dispatch(meth)` attribute (represented by the variable `?RealMeth`) is obtained by value inheritance. As explained above, this value is `foo` when `?X` is bound to `bb` and `bar` when `?X = aa`. Finally, the auxiliary method whose name is obtained by value inheritance is invoked in the context of the class instance `?X`. One can easily verify the following results:

```
ergo>  aa[meth(4) -> ?Z].
?Z = 8

ergo>  bb[meth(4) -> ?Z].
?Z = 9
```

This is exactly what would have happened in Java if `aa` inherited the instance method whose code is that of `bar/1` and if `bb` inherited the code of `foo/1`.

21 Custom Module Semantics

$\mathcal{E}$RGO enables the user to choose the desired semantics for any user module. This is done with the help of the following directive:

```
:- setsemantics{Option1, Option2, ...}.
```

The following options are allowed:

Equality: `equality=none`, `equality=basic`, where `equality=none` is the default.

Inheritance: `inheritance=none`, `inheritance=flogic` `inheritance=monotonic`, where `inheritance=flogic` is the default.

Tabling: `tabling=reactive` (default), `tabling=passive`, `tabling=variant` (default), and `tabling=subsumptiv`

Custom: `custom=none`, `custom=filename`, where `custom=none` is the default.

These options are described in more detail in the following subsections. Within each group only one choice can be present or else an error will result. It is not required that all options be present — defaults are substituted for the missing options.

The compiler directive described above determines the initial semantics used by the module in which the instruction occurs. However, it is also possible to change the semantics at run time using the *executable directive*:

```
?- setsemantics{Option1, Option2, ...}.
```

Note the use of `?-` here: the symbol `:-` in the first directive designates the directives that are used at compile time only. Executable directives, on the other hand, can occur in any query or rule body. It is also possible for one module to change the semantics in another module. Typically this

is needed when one module creates another. In this case the new module is created with the default semantics, and the `setsemantics` executable directive makes it possible to change the semantics of such a module.

The following options are available only with the executable version of `setsemantics`, while the previously mentioned options can be used both at compile and run time.

Subclassing: `subclassing=strict`, `subclassing=nonstrict`, with *strict* being the default.

Class expressions: `class_expressions=on`, `class_expressions=none`; default: *none*.

Here is an example:

```
?- setsemantics{equality=basic, custom='a/b/c'}.
```

The order of the options in the directive does not matter.

Changing module semantics — precautions. Changing module semantics on the fly at run-time is a rather drastic operation. It is therefore *not* recommended to do this in the body of a rule, especially if the rule defines a tabled HiLog predicate or a frame. The only safe way to execute `setsemantics` is in a query at the top level. For instance,

```
?- setsemantics{...}.
```

21.1 Equality Maintenance

User-defined equality. $\mathcal{E}$RGO users can define equality *explicitly* using the predicate `:=:`. For instance,

```
John:=:Batman.
?X:=:?Y :- ?X[similar->?Y].
```

Once two oids are established to be equal with respect to `:=:`, whatever is true of one object is also true of the other. Note that `:=:` is different from the built-in `=`. The latter is a predefined primitive, which cannot occur in facts or in rule heads. Since `=` is understood as unification, ground terms can be =-equal only if they are identical. Thus, `a=a` is always true and `a=b` is always false. In contrast, the user can assert a fact such as `a:=:b`, and from then on the object `a` and the object `b` are considered the same (modulo the equality maintenance level, which is described below).

Equality maintenance levels. Once an equality between terms is derived, this information may need to be propagated to all F-logic structures, including the subclass hierarchy, the ISA hierarchy, etc. For instance, if `x` and `y` are equal, then so must be `f(x)` and `f(y)`. If `x:a` has been previously derived then we should now be able to derive `y:a`, etc. Although equality is a powerful feature, its maintenance can slow down the execution quite significantly. In order to be able to eat the cake and have it at the same time, $\mathcal{E}$RGO allows the user to control how equality is handled, by providing the following three compiler directives:

```
:- setsemantics{equality=none}.  // default
:- setsemantics{equality=basic}.
```

The first directive, `setsemantics{equality=none}`, does not maintain any equality and `:=:` is just a symmetric transitive relation that includes the identity. However, the congruence properties of equality are not supported (for instance, `p(a)` and `a:=:b` do not imply `p(b)`). The directive `setsemantics{equality=basic}` guarantees that `:=:` obeys the usual rules for equality, i.e., transitivity, reflexivity, symmetry, and (limited) congruence.

If a $\mathcal{E}$RGO module does not define facts of the form `a:=:b`, which involve the equality predicate `:=:`, then the default equality maintenance level is `none`. If the knowledge base does have such facts, then the default equality maintenance level is `basic`, because it is assumed that the use of `:=:` in the source is not accidental. In any case, the explicit `equality=...` option overrides the default.

Locality of equality. Equality in $\mathcal{E}$RGO is always local to the module in which it is derived. For example, if `a:=:b` is derived by the rules in module `foo` then the query

```
?- (a:=:b)@foo.
```

will succeed, but the query

```
?- (a:=:b)@bar.
```

will fail (unless, of course, `a:=:b` is also derived by the rules in module `bar`).

Since equality information is local to each module, the directives for setting the equality level affect only the particular user modules in which they are included. Thus, equality can be treated differently in different modules, which allows the knowledge engineer to compartmentalize the performance problem associated with equality and, if used judiciously, can lead to significant gains in performance.

Portions Copyright © 2013–2017 Coherent Knowledge

Run-time changes to the equality maintenance level. In $\mathcal{E}$RGO, the desired level of equality maintenance can also be changed at run time by executing a goal such as

```
?- setsemantics{equality=basic}.
```

Furthermore, $\mathcal{E}$RGO allows one user module to set, at run time, the level of equality maintenance in another user module:

```
?- setsemantics{equality=basic}@foobar.
```

This might be useful for *dynamic* modules, i.e., modules that are not associated with any files and whose content is generated completely dynamically. (See Section 25.)

Using the preprocessor to avoid the need for equality maintenance. One final bit of advice regarding equality: In many cases, knowledge engineers tend to use equality as an aliasing technique for long messages, numbers, etc. In this case, we recommend the use of preprocessor commands, which achieve the same result without loss of performance. For instance,

```
#define YAHOO  'http://yahoo.com'

?- YAHOO[fetch -> ?X].
```

Assuming that `fetch` is a method that applies to strings that represent Web sites and that it fetches the corresponding Web pages, the above will fetch the page at the Yahoo site, because the $\mathcal{E}$RGO compiler will replace YAHOO with the corresponding string that represents a URL.

Limitations of equality maintenance in $\mathcal{E}$RGO. The implementation of equality in $\mathcal{E}$RGO supports only a limited version of the *congruence axiom* due to the overhead associated with such an implementation. A congruence axiom states that if α `=` β then β can be substituted for any occurrence of α in any term. For instance, $f(x, \alpha) = f(x, \beta)$. In $\mathcal{E}$RGO, however, the query

```
a :=: b.
?- g(a) :=: g(b).
```

will fail. However, equal terms can be substituted for the arguments of frames and HiLog predicates. For instance, the queries

```
a:=:b.
a[f->c].
p(a,c).
?- b[f->c].
?- p(b,c).
```

will succeed.

21.2 Choosing a Semantics for Inheritance

As mentioned earlier, the `setsemantics` directive accepts three options: `inheritance=none`, `inheritance=flogic` and `inheritance=monotonic`. The default is `flogic`; this type of inheritance is described in Section 20.

With `inheritance=none`, behavioral inheritance is turned off in the corresponding module. This can significantly improve performance in cases when inheritance is not needed.

Note that `inheritance=none` does *not* turn off inheritance of signatures. Inheritance of signatures can be used for run-time type checking and it makes no good sense to disable it. Preserving inheritance of signatures does not affect the performance either.

Monotonic inheritance `inheritance=monotonic` is also sometimes appropriate—mostly in situations when information from superclasses is to be propagated and accumulated by subclasses and members without overriding. This type of inheritance is also significantly cheaper resource-wise than the F-logic inheritance, although it is more expensive than turning inheritance off completely.

21.3 Choosing a Semantics for the Subclass Relationship

The default semantics for the subclass relationship `::` in $\mathcal{E}$RGO is *strict*. This means that there can be no loops in the subclass hierarchy. If $\mathcal{E}$RGO detects a cycle at run time, it issues an error. The user can change this semantics by executing the runtime directive

```
?- setsemantics{subclassing=nonstrict}.
```

Note that the above sets the nonstrict semantics for subclassing in the current module only. To change back to the strict semantics, one can execute

```
?- setsemantics{subclassing=strict}.
```

and to change the semantics in a different module one can execute

```
?- setsemantics{subclassing=nonstrict}@foo.
```

One can gind out the subclassing semantics in effect in the current module by executing the following query:

```
?- semantics{subclassing=?Sem}.
```

Note: the `subclassing` option is not available as a static `setsemantics` directive (i.e., it works only with " ?– " and not with ":–").

21.4 Choosing a Semantics for Tabling

The semantics for tabling can be specified along two different dimensions: *reactivity* and *identification*. The options for reactivity are `tabling=reactive` and `tabling=passive`, which is explained in Section 25.3. The options for the identification dimension are `tabling=variant` (the default) and `tabling=subsumptive`. Subsumptive tabling currently works only in conjunction with passive tabling, and it is experimental at present. In some cases, subsumptive tabling may significantly speed up computation time and save memory.

21.5 Class Expressions

$\mathcal{E}$RGO defines a number of set-theoretic operations on classes. For instance, $(\mathtt{a}, \mathtt{b})$ is the intersection class, $(\mathtt{a}; \mathtt{b})$ is the union class, and $(\mathtt{a} - \mathtt{b})$ is the difference class. More precisely, $(\mathtt{a}, \mathtt{b})$ is the *maximal* subclass of `a` and `b` in the class hierarchy, and its extension is precisely the intersection of the extensions of `a` and `b`. The class $(\mathtt{a}; \mathtt{b})$ is the smallest superclass of `a` and `b`. The class $(\mathtt{a} - \mathtt{b})$ is the maximal subclass of `a` with extension that contains all the elements of `a` that are not in the extension of `b`.

The class expressions feature is *off* by default and must be enabled at runtime with

```
?- setsemantics{class_expressions=on}.
?- setsemantics{class_expressions=on}@somemodule.
```

It can also be disabled via

```
?- setsemantics{class_expressions=none}.
?- setsemantics{class_expressions=none}@somemodule.
```

Suppose the following information is given:

```
a, b, c  in class1
c        in class2
e        in class3
```

Then (`class1` − `class2`); `class3` has the extension of `a`, `b`, `e`.

We call the above combinations of types **class expressions**. Type expressions can occur in signature expressions as shown below:

```
cl[attr => ((c1 - c2) ; c3)].
cl[|attr => ((c1,c2) ; c3)|].
```

In the first case, the type is specified for `c1` as an individual object only. The second statement is about `c1` as a *class*, so this type specification is inherited by each member of the class `c1` and each subclass.

$\mathcal{E}$RGO also defines a number of subclass relationships among class expressions as follows.

1. If `c::c1` and `c::c2` then `c::(c1,c2)`, i.e., `(c1,c2)` is the greatest lower bound of `c1` and `c2` in the class hierarchy.
2. If `c1::c` and `c2::c` then `(c1;c2)::c`, i.e., `(c1;c2)` is the lowest upper bound of `c1` and `c2` in the class hierarchy.
3. Any class, `c`, is considered a superclass of `(c,?_)` and `(?_,c)`. In particular, `(c,c)::c`. At present, $\mathcal{E}$RGO does not enforce the equality `c:=:(c,c)`.
4. Any class, `c`, is considered a subclass of `(c;?_)` and `(?_;c)`. In particular, `c::(c;c)`. At present, $\mathcal{E}$RGO does not enforce the equality `c:=:(c;c)`.
5. Any class, `c`, is considered a superclass of `c-d` for any class `d`.

Unfortunately, these subclass relationships may adversely affect the performance of user knowledge bases, and $\mathcal{E}$RGO provides an optimization option that allows the user to disable these relationships for situations that do not need them, which is why this feature is off, by default.

Note: Type expressions introduce a potential for infinite answers for seemingly innocuous queries and so this feature is disabled by default, as explained earlier in this section. For instance, suppose that `a:c` is true. Then also `a:(c,c)`, `a:(c;c)`, `a:(c,(c,c))`, `a:(c;(c;c))`, etc. So, the query `?- a:?X.` will not terminate. To mitigate this problem, when class expressions are involved $\mathcal{E}$RGO guarantees to provide sound answers to queries about class membership and subclasses only when the arguments are ground; it does not guarantee that all class expressions will be returned to queries that involve open calls to "::" and ":". □

21.6 Ad Hoc Custom Semantics

The `setsemantics` directive allows the user to include additional axioms that define the semantics of a particular module. These axioms should be stored in a file and included into the module using the compiler or executable directive

```
:- setsemantics{custom=filename}.
```

However, the default is `custom=none`.[12] To take advantage of this feature, the user must write the axioms using the same API that is used for $\mathcal{E}$RGO trailers, which are located in the `closure` directory of the distribution. This API will be described at a later date.

21.7 Querying the Module Semantics

In addition to the ability to change the semantics of a module, $\mathcal{E}$RGO also lets the user *query* the semantics used by any given module through the `semantics` primitive. The syntax is similar to the `setsemantics` directive:

```
?- semantics{ Option1,  Option2, ...}.
?- semantics{ Option1,  Option2, ...}@modulename.
```

The options are the same as in the case of the `setsemantics` directive, but variables are allowed in place of the specific semantic choices, *e.g.*, `equality=?X`. The options unify with the current semantic settings in the module, so queries such as

```
?- semantics{equality=?X, custom=none}.
?- semantics{inheritance=flogic, equality=?X, custom=?Y}@foo.
```

are allowed. The order of the options in a `semantics`-query does not matter.

The `@module` part in the `semantics` primitive must be bound to a module name at the time the query is executed. However, it is still possible to determine which modules have a given combination of semantic options by examining every loaded module via the `isloaded{...}` primitive and then posing the desired `semantics{...}` query.

[12] Which implies that if the file has the name `none` then a full path name should be specified — just "none" implies no custom file.

22 ℰRGO and Tabling

22.1 Tabling in a Nutshell

Tabling is a fundamental deduction technique that augments query evaluation with a mechanism that remembers previously inferred conclusions. The result is a very efficient deductive engine.

ℰRGO automatically tables frames and HiLog predicates, but allows the user to have non-tabled predicates as well. Such predicates are called *transactional* and are mostly used for side-effects, such as writing to or reading from files (or to/from the screen) and to insert and delete facts and rules. To indicate that a HiLog literal is non-tabled, it must be preceded with the "%" sign.

For instance, in the following rules, `tc/2` is tabled but `%edge/2` is not tabled.

```
tc(X,Y) :- %edge(?X,?Y).
tc(X,Y) :- %edge(?X,?Y), tc(?Y,?Z).
```

A predicate with the `%` prefix is logically unrelated to the predicate without the `%` prefix. Thus, `p(a)(b)` being true does not imply anything about `%p(a)(b)`, and vice versa. However, ℰRGO issues a warning in case of such a double-use.

Symbols that are prefixed with the "%" sign can appear only as predicate formulas, predicate names, Boolean method names, or variables. A variable prefixed with "%" cannot be a stand-alone formula, unless it is associated with a module specification. The following occurrences of "%" are legal

```
?- insert{%p(a)}, %?(?X).  // %? is a variable ranging over non-tabled
                           // predicate names
?- a[%b(c)], a[%?Y].       // %b and %?Y are transactional Boolean methods
?- %?X@?M ~ %p(a).         // %p - a non-tabled predicate
```

but the following are not:

```
?- p(%a).                  // %a appears as a term, not a formula
?- ?X = %a.                 // %a appears as a term, not a formula
?- %?X = a.                 // %?X appears as a term, not a formula
?- a[%b(c)->d].            // %b is not a Boolean method
?- %?X ~ %p(a).             // %?X as a stand-alone formula
```

The first formula is illegal because `%a` occurs as a term and not as a predicate (it can be made legal by reifying the argument: `p(${%a})`). In the second and third formulas `%a` and `%?X` also appear as

unreified arguments. The fourth formula is illegal because `%b(c)` is not a Boolean method. The last one is illegal because `%?X` can not be a stand-alone formula (it can be made legal by associating a module with it).

Occurrences of variables that are prefixed with `%` are treated specially. First, it should be kept in mind that `%?X` and `?X` represent the same variable. If `?X` is *already bound* to something then both of them mean the same thing. However, `?X` itself can range not only over predicates but also terms, conjunctions/disjunctions of predicates, and even rules. In contrast, `%?X` with module specification can be bound only to non-tabled formulas and `?X` with module specification can be bound only to tabled formulas. Thus error messages will be issued for the following two queries:

```
?- ?X ~ p(a), %?X@?M ~ p(a).
?- ?X ~ a[%b], ?X@?M ~ a[%b].
```

The following query fails because `%?X` and `?X` represent the same variable: the first conjunct determines the binding for `?X`, and this binding does not match the expression on the right side of `~` in the second conjunct.

```
?- %?X@?M ~ %p(a), ?X ~ p(a).
```

In the query, `?X` is bound to the non-tabled formula `%p(a)`, and this does not meta-unify with the tabled formula `p(a)`.

When a bound variable occurs with an explicit module specification, then the following rules apply:

- If the idiom `?X@module` is used, `?X` can be bound only to a tabled predicate, a tabled molecular formula, or a HiLog *term* (not a predicate). Otherwise, an error is issued. If `?X` is already bound to a tabled predicate or molecular formula, then the explicit module specification (`@module`) is discarded. When `?X` is bound to a HiLog term, e.g., `p(a)(?Z)`, `?X@module` represents the tabled predicate `p(a)(?Z)@module`.
- If the idiom `%?X@module` is used, `?X` can be bound to only a non-tabled predicate, a non-tabled molecular formula, or a HiLog *term*. If `?X` is already bound to a non-tabled predicate or molecular formula, the explicit module specification is discarded, as before. If `?X` is bound to a HiLog term, then `%?X@module` represents the non-tabled predicate `p(a)(?Z)@module`.

Due to these rules, the first query below succeeds, while the second fails and the third causes an error.

```
?- ?X = p(a), %?X@?M ~ %p(a), ?X@?N ~ p(a)@foo.
?- ?X ~ p(a), ?X@?M ~ p(a)@foo.
?- ?X ~ p(a), %?X@?M ~ %p(a)@foo.
```

The first query succeeds because ?X is bound to the term p(a), which %?X@?M promotes to a non-tabled predicate with a yet-to-be-determined module. The meta-unification that follows then binds ?M to main. Similarly ?X@?N promotes the term p(a) to a tabled predicate with a yet-to-be-determined module, and meta-unification binds ?N to foo. The second query fails because ?X is already bound to a tabled predicate and therefore ?X@?M represents p(a)@main, which does not meta-unify with p(a)@foo. The third query gives an error because ?X is bound to a tabled predicate, while %?X@?M expects a non-tabled predicate or a HiLog term.

When ?X and %?X occur with explicit module specifications and are *unbound* then the occurrences of %?X indicate that ?X is expected to be bound to predicate names, Boolean method names, or predicate/molecular formulas that correspond only to non-tabled methods or predicates. Likewise, an occurrence of an unbound ?X indicates that ?X is expected to be bound to predicate names or predicate/molecular formulas that correspond to tabled methods or predicates.

Transactional (%-prefixed) literals and meta-programming. In meta-unifications, update operations and the clause construct, variables that are prefixed with a "%" to indicate non-tabled occurrences must have explicit module specifications. An unprefixed variable without a module specification, such as ?X, can meta-unify with both tabled and non-tabled predicates. However, when an explicit module specification is given, such as in ?X@main, unprefixed variables can be bound only to tabled predicates. For example, all of the following queries succeed without errors.

```
?- ?X ~ %p(a).
?- ?X ~ p(a).
?- ?X ~ a[b->c]@foo.
?- ?X ~ a[%b]@?M.
?- ?X@?M ~ p(a).
?- %?X@foo ~ a[%b]@?M.
```

In the context of update operations, ERGO uses the same rules for variables of the form %?X and ?X. Therefore, the following operations will succeed:

```
?- insert{p(a),%q(b)}.     // Yes
?- delete{?X@\@}.          // Yes, with ?X bound ${p(a)}
?- delete{%?X@\@}.         // Yes, with ?X bound ${%q(b)}
?- insert{p(a),%q(b)}.     // Yes
?- delete{?X}.             // Yes, ?X bound to ${p(a)} or ${%q(b)}
```

These rules also apply to queries issued against rule bases using the clause primitive (see Section 27 for the discussion of this primitive) or to deletion of rules with the deleterule primitive.

```
?- insertrule{p(?X) :- q(?X)}.
?- insertrule{%t(?X) :- %r(?X)}.
?- insertrule{pp(?X) :- q(?X), %r(?X)}.
?- clause{?X,?Y}.          // all three inserted rules above would be retrieved
?- clause{%?X@\@,?Y}.      // ?X = %t(?_var) and ?Y = %r(?_var)
?- clause{?X@\@,?Y@\@}.    // ?X = p(?_var) and ?Y = q(?_var)
?- clause{?X@\@,?Y}.       // the first and the third rules would be retrieved
```

22.2 Transactional Methods

Because tabling is not integrated with the update mechanism in Prolog, it can have unintended effects when used in conjunction with predicates that have non-logical "side effects" (e.g., writing or reading a file) and predicates that change the state of the database. If a tabled predicate has a side effect, the first time the predicate is called the side effect is performed, but the second time the call simply returns with success or failure (depending on the outcome of the first call), because Prolog will simply look it up in a table. Thus, if the predicate is intended to perform the side effect each time it is called, it will not operate correctly.

Object-oriented programs often rely on methods that produce side effects or make updates. In ℰRGO we call such methods *transactional*. Because by default ℰRGO tables everything that looks like a frame, these transactional methods are potentially subject to the aforesaid problem.

To sidestep this issue, ℰRGO introduces a new syntax to identify transactional methods — by allowing the "`%`" sign in front of a transactional method. For instance, the following rule defines an output method that, for every object, writes out its oid:

```
?O[%output] :- write(?O)@\prolog.
```

Like boolean methods, transactional methods can take arguments, but do not return any values. The only difference is that transactional methods are *not* tabled, while boolean methods are.

Transactional signatures. Transactional methods can have signatures like other kinds of methods, which are specified as follows:

```
Obj[=>%Meth]
Class[|=>%Meth|]
```

ℰRGO does not support transactional methods specified as defaults at the class level. However, as seen from the second statement above, class-level signatures for transactional methods are supported.

22.3 Re-examining Query Results: The peek{...} primitive

ℰRGO also provides a way to browse the results of the previous tabled (non-transactional) queries, called *peeking*, and is provided via a special primitive

`peek{`*frame-or-predicate-goal*`}`

The result is that the results of the query are returned without re-executing the query.

At first glance, it may seem that this is no different from tabling, which also returns the results without re-executing, but there is an important and subtle difference: `peek{Goal}` returns the answers to all *previously executed* queries that unify with *Goal*. There could be several such previous queries or none. If none, then the above primitive returns no answers; if several, the primitive returns the answers to all those queries. In contrast, if the query were just `Goal` then, if this exact query was asked before then its answers will be returned without re-evaluation. The answers to other queries that unify with `Goal` will *not* be returned. If `Goal` was not asked before then it will be evaluated and the results returned.

To illustrate, consider the following simple example:

```
p(1,a), p(1,b).
p(2,c), p(2,d).
p(3,e). p(3,f).
?- p(1,?), p(2,?).
```

If we now ask the query `peek{p(?X,?Y)}` then we get the following four answers:

```
?X = 1
?Y = a

?X = 1
?Y = b

?X = 2
?Y = c

?X = 2
?Y = d
```

Note that `?X=3` and `?Y=e`, `?Y=f` are not returned because the query `p(3,?Y)` was never asked. In contrast, the query `p(?X,?Y)` will return all six answers (and it will be evaluated, as it is not a `peek`-query) and after asking `p(?X,?Y)` the subsequent query `peek{p(?X,?Y)}` will also return all six answers (because it will pick up the answers produced by `p(?X,?Y)`).

Another important use of the `peek` primitive is that it enables one to look into the partially computed query results during pauses in the computation. For instance, start ℰRGO and run `demo{owl_benchmark}` at the prompt. After a few seconds, type `Ctrl-C` to pause the execution and at the prompt ask the query `?X::?Y`. Since non-transactional queries are currently not allowed during these pauses (see Section 43.9), one would see an error message like this:

```
++Abort[Ergo]> you are trying to query or modify the knowledge base while
a previous command is paused; only informational queries regarding
the system runtime state are permitted at this point
```

However, `peek{?X::?Y}` *is* allowed, and it will give the answers to the query that have been computed so far.

22.4 Operational Semantics of ℰRGO

Although ℰRGO is a declarative language, it provides primitives, such as input/output, certain types of updates, cuts, etc., which have no logical meaning. In such cases, it is important to have an idea of the *operational semantics* of ℰRGO. This operational semantics is essentially the same as in XSB and when no tabled predicates or frames are involved, the behavior is the same as in Prolog. However, when tabled HiLog predicates or frames (other than transactional methods) are used, the knowledge engineer must have some understanding of the way XSB evaluates tabled predicates.

Unlike Prolog, which computes answers one-at-a-time, ℰRGO computes answers to the entire clique of inter-dependent predicates are computed before the computation proceeds to the next literal in a rule body. The following little example illustrates the difference:

```
a:b.
d:b.
c:b.

?X[foo(?Y)] :- ?X:?Y, writeln(?X)@\prolog.
%q(?X,?Y) :- ?X:?Y, writeln(?X)@\prolog.

?- ?X[foo(?Y)], writeln(done)@\prolog.
?- %q(?X,?Y), writeln(done)@\prolog.
```

The two queries are essentially the same. The first is a frame and so it is implemented internally as a tabled XSB predicate. The second query is implemented as a non-tabled predicate. Thus, despite the fact that the two queries are *logically equivalent*, they are not *operationally equivalent*. Indeed, a simple experiment shows that the answers to the above two queries are produced in different orders (as seen by the order of execution of the print statements). In the first query, `?X[foo(?Y)]` is evaluated completely before proceeding to `writeln(done)@\prolog` and thus the executions of `writeln(?X)@\prolog` are grouped together. In the second case, executions of `writeln(?X)@\prolog` and `writeln(done)@\prolog` alternate, because `q/2` is not tabled and thus its evaluation follows the usual Prolog semantics.

On the other hand, if we have

```
?X[foo(?Y)] :- ?X:?Y, writeln(?X)@\prolog.
q(?X,?Y) :- ?X:?Y, writeln(?X)@\prolog.

?- ?X[foo(?Y)], writeln(done)@\prolog.
?- q(?X,?Y), writeln(done)@\prolog.
```

then the two queries will behave the same, as both `q/2` and `?X[foo(?Y)]` would then be implemented internally as tabled predicates. Likewise, if we replace `foo` with `%foo` then the corresponding frame would be represented internally as a non-tabled predicate. Thus, the two queries, below,

```
?X[%foo(?Y)] :- ?X:?Y, writeln(?X)@\prolog.
%q(?X,?Y) :- ?X:?Y, writeln(?X)@\prolog.

?- %?X[foo(?Y)], writeln(done)@\prolog.
?- %q(?X,?Y), writeln(done)@\prolog.
```

will produce the same result where `a, b, c` and `done` alternate in the output.

22.5 Tabling and Performance

It is important to keep in mind that Prolog does not reorder frames and predicates during joins. Instead, all joins are performed left-to-right. Thus, rules and queries must be written in such a way as to ensure that smaller predicates and classes appear early on in the join. Also, even though XSB tables the results obtained from previous queries, the current tabling engine has several limitations. In particular, when a new query comes in, XSB tries to determine if this query is "similar" to one that already has been answered (or is in the process of being evaluated). Unfortunately, the default notion of similarity used by XSB is fairly weak, and many unnecessary recomputations might result. XSB has partial support for a novel technique called *subsumptive tabling*, and it is known

that subsumptive tabling can speed up certain queries by an order of magnitude. However, XSB's implementation of subsumptive tabling does not support active tabling, which prevents most of the uses of this kind of tabling in ℰRGO.

22.6 Cuts

No discussion of a logic programming language is complete without a few words about the infamous Prolog cut (!). Although Prolog cut has been (mostly rightfully) excommunicated as far as Database Query Languages are concerned, it is sometimes indispensable when doing "real work", like pretty-printing ℰRGO knowledge bases or implementing a pattern matching algorithm. To facilitate this kind of tasks, ℰRGO lets the user use Prolog-like cuts.

Cuts across tables. The current implementation of XSB has a limitation that Prolog cuts cannot "cut tabled predicates." If you get an error message saying something about cutting across tables — you know that you have cut too much!

The basic rule that can keep you out of trouble is: do not put a cut in the body of a rule *after* any frame or tabled HiLog predicate. However, it is OK to put a cut before any frame. It is even OK to have a cut in the body of a rule that *defines* a frame (again, provided that the body has no frame to the left of that cut). If you need to use cuts, plan on using transactional methods or non-tabled predicates.

The Prolog cut operator (!) in rule bodies may cause problems because XSB does not allow the cut to appear in the middle of a computation of a tabled predicate. For instance,

```
?X[%foo(?Y)] :- ?Z[moo->?W], ?W:?X, !, rest.
```

will not cause problems, but

```
?X[foo->?Y] :- ?Z[moo->?W], ?W:?X, !, rest.
```

will likely result in a runtime error. The reason is that in the first case the frame `?X[%foo(?Y)]` is implemented using a non-tabled predicate, so by the time the evaluation reaches the cut, both `?Z[moo->?W]` and `?W:?X` will be evaluated completely and their tables will be marked as "complete." In contrast, in the second example, `?X[foo->?Y]` is implemented as a tabled predicate, which is interdependent with the predicates that are used to implement `?Z[moo->?W]` and `?W:?X`. Thus, the cut would occur in the middle of the computation of the tabled predicate `?X[foo->?Y]` and an error will result.

Portions Copyright © 2013–2017 Coherent Knowledge

In a future release, XSB might implement a different tabling schema. While cutting across tables will still be prohibited, it will provide an alternative mechanism to achieve many of the goals a cut is currently used for.

Cuts and facts. Prolog programmers are accustomed to treat facts as rules, and so Prolog programs with cuts heavily rely on the order of the rules. For instance, in

```
p(X,Y) :- q(X), !, r(X,Y).
p(3,4).
q(1).
r(1,2).
```

it is expected that the open query `p(X,Y)` will succeed from the first rule and will return `p(1,2)`. The query will not match the fact `p(3,4)` due to the cut because that fact comes after the first rule.

This kind of reasoning will not work in $\mathcal{E}$RGO because facts are stored in an internal database in an unordered fashion and one cannot know when they are going to be matched against the query. So, in $\mathcal{E}$RGO an open call `p(?X,?Y)` would return goth `p(1,2)` and `p(3,4)`. A user who wants to rely on rule ordering should convert the relevant facts into rules, as follows:

```
p(?X,?Y) :- q(?X), !, r(?X,?Y).
p(3,4) :- \true.
q(1).
r(1,2).
```

In that case, `p(?X,?Y)` will return only one answer.

23 User Defined Functions

User-defined functions (abbr., *UDF*) are a syntactic extension that permits the users to enjoy certain aspects of functional programming in the framework of a logic-based language.

23.1 Syntax

To define a UDF in $\mathcal{E}$RGO, one uses the following syntax:

```
\udf foo(t1,...,tn) := Expr \if Body.
```

`Expr` is a term and `Body` can be any formula that can be used as a rule body. The "`\if Body`" part is optional. The arguments of the UDF `foo` are terms, which usually are distinct variables.

`Expr` and *Body* can contain occurrences of other UDFs, but those UDFs must be defined previously. At first, this suggests that mutually-recursive UDFs cannot be defined. However, this is not the case, as will be explained shortly.

Instead of "if", $\mathcal{E}$RGO also allows the use of the rule connective `:-`:

```
\udf foo(t1,...,tn) := Expr :- Body.
```

Example. The following simple example defines `father/1` as a function.

```
\udf father(?x):=?y \if father(?x,?y).
father(mary,tom).
father(john,sam).
```

So, instead of writing `father(John,?y)` and then using `?y` one can simply write `father(?John)`:

```
?- ?y=father(?x).
```

will return

```
?x=mary
?y=tom

?x=john
?y=sam
```

The query

```
?- writeln(father(mary))@\io.
```

will output

```
tom
```

Portions Copyright © 2013–2017 Coherent Knowledge

Mutual recursion in UDF definitions. It was mentioned that any UDF that occur in a definition of another UDF must be defined previously. However, this does not mean that mutually recursive UDFs cannot be defined: one should just exercise care.

First, we should make it clear that recursive UDFs by themselves do not pose problems and can be used freely. For example, in

```
\udf foo(...) := Expr \if Body.
```

A more complex situation arises when we have mutual recursion, as in

```
\udf foo(...) := Expr-cannot-contain-bar \if Body-contains-bar.
\udf bar(...) := Expr-can-contain-foo \if Body-might-contain-foo.
```

This is not allowed, because `bar` is used in `Body-contains-bar` before it was defined. But there is a simple workaround: we can introduce a temporary predicate as follows:

```
\udf foo(...) := Expr-cannot-contain-bar \if tempfoo(...).
\udf bar(...) := Expr-can-contain-foo \if Body-can-contain-foo.
tempfoo(...) :- Body-contains-bar.
```

In this latter case, all uses of UDFs are defined previously, and UDFs are still mutually recursive. If `foo` has more than one `\udf`-clause and one of the clauses does not involve `bar` then one does not even need to introduce temporary predicates. For instance, the following is legal:

```
\udf foo(...) := Expr-contains-no-bar \if Body-contains-no-bar.
\udf bar(...) := Expr-can-contain-foo \if Body-can-contain-foo.
\udf foo(...) := Expr-can-contain-bar \if Body-can-contain-bar.
```

Here the first clause declares `foo` as a UDF, so the compiler is able to handle `foo` correctly in the second clause despite the fact that `foo` has not been fully defined yet at that point.

UDFs and modules. A UDF definition is a purely compile-time directive. It has an effect from the moment the function is defined till the end of the file (which extends to all the files included using the `#include` preprocessor directive). This implies that UDFs that occur inside calls to other $\mathcal{E}$RGO modules are subject to the directives of the module where the functions occur, not in the module that is being called. For instance, consider

```
\udf func1(?x) := ?y \if pred1(?x,?y).
p(?x) :- q(func1(?x))@bar.
pred1(c,a).
```

Here `func1` will be compiled according to the function definition shown above. If the module `bar` has another definition for `func1`, that definition has no effect on the above occurrence of `func1(?x)`. Thus, the query

```
?- p(c).
```

results in a call to `q(func1(c))@bar`, and since `func1(c)` is defined to be `a` this is tantamount to executing the query

```
?- q(a).
```

Note that, since function definitions can contain anything that constitutes a rule body, calls to other modules are allowed. For instance,

```
\udf foo(?x):=?y \if pred(?x,?y)@bar, pred(?x,?y,?z)@moo, q(?z).
```

UDFs and updates. At present, only non-transactional UDFs are supported. If a UDF definition contains an update then a runtime error will result. For instance, if a UDF `foo(...)` has such a definition then the following error will likely occur at runtime:

```
++Error[ℰRGO]> non-transactional head-literal foo(...) := ?A depends on an update
```

Stealth updates (Section 25.6) *are* allowed in UDF definitions (in ℰRGO only). Example:

```
%fooPRED(?val, =?Res+1) :- insert{stealth{%p(?val)}}, %p(?Res).
\udf fooUDF(?form) := ?res \if %fooPRED(?form, ?res).
```

```
?- ?A = fooUDF(1).
```

```
?A = 2
```

```
?- ?A = fooUDF(2).
```

```
?A = 2
?A = 3
```

Advanced examples. One of the most interesting examples of the use of UDFs arises when we define arithmetic operations as functions. For instance, normally one would write

```
?- ?x \is 1+2, writeln(?x)@\io.
```

but with UDFs we can define "+" as a function:

```
\udf ?x+?y := ?z }\if ?z \is ?x+?y.
```

and then issue the following query:

```
?- writeln(1+2)@\io.
```

The following example shows how we can define the Fibonacci function:

```
\udf ?x+?y := ?z \if ?z \is ?x+?y.
\udf ?x-?y := ?z \if ?z \is ?x-?y.
\udf fib(0) := 0.
\udf fib(1) := 1.
\udf fib(?x) := fib(?x-1)+fib(?x-2) \if ?x>1.
```

We can now ask the query

```
?- writeln(fib(34))@\io.
```

Note that the above example also illustrates that a definition of a UDF can consist of multiple `\udf`-statements—just like a definition of a predicate can have multiple rules.

Querying UDFs. $\mathcal{E}$RGO maintains meta-information about UDFs and they can be queried at runtime using the primitives `clause{...}` and `@!{...}`. For instance,

```
\udf foo(?X,?Y) := p(?Z)  bar(?X,?Y,?Z).

ergo>  clause{@!{?X[type->?T]} ?H,?B}.

?X = 4      // 4 happens to be the Id of that UDF
?T = udf
?H = foo(?_h1642,?_h1643) := _h1637
?B = (${bar(?_h1642,?_h1643,?_h1660)@main}, ?_h1637 = p(?_h1660))

ergo>  clause{@!{?X[type->?T]} (?UDF(?,?) := ?s) ,?B}.

?X = 4
?T = udf
```

```
?UDF = foo
?B = ($bar(?_h5733,?_h5734,?_h5735)@main, ?_h5742 = p(?_h5735))

ergo>  @!{?X[type->?T]}.

?X = 4
?T = udf
```

Note from the second query above that the head of a UDF can be queried using the (*udf* `:=` *result*) idiom. The parentheses around this idiom are important in that context. This is because the expression `?UDF(?,?):= ?,?B` inside clause{...} above is parsed as `?UDF(?,?):= (?,?B)` and not as `(?UDF(?,?):= ?), ?B`.

23.2 Higher-order UDFs

UDFs can use arbitrary HiLog terms. These terms can be used both in the head of the UDF definition and in its body. For instance,

```
\udf abc(?x):=?x(a,b).
\udf ?x(?y,b)(?y):=t(?z) \if pred(?x,?y,?z).
\udf t(c):=d.
pred(foo,a,c).
q(d).
?- q(abc(?x)(?y)).
```

Perhaps it is not obvious, but the query succeeds with the answers `?x = foo`, `?y = a`. Indeed, `abc(?x)` rewrites to `?x(a,b)` by the first UDF definition, so the query becomes `q(?x(a,b)(?y))`. By the second (higher-order!) UDF, this rewrites to `q(t(c))` with the bindings `?x=foo` and `?y=a`. Finally, the query rewrites to `q(d)`, by the third UDF, and succeeds.

23.3 User-defined Functions and Multifile Modules

Recall from Section 15.5 that sets of rules can be added to an already loaded module using the `add`{*file*} or `[+`*file*`]` commands. If the already loaded module has UDF definitions, say, for `foo/2` and `bar/3`, and if the file `file.ergo` uses those functions, then it is necessary to tell $\mathcal{E}$RGO *in the file* `file.ergo` that these are UDFs and not regular HiLog function symbols. If this is not done, the things are not unlikely to work correctly and it would be very hard to find out why. To prevent

this type of hard-to-find mistakes, the system will issue an error at loading time.[13]

To deal with such situations, $\mathcal{E}$RGO provides the `useudf` directive. For instance, if one puts the following at the top of `file.ergo`

```
:- useudf{foo/2, bar/3}.
```

then `foo/2` and `bar/3` will be compiled correctly in `file.ergo` and no errors will be issued.

Note: UDF definitions and the `useudf` directives are intended to reside in *different* files (for the same UDF): it makes *no* good sense to have both a function definition and a `useudf` directive for the *same* UDF in the *same* file. However, $\mathcal{E}$RGO is tolerant to this kind of misspecification. □

To simplify the use of UDFs, $\mathcal{E}$RGO supports *implicit* `useudf` directives in certain cases. Namely, if a file with a UDF definition is loaded into a module then any file that will be *compiled* and then *added* to that same module in that $\mathcal{E}$RGO session will inherit the existing function definitions. In this case, there is no need to specify the `useudf` directive explicitly in `foo.ergo`. However, if the file being added, say `foo.ergo`, is compiled in a different $\mathcal{E}$RGO session (one that does not have the appropriate UDF definitions) then an explicit `useudf` directive would be required. Otherwise, a runtime error will be issued the next time `foo.ergo` is added if there is a mismatch between the UDF definitions in these two different $\mathcal{E}$RGO sessions.

23.4 Semantics

Semantically, the functional notation is just a syntactic sugar. A function definition of the form

```
\udf foo(t1,...,tn) := Expr if Body
```

is converted into the rule

```
newpred_foo(t1,...,tn,Expr) :- Body.
```

Then every occurrence of the function in a rule head

```
head(...,foo(s1,...,sn),...) :- rulebody.
```

is rewritten into

[13] It is impossible to catch this mistake at compile time, since one cannot know in advance that `file.ergo` is intended to be loaded into a module with existing UDF definitions. The `useudf` directive is used to supply precisely this kind of information to the compiler.

```
head(...,?newvar,...) :- newpred_foo(s1,...,sn,?newvar), rulebody.
```

In the rule body, the occurrences of `foo(s1,...,sn)` are rewritten as follows:

```
... :- ..., pred(...,foo(s1,...,sn),...), ...
```

is changed to

```
... :- ..., newpred_foo(s1,...,sn,?newvar), pred(...,?newvar,...), ...
```

It is important to keep in mind that UDFs are *partial* functions and that any argument bindings for which an UDF is undefined falsifies the containing predicate or frame. For instance,

```
p(?).
\udf f(?x) := ?y \if cond(?x,?y).
cond(1,2).
?- p(f(1)).
Yes
?- p(f(2)).
No
```

Here `f(?x)` is a partial function that is defined only for `?x=1`. Since `f(1)=2` and `p(...)` is true of everything, `p(f(1))` evaluates to true. However, `p(f(2))` evaluates to false because `f(2)` is undefined and this falsifies `p(f(2))`. The rationale is that `p(...)` is true of every *defined* argument, but is false when an argument cannot be evaluated to a bona fide term.

23.5 Representing equality with user defined functions

An important application of UDFs is simulation of certain cases of equality. Simulating equality using UDFs is *orders of magnitude more efficient*. The equality that can be simulated in this way must be such that all terms that one wants to equate must be explicitly given at compile time. For instance, a term or its main function symbol cannot be represented by a variable (e.g., `?X` or `?X(?y,?z)`) and determined only at run time. Also, if the system of UDF definitions is not *confluent* (confluence is also known as the Church-Rosser property) then the fundamental property of substitutivity of equals by equals cannot be guaranteed.

For instance, equality of the following kind $?\mathtt{x} = \mathtt{abc}, ?\mathtt{x}(\mathtt{pqr}) :=: \mathtt{cde}$ cannot be simulated because the term `abc(pqr)` that is equated to `cde` here is not explicitly given at compile time.

It is important to also keep in mind that UDFs provide only **unidirectional** rewriting. For instance, in

```
\udf a := b.
\udf a := c.
q(c).
?- q(b).
```

the query fails, since `b` is **not** rewritten back into `a` (and then into `c`).

To illustrate the cases where simulation is possible, suppose we want to state that any occurrence of `foo(?x)` should be equated to (rewritten into) `bar(?x)` if `pred(?x)` is true and `?x>2`:

```
\udf foo(?x):=bar(?x) if  pred(?x), ?x>2.
pred(1).
pred(2).
pred(3).
pred(5).
```

If we ask query

```
?- foo(?x)=bar(?x)
```

The answers will be

```
?x=2
?x=3
```

Here are some very simple, but useful examples of equality: rather than writing things like

```
foo(1) :=: 3.
foo(4) :=: 7.
```

we can instead define

```
\udf foo(1):=3.
\udf foo(4):=7.
```

The latter does not involve complex equality maintenance and is much more efficient.

The following example illustrates the issue of non-confluence of the equality through UDFs. Consider the following definitions:

```
\udf a := b.
\udf c := d.
\udf a := c.
```

and suppose `p(d)` is true. Then we can derive that `p(a)` and `p(c)` is true, but not that `p(b)` is true.

Note that if instead of `p(d)` we had `p(a)` then all these would be true: `p(b)`, `p(c)`, `p(d)`. This shows that one can go a long way with simulation of equality through UDFs, but care must be exercised. For example, if the set of UDF definitions is not confluent then substitution of equals by equals is not always guaranteed.

24 Controlling Context of Symbols

Using the same symbol in different contexts in the same knowledge base can be useful, but often can be the result of a typo, and such errors is very hard to catch. Examples of different contexts include the use of the same symbol as a function symbol and as a predicate, as a regular function symbol and a UDF, and even the uses of the same symbol with different arities.

$\mathcal{E}$RGO checks if the same symbol appears in different contexts and issues warnings if it finds such uses. For instance, in

```
p(?x,?y,?z):-q(?x,?y,?z).
p(?x):-t(?x).
```

the predicate `p` is used with arity 3 and 1, which may or may not be a mistake. In the following,

```
p(?x):-q(?x).
f(p(?x,?y)):-t(?x).
```

`p` occurs both as a predicate and a function symbol, and with different arities on top. In the next example we encounter the symbol `f` in the role of a UDF and also as a predicate:

```
\udf f(?x):=?y if ff(?x,?y).
t(?x):-f(?x).
```

and, in the next example, `tp` has both transactional and non-transactional occurrences.

```
%tp(?x,?y):-s(?x,?y).
tp(?x,?y,?z):-q(?x,?y,?z).
```

Portions Copyright © 2013–2017 Coherent Knowledge

UDFs are also not allowed to have the same name as $\mathcal{E}$RGO builtins, such as `isatom...`, `isground...` and the like.

In all these cases, warnings will be issued. However, the user can turn off the warnings that she believes not to indicate an error. This is done with the help of the directive `symbol_context`. The syntax is the following

`:- symbol_context{`*comma-separated-list-of-terms*`}`

The terms in the aforementioned list can be function symbols, predicates, or UDFs. For each type there is a separate idiom to indicate what is to be permitted. The legal specifications are as follows:

1. `termName(?,...,?)@?`, `termName(?,...,?)@\@`, `termName(?,...,?)@moduleName` — indicates that `termName` is a predicate of the given arity. The first form applies to all modules, the second applies to predicates that appear without any module specification (e.g., `abc(1,2)`), and the third applies to predicates in a specific module, `moduleName`.
2. `termName(?,...,?)` — indicates that `termName` is a term (not a predicate).
3. `udf termName(?,...,?)` — indicates that `termName` is a UDF.
4. `termName(*)@moduleName`, `termName(*)`, and `udf termName(*)` — these are wildcards that have the same meaning as above except that the warnings for the terms of the form `termName(...)` are suppressed for all arities in the corresponding contexts (predicate, HiLog term, UDF).
5. `termName/N`, where `N` is a non-negative integer — suppresses warnings for terms of the form `termName(`$\mathtt{arg}_1$`,...,`$\mathtt{arg}_N$`)` in *any* context (predicate, HiLog term, or a UDF).
6. `termName/?` — like `termName/N` above but suppresses warnings for all arities. Thus, warnings will not be issued for `termName(...)` in all contexts and for all arities.
7. `warnoff` — suppresses symbol context warnings, but not errors.
8. `constoff` — suppresses warnings due to constants appearing in multiple contexts. This not only affects warnings, but also performance. In very large fact bases, the number of constants is typically very large. For each constant, $\mathcal{E}$RGO would normally create an entry in the `.fls` file, which is used to check for multiple contexts. For a large file, such an `.fls` file can be much larger than the original set of facts. With the `constoff` option, entries describing constants will not be saved in the `.fls` file, which can significantly improve the compilation and load time.

In the above, the arguments must be anonymous variables and `moduleName` can be either a module name or an anonymous variable. Examples:

```
:- symbol_context{p(?,?), q(?)@?, r(?,?)@bar, udf foo(?)}.
```

Once this directive is issued, the compiler will no longer warn about the occurrences of `p` as a term with two arguments, `q` as a predicate with one argument in *any* module, `r` as a binary predicate in module `bar`, and `foo` as a unary UDF.

The `symbol_context` directive also has special syntax for sensors and predicates declared using the `:- prolog{...}` directive:

```
:- symbol_context{sensor foo(?,?), prlg bar(?,?,?)}.
```

Note that $\mathcal{E}$RGO issues warnings only on finding a *second* inconsistent use of the same symbol. Since most symbols are used in just one context, there is no need to supply the above directive in most cases. For instance, in the following

```
:- symbol_context{p(?,?)}.
p(?x):-q(?x), t(?x).
f(p(?x,?y)):-t(?x).
```

one needs to notify the compiler only about the use of `p(?,?)` as a function symbol, since there is prior use of `p/1` as a predicate. The other symbols are used consistently and there is no need to do anything special for them.

25 Updating the Knowledge Base

$\mathcal{E}$RGO provides primitives to update the runtime database. Unlike Prolog, $\mathcal{E}$RGO does not require the user to define a predicate as dynamic in order to update it. Instead, every predicate and object has a *base part* and a *derived part*. Updates directly change only the base parts and only indirectly the derived parts.

Note that the base part of a predicate or of an object contains *both* the facts that were *inserted explicitly* into the knowledge base and the facts that are specified implicitly, via rules. For instance, in

```
p(a).
a[m->b].
```

the fact `p(a)` will be placed in the base part of the predicate `p/1` and it can be deleted by the `delete` primitive. Likewise, the fact `a[m->b]` is updatable. If you do not want some facts to be updatable, use the following syntax:

```
p(a) :- \true.
a[m->b] :- \true.
```

$\mathcal{E}$RGO updates can be *non-transactional,* as in Prolog, or *transactional*, as in Transaction Logic [2, 1]. We first describe non-transactional updates.

25.1 Non-transactional (Non-logical) Updates

The effects of non-transactional updates persist even if a subsequent failure causes the system to backtrack.

$\mathcal{E}$RGO supports the following non-transactional update primitives: `insert`, `insertall`, `delete`, `deleteall`, `erase`, `eraseall`. These primitives use special syntax (the curly braces) and are *not* predicates. Thus, it is allowed to have a user-defined predicate such as `insert`.

Insertion. The syntax of an insertion is as follows (note the {,}s!):

insop`{`*literals* `[|` *query*`]}`

where *insop* stands for either `insert` or `insertall`. The *literals* part represents a comma separated list of literals, which can include predicates and frames. The optional part, `|`*query*, is an additional condition that must be satisfied in order for *literals* to be inserted. The semantics is that *query* is posed first and, if it is satisfied, *literals* is inserted (note that the query may affect the variable binding and thus the particular instance of *literals* that will be inserted). For instance, in

```
?- insert{p(a),Mary[spouse->Smith,children->Frank]}
?- insert{?P[spouse->?S] | ?S[spouse->?P]}
```

the first statement inserts a particular frame. In the second case, the query `?S[spouse->?P]` is posed and one answer (a binding for `?P` and `?S`) is obtained. If there is no such binding, nothing is inserted and the statement fails. Otherwise, the instance of `?P[spouse->?S]` is inserted for that binding and the statement succeeds.

The insert statement has two forms: `insert` and `insertall`. The difference between `insert` and `insertall` is that `insert` inserts only one instance of *literals* that satisfies the formula, while `insertall` inserts *all* instances of the literals that satisfy the formula. In other words, *query* is posed first and *all* answers are obtained. Each answer is a tuple of bindings for some (or all) of the variables that occur in *literals*. To illustrate the difference between `insert` and `insertall`, consider the following queries:

```
?- p(?X,?Y), insert{q(?X,?Y,?Z)|r(?Y,?Z)}.
?- p(?X,?Y), insertall{q(?X,?Y,?Z)|r(?Y,?Z)}.
```

In the first case, if `p(x,y)` and `r(y,z)` are true, then the fact `q(x,y,z)` is inserted. In the second case, if `p(x,y)` is true, then the update means the following:

> For each `z` such that `r(y,z)` holds, `insert q(x,y,z)`.

The primitive `insertall` is also known as a bulk-insert operator.

Unlike `insert`, the operator `insertall` always succeeds and it always leaves its free variables unbound.

The difference between `insert` and `insertall` is more subtle than it may appear from the above discussion. In the all-answers mode, the above two queries will actually behave the same, because $\mathcal{E}$RGO will try to find all answers to the query `p(?X,?Y), r(?Y,?Z)` and will do the insertion for each answer. The difference becomes apparent if $\mathcal{E}$RGO is in one answer at a time mode (because `\one` was executed in a preceding query) or when the all-answers mode is suppressed by a cut as in

```
?- p(?X,?Y), insert{q(?X,?Y,?Z)|r(?Y,?Z)}, !.
?- p(?X,?Y), insertall{q(?X,?Y,?Z)|r(?Y,?Z)}, !.
```

In such cases, the first query will indeed insert only one fact, while the second will insert all.

Note that literals appearing inside an insert primitive (to the left of the `|` symbol, if it is present) are treated as facts and should follow the syntactic rules for facts and literals in the rule head. In particular, path expressions are not allowed. Similarly, module specifications inside update operators are illegal. However, it is allowed to insert facts into a different module so module specifications are permitted in the literals that appear in the `insert{...}` primitive:

```
?- insert{(Mary[children->Frank], John[father->Smith]) @ foomod}
```

The above statement will insert `Mary[children->Frank]` and `John[father->Smith]` into module `foomod`.

Note that module specifications are also allowed in the *condition* part of an update operator (to the right of the `|` mark):

```
?- insert{Mary[children->?X]@foobar | adult(?X)@infomod}
```

Updates to Prolog modules is accomplished using the usual Prolog `assert/retract`:

```
?- assert(foo(a,b,c))@\prolog.
```

The following subtleties related to updates of Prolog modules are worth noting. Recall Section 16.4 on the issues concerning the difference between the HiLog representation of terms in $\mathcal{E}$RGO and the one used in Prolog. The problem is that `foo(a,b,c)` is a HiLog term that Prolog does not understand and will not associate it with the predicate `foo/3` that it might have. To avoid this problem, use explicit conversion:

```
?- p2h{?PrologRepr,foo(a,b,c)}, assert(?PrologRepr)@\prolog.
```

This will insert `foo(a,b,c)` into the default XSB module called `usermod`.

If all this looks too complicated, $\mathcal{E}$RGO provides a higher-level primitive, `@\prologall` (equivalently `@\plgall`), as described in Section 15.7. This module specifier does automatic conversion of terms to and from Prolog representation, so the above example can be written much more simply:

```
?- assert(foo(a,b,c))@\prologall.
```

Another possible complication might be that if `foo/3` is defined in another Prolog module, `bar`, and is imported by `usermod`, then the above statement will not do anything useful due to certain idiosyncrasies in the XSB module system. In this case, we have to tell the system that `foo/3` was defined in Prolog module `bar`. Thus, if `foo/3` was defined as a dynamic predicate in the module `bar`, we have to write:

```
?- assert(foo(a,b,c)@\prolog(bar))@\prolog.
```

Note that if we want to assert a more complex fact, such as `foo(f(a),b,c)`, we would have to use either

```
?- assert(foo(f(a)@\prolog(bar),b,c)@\prolog(bar))@\prolog.
```

or `@\prologall`:

```
?- assert(foo(f(a),b,c)@\prologall(bar))@\prolog.
```

We should also mention one important difference between insertion of facts in $\mathcal{E}$RGO and Prolog. Prolog treats facts as members of a *list*, so duplicates are allowed and the order matters. In contrast, $\mathcal{E}$RGO treats the database as a *set* of facts with no duplicates. Thus, insertion of a fact that is already in the database has no effect.

Deletion. The syntax of a deletion primitive is as follows:

```
delop{literals [| query]}
```

where *delop* can be `delete`, `deleteall`, `erase`, and `eraseall`. The *literals* part is a comma separated list of frames and predicates. The optional part, `|`*query*, represents an additional constraint or a restricted quantifier, similar to the one used in the insertion primitive.

For instance, the following predicate:

```
?- deleteall{John[?Year(?Semester)->?Course] | ?Year < 2000}
```

will delete John's course selection history before the year 2000.

Note that the semantics of a `delete`{*literal*`|`*query*} statement is that first the query *literal* $\wedge$ *query* should be asked. If it succeeds, then deletion is performed. For instance, if the database is

```
p(a). p(b). p(c). q(a). q(c).
```

then the query below:

```
?- deleteall{p(?X)|q(?X)}
```

will succeed with the variable `?X` bound to `a` and `c`, and `p(a)`, `p(c)` will be deleted. However, if the database contains only the facts `p(b)` and `q(c)`, then the above predicate will succeed (`deleteall` always succeeds) and the database will stay unchanged.

$\mathcal{E}$RGO provides four deletion primitives: `delete`, `deleteall`, `erase`, and `eraseall`. The primitive `delete` removes at most one fact at a time from the database. The primitives `deleteall` and `eraseall` are **bulk delete** operations; `erase` is kind of a hybrid: it starts slowly, by deleting one fact, but may go on a joy ride and end up deleting much of your data. These primitives are described below.

1. If there are several bindings or matches for the literals to be deleted, then `delete` will choose only one of them nondeterministically, and delete it. For instance, suppose the database contains the following facts:

   ```
   p(a). p(b). q(a). q(b).
   ```

   then

```
?- delete{p(?X),q(?X)}
```

will succeed with `?X` bound to either `a` or `b`, depending on the ordering of facts in the database at runtime.

However, as with insertion, in the all-answers mode the above deletion will take place for each binding that makes the query true. To avoid this, use one answer at a time mode or a cut.

2. In contrast to the plain `delete` primitive, `deleteall` will try to delete all bindings or matches. Namely, for each binding of variables produced by *query* it deletes the corresponding instance of *literal*. If *query* $\wedge$ *literal* is false, the `deleteall` primitive fails. To illustrate, consider the following:

   ```
   ?- p(?X,?Y), deleteall{q(?X,?Y,?Z)|r(?Y,?Z)}.
   ```

   and suppose `p(x,y)` is true. Then the above statement will, for each `z` such that `r(y,z)` is true, delete `q(x,y,z)`.

   For another example, suppose the database contains the following facts:

   ```
   p(a). q(b). q(c).
   ```

   and the query is `?- deleteall{p(a),q(?X)}`. The effect will be the deletion of `p(a)` and of all the facts in `q`. (If you wanted to delete just one fact in `q`, `delete` should have been used.)

   Unlike the `delete` predicate, `deleteall` *always* succeeds. Also, `deleteall` leaves all variables unbound.

3. `erase` works like `delete`, but with an object-oriented twist: For each F-logic fact, f, that it deletes, `erase` will traverse the object tree by following f's methods and delete all objects reachable in this way. It is a power-tool that can cause maiming and injury. Safety glasses and protective gear are recommended.

   Note that only the base part of the objects can be erased. If the object has a part that is derived from the facts that still exist, this part will not be erased.

4. `eraseall` is the take-no-prisoners version of `erase`. Just like `deleteall`, it first computes *query* and for each binding of variables it deletes the corresponding instance of *literal*. For each deleted object, it then finds all objects it references through its methods and deletes those. This continues recursively until nothing reachable is left. This primitive always succeeds and leaves its free variables unbound.

25.2 Transactional (Logical) Updates

The effects of transactional updates are *undone upon backtracking*, i.e., if some post-condition fails and the system backtracks, a previously inserted item will be removed from the database, and a previously deleted item will be put back.

The syntax of transactional update primitives is similar to that of non-transactional ones and the names are similar, too. The syntax for transactional insertion is:

t_insop{*literals* [| *formula*]}

while the syntax of a transactional deletion is:

t_delop{*literals* [| *query*]}

where `t_insop` stands for either `t_insert` or `t_insertall`, and `t_delop` stands for either of the following four deletion operations: `t_delete`, `t_deleteall`, `t_erase`, and `t_eraseall`. The meaning of *literals* and *query* is the same as in Section 25.1.

The new update operators `t_insert`, `t_insertall`, `t_delete`, `t_deleteall`, `t_erase`, and `t_eraseall` work similarly to the non-transactional `insert`, `delete`, `deleteall`, `erase`, and `eraseall`, respectively, except that the new operations are *transactional*. Please refer to Section 25.1 for details of the non-transactional update operators.

The keywords `tinsert`, `tinsertall`, `tdelete`, `tdeleteall`, `terase`, and `teraseall` are also understood and are synonymous to the `t_*` versions of the transactional operators.

To illustrate the difference between transactional and non-transactional updates, consider the following execution trace immediately after the $\mathcal{E}$RGO system starts:

```
ergo>  insert{p(a)}, \false.

No

ergo>  p(a).

Yes

ergo>  t_insert{q(a)}, \false.   // or tinsert{q(a)},

No
```

```
ergo>  q(a).

No
```

In the above example, when the first `\false` executes, the system backtracks to `insert{p(a)}` and does nothing. Thus the insertion of `p(a)` persists and the following query `p(a)` returns with `Yes`. However, when the second `\false` executes, the system backtracks to `t_insert{q(a)}` and removes `q(a)` that was previously inserted into the database. Thus the next query `q(a)` returns with `No`. This behavior is similar to database transactions, hence the name "transactional" update.

Notes on working with transactional updates. Keep in mind that some things that Prolog programmers routinely do with `assert` and `retract` go against the very concept of transactional updates.

- `fail`-loops are not going to work (will leave the database unchanged) for obvious reasons. The loop forms `\while` and `\until` should be used in such situations.
- Tabled predicates or methods must never depend on transactional updates. First, as explained on page 145, tabled predicates should not depend on any predicates that have side effects, because this rarely makes sense. Second, when evaluating tabled predicates, XSB performs backtracking unbeknownst to the knowledge engineer. Therefore, if a tabled predicate depends on a transactional update, backtracking will happen invisibly, and the updates will be undone. Therefore, in such situations transactional updates will have no effect.
- As before, `t_insertall`, `t_deleteall`, and `t_eraseall` primitives always succeed and leave the free variables unbound. Likewise, in the all-answers mode, the primitives `t_insert`, `t_delete`, and `t_erase` behave similarly to the `t_*all` versions in other respects, i.e., they will insert or delete facts for every answer to the associated query. This can be prevented with the use of a cut or the `\one` directive.

25.3 Updates and Tabling

Changing predicates on which tabled predicates depend. By default, $\mathcal{E}$RGO uses an advanced form of tabling, called *reactive* tabling. To understand what it is, consider the following knowledge base.

```
p(?X) :- q(?X).
q(1).
?- p(?X).
```

Loading this little example into $\mathcal{E}$RGO will yield the answer `?X = 1`, as expected. Next, suppose we add `q(2)` to the knowledge base and ask the same query, `p(?X)`. This time, the answer will be both `?X = 1` and `?X = 2` — also as expected. Although one tends to take this behavior for granted, it is useful to understand what is going on here.

After the first query, the query itself and its single answer is recorded in a table. All subsequent calls to that query are supposed to be answered just by a table lookup, without using the rules that define `p(?X)`. However, when `q(2)` is added, the system *reacts* and updates the answer table for our query. This is why we got the correct result when the query was issues the second time. This is achieved through the mechanism of *reactive tabling* of the underlying XSB inference engine. Natural as it is, implementing this type of tabling is very hard and XSB is the *only* logic programming system that supports it. Reactive tabling is similar to materialized view maintenance in commercial database systems. However, commercial DBMS have an easier job, as none of them does maintenance for recursive views. Even for non-recursive views, maintaining materialized views is not straightforward from the user point of view in such systems.

Passive (non-reactive) tabling mode. Although reactive tabling is what one usually needs, it is more expensive computationally. In some cases, a knowledge base might not be making any updates at all and paying the computational price of reactive tabling would be unjustified. For this case, $\mathcal{E}$RGO allows the user to request *passive tabling* for any module.

In $\mathcal{E}$RGO modules that use passive tabling, tables are not updated after they are constructed in response to queries. Therefore, subsequent updates followed by further queries in such modules may return incorrect (stale) answers. For instance, in the above example, the second query `p(?X)` will return a stale answer in which `?X = 2` will be missing.

To request passive tabling in a module at compile time, use this directive:

```
:- setsemantics{tabling=passive}.
```

Passive tabling can also be requested at run time, in which case the mode will change from reactive to passive on-the-fly:

```
?- setsemantics{tabling=passive}.
```

However, runtime changes to the tabling mode is not recommended and there are certain restrictions. For instance, this can be done only as a top-level query, not as part of any other query.

Interaction between passive and reactive tabling. In $\mathcal{E}$RGO, passive modules can issue calls to reactive modules and vice versa. However, inter-module queries (between passive and reactive

and between passive and passive modules) will not be updated reactively in this case. Consider the following example:

```
// Module m1 (reactive, by default)          // Module m2 (passive)
                                            :- setsemantics{tabling=passive}.
r(a).
s(?X) :- r(?X).                             p(?X) :- s(?X)@m1.
%upd :- insert{r(b)}.
```

If the query `p(?X)@m2` is issued first, one gets the answer `?X=a`. Suppose next `%upd@m1` is invoked to update `r/1`. If then `p(?X)@m2` is asked again, the answer will still be `?X=a` even though `s(?X)@m1` would be reactively updated to include the answer `?X=b`. The reason is that `p(?X)@m2` is maintained passively, and it will not be alerted to the change in `s(?X)@m1`. Similarly, if `m1` were passive and `m2` reactive, the query `p(?X)@m2` will still not be updated, but for a different reason: here `s(?X)@m1` will not get updated, since in that case it would be maintained passively.

More generally, passively tabled queries invoked anywhere in the call-chain will stop update propagation up the chain. To illustrate, consider the following example:

```
// Module m1 (reactive, by default)          // Module m2 (passive)
                                            :- setsemantics{tabling=passive}.
r(a).
s(?X) :- r(?X).                             q(?X) :- s(?X)@m1.
p(?X) :- q(?X)@m2.
%upd :- insert{r(b)}.
```

Here module `m1` calls `m2` and the latter calls back. If the queries `s(?X)@m1` and `p(?X)@m1` are issued then one gets `?X=a` as an answer in both cases. Suppose now that `%upd@m1` is invoked thereby changing `r/1`. Since `m1` is reactive, the query `s(?X)@m1` will now return both `?X=a` and `?X=b`. On the other hand, `p(?X)@m1` will still return only the first answer because it calls `q(?X)@m2` and the change to `r/1` is not propagated to `q(?X)@m2`, since the latter is defined in a passive module.

Explicit refresh of passive tables. What if updates in certain modules are very rare, but queries are frequent and time is money? In this case, passive tabling might still make sense. $\mathcal{E}$RGO provides a partial redress in the form of the `refresh{...}` operator. This operator lets the knowledge engineer explicitly remove stale answers from tables. For instance, in the above example we could do the following:

```
ergo>  refresh{p(?)}, p(?X).
```

and get the right result. In general, `refresh{...}` can take a comma-separated list of facts to be purged from the tables, and these facts can even contain unbound variables. In the latter case, any stale call that unifies with the given facts will be refreshed. For instance,

```
ergo>  refresh{a[b->?X], c:?Y, p(z,?V)@foo}.
```

will refresh the tables for `a[b->?X]` and `c:?Y` in module `main`, and for `p(z,?V)` in module `foo`.

Sometimes it is desirable to completely get rid of all the information stored in tables—for instance, when it is hard to track down all the facts that might depend on the changed base facts. In such a case, the command

```
ergo>  \abolishtables.
```

can be used. However, this command should be executed only as a *standalone* query. Also, neither `refresh{...}` nor `\abolishtables` can occur under the scope of the negation operator `\naf` (either directly or indirectly).

Note that both `refresh` and `\abolishtables` can be also used in modules that use reactive tabling. However, in this case, `refresh` has no effect and `\abolishtables` will affect only the passive modules, if there are any.

Tabled predicates that depend on update operations. A related issue is that a tabled predicate (or a frame literal) might occur in the head of a rule that has an update operation in its body, or it may be transitively dependent on such an update. Note that this is different from the previous issue, where tabled predicates did not necessarily depend on update operations but rather on other predicates that were modified by these update operations.

In this case, the update operation will be executed the first time the tabled predicate is evaluated. Subsequent calls will return the predicate truth value from the tables, without invoking the predicate definition. Moreover, if the update statement is non-logical (i.e., non-transactional), then it is hard to predict how many times it will be executed (due to backtracking) before it will start being ignored due to tabling.

If $\mathcal{E}$RGO compiler detects that a tabled literal depends on an update statement, a warning is issued, because such a dependency is most likely a mistake. This warning is issued also for transactional methods (i.e., Boolean methods of the form `%foo(...)`) when a tabled literal depends on them. Moreover, because non-tabled HiLog predicates are regarded as having transactional side-effect by default, this warning is also issued when a tabled literal depends on non-tabled HiLog predicates. $\mathcal{F}$LORA-*2* also imposes restrictions on the use of updates that affect tabled facts: it does not allow such dependency to occur in modules with reactive tabling. $\mathcal{E}$RGO does allow such dependency by providing the special feature of "stealth mode" updates.

See the `stealth{...}` primitive described in Section 25.6.

Here is an example of a situation where dependency on an update makes perfect sense. For instance, we might be computing a histogram of some function by computing its values at every point and then adding it to the histogram. When a value, $f(a)$, is computed first, the histogram is updated. However, subsequent calls to $f(a)$ (which might be made during the computation of other values for f) should not update the histogram. In this case it makes sense to make $f/1$ into a tabled predicate, whose definition will include an update operator. For this reason, a compiler directive `ignore_depchk` is provided to exempt certain predicates and methods from such dependency checks and thus silence the warnings.

The example below shows the usage of the `ignore_depchk` directive.

```
:- ignore_depchk{%ins(?), ?[%?]@?}.
t(?X,?Y) :- %ins(?X), ?Y[close]@\io.
%ins(?X) :- insert{?X}.
```

No dependency warning is issued in this case. However, without the `ignore_depchk{...}` directive, three warnings would be issued saying that tabled literal `t(?X,?Y)` depends on `%ins(?X)`, `?Y[%close]`, and `insert`. Notice that `ignore_depchk{%ins(?_)}` tells the compiler to ignore not only dependencies on `%ins/1`, but also all dependencies that have `%ins/1` in the path.

The `ignore_depchk{...}` directive can also be used to ignore direct dependencies on updates. For example,

```
:- ignore_depchk{insert{?_,?_|?_}}.
```

ignores dependencies on conditional insertions which insert two literals such as `insert{a,b|c,d,e}`. And

```
:- ignore_depchk{insert{?}}.
```

ignores dependencies on unconditional insertions which insert exactly one literal such as `insert{p(a)}` but not `insert{p(a),p(b)}`.

We should also mention here that executing the command `warnings{compiler}` in the $\mathcal{E}$RGO shell will turn the dependency checking off globally. In some cases, this can reduce the compilation time, but is discouraged except when debugging is finished.

25.4 Updates and Meta-programming

The update operators can take variables in place of literals to be inserted. For instance,

```
?- ?X ~ a[b->c], insert{?X}.
```

Here ~ is a meta-unification operator described in Section 16.1. One use for this facility is when one module, `foo`, provides methods that allow other modules to perform update operations on objects in `foo`. For instance, `foo` can have a rule

```
%update(?X,?Y) :- delete{?X}, insert{?Y}.
```

Other modules can then issue queries like

```
?- John[salary->?X]@foo, ?Y \is ?X+1000,
        %update(John[salary->?X],John[salary->?Y])@foo.
```

25.5 Updates and Negation

Negation applied to methods that have side effects is typically a rich source of trouble and confusion.

First of all, applying negation to frames that involve non-transactional updates does not have logical semantics, and this requires the knowledge engineer to have a certain understanding of the *operational* semantics of $\mathcal{E}$RGO (Section 22.4). In this case, the semantics of negation applied to methods or predicates that produce side-effects through updates or I/O is that of Prolog negation (Section 17.1).

When only transactional updates are used, the semantics is well defined and is provided by *Transaction Logic* [2, 1]. In particular, negation is also well defined. However, simply negating an update, A, is useless, since logically it means jumping to some random state that is not reachable via execution of A. As explained in [2, 1], negation is typically useful only in conjunction with $\wedge$, where it acts as a constraint, or with the hypothetical operator of possibility `<>`. In most cases, when the knowledge engineer wants to apply negation to a method that performs logical (transactional) updates, he has ~`<>`*method* in mind, i.e., a test to verify that execution of *method* is not possible. Both <> and ~`<>` (equivalently `\naf<>`) are supported in $\mathcal{E}$RGO.

The idiom `\neg` %*method* is considered to be an error, but `\+` %*method* and `\naf` %*method* (the latter interpreted as `\+` %*method*) are permitted. However, they yield meaningful results only if %*method* has no side effects and is non-recursive.

A transaction of the form ?– `<>`*expression* is true in the current state if the transaction ?– *method can* be executed in the current state. The state, however, is not changed even if ?– *method* does change the state during its execution. In other words, ?– `<>`*expression* only *tests* if execution is possible. The transaction ?– ~ `<>`*expression* tests if execution of *expression* is *im*possible. It is true if ?– *expression* fails and false otherwise. However, regardless of whether ?– *expression* fails or not, the current state underlying the knowledge base does not change. Note

that `<>` and sin`<>` can apply to a group of subgoals, if they are enclosed in parentheses. If the expression does not change the underlying state of the knowledge base then ?– `<>` *expression* and ?– ∼`<>` *expression* reduce to ?– *expression* and ?– `\naf` *expression*, respectively. Examples:

```
?- <> %p(?X).
?- <> (%p(?X), q(?X,?Y), %r(?Y)).
?- ~<> (%p(?X), q(?X,?Y), %r(?Y)).
?- ~<> (%p(?X), q(?X,?Y), <> %r(?Y)).
```

25.6 Stealth Updates

Issuing an update command while still evaluating a non-transactional subgoal in a reactive tabling module will cause a runtime error:

```
p(?X) :- insert{%abc}.
?- p(1).

++Error[Ergo]> tabled predicate/method depends on an update.
```

This happens because, while computing the truth value of `p(1)`, we changed the underlying knowledge base, which could have affected the result. In this particular case, inserting `%abc` cannot affect `p(1)` and the inference engine could have noticed this. However, maintaining precise dependencies among the literals is expensive and is usually not worth the effort and the computational cost. In general, of course, inserting `%abc` could have affected the truth value of `p(1)`, which makes `p(1)` even harder to evaluate, whence the error.

As discussed earlier, the above type of dependency is bad style and issuing an error makes perfect sense. However, in some situations, like the example of computing histograms on page 146, updates issued while evaluating tabled subgoals can be useful. Another situation when this facility could be used is when one needs to do some kind of bookkeeping such as recording if a particular tabled literal has ever been called.

To this end, $\mathcal{E}$RGO provides the `stealth{...}` primitive, which can be used in insert and delete commands to make literals being inserted/deleted invisible to the reactive tabling mechanism, thus avoiding runtime errors. For instance,

```
p(?_X) :- insert{stealth{%abc,%p(2),a[%b]}}.
?- p(1).

Yes
```

And yes, only transactional literals, as in the above example, can be inserted/deleted with the help of the `stealth{...}` primitive. Without this primitive, even transactional literals cannot be used in updates while evaluating a tabled subgoal in a reactive module.

Note that transactional variables *are* allowed in stealth updates. For example,

```
ergo> ?X = ppp, insert{stealth{%?X(1)}}.
?X = ppp

ergo> %ppp(?X).
?X = 1
```

Another interesting use of stealth updates is defining UDFs (see Section 23) that have side effects like database updates. For instance, suppose we want a UDF to insert something into the database and return some result:

```
%fooPRED(?val, =?Res+1) :- insert{%abc(?val)}, ?Res = ?val.
\udf fooUDF(?form) := ?res \if %fooPRED(?form, ?res).
```

Unfortunately, this won't work because if one asks the query `?- ?A=fooUDF(1)` then an error will result for the same reason as earlier in this section. However, changing `insert{%abc(?val)}` to `insert{stealth{%abc(?val)}}` saves the day.

25.7 Counter Support

It is often necessary to maintain a global counter, which can be set, queried, and updated. In principle, this can be done by designating a certain unary predicate and updating its content as necessary. However, this is a bit cumbersome, and $\mathcal{E}$RGO provides a more efficient way. A counter is just an $\mathcal{E}$RGO symbol that can be accessed via the following operations:

- `counter{Name := Integer}` — set counter
- `counter{Name := ?Var}` — query counter
- `counter{Name + Integer}` — increment counter
- `counter{Name − Integer}` — decrement counter

For example,

```
?- counter{abc:=3}, counter{abc+5}, counter{abc=?X}.
?X = 8.
```

26 Insertion and Deletion of Rules

$\mathcal{E}$RGO supports *non-transactional* insertion of rules into modules as well as deletion of inserted rules. A module in $\mathcal{E}$RGO gets created when a file is loaded into it, as described in Section 4, or it can be created using the primitive `newmodule`. Subsequently, rules can be added to an existing module. Rules that are inserted via the `insertrule` and `add{...}` commands are called *dynamic* and the rules loaded using the `load{...}` or `[...]` commands are called *static* or *compiled*. Dynamic rules can be deleted via the `deleterule` command. As mentioned in Section 25, $\mathcal{E}$RGO predicates and frames can have both static and dynamic parts and no special declaration is required to make a predicate dynamic. The same frame or a predicate can be defined by a mixture of static or dynamic rules.

In this section, we will first look at the syntax of creating new modules. Then we will describe how to insert rules and delete rules. Finally, we address other related issues, including tabling, indexing, and the cut.

26.1 Creation of a New Module and Module Erasure at Run-time

The syntax for creating a new module is as follows:

newmodule{modulename}

This creates a blank module with the given name and default semantics. If a module by that name already exists, an error results. A module created using `newmodule` can be used just as any module that was created by loading a user knowledge base.

A dual operation to module creation is *erasemodule* with the following syntax:

erasemodule{modulename}

26.2 Insertion of Rules

Dynamic rules can be inserted before all static rules, using the primitive `insertrule_a`, or after all static rules, using the primitive `insertrule_z` or just `insertrule`. The reason for having three different insertion commands is the same as in Prolog: the position of a rule with respect to other rules may sometimes have an affect on performance or query termination.

Several rules can be inserted in the same command. The syntax of inserting a list of rules is as follows:

insruleop{rulelist}

where *insruleop* is either `insertrule_a`, `insertrule_z`, or `insertrule`, and *rulelist* is a comma-separated list of rules. The rules being inserted *should not* terminate with a period (unlike the static rules):

```
?- insertrule_a{?X:student :- %enroll(?X,?_T)}.
```

The above inserts the rule `?X:student :- %enroll(?X,?_T)` in front of the current module.

If a rule is meant to be inserted into a module other than the current one, then the rule needs to be parenthesized *and* the module name must be attached using the usual module operator `@`. If *several* rules need to be inserted using the same command, each rule must be parenthesized. For example, the following statement inserts the same rule into two different modules: the current one and into module `mod1`.

```
?- insertrule_a{(?X:student :- %enroll(?X,?_T)),
                (?X:student :- %enroll(?X,?_T))@mod1}.
```

As a result, the rule `?X:student :- %enroll(?X,?_T)` will be inserted in front of each of these two modules. For this to be executed successfully, the module `mod1` must already exist.

Note: rule Ids and other meta-data (see Sectin 34) can be supplied with the insert operator:

```
?- insertrule{@!{abc[foo->bar]} ?X:student :- %enroll(?X,?_T)}.
```

Meta information can also be supplied with the `deleterule` operator.

26.3 Deletion of Rules

Rules inserted dynamically using `insertrule_a` can be deleted using the primitive `deleterule_a`, and rules inserted using `insertrule_z` can be deleted using the primitive `deleterule_z`. If the user wishes to delete a rule that was previously inserted using either `insertrule_a` or `insertrule_z` then the primitive `deleterule` can be used. Similarly to rule insertion, several rules can be deleted in the same command:

delruleop`{`*rulelist*`}`

where *delruleop* is either `deleterule_a` or `deleterule_z` and *rulelist* is a comma-separated list of rules. Rules in the list must be enclosed in parentheses and *should not* terminate with a period.

To delete the rules inserted in the second example of Section 26.2, we can use

```
?- deleterule_a{(?X:student :- %enroll(?X,?_T)),
                 (?X:student :- %enroll(?X,?_T))@mod1}.
```

or

```
?- deleterule{(?X:student :- %enroll(?X,?_T)),
                 (?X:student :- %enroll(?X,?_T))@mod1}.
```

$\mathcal{E}$RGO provides a flexible way to express rules to be deleted by allowing variable rule head, variable rule body, and variable module specification. For example, the rule deletions below are all valid:

```
?- deleterule{(?H:-q(?X))@foo}.
?- deleterule{(p(?X):-q(?X))@?M}.
?- deleterule{?H:-?B}.
```

The last query attempts to delete every dynamically inserted rule. So, it should be used with great caution.

We should note that a rule with a composite head, such as

```
o[b->?V1,c->?V2] :- something(?V1,?V2).
```

is treated as a pair of separate rules

```
o[b->?V1] :- something(?V1,?V2).
o[c->?V2] :- something(?V1,?V2).
```

Therefore

```
?- deleterule{o[b->?V1] :- something(?V1,?V2)}.
```

will succeed and will delete the first of the above rules. Therefore, the following action will subsequently fail:

```
?- deleterule{o[b->?V1,d->?V2] :- ?Body}.
```

The problem of Cuts What is behind rule insertion is pretty simple. As we know from Section 25, every predicate and object has a base part and a derived part. Now we further divide the derived part into three sub-parts: the *dyna sub-part* (the part that precedes all other facts in the predicate), the static sub-part, and the *dynz sub-part*. All rules inserted using `insertrule_a` go into the dyna sub-part; all the rules in the file go into the static sub-part; and all the rules inserted using `insertrule_z` go into the dynz sub-part.

This works well when there are no cuts in rules inserted by `insertrule_a`. With cuts, the knowledge base might not behave as expected. For example, if we have the following rules:

```
p(?X) :- r(?X).
r(a).
q(b).
?- insertrule_a{p(?X) :- q(?X),!}.
?- p(?X).
```

we would normally expect the answer to be `b` only. However, $\mathcal{E}$RGO will return two answers, `a` and `b`. This is because the cut affects only the dynamic part of p(?X), instead of all the rules for `p/1`.

27 Querying the Rule Base

Rule bases in $\mathcal{E}$RGO can be queried in two ways. For basic querying, $\mathcal{E}$RGO provides the `clause{...}` primitive, which is severely limited: to query rules using rule heads the user must know the exact structure of thees rule heads after the omnification transformation (which is not always obvious—see Section 19); to query via the subformulas in the rule body, the user must know the structure of the body or to write programs for parsing rule bodies, which is fairly complicated. To simplify this task, $\mathcal{E}$RGO maintains the structural database of the rules, which provides flexible interface for querying the rules.

27.1 Basic Querying of Rules via `clause{...}`

The rule base can be queried using the primitive `clause`. The syntax of `clause` is as follows:

`clause{`*head,body*`}`

where *h*ead can be anything that is allowed to appear in a rule head and *b*ody can be anything that can appear in a rule body. In addition, explicit module specifications are allowed in rule heads in the `clause` primitive. Both *head* and *body* represent templates that unify with the actual rules and those rules that unify with the templates are returned.

The following example illustrates the use of the `clause` primitive. Suppose we have previously inserted several rules:

```
?- insertrule_a{tc(?X,?Y) :- e(?X,?Y)}.
?- insertrule_a{tc(?X,?Y) :- tc(?X,?Z), e(?Z,?Y)}.
?- newmodule{foo}.
?- insertrule_a{(tc(?X,?Y) :- e(?X,?Y)@\@)@foo}.
?- insertrule_a{(tc(?X,?Y) :- tc(?X,?Z), e(?Z,?Y)@\@)@foo}.
```

Then the query

```
?- clause{?X,?Y}.
```

will list all the inserted rules. In this case, four rules will be returned. To query specific rules in a specific module — for example, rules defined for the predicate `tc/2` in the module `foo` — we can use

```
?- clause{tc(?X,?Y)@foo,?Z}.
```

We can also query rules by providing patterns for their bodies. For example, the query

```
?- clause{?X, e(?_,?_)}.
```

will return the first and the third rules.

Querying rules with composite heads involves the following subtlety. Recall from Section 26.3 that a rule with a composite head, such as

```
o[b->?V1,c->?V2] :- something(?V1,?V2).
```

is treated as a pair of rules

```
o[b->?V1] :- something(?V1,?V2).
o[c->?V2] :- something(?V1,?V2).
```

Therefore, if we delete one of these rules, for instance,

```
?- deleterule{o[b->?V1] :- something(?V1,?V2)}.
```

then a query with a composite head that involves the head of the deleted rule will fail (unless there is another matching rule). Thus, the following query will fail:

```
?- clause{o[b->?V1,d->?V2], ?Body}.
```

The `clause` primitive can be used to query static rules just as it can be used to query dynamic rules. The normal two-argument primitive queries all rules. If one wants to query only the static (compiled) rules or only dynamic (inserted) rules, then the three-argument primitive can be used.

`clause{`*type,head,body*`}`

For example,

```
?- clause{static,?X,?Y}.
?- clause{dynamic,?X,?Y}.
```

Within the dynamic rules, one can separately query just the dynamic rules that precede all the static rules (using the flag `dyna`) or just those dynamic rules that follow all the static ones (with the `dynz` flag):

```
?- clause{dyna,?X,?Y}.
?- clause{dynz,?X,?Y}.
```

Due to a limitation of the underlying Prolog system, the `clause` primitive cannot query rules whose size exceeds a limit imposed by the Prolog system. A warning message is issued when a rule exceeds this limit and thus cannot be retrieved by `clause`. The only way to remedy this problem is to split the long rule into smaller rules by introducing intermediate predicates.

The `clause` statement can also be used to query the *labels* and meta-data associated with $\mathcal{E}$RGO rules. This form of the clause statement is described in Section 34.

Note that in $\mathcal{E}$RGO facts are different from rules. They cannot be queried using the `clause` primitive. Instead, the primitive `isbasefact{...}` should be used. For instance,

```
a:b.
?- isbasefact{a:b}.

Yes

?- clause{a:b,?}.

No
```

27.2 Querying Rules via the Rule Structure Database

The *rule structure database* provides a simple yet powerful interface for querying the rule structure of $\mathcal{E}$RGO knowledge bases. In many respects, this interface goes beyond what `clause{...}` can do, but it does not replace `clause{...}`. The major advance over `clause{...}` is that it allows one to find all rules *containing* a particular atomic formula—either in the head or in the body. This is possible even if the formula occurs under `\neg` and/or `\naf`, under a quantifier, aggregate, and other constructs. In addition, the query can have complex goals such as those allowed under the Lloyd-Topor and omniformity extensions described in Sections 18 and 19.

Requesting the structural database feature. $\mathcal{E}$RGO does not build the structure database by default because of a slight overhead in compile time, load time, and additional demands on RAM. The user can request that $\mathcal{E}$RGO builds a structure database for a file by placing the directive

```
:- use_rule_structure_db.
```

in the file (or in a file `#include`'ed in that file). If the above directive is not present, the calls to the structural query API will simply fail (will return false).

If, so some reason, it is preferable to not include the above directive in a file, structure database may be requested at *runtime* by executing the primitive `use_rule_structure_db{Module,Flag}` as a query. For instance,

```
?- use_rule_structure_db{foo,on}.
```

From then on, any file compiled for loading into the module `foo` will have the rule structure database available. To turn this mode off, use

```
?- use_rule_structure_db{foo,off}.
```

Note, these commands *affect only the files that need recompilation* (because they were never compiled or because they were changed since their last compilation). If a file was compiled *before* the change in the default mode for rule structure databases (from `on` to `off` or vice versa) then reloading such a file will not cause a recompilation and thus the availability of that database with respect to the aforesaid file will not change.

Structural queries. The structural database provides special primitives for querying the rule structure:

- `structdb{`*?Goal*,*?SearchSpec*,*?Id*,*?Module*,*?File*`}`
  Here *?Goal* and *?SearchSpec* must be bound. The *?Goal* argument is a formula that is allowed in the head or the body of a rule; it does not need to be reified, as $\mathcal{E}$RGO knows what to expect here. The *?Goal* formula is decomposed into a list of atomic frames or predicates and then the rule base is searched according to *?SearchSpec*. The latter can be as follows:

  - `any` — means that a rule satisfies the search condition if it contains *any* of the aforesaid atomic formulas (in the head or body)
  - `all` — means a rule satisfies the search condition if it matches *all* of the aforesaid atomic formulas (whether they appear in the head or body)
  - $X+Y$, where X can be `any` or `all` and Y can be `head`, `body` or a variable. For instance, `all+head` means that for a rule to satisfy the search condition it must contain all of the aforesaid atomic formulas in the head; `all+body` is similar except that all the formulas must occur in the body of that rule. If `any+head` is given, it means that it is enough to have just one of the aforesaid atomic formulas in the rule head, and so on.

  The output variables are *?Id*, *?Module*, and *?File*. Why these three? Recall from Section 34 that a rule Id is composed out of these three components and knowing them opens the door to the host of rule manipulation meta-facilities described in this manual.

- `structdb{`*?Id*,*?Module*,*?File*,*?Goal*,*?LocationInRule*,*?Context*`}`
  Here *?Id* and *?Module* must be bound before issuing a call to this primitive. That is, given a partial (or complete) rule Id, this primitive returns the atomic formulas that occur in the corresponding rules (via the *?Goal* variable). In addition, for each such formula, it tells us (via the `?LocationInRule` variable) whether that formula occurs in the rule head or in its body. Finally, the semantic context of the formula is also returned as a list of semantic attributes. The relevant semantic attributes are `\naf`, `\neg`, `\+`, and all the aggregate functions `avg`, `sub`, `setof`, and the others. For instance, if, say, $max\{?X \mid p(?X,?Y),$ `\naf \neg` $q(?Y)\}$ appears in the rule then the semantic context of $p(?X,?Y)$ would be the list `[max]` while for $q(?Y)$ it will be `[\neg,\naf,max]` (the semantic attributes are listed in the reverse order of their appearance).

Here are some examples that illustrate the use of the above primitives.

```
:- use_rule_structure_db.
@!{r1} p(?X) :- q(?X), r(?X).
@!{r2} p(?X)==>r(?X) :- q(?X), ?X=avg{?V|\neg t(?V)}.
@!{r3} \neg q(?X) :- \neg p(?X).
?- structdb{(p(?)==>r(?)),all+head,?Id,?Mod,?File}.
?- structdb{(p(?)==>r(?)),all,?Id,?Mod,?File}.
```

```
?- structdb{(p(?)==>r(?)),any,?Id,?Mod,?File}.
?- structdb{r2,main,?,?P,?H,?C}.
```

Then the first query returns

```
?Id = r2
?Mod = main
?File = 'foo.ergo'
```

because `p(?)` and `r(?)` both occur in the rule *head* only in rule `r2`. The second query returns

```
?Id = r1
?Mod = main
?File = 'foo.ergo'

?Id = r2
?Mod = main
?File = 'foo.ergo'
```

because `p(?)` and `r(?)` both occur in *some* part of a rule (head or body) in rules `r1` and `r2`. The third query returns information about all three rules because it is looking for rules where either `p(?)` or `r(?)` occur (in the head *or* body). Finally, the last query (which uses the 6-argument version of the `structdb` primitive) returns the information about all the literals found in rule `r2`. For each literal it says where the literal occurs and in what context:

```
?P = ${p(?_h2847)@main}
?H = body
?C = []

?P = ${p(?_h2875)@main}
?H = head
?C = ['\\neg']

?P = ${q(?_h2819)@main}
?H = body
?C = []

?P = ${r(?_h2729)@main}
?H = head
?C = []
```

```
?P = ${r(?_h2757)@main}
?H = body
?C = ['\\neg']

?P = ${t(?_h2787)@main}
?H = body
?C = ['\\neg', avg]

?P = ${\neg p(?_h2963)@main}
?H = head
?C = []

?P = ${\neg r(?_h2905)@main}
?H = body
?C = []

?P = ${\neg t(?_h2933)@main}
?H = body
?C = [avg]
```

Of particular interest here are the answers such as

```
?P = ${r(?_h2757)@main}
?H = body
?C = ['\\neg']

?P = ${t(?_h2787)@main}
?H = body
?C = ['\\neg', avg]

?P = ${\neg t(?_h2933)@main}
?H = body
?C = [avg]
```

The first says that `r(...)` occurs in the body under negation `\neg`, the second says that `t(...)` occurs in the body under first `\neg` and then the aggregation `avg`. The last answer rephrases the second one, saying that the literal `\neg t(...)` occurs under aggregation `avg`.

Finally, we give an example of a 5-argument `structdb` query where the mode has form `all+?Y` or `any+?Y`:

```
?- structdb{(p(?)==>r(?)),all+?H,?Id,?Mod,?File}.

?H = body
?Id = r2
?Mod = main
?File = 'foo.ergo'

?H = head
?Id = r2
?Mod = main
?File = 'foo.ergo'
```

We do know that `r2` is the right rule here, but one might wonder why in the first answer `?H` is bound to `body` while in the second to `head`. On the surface, the first answer seems incorrect since in rule `r2` both `p(?)` and `r(?)` seem to appear in the head only. The explanation is that that rule `p(?X)==>r(?X) :- q(?X), ?X=avg{?V|\neg t(?V)}` has a composite head, which gets transformed by the omniformity transformation (Section 19) into two rules:

```
r(?X) :- q(?X), ?X=avg{?V|\neg t(?V)}, p(?X).
\neg p(?X) :- q(?X), ?X=avg{?V|\neg t(?V)}, \neg r(?X).
```

Thus, in reality, both `r(?)` and `p(?)` appear in the head as well as in the body of the rules created out of the original rule.

28 Aggregate Operations

Aggregate operators play important role in data analytics applications and are commonly used in database languages. This section describes the aggregate operators available in $\mathcal{E}$RGO.

28.1 Syntax of Aggregate Operators

An aggregate query in $\mathcal{E}$RGO has the following form:[14]

`agg{?X` | *query*`}`
`agg{?X[?Gs]` | *query*`}`

[14] The syntax for aggregates is similar to that used in the FLORID system http://www.informatik.uni-freiburg.de/~dbis/florid/.

where

- `agg` is the *aggregate operator*; it can be one of these: `min`, `max`, `count`, `countdistinct`, `sum`, `sumdistinct`, `avg`, `avgdistinct`, `setof`, `bagof`.
- `?X` is called the *aggregation variable*
- `?Gs` is a list of comma-separated *grouping variables* `bagoff` aggregate operations are to be sorted.
- *query* is a logical formula that specifies the query conditions; it has the form of a rule body formula (including conjunctions, disjunctions, quantifiers, and even nested aggregation).

All the variables appearing in *query* but not in `?X`, `?Gs`, and not appearing prior to the aggregation are considered to be existentially quantified. Furthermore, the syntax of an aggregate must satisfy the following conditions:

1. All names of variables in both `?X` and `?Gs` must appear in *query*;
2. `?Gs` should not contain `?X`.

The `setof` and `bagof` aggregates have three additional forms:

```
agg{?X(SortSpec) | query}
agg{?X[?Gs](SortSpec) | query}
agg{?X[?Gs](SortSpec1,SortSpec2) | query}
```

where `SortSpec` specifies how the result of the `setof`. The exact syntax for this is described in Section 28.4.

28.2 Evaluation of Aggregates

Aggregates are evaluated as follows: First, the query condition specified in *query* is evaluated to obtain all the bindings for the template of the form `<?X, ?Gs>`. Then, these tuples are grouped according to each distinct binding for `<?Gs>`. Finally, for each group, the aggregate operator is applied to the list of bindings for the aggregate variable ?`X`.

The following aggregate operators are supported in $\mathcal{E}$RGO: `min`, `max`, `count`, `countdistinct`, `sum`, `sumdistinct`, `avg`, `avgdistinct`, `setof`, and `bagof`.

The operators `min` and `max` can apply to any list of terms. The order among terms is defined by the Prolog operator `@=<`. In contrast, the operators `sum`, `avg`, `sumdistinct`, and `avgdistinct`

can take numbers only. If the aggregate variable is instantiated to something other than a number, these operators will discard it and produce a runtime warning message.

The operator `bagof` collects all the bindings of the aggregation variable into a list. The operator `setof` works similarly to `bagof`, except that all the duplicates are removed from the result list and the resulting list is sorted lexicographically. Note that these operators create ordered lists, not sets, unlike what as the operator names might suggest.

Note: The aggregates `min`, `max`, `avg`, and `avgdistinct` fail if *query* produces no answers. In contrast, `sum`, `count`, `sumdistinct`, and `countdistinct` return 0 in such a case, and the operators `bagof` and `setof` return the empty list.

The difference between `sum`, `count`, and `avg` on the one hand and `sumdistinct`, `countdistinct`, and `avgdistinct` on the other is that the latter eliminate duplicates from the bindings produced by *query*. Thus, for example, if *query* binds the aggregate variable to, say 31, more than once then `countdistinct` will count this only once (and `sumdistinct`, `avgdistinct` will consider this binding only once also), while `count` (respectively, `sum` and `avg`) will consider 31 as many times as it was produced by the query. Example:

```
p({31,45}).
p(?X) :- ?X=31.
```

Here `count{?X|p(?X)}` yields 3 and `sum{?X|p(?X)}` evaluates to 107, while `countdistinct{?X|p(?X)}` yields 2 and `sumdistinct{?X|p(?X)}` produces 76. This is because the answer `p(31)` will be derived twice: once via one of the above facts and once via the rule. (One might not realize this by posing the top-level query `?- p(?X)`, since $\mathcal{E}$RGO eliminates duplicate answers before showing them to the user).

In general, aggregates can appear wherever a number or a list is allowed. Therefore, aggregates can be nested. The following examples illustrate the use of aggregates (some borrowed from the FLORID manual):

```
?- ?Z = min{?S|John[salary(?Year)->?S]}.
?- ?Z = count{?Year|John.salary(?Year)<max{?S|John[salary(?Y)->?S], ?Y<?Year}}.
?- avg{?S[?Who]|?Who:employee[salary(?Year)->?S]} > 20000.
```

If an aggregate contains grouping variables then this aggregate would backtrack over such grouping variables. In other words, grouping variables are considered to be existentially quantified (and the scope of that quantifier is the entire rule body). For instance, in the last query above, the aggregate will backtrack over the variable `?Who`. Thus, if `John`'s and `Mary`'s average salary is greater than `20000`, this query will backtrack and return both `John` and `Mary`.

The following query returns, for each employee, a list of years when this employee had salary less than 60. This illustrates the use of the `setof` aggregate.

```
ergo>  ?Z = setof{?Year[?Who]|?Who[salary(?Year)->?X], ?X < 60}.
?Z = [1990,1991]
?Who = Mary

?Z = [1990,1991,1997]
?Who = John
```

28.3 Notes on Scoping for Variables in Aggregate Operators

Another very important aspect is that the aggregation variable, `?X`, has the scope that is restricted to *query*. The scope of the grouping variables, `?G`, is not limited to the query, but it is usually *meaningless* for them to occur *to the left* of that scope. (The semantics of the aggregates will be explained shortly.) Such an occurrence will cause a compile-time warning and, possibly, a run-time error. To illustrate, consider the following rules:

```
head :- p(?X), r(?X), ?P = sum{?X | q(?X)}.                  // rule 1
head :- p(?X,?W), r(?X), ?P = avg{?X[?V,?W] | q(?X,?V,?W)}.  // rule 2
head :- r(?X), ?P = min{?X[?V,?W] | q(?X,?V,?W)}, p(?X,?W).  // rule 3
```

In the first rule, the first two occurrences of `?X` are outside of the scope of `sum{...}` where `?X` is used as an aggregation variable. Therefore, the first two occurrences represent variables that are distinct from the variable used in the aggregation. This is similar to scoping of variables by universal and existential quantifiers. In the second rule, the first two occurrences of `?X` are likewise unrelated to the last two occurrences of `?X` in `avg{...}`. The occurrence of `?W` to the left of the aggregation is likely a logical error caused by a misunderstanding of grouping (or, at best, is meaningless). The user probably needed this instead:

```
head :- p(?X,?W), r(?X), ?P = avg{?X[?V] | q(?X,?V,?W)}.     // rule 2'
```

If the user thought `?W` was really needed among the grouping variables, then a different behavior must have been sought. $\mathcal{E}$RGO is zealous about this and will issue an stern warning, if it finds such occurrences.[15] Moreover, if, in the second rule above, p/2 binds `?W` to a ground term then a runtime error will be issued because it makes little sense to keep a constant in a grouping list. It the user really understands what he is doing, he can avoid both the error and the warning by rewriting the second rule as

```
head :- p(?X,?WW),r(?X),?P = avg{?X[?V,?W] |q(?X,?V,?W),?W=?WW}. // rule 2''
```

[15] In some cases (e.g., in a disjunction like `p(?X) ; ?Z = sum{?P[?X]|q(?P,?X)}`), $\mathcal{E}$RGO might issue a false warning. The user can circumvent this by renaming one of the occurrences of `?X`.

but most likely the intent was rule 2' above where `?W` was simply removed from the grouping list.

In contrast, the third rule above is proper (since `?W` occurs to the right, not left, of the aggregation) but, perhaps, renaming the first `?X` would have been an improvement.

28.4 Aggregation and Sorted Results

The aggregate functions `setof` and `bagof` also have other forms, which supports sorting of the output of these aggregates:

```
agg{?X(SortSpec) | query}
agg{?X[?Gs](SimpleSortSpec) | query}
agg{?X[?Gs](SortSpec1,SortSpec2) | query}
```

where `agg` can be `setof` or `bagof`.

A *SortSpec* specifies how the output is to be sorted. It is either the constant `asc` (ascending), `desc` (descending), or a list of the form `[`*spec1*`,...,`*specN*`]`, where each component has the form `asc(`*N*`)` or `desc(`*N*`)` for some positive integer *N*. A *SimpleSortSpec* is either `asc` or `desc`.

The first form of the `setof/bagof` operator does not have grouping and the sort specification controls the order in which the values of `?X` that satisfy *query* will appear in the result of the aggregation. As the names suggest, `asc` means ascending lexicographic ordering and `desc` means that the order is descending. If the sort specification has the form `[`*spec1*`,...,`*specN*`]`, the control of the sorted order is finer. The elements of the aggregation result will be sorted first according to `spec1`, then `spec2`, etc.

Recall that each *spec-i* has the form `asc(`*N*`)` or `desc(`*N*`)`, where *N* refers to a component of each member of the aggregation result. In HiLog predicates, 1 refers to the predicate term (the predicate name), 2 refers to the first argument, etc. In a frame `obj[prop->val]`, 1 refers to `obj`, 2 to `prop`, and 3 to `val`. In binary formulas, such as `L:=:R`, `L:R`, or `L::R`, 1 refers to `L` and 2 to `R`. For lists, 1 refers to the first list element, 2 to the second, etc. Note that if *N* refers to a non-existing component of a formula, an error will be issued from the Prolog level.

The second form of the above aggregates uses a simplified form of the sort specification (i.e., `asc` or `desc`). It provides a simple way of controlling the ordering for both the instantiations of the grouping variables and the aggregation result. For example, if the sorting specification is `asc` then first we will get the results that are grouped around the smallest instance of `Gs`, and those results themselves will be ordered in the ascending order. On backtracking, we get the results grouped around the next-smallest instantiation of `Gs` and, again, those results will be ordered in the ascending order.

The third form of the above `setof/bagof` aggregates provides the most general means of con-

trolling the ordering for the case when grouping is used. Here *SortSpec1* is used to control the results and *SortSpec2* controls the grouping order. That is, the instances of `Gs` are first ordered according to *SortSpec2*, and then the results associated with the first grouping in the `SortSpec2`-ordering are returned. These results are ordered according to *SortSpec1*. Then the next grouping in the *SortSpec2*-ordering is chosen and its associated results are returned according to *SortSpec1*-ordering, etc. Both *SortSpec1* and *SortSpec2* can be the simple sort specs `asc/desc` or they can be lists of the fine-grained specs of the form `asc(`*i*`)/desc(`*i*`)`.

Example:

```
q(a,a,r(z,1)). q(a,a,r(g,2)). q(b,b,r(b,2)). q(b,b,r(d,3)).
q(b,b,r(d,3)). q(b,c,r(d,3)). q(b,c,r(e,1)). q(e,e,r(k,3)).
q(e,e,s(k,4)). q(e,e,s(p,5)). q(e,e,s(p,5)). q(e,e,s(p,5)).

?- ?Y = setof{?Z[?X,?Q]([asc(1),desc(2),desc(3)],[desc(1),asc(2)])|q(?X,?Q,?Z)}.
```

will produce this result:

```
?Y = [r(d,3), r(b,2)]
?X = b
?Q = b

?Y = [r(e,1), r(d,3)]
?X = b
?Q = c

?Y = [r(k,3), s(p,5), s(k,4)]
?X = e
?Q = e

?Y = [r(z,1), r(g,2)]
?X = a
?Q = a
```

28.5 Aggregation and Set-Valued Methods

Aggregation is often used in conjunction with set-valued methods, and $\mathcal{E}$RGO provides several shortcuts to facilitate this use. In particular, the operator `->->` collects all the values of the given method for a given object in a set. The semantics of these operators can be expressed by the following rules:

```
?O[?M->->?L] :- ?L=setof{?V|?O[?M->?V]}
```

```
?O[|?M->->?L|] :- ?L=setof{?V|?O[|?M->?V|]}
```

Note that `?L` in `?O[?M->->?L]` and `?O[|?M->->?L|]` is a list of oids.

The special meaning for `->->` implies that this construct *cannot* appear in the head of a rule. One other caveat: recursion through aggregation is not supported and can produce incorrect results.

Sets collected in the above manner often need to be compared to other sets. For this, $\mathcal{E}$RGO provides another the primitive `+>>`. A statement of the form `o[m+>>s]` is true if the list of all values of the attribute `m` for object `o` *contains* every element in the list `s`. A statement of the form `o[|m+>>s|]` is true if `s` is the list of all values of `m` on the object `o`, which are obtained by inheritance from the superclasses of `o`.

For instance, the following query tests whether all Mary's children are also John's children:

```
?- Mary[children->->?L], John[children+>>?L].
```

As with `->->`, the use of `+>>` is restricted to rule bodies.

29 Control Flow Statements

$\mathcal{E}$RGO supports a number of control statements that are commonly used in procedural languages. These include `if - then - else` and a number of looping constructs.

29.1 If-Then-Else

This is the usual conditional control flow construct supported by most programming languages. For instance,

```
?- \if (foo(a),foo2(b)) \then (abc(?X),cde(?Y)) \else (qpr(?X),rts(??Y)).
```

Here the system first evaluates `foo(a),foo2(b)` and, if true, evaluates `abc(?X),cde(?Y)`. Otherwise, it evaluates `qpr(?X),rts(?Y)`. Note that `\if`, `\then`, and `\else` bind stronger than the conjunction ",", the disjunction ";", etc. This is why the parentheses are needed in the above example.

The abbreviated `if-then` construct is also supported. However, it should be mentioned that $\mathcal{E}$RGO gives a different semantics to `if-then` than Prolog does. In Prolog,

```
..., (Cond -> Action), Statement, ...
```

fails if `Cond` fails and `Statement` is not executed. If the knowledge engineer wants such a conditional to succeed even if `Cond` fails, then `(Cond->Action; \true)` must be used. Our experience shows, however, that it is the latter form that is used in most cases in Prolog programming, so in $\mathcal{E}$RGO the conditional

```
..., \if Cond \then Action, Statement, ...
```

succeeds even if `Cond` fails and `Statement` is executed next. To fail when `Cond` fails, one should explicitly use `else`: `\if Cond \then Action \else \false`. More precisely:

- `\if Cond \then Action` fails if and only if `Cond` succeeds but `Action` fails.
- `\if Cond \then Action \else Alternative` succeeds if and only if `Cond` and `Action` both succeed or `Cond` fails while `Alternative` succeeds.

The form `if-then` also has the following alternative forms: `~~>`and `<~~`.

Note that the `if`-statement is friendly to transactional updates in the sense that transactional updates executed as part of an `if`-statement would be undone on backtracking, unless the changes done by such updates are explicitly committed using the `commit` method of the system module `\db` (see Section 46.2).

29.2 Loops

unless-do. This construct is an abbreviation of `\if Cond \then \true else Action`. If `Cond` is true, it succeeds without executing the action. Otherwise, it executes `Action` and succeeds or fails depending on whether `Action` succeeds or fails.

29.2.1 The while-do and do-until Loops

These loops are similar in intent to those in C, Java, and other procedural languages. In `\while Condition \do Action`, `Condition` is evaluated before each iteration. If it is true, `Action` is executed. This statement succeeds even if `Condition` fails at the very beginning. The only case when this loop fails is when `Condition` succeeds, but `Action` fails (for all possible instantiations).

The loop `\do Action \until Condition` is similar, except that `Condition` is evaluated after each iteration. Thus, `Action` is guaranteed to execute at least once.

These loops work by backtracking through `Condition` and terminate when all ways to satisfy them have been exhausted (or when `Action` fails). The loop condition should *not* be modified inside the loop body. If it is modified (*e.g.*, new facts are inserted in a predicate that `Condition`

uses), it is not guaranteed that the changes will be seen during backtracking and thus the result of such a loop is indeterminate. If you need to modify `Condition`, use the statements `while-loop` and `loop-until` described below. Examples:

```
p({1,2,3}).
?- \while p(?X) \do writeln(?X)@\plg.
?- \do (p(?X),writeln('loop executed for'=?X)@\io) \until \naf p(?X).
```

For the second loop above, note that `?X` must be bound at the time the until-condition is checked. When negating conditions under `\until`, caution should be exercised. For instance, something like

```
?- \do writeln('loop executed')@\io \until \naf p(?X).
```

would not have worked: the loop would execute only once because `\naf p(?X)` will succeed on the first check (since `?X` is unbound, `\naf p(?X)` is undefined, and in loop conditions "undefined" is treated as a successful test).

The above loop statements have special semantics for transactional updates. Namely, changes done by these types of updates are *committed* at the end of each iteration. Thus, if `Condition` fails, the changes done by transactional updates that occur in `Cond` are undone. Likewise, if `Action` fails, backtracking occurs and the corresponding updates are undone. However, changes made by transactional update statements during the previous iteration remain committed. If the current iteration finishes then its changes will also remain committed regardless of what happens during the next iteration.

Subtleties related to the loop conditions. In both of the above loops, `Condition` should normally be user-defined backtrackable predicate. The use of non-backtrackable predicates as conditions requires special care, as described below.

First, not all non-backtrackable Prolog predicates fail when their work is done. For instance, `file_read_line_atom/2`, which reads files line-by-line, fails at the end, but `read/1` and `read/2` do not. Instead, they return the atom `end_of_file` when done. Therefore, `read(?X)@\prolog` or `read(?X)@\io` can never be a condition in the above loops; use `(read(?X)@\prolog, ?X \== end_of_file)` instead.

Second, non-backtrackable predicates, like `file_read_line_atom` and `read`, will fail if they are backtracked over. Therefore, if they are used as conditions in the `while-do` or `do-until` loops, they will execute only once. Fortunately, $\mathcal{E}$RGO provides a way to use such predicates in the above loops by wrapping them with the `\repeat` hint. For instance (where we use $\mathcal{E}$RGO's equivalents for `file_read_line_atom`, `open`, and `close`),

```
?- myfile[open(read,?Stream)]@\io,
    \while \repeat(?Stream[readline(atom,?Line)]@\io) \do writeln(?Line)@\io,
    ?Stream[close]@\io.
```

Third, for the `do-until` loop, the loop quits immediately after the condition becomes true. Therefore, if one wants to use non-backtrackable predicates like `file_read_line_atom` or `read` (or their $\mathcal{E}$RGO equivalents like `?Stream[readline(atom,?Line)]@\io`) then they must be negated. For instance,

```
q({1,2,3,4,5}).
?- myfile[open(read,?Stream)]@\io,
    \do (q(?X),writeln('loop executed for X'=?X)@\io)
    \until \repeat(\+ (?Stream[readline(atom,?Line)], writeln(?Line))@\io),
    ?Stream[close]@\io.
```

If `myfile` has 5 or more lines, this loop will execute 5 times and then fail. If `myfile` has less than 5 lines, the loop will execute once per line and then succeed.

Observe that `\naf` should *never* be used with non-logical conditions like the above, especially if these conditions are called with unbound variables.

29.2.2 The while-loop and loop-until Loops

This pair of loop statements is similar to `while-do` and `do-until`, except that transactional updates are *not* committed after each iteration. Thus, failure of a statement following such a loop can cause all changes made by the execution of the loop to be undone. In addition, `while-loop` and `loop-until` do not work through backtracking. Instead, they execute as long as `Condition` stays true in `while-loop` loops and until it becomes true in `loop-until` loops. Therefore, the intended use of these loop statements is that `Action` in the loop body must modify `Condition` and, eventually, make it false (for instance, by deleting objects or tuples from some predicates mentioned in `Condition`) or true, in case of `loop-until`.

As in the case of the previous two loops, `while-loop` and `loop-until` succeed even if `Condition` is false (`while-loop`) or is true (`loop-until`) right from the outset. The only case when these loops fail is when `Action` fails — see below for the ways to avoid this (i.e., to continue executing the loop even when `Action` fails) and the possible pitfalls.

The statements `while-loop` and `loop-until` are more expensive (both time- and space-wise) than `while-do` and `do-until`. Therefore, they should be used only when truly transactional updates are required. In particular, such loops are rarely used with non-transactional updates.

Portions Copyright © 2013–2017 Coherent Knowledge

Subtleties related to the while-loop and loop-until statements. Observe that `while-loop` and `loop-until` assume that the condition in the loop is being updated inside the loop body. Therefore, the condition must *not* contain tabled predicates. If such predicates are involved in the loop condition, the loop is likely to continue forever.

Also, keep in mind that in any of the four loop statements, if `Action` fails, the loops terminate and are evaluated to *false*. Therefore, if the intention is that the loop should continue even if `Action` fails, use the

```
(Action ; \true)
```

idiom in the loop body. In case of `while-do` and `do-until`, continuing execution of the loop is not a problem, because these loops work by backtracking through `Condition` and the loop will terminate when there are no more ways to backtrack. However, `while-loop` and `loop-until` have a potential pitfall. The problem is that these loops will continue as long as there is a way to satisfy `Condition`. If condition stays true, the loop runs forever. Therefore, as mentioned above, the `while-loop/loop-until` loops must make sure that `Condition` is modified by `Action`. Thus, if `Action` has non-transactional updates, the user must arrange that if `Action` fails then `Condition` is modified appropriately anyway for, otherwise the loop will never end. If `Action` is fully transactional and it fails, then using the `(Action ; \true)` idiom in the loop body will *definitely* make the loop not terminate, so the use of this idiom in the body of `while-loop` and `loop-until` is dangerous if there is a possibility that `Action` will fail, and this idiom is useless if the action is expected to always succeed.

30 Constraint Solving

$\mathcal{E}$RGO provides an interface to constraint solving capabilities of the underlying Prolog engine. Currently XSB supports linear constraint solving over the domain of real numbers (CLPR). To pass a constraint to a constraint solver in the body of a $\mathcal{E}$RGO rule (or query), simply include it inside curly braces.

Here is a 2-minute introduction to CLPR. Try the following:

```
?- insert{p(1),p(2),p(3)}.
?- ?X>1, ?X<5, p(?X).
```

Traditional logic languages, like Prolog, would give an error in response to this query. $\mathcal{E}$RGO is actually pretty good in this respect, as it will delay the inequalities until they can be solved. So, it will return two answers: 2 and 3. But what if we ask $\mathcal{E}$RGO to solve an equation:

```
?- insert{p(1),p(2),p(3),p(9)}.
?- ?X=?Y*?Z, ?Y>1, ?Z>2, p(?X), p(?Y).
```

That $\mathcal{E}$RGO cannot do without the help of *Constraint Logic Programming*. Constraint logic programming takes the view that `?X=?Y*?Z, ?Y>1, ?Z>2` is a a *constraint* on the set of solutions of the query `p(?X)`. This approach allows Prolog to return meaningful answers to the above query by solving this constraint. However, the user must explicitly tell $\mathcal{E}$RGO which view to take: the traditional view that treats arithmetic built-ins as infinite predicates or the one that treats them as constraints. This latter view is indicated by enclosing constraints in curly braces. Thus, the above program becomes:

```
?- insert{p(1),p(2),p(3),p(9)}.
?- {?X=?Y*?Z, ?Y>1, ?Z>2}, p(?X), p(?Y).

?X = 9
?Y = 2
?Z = 4.5000

?X = 9
?Y = 3
?Z = 3.0000
```

Note the use of curly braces in the above example: they are essential in order to tell the system that you want constraints to be *solved* rather than *checked*.

31 Low-level Predicates

Unadulterated Prolog predicates. Sometimes it is useful to define predicates that are handled *directly* by the underlying Prolog engine. Such predicates would be represented as Prolog, not HiLog predicates. They are visible to HiLog queries, are not tabled automatically, and they are indexed as any other Prolog predicate. One use of such predicates, as *sensors*, is described in Section 32. To support this feature, $\mathcal{E}$RGO provides two directives:

```
:- prolog{predname/arity, predname/arity, ...}.
:- table{predname/arity, predname/arity, ...}.
```

The first directive defines a predicate of a certain arity as a Prolog predicate to be handled directly by Prolog. If this predicate is defined recursively, it might need to be tabled to help with termination

or to reduce computational complexity. This is accomplished with the help of the second directive. The `table` directive implies the `prolog` directive, however, so there is no need for the former in that case.

As mentioned above, predicates declared with the `prolog` directive are not visible to HiLog queries. For instance, in

```
:- prolog{foo/2}.
?- ?X(?,?).
```

the variable `?X` will *not* be bound to `foo`.

The names of the predicates defined via the `prolog` directive are scrambled so there is no fear of a clash with Prolog builtins or `:- prolog{...}`-defined predicates in other $\mathcal{E}$RGO modules.

At present prolog predicates are automatically exported and there is no way to encapsulate them in a module. To refer to a prolog predicate defined in a different module, e.g., predicate `foo/1` in module `bar`, the following idiom can be used:

```
:- prolog{foo/1}.
... ... ...
... :- ..., foo(?X)@bar, ...
```

Note that there are no *executable* `prolog` or `table` directives (i.e., invocable via ?– or from the $\mathcal{E}$RGO shell).

The `prolog` and `table` directives in multi-file modules. Recall from Section 15.5 that a module can consist of several files: the first file would be normally loaded into the module and the subsequent files are added. Once a `prolog` or `table` directive is issued, it affects the compilation of the corresponding predicate symbols. The files added after the file that contains the initial `prolog/table` directives *must* include the necessary `prolog` directives, *if* that uses the affected predicate symbols. If this is not done, the system would issue an error, to prevent hard-to-find error from creeping into the knowledge base.[16] It should be noted, however, that if a file with `prolog` declarations (e.g., `:- prolog{p/1}` or `:- table{p/1}`) is loaded into a module, `foo`, that declaration is automatically made available to the $\mathcal{E}$RGO shell. So, such predicates are readily accessible from the shell (e.g., as `p(?X)@foo`).

The `\nontabled_module` directive. For simple modules that *do not have recursive rules*, significant amount of main memory can be saved by telling $\mathcal{E}$RGO to not table predicates and methods.

[16] Note: subsequently added files need only the `prolog` directives—even for predicates that are declared with a `table` directive.

This is especially useful for modules that process large amounts of data and can be done by placing the following directive in the module file:

```
:- \nontabled_module.
```

32 Sensors: Predicates with Restricted Invocation Patterns

Sometimes it is useful to define predicates that have fixed invocation patterns: the requirements that certain arguments must be bound (non-variable) or be ground. $\mathcal{E}$RGO provides special support for this kind of predicates, which are called ***sensors***. Namely, if a predicate is registered as a sensor, $\mathcal{E}$RGO will monitor the binding pattern of that predicate and, if the predicate is called before the binding conditions are fulfilled, it will delay the predicate until the conditions are fulfilled. If at some point $\mathcal{E}$RGO determines that the binding condition *cannot* be satisfied, $\mathcal{E}$RGO will call the sensor anyway. The sensor's implementation can then examine the state of the argument bindings and issue an error, if appropriate.

The overall scenario for sensor use is as follows:

- Sensors are *used* and *defined* in separate files. Sensors must also be *declared*. Declaration is done via the `defsensor` directive; definitions are done by means of the regular rules.
- A sensors can be defined in a `.P` file using the Prolog syntax (P-sensors) or in `.ergo` files using the $\mathcal{E}$RGO syntax (F-sensors). P-sensors can also be defined as external modules in the C language. Usually all sensor declarations are collected in one `.ergo` file, which should be loaded into a separate $\mathcal{E}$RGO module. That `.ergo` file should also contain the rules that define F-sensors.
- $\mathcal{E}$RGO files containing sensor definitions *cannot* be added to another $\mathcal{E}$RGO module (e.g., loaded using the `add{...}` command).
- There are no restrictions on how sensors are to be defined. For instance, they can be recursive. However, there are certain conventions to abide by for P-sensors.
- Sensors are *used* in modules *other* than those where they are defined. To *use* a sensor in a file, it has to be declared in that file with the `usesensor` directive.

  However, if a file with a `usesensor` declaration is loaded into some module, all files that are *compiled* in the *same* $\mathcal{E}$RGO session and then *added* into the same module will inherit that declaration. In that case, no explicit `usesensor` declaration is needed. However, if the file being added is compiled in a different $\mathcal{E}$RGO session (which does not have the requisite `usesensor` declarations loaded) then explicit `usesensor` declarations are required.

- $\mathcal{E}$RGO also provides executable versions of the directives `defsensor` and `usesensor`.

In most cases, the developer would choose F-sensors. There are two slight advantages in choosing to define a sensor as a P-sensor, however: the encapsulation provided by Prolog modules or if the sensor is implemented completely in C.

Declaring a sensor. There are two forms of the `defsensor` directive. The first is used for P-sensors (defined in `.P` files in Prolog or in `.c` files in C); the second is used for F-sensors (defined using $\mathcal{E}$RGO syntax in `.ergo` files).

```
:- defsensor{sens1(?Y,?X), sensorfoo, (nonvar(?X),ground(?Y))}. // P-sensor
:- defsensor{sens2(?Y,?X), (ground(?X),ground(?Y))}.            // F-sensor
```

The first argument is a sensor invocation template. The last argument is a *guard*: a Boolean combination of `nonvar/1` and `ground/1` predicates applied to the input variables of the sensor. The sensor will be *delayed* by the $\mathcal{E}$RGO engine until the guard is satisfied (or until the engine determines that the guard cannot be satisfied).

In the first `defsensor` directive, which applies only to P-sensors, the middle argument is the name of the Prolog module in which the P-sensor is defined. In our case, `sens1` is declared as a P-sensor in the Prolog module `sensorfoo`. In that case, $\mathcal{E}$RGO will expect those rules to be in a file named `sensorfoo.P` (or `sensorfoo.c`, if the sensor is defined as an external C module) and that file should be found somewhere on the $\mathcal{E}$RGO search path (e.g., in the current directory).

Defining an F-sensor. The F-sensor `sens2` above is declared using a two-argument `defsensor` directive, so its definition is expected to be in a `.ergo` file and then loaded into a separate $\mathcal{E}$RGO module. Here is an example of such a definition:

```
sens2(abc,cde) :- !, writeln(details=\?F+\?L)@\io.
sens2(?X,?Y) :-
     \if \+ground(?X) // only ?X must be ground
     \then
        abort(['In file ', \?F, ', line ', \?L, ': ',
                'Instantiation error in arg 3 in sens2'])@\sys,
     ?Y \is ?X/2,
     writeln(Y=?Y)@\prolog,
     sens2(abc,cde).
```

The above rules are quite standard except for the quasi-variables `\?F` and `\?L`. When these quasi-variables occur in the body of a sensor definition, `\?F` is replaced with the file name from which the

sensor was called; the quasi-variable `\?L` is replaced with the line number in the file where the call to the sensor occurred. A typical use of these quasi-variables is to report errors or issue warnings in case the sensor guard is not satisfied at the time the sensor is called, as illustrated in the above example. If these quasi-variables are used outside of a sensor definition, they are treated as new unbound variables.

Defining an P-sensor. From the perspective of the *user*, the above P-sensor `sens1` is a binary predicate. However, from the perspective of the *developer*, a P-sensor has two "hidden" extra arguments. These arguments are automatically pre-pended by the compiler to the list of the arguments that the user specified in a `defsensor` directive.

In any actual call (in a .ergo file) to a P-sensor, the first argument will *always* be bound (at compile time) to the file in which the sensor is used and the second argument will be bound to the line number on which the call to the sensor occurs. This is done to make it possible to define sensors so that they would issue useful runtime errors to enable the user to quickly locate the offending call. Thus, in the actual rules that define a P-sensor, the first two arguments to the sensor must be distinct variables, and this is the responsibility of the P-sensor developer. The remaining arguments are up to the developer to choose.

The upshot of the above is that to define `sens1` one uses a predicate of arity 4, not 2. Here is an example of a definition for `sens1` in a Prolog file `sensorfoo.P`. As we have just explained, the actual predicate to be defined here must be `sens1` and its first two arguments must be distinct variables reserved for file names and line numbers (Prolog variables `F` and `L`).

```
:- export sens1/4.
sens1(F,L,X,Y) :-
     (var(X)       // X must be nonvar
     ->  abort(['In file ', F, ', line ', L, ': ',
                'Instantiation error in arg 1 in in sens1/2'])
     ; \+ground(Y) // Y must be ground
     ->  abort(['In file ', F, ', line ', L, ': ',
                 'Instantiation error in arg 2 in sens1/2'])
     ),
     Z \is X+Y,
     writeln(z=Z).
```

It should be noted that F-sensors are also internally represented by predicates that have two extra arguments. However, the transformations that add these arguments are done by the $\mathcal{E}$RGO compiler completely transparently to both the users and the developers of F-sensors.

Using a sensor. When sensors are used in $\mathcal{E}$RGO modules (which are different from the modules where the sensors are defined), they must be declared using the `usesensor` directive. For instance,

```
:- usesensor{sens1/2, sens2/2}.
?- sens1(?X,?Y), ?X=123, ?Y=aaaaa, sens2(?X,?Y), ?Y=345, ?X=bbbbb.
```

Note that if the definition of a sensor is recursive, tabling might be needed in order to ensure termination. In a `.P` file, this is done using the usual Prolog `table` directive. $\mathcal{E}$RGO provides a similar directive for sensors defined in `.ergo` files:

```
:- table{sens2/2}.
```

This directive is to be placed in the file that contains the rules *defining* the sensor and the number of arguments must match the number used in the F-sensor declaration. It must come *after* the `defsensor` directive. For P-sensors, the arity must match the number of arguments used in the sensor definition, i.e.,

```
:- table{sens1/4}.
```

Sensors are viewed as low-level non-logical predicates, analogous to builtins, so they are not visible to HiLog queries. For instance, neither

```
?- ?X(?,?,?,?).
```

nor

```
?- ?X(?,?).
```

would bind `?X` to `sens1` or `sens2` (cf. Section 31).

33 Rearranging the Order of Subgoals at Run Time

Sometimes certain subgoals should be executed only if certain arguments are bound or ground and should be delayed otherwise. The reason for this might be correctness or efficiency of the execution. Since it may be impossible to know at compile time when the requisite arguments become bound, one might want to place the relevant subgoals as early as possible to the left and then wait until the binding conditions become true. Once they become true, the affected subgoals can be executed. It the binding condition never becomes true, then we can have two possible actions: abort or execute the subgoals anyway.

$\mathcal{E}$RGO supports this mode of goal rearranging via *delay quantifiers*. A delay quantifier has one of these forms:

`must`(*Condition*)
`wish`(*Condition*)

where *Condition* consists of the terms of the form `ground`(? *Var*) and `ground`(? *Var*) connected with "," (or `and`) and ";" (or `or`). If parentheses are not used in *Condition*, then the comma is considered to bind stronger than the semicolon.

Delay quantifiers are connected via the operator ^ to the actual goal that is to be executed with possible delay. The goal can be a simple frame or a predicate, or it can be a more complex goal of any kind that is allowed to appear in the rule bodies. In the latter case, the goal has to be enclosed inside parentheses. A subgoal with an attached delay quantifier is called a *delayable subgoal* and the regular subgoal part in a delayable subgoal is said to be *controlled* by the quantifiers. So, the general syntax is

delay-quantifier^*delay-quantifier*^...^*Goal*

Delayable subgoals can be nested and multiple delay quantifiers can be attached to the same goal. For instance,

```
must(ground(?X) or nonvar(?Y))^?X[foo->?Y]
wish(ground(?X) and ground(?Y) ; nonvar(?Z))^(p(?X), foo[bar(?Y)->?Z])
wish(ground(?X) or ground(?Y))^must(nonvar(?Z))^
           (p(?X), wish(nonvar(?W))^q(?W), foo[bar(?Y)->?Z])
```

When a delayable subgoal is to be executed, the attached delay quantifiers are checked. If at least one the quantifiers is not satisfied (i.e., if its condition is not satisfied) then the goal is *not* executed, but is delayed instead until such time that all the quantifiers are satisfied or the engine determines that satisfying all the quantifiers is impossible. In the above example, the first goal would be delayed is `?X` is non-ground and `?Y` is an unbound variable. The second delayable subgoal has a more complex execution condition and it controls a more complex subgoal, the conjunction of `p(?X)` and `foo[bar(?Y)->?Z]`. This subgoal will be delayed if either `?X` or `?Y` is non-ground and `?Z` is an unbound variable. The third delayable subgoal is even more complex. First, it involves two delay quantifiers; second, it controls a complex subgoal that is a conjunction of three other subgoals, and one of them is itself a delayable subgoal. This complex delayable subgoal will be executed immediately only if `?X` or `?Y` are ground and `?Z` is not a variable. Otherwise, it will be delayed. But when the controlled subgoal is eventually ready for execution, its middle part, `wish(nonvar(?W))^q(?W)`, might still be delayed if `?W` is an unbound variable.

When a delayable subgoal is delayed, its attached quantifiers are periodically checked for satisfaction. If the first (leftmost) quantifier is satisfied, the controlled subgoal is executed. If this subgoal still has a controlling delay quantifier, then this quantifier's condition will be checked, etc.

At some point during the execution, the inference engine might stumble upon a delayed subgoal and *determine* (provably!) that its first delay quantifier is *not satisfiable* any more, i.e., subsequent execution will not be able to bind the variables in that quantifier in the required way. In this case, the cause of action depends on the type of the quantifier. If the unsatisfied quantifier is a `wish`-quantifier, then the controlled subgoal is executed anyway: the unsatisfied delay condition was only a "wish." If, however, the unsatisfied quantifier is a `must`-quantifier, the execution of the subgoal is aborted and the error message will indicate which `must`-condition was at fault. To illustrate this on an example, consider the following delayable subgoal issued as a query against a knowledge base that contains the listed facts:

```
p(1).
foo[bar(2)->3].
?- must(ground(?X) or ground(?Y))^wish(nonvar(?Z))^
           (p(?X), must(nonvar(?W))^q(?W), foo[bar(?Y)->?Z]),
   ?X=1.
```

When the subgoal is first encountered, it is delayed because `?X` and `?Y` are not ground. Then the engine executes the subgoal `?X=1` and `?X` gets bound to 1. The delay condition becomes satisfied, so the engine considers the subgoal

```
wish(nonvar(?Z))^(p(?X), must(nonvar(?W))^q(?W), foo[bar(?Y)->?Z])
```

Clearly, `?Z` will not be bound if the engine continues delaying the controlled subgoal, so it must decide on the final disposition for this delayable subgoal. Since the delayable subgoal is controlled by a `wish`-quantifier, the subgoal is executed despite the fact that the delay condition does not hold. During the execution, `p(1)` is satisfied, `must(nonvar(?W))^q(?W)` gets delayed, and then `foo[bar(?Y)->?Z]` is satisfied. Now the disposition of the remaining delayed subgoal has to be decided. Since `?W` cannot be bound any more and the controlling quantifier is of the `must`-variety, the engine will throw an error:

```
++Abort[Ergo]> in file foo.ergo, line 3: unsatisfied must-condition.
Goal: q(?A)
Condition: nonvar(?A)
```

Forcing immediate execution of delayed subgoals. Sometimes it is necessary to tell the system to stop delaying subgoals and force their immediate execution. For instance, in

```
?- must(ground(?X) or ground(?Y))^wish(nonvar(?Z))^
```

```
                    (p(?X), wish(nonvar(?W))^q(?W), foo[bar(?Y)->?Z]),
    ...,
    ...,
    test(?X,?Y,?Z,?W).
```

the query `test` might be executed before the queries `p(?X)`, `q(?W)`, and `foo[bar(?Y)->?Z])`, *if* some of the delay quantifiers remain unsatisfied. If instead we want to make sure that those queries are executed before `test(?X,?Y,?Z,?W)`, $\mathcal{E}$RGO provides the *immediate execution operator*, `!!`, which forces immediate execution of all the "reachable" delayed subgoals:

```
?- must(ground(?X) or ground(?Y))^wish(nonvar(?Z))^
                    (p(?X), wish(nonvar(?W))^q(?W), foo[bar(?Y)->?Z]),
    ...,
    ...,
    !!,
    test(?X,?Y,?Z,?W).
```

In this example, the aforesaid queries will be executed before `test(?X,?Y,?Z,?W)`.

A delayed subgoal, G, is *reachable* by an occurrence of the immediate execution operator `!!` if G is suspended on a ground/nonvar condition that involves a variable that appears in the the scope of the rule or the query where the aforesaid occurrence of `!!` is found. What it means to be suspended on a variable is a little trickier. In the above example, if neither `?X` nor `?Y` is ground, then `p(?X)`, `q(?W)`, and `foo[bar(?Y)->?Z])` would be suspended on either `?X` or `?Y`. If one of these variables becomes ground, then these queries would be suspended on `?Z`, if `?Z` is still not bound. Once `?Z` gets bound, then `p(?X)` and `foo[bar(?Y)->?Z])` will no longer be suspended, but `q(?W)` will still be suspended on `?W` unless `?W` is already bound by that time.

Note, however, that `!!` will not have any effect on subgoals that are suspended on variables that are not in the scope of the rule or query containing this immediate execution operator. For instance, in

```
p(?X) :- must(nonvar(?X))^foo(?X,?Y), wish(ground(?Z))^bar(?Z), moo(?Y,?X,?Z).
moo(?Y,?X,?Z) :- ..., !!, ...
?- p(?W), !!, writeln(?W)@\plg.
```

the variable `?W` is in scope for the query. Since `p(?W)` unifies with the head of the first rule, `?W` and `?X` become the same variable. This means that `foo(?X,?Y)` is reachable from `!!` and will be forced to execute by that operator. In contrast, `bar(?Z)` is suspended on the variable `?Z`, which is not in the scope of the query. Therefore, `bar(?Z)` will not be forced to execute by the operator !! in the above query. On the other hand, `?Z` is in the scope of the second rule above, so the occurrence of `!!` in *that* rule will force the execution of `bar(?Z)`.

34 Rule Ids and Meta-information about Rules

Every rule in $\mathcal{E}$RGO—whether it appears in a file, is inserted at run time, or is reified and lives as an object it its own right—is assigned a unique *rule Id* and the user can also supply additional meta-data for it. The rule Id can be either given explicitly by the knowledge engineer or, if not given explicitly, is assigned by the system (at compile time for rules that live in files and at run time for inserted and reified rules).

A rule Id is a triple of the form
$(local_id, file_name, module)$, where

- *local_id* is a term that is either explicitly given to the rule by the author or is generated by the compiler or loader.
- *file_name* is the file name where the rule occurs. The name does not include the directory part, but it does include the name extensions (e.g., `foo.ergo`). The `#include` commands are taken into account, so if a rule is found in file `foo` that is `#include`d in file `bar` then the rule Id will use `foo`, not `bar`.
- *module* is the module into which the rule is loaded.

For rules that are inserted dynamically at run time via an `insert{...}` or `insertrule{...}` statement, their file name is to be specified as `dynrule(`*containing-file*`)`, where *containing-file* is the file that contains the insert statement in question.

$\mathcal{E}$RGO supports the use of rule *descriptors* of the form

- *Rule Id descriptor*: `@!{id[frame]}` or just `@!{id}`,
  where `id` is a term and the frame information is the same as what is allowed in a simple (not nested) frame. Some properties in the frame have special meaning as detailed below.

  The rule Id descriptor is used to explicitly specify *only* the *local id* part of the rule Id, as described above. The other two components are assigned by the system. If no explicit rule Id is given, the system will generate one.
- `Tag descriptor`: The property `tag` indicates that the values of that property are rule tags, which are used in defeasible reasoning (Section 38). For instance,

  ```
  @!{abc[tag->{tag1,tag2},author->'kifer@cs.stonybrook.edu']} head :- body.
  ```

  Here the rule is given an explicit rule Id, `abc`, and two tags, the terms `tag1` and `tag2`. The property `author` (as most others) is not special; it can be used for whatever purpose the user

decides to use it.
Because rule tags are very common pieces of metadata, $\mathcal{E}$RGO provides a convenient shortcut, which is used as a standalone descriptor. For instance, the following is equivalent to the previous descriptor:

```
@{tag1,tag2} @!{abc[author->'kifer@cs.stonybrook.edu']} head :- body.
```

- *Defeasibility descriptor*: Defeasibility descriptors are used for defeasible reasoning, as described in Section 38. The descriptor `strict` means that the rule cannot be defeated by a higher-priority rule, and `defeasible` means it can be. These are Boolean properties in the descriptor frame and, as one may guess, only one of them can be specified for any given rule. For instance,

```
@{tag1,tag2}
@!{abc[author->'kifer@cs.stonybrook.edu', defeasible]} head :- body.
```

Defeasibility descriptors are also fairly common and $\mathcal{E}$RGO provides shortcuts for them as well. As with tags, these shortcuts are specified as standalone descriptors. For instance,

```
@{tag1,tag2}
@@{defeasible}
@!{abc[author->'kifer@cs.stonybrook.edu']}
head :- body.
```

Of course, we did not save any typing with the above defeasibility shortcut. However, this shortcut does improve the visibility of the fact that a particular rule is defeasible (or strict). Also, when no explicit Id and frame properties are specified and the user opts for the default rule Id, the defeasibility and the tag shortcuts save significant amount of typing. For instance,

```
@{tag1,tag2} @@{defeasible} head :- body.
```

Without these shortcuts, the user would have to use the *current rule Id quasi-constant*, `\@!`, which gets substituted for the Id of the rule in which it occurs:

```
@!{\@![tag->{tag1,tag2}, defeasible]} head :- body.
```

In this case, `\@!` gets replaced by the default Id, which the system assigns to this rule.

The descriptors extend the rule syntax as follows:

```
rule-descriptors Head :- Body.
rule-descriptors Fact.
```

Any number of tag descriptors is allowed but only one rule Id and one defeasibility descriptor. In other places, rule descriptors are either ignored or will cause syntax errors.

Rule tags and Ids are terms and so can have variables. For instance,

```
@{allow(?Device,?Person)}  authorized(?Person,?Device) :- \naf abused(?Person,?Device).
```

If no Id is given, a system-generated rule Id is used. If no tag is given, the Id of the rule also serves as a tag.

If no defeasibility descriptor is given then a rule is considered defeasible by default if an explicit tag is specified. Otherwise, the rule is considered strict. However, for defeasibility to have any effect, the rule must be in a *defeasibility-capable* document (local file or a remote document), i.e., the document must have the directive `:- use_argumentation_theory` at the top (see Section 38). Also, rules whose heads are transactional predicates or methods are currently never defeasible—no matter what the defeasibility descriptor says.

34.1 The Current Rule Id Quasi-constant

We have already seen a use of this quasi-constant earlier. In general, rules can reference their Ids even if the Id is not given explicitly by the author and thus cannot be determined by just looking at the rule. This is accomplished with the help of the special "current rule Id" quasi-constant `\@!`. This quasi-constant gets replaced by the *local part* of the Id of the rule where the quasi-constant occurs. The replacement happens at compile time for the static rules and at run time for dynamic or reified rules. To get the file name part, use `\@F` and the module part use `\@` — the quasi-constants introduced in Section 7.6.

34.2 Enabling and Disabling Rules

Rule Ids make it possible to *enable* and *disable* rules, producing the effect of deleting and re-inserting rules. The difference is, however, that enabling and disabling do not actually change the rules in the system (so they are very fast compared to `insert` and `delete`) and, most importantly, they apply to static rules, not just the dynamic rules. To specify a rule, one needs to specify the local Id of the rule in question, the file name of the rule, and its module. For dynamic rules, the file name is `dynrule(`*containing-file*`)`, where *containing-file* is the file that contains the `insert...` or `insertrule...` statement that inserted the dynamic rule. The following primitives are provided:

- `enable{`*LocalId,File,Module*`}`
- `disable{`*LocalId,File,Module*`}`
- `tenable{`*LocalId,File,Module*`}`
- `tdisable{`*LocalId,File,Module*`}`
- `isenabled{`*LocalId,File,Module*`}`
- `isdisabled{`*LocalId,File,Module*`}`

In the `enable`/`tenable` primitives, all arguments must be bound. The first two primitives are non-transactional, i.e., these operations will not be undone if the query in which they occur fails. The second two primitives, `tenable` and `tdisable`, *are* transactional. Their effects are undone on backtracking. For instance,

```
bbbb.
@!{r1} aaaa:-bbbb.
?- tdisable{r1,\@F,\@}, aaaa.
No
?- aaaa.
Yes
```

Here the first query disables the rule `r1` so `aaaa` becomes false and thus the entire query fails. Due to that, the disabling action is undone, so the rule remains enabled. Therefore, the second query succeeds.

The last two primitives are queries that tell us whether a particular rule is enabled or disabled. To enable or disable a rule, the rule must already exist as a static or dynamic rule. Enabling and disabling operations are idempotent: they always succeed and applying `enable` to an enabled rule or `disable` to a disabled rule has no effect. If a rule with the specified Id and module does not exist in the system, both `enable` and disable fail and so do both of the above queries.

All of the above primitives have a one-argument form, which is a shortcut for the three-argument version, where argument 2 is `\@F` and argument 3 is `\@`. For instance, `tenable{foo}` instead of `tenable{foo,\@F,\@}`.

The difference between disabling and deleting a rule. It is important to realize that disabling a rule, as described here, and deleting a rule, described in Section 26.3, are two very different operations, although there are similarities. For similarities, both operations take the affected rule(s) out of the reasoning process and the effect is as if the rule is "not there." The major differences are:

1. Rule deletion works only for dynamic rules (those added with `insertrule`), while disabling works for all rules.

2. A deleted rule is physically removed from the system. A disabled rule is only marked as disabled. It can be re-enabled by supplying its Id and module as arguments to the `enable` primitive. In contrast, to re-instate a deleted rule, it must be re-inserted, which requires supplying the entire original rule as an argument. Disabling/enabling are very fast operations, while deleting/inserting a rule is much more expensive.

3. For *multi-headed* rules, there is another important difference. A multi-headed rule is one that has more than one literal in the head, which is equivalent to several single-headed rules. Such rules typically arise when a complex frame is put in a rule head. With a *deleterule* operation, one can delete just one of the heads, leaving the remainder of the rule intact. For instance, in

   ```
   ?- insertrule{?X:person[name->?N] :- pred(?X,?N)}.
   ?- deleterule{?X:person :- pred(?X,?N)}.
   ```

   the `person[name->?N] :- pred(?X,?N)` part of the rule remains. In contrast, in

   ```
   ?- insertrule{@!{abc} ?X:person[name->?N] :- pred(?X,?N)}.
   ?- disable{abc,\@}.
   ```

   the `disable`-primitive takes all parts of the rule out of circulation, while a subsequent matching `enable`-operation would reinstate the entire rule.

34.3 Changing the Defeasibility at Run Time

The defeasibility status of a rule can be changed via the following statements:

- makedefeasible{Id,File,Module}
- makestrict{Id,File,Module}
- isdefeasible{Id,File,Module}
- isstrict{Id,File,Module}

The last two primitives are queries. The primitive `makedefeasible` and `makestrict` are idempotent and always succeed. The queries `isdefeasible` and `isstrict` tell us whether the rule in question is currently defeasible or strict. If the rule with the given Id does not exist in the given module, both of these queries fail.

Portions Copyright © 2013–2017 Coherent Knowledge

All of these primitives have a short, one-argument form, which is a shortcut for the three-argument version, where argument 2 is `\@F` and argument 3 is `\@`. For instance, `makedefeasible{foo}` instead of `makedefeasible{foo,\@F,\@}`.

34.4 Querying Rule Descriptors

Rule Ids and other rule descriptor data can be queried using the `clause{...}` and `@!{...}` constructs. The `clause{...}` construct was already introduced in Section 27. Now we introduce two new forms of this statement, which extend the statements defined in Section Section 27:

`clause{`*descriptors head,body*`}`
`clause{`*descriptors type,head,body*`}`

Here descriptors are rule descriptors as above and *type*, *head*, and *body* are as in Section 27. For instance,

```
?- clause{@{?Tag} @@{defeasible} foo[bar->?Val],?Body}.
```

This query will retrieve all defeasible rules whose head matches `foo[bar->?Val]`. The variable `?Tag` will be bound to the tag of the rule.

Note: The `clause{...}` primitive queries the rules as they were created by the author. Disabling a rule or changing its defeasibility status won't affect `clause...`. For instance, if we had a rule `@!{abc} @{defeasible} p:- q` and then this rule was disabled and/or made strict, `clause{@@{defeasible} p,q}` will still be true and `clause{@@{strict} p,q}` false. In this way, one can investigate the changes made to the status of the various rules in the course of time. □

The `@!{...}` construct is more convenient than the `clause`-construct and much more efficient, if one needs to query the *descriptor data only*, not the rule base. Inside the braces, this construct expects a usual frame formula of the form `ruleId[`$prop_1$`,`$prop_2$`,...]`, where $prop_i$ have the form `attr->val` or `term`. In the first case, the property is a regular attribute-value pair and in the second it is a Boolean property. Nested or composite frames are not allowed, but module specifications are supported, so one can query rule descriptors in another module. For instance,

```
?- @!{?R[foo->bar, tag->low, defeasible]}.
?- @!{rule123[]@foo}.
?- @!{?R[tag->high,strict,author->Bob]@foo}.
```

Recall that in all the above examples, only the local part of the rule Ids was queried (cf. `?R` in the queries above). To enable querying the file name and the module components of rule Ids, $\mathcal{E}$RGO also provides the `file` and `module` properties. For instance,

```
?- @!{?R[foo->bar, tag->low, defeasible, file->?F, module->?M]}.
?- @!{rule123[file->?F, module->?M]@foo}.
?- @!{?R[tag->high,strict,author->Bob, file->?F, module->?M]@foo}.
```

In addition, the `type` property can be used to find out whether a particular annotated statement is a rule or a latent query (see Section 35):

```
?- @!{statement43[type->?T, module->main]}.
```

If the statement is a rule then `?T` will get bound to `rule`; if it is a latent query then this variable gets bound to `query`.

34.5 Reserved Properties in Rule Descriptors

The following property names in rule descriptors or in rule descriptor queries (described in Section 34.4) are reserved:

- `tag`
- `module`
- `file`
- `type`
- `defeasible`
- `strict`

The properties `module`, `file`, and `type` are not allowed in rule descriptors (i.e., in `@!{...}` in rule heads)—only in rule descriptor queries (i.e., in `@!{...}` and `clause{...}` in rule bodies)—because they are assigned at compile or load time.

35 Latent Queries

Latent queries are a cross between queries and rules. From queries they borrow the main purpose: to query the knowledge base. From rules they borrow the ability to use descriptors (so, latent queries have Ids and, possibly, other meta-data) and the fact that such queries are not executed immediately. Instead, they are saved to be called on-demand some other time. Like rules, latent queries can also be inserted, deleted, enabled, disabled, and queries.

Latent queries are useful as integrity constraints and also in GUI development, but because these queries can have arbitrary properties attached to them via descriptors, they can be used for various other, specialized needs (e.g., as standing queries that are called periodically to update user views). At the same time, latent queries is an older mechanism, which is primarily retained for $\mathcal{F}$LORA-*2* and $\mathcal{E}$RGO^{Lite} users. $\mathcal{E}$RGO has newer, much more powerful mechanisms of *integrity constraints* and *alerts*.

A latent query has the following form:

descriptor `!-` *query-body*`.`

The descriptor part is mandatory and has the same form as the rule descriptors introduced in Section 34. The *query-body* part is the same as the body of a query. Thus, from regular queries the latent queries differ in that they have descriptors and they use `!-` instead of `?-`. From rules, this new type of queries differs in that they have no heads and use `!-` instead of `:-`.

The main difference with queries, however, is the fact that latent queries, when they appear in a file, are not executed right away. Instead, they are saved and must be called explicitly by their query Id using the `query{...}` primitive. In order to get results from such a query, the query Id must have variables. For instance,

```
p(1,2), p(2,3), p(3,4), p(2,4).
@!{test(?X,?Y)} !- p(?X,?Z), p(?Z,?Y).
?- query{test(?X,4),\@F,\@}.

?X = 1

?X = 2
```

(The quasi-constants `\@F`, `\@`, and others were introduced in Section 7.6.) Observe that latent queries are executed using their Ids. This is different from rule invocation, which is done using the rule-head as a query. In the above, `test(?X,4)` is *not* a predicate and it does not interfere with any other rule (even with the ones that have `test(?X,?Y)` as head- or body-predicates!).

The following two forms are equivalent:

```
?- query{test(?X,4),\@F,\@}.
?- query{test(?X,4)}.
```

However, one must keep in mind that `\@F` occurring in the $\mathcal{E}$RGO shell is different from `\@F` occurring in a file (Section 7.6) and that the file component of the rule Id of a dynamically inserted rule has its own representation (Section 34).

Latent queries can be inserted, deleted, enabled, and disabled using the same primitives as the ones used for rules:

```
?- insertrule@!rr2(?X)[bar->4] !- q(?X), ?X<4.
?- deleterule@!rr2(?)[bar->?X] !- ?.
?- enableqq1(?),'foobar.ergo',.
?- tenableqq1(?),?F,?M.
?- disableqq2(?),F,.
?- tdisableqq2(?).
```

Finally, like rules, latent queries can be themselves queried using the primitive `clause...` and the meta-query @!.... In the clause... primitive, the head part is ignored.

```
?- clause{@!{?X[type->query]} ?,?B}.

?X = rr1(?_h4115)
?B = (${p(?_h4115)@main}, ?_h4115 < 3)

?X = rr2(?_h4156)
?B = (${q(?_h4156)@main}, ?_h4156 < 4)

ergo>  @!{?X[type->query]}.

?X = rr1(?_h3033)

?X = rr2(?_h3046)
```

Note the use of the `type` property in the descriptors of the above two queries. This is a special builtin descriptor property, which can be used only in rule bodies. The value of that property is `query` for latent queries and `rule` for rules.

36 Integrity Constraints

Database systems have a very useful notion, called *integrity constraint*. An integrity constraint is a logical statement that specifies a condition that must be maintained at all times regardless of the changes made to the knowledge base. We have seen two examples of integrity constraints in $\mathcal{E}$RGO: type constraints and cardinality constraints. These particular constraints are dealt with in greater detail in Sections 43.2 and 43.3, but those sections deal with *type-checking*, i.e., with verifying that the knowledge base satisfies typing and cardinality constraints and with finding out what

violates them. The primitives described in those sections must be invoked by the user *explicitly* each time a type-check is desired and then the knowledge base must be repaired manually. In this section, however, we are concerned with *automatic* maintenance of constraints: timely warning the user when a violation occurs in the cause of an update transaction and, if possible, rolling back the violating transaction. The primitives introduced in this section can be used in tandem with Sections 43.2 and 43.3 to enable automatic maintenance of typing and cardinality constraints.

Importantly, integrity constraints are *not* to be confused with *constraint solving* described in Section 30.

Conceptually, an integrity constraint is a query that is supposed to be *false* at all times and this fact is ascertained automatically after every transaction in $\mathcal{E}$RGO. If the query is true for some bindings of its arguments, each of those bindings should be viewed as a witness of an integrity violation. (Yes, normally people think of constraints as something that *must* always be true, not as something that must always be false, but constraints-as-false-statements—i.e., requiring that something must *not* happen—are more useful, since this produces the culprits responsible for constraint violation.)

To define an integrity constraint, one must first define a suitable predicate or a frame to be used as a constraint. For instance, suppose that it is not allowed to both receive a salary from a company and also be a consultant for it:

```
// A Conflict Of Interest (COI) constraint:
//    it should NOT be the case that, for some binding p for ?Person,
//    some binding c for ?Company, and some binding s for ? (the salary
//    argument) both salary(p,c,s) and consults(p,c) are true.
COI(?Pers,?Company) :- salary(?Pers,?Company,?), consults(?Pers,?Company).
```

If it so happens that someone called John both works and consults for Acme, Inc., i.e., that something like `salary(John,Acme,100000)` and `consults(John,Acme)` both become true, we would like to know this immediately and, if possible, the transaction that introduced this violation should be rolled back.

If automatic maintenance of this constraint is what we want, the second step is to tell $\mathcal{E}$RGO about this want by executing this command:

```
?-  +constraint{COI(?,?)}.  // activate constraint
```

The plus sign says that we want to *activate* the aforesaid query as an integrity constraint; the minus sign is used when we want to *deactivate* a constraint:

```
?-  -constraint{COI(?,?)}.  // deactivate constraint
```

The first argument of `constraint{...}` must be a HiLog predicate or a frame. It can also be `\neg` of a predicate or a frame, but it cannot be a `\naf`, conjunction, disjunction, or any kind of more complex subgoals.

Once a constraint is activated, it is first checked against the current knowledge base. If it is not violated (i.e., the constraint query is false), everything is good. Otherwise, a warning is issued. For instance, suppose John and Mary *do* work and also consult for Acme. Then the following warning will be triggered after the above `+constraint{...}` statement is executed:

```
*** A violation of the constraint activated on line 5 in file mykb.ergo
    existed prior to adding this constraint
    The offending instances of the constraint are:
        COI(John,Acme)
        COI(Mary,Acme)
```

In this case, the violation existed prior to the moment the constraint was activated, so there is little one can do automatically. Although a number of theoretical approaches to resolving such situations exist, their practical value is questionable. The practical approach offered by $\mathcal{E}$RGO is to use Studio and ask for an explanation for each of the instances of the query that indicates a violation. Then one can decide what to do (perhaps the fact that John consults for Acme was supposed to be, but was not, deleted once he was hired by Acme as an employee). If, on the other hand, it is *known* how to resolve a particular constraint automatically then constraints with callbacks (described later) might be the way to go. Another way to resolve constraints automatically is to *always* use transactional updates (Section 25.2), as shown next.

Suppose the knowledge base has only the fact `consults(Bob,GroupLTD)` but not any fact of the form `salary(Bob,GroupLTD,...)`, i.e., `mykb.ergo` is like this:

```
COI(?Pers,?Company) :- salary(?Pers,?Company,?), consults(?Pers,?Company).
consults(Bob,GroupLTD).
?-  +constraint{COI(?,?)}.
```

Suppose that next we execute the transaction

```
?- tinsert{salary(Bob,GroupLTD,200000)}.
```

If this insertion is to be performed, the knowledge base would get a constraint violation and so the following warning is going to be issued:

```
*** A violation of the constraint activated on line 3 in file mykb.ergo
    is detected after the transaction t_insert{salary(Bob,GroupLTD,100000)}
```

```
The offending instances of the constraint are:
    COI(Bob,GroupLTD)
```

If we now check the query `salary(Bob,GroupLTD,?Salary)` then we will see that it has *no* answers, i.e., the offending insertion was rolled back! This magic happened because we used *transactional* insertion, `tinsert` (or, synonymously, `t_insert`) — see Section 25.2 for the details. If we used a *non*-transactional insertion, such as `insert{salary(Bob,GroupLTD,100000)}`, then the above warning would also appear, but the insertion would stick and the user would have to resolve the issue manually afterwards.

If a constraint is no longer needed, it can be deactivated:

```
?- -constraint{COI(?,?)}.
```

After this, John, Bob, and Mary can both work and consult for Acme and other companies at the same time—with no conflict of interest questions asked.

Unresolved constraint violations. Note that if constraint violations are not resolved, $\mathcal{E}$RGO will keep reminding about them when new update transactions are executed whether these transactions add new violations or not. In contrast, queries that produce no change to the underlying state of the knowledge base will *not* trigger such warnings, even if violations exist prior to the query execution—to reduce the annoyance factor.

For instance, suppose that Mary and John violate the conflict of interests constraint and a new transaction added a COI for Bob with respect to GroupLTD. Then the warning will be

```
... ... ...
The offending instances of the constraint are:
    COI(John,Acme)
    COI(Mary,Acme)
    COI(Bob,GroupLTD)
```

If the user does not want to be reminded about the old violations then, perhaps, an integrity constraint is not the right tool for the situation at hand. After all, what is the purpose of an integrity *constraint* if the violations of the constraint are not being resolved? To address this type situations, $\mathcal{E}$RGO provides a different mechanism, called *truth*-type *alerts*, described in Section 37. This is a more computationally expensive mechanism than constraints and so it should be used only if resolving constraint violations is not always desirable or when the conditions being checked are not conceptualized as integrity constraints that must not be violated.

Constraints with callbacks. Sometimes, the knowledge engineer might need to take a programmatic action in response to a constraint violation. In most of the previous examples, constraint violations were reported in print and manual intervention was needed to take a corrective action. To let the user specify programmatic corrective actions, $\mathcal{E}$RGO provides a 2-argument version of the `+constraint{...}` primitive:

`+constraint{`*ConstraintQuery*,*CallbackPredicate*`}.`

The callback predicate must be a HiLog predicate with at least one argument; it can be also a `\neg` of such a predicate.

If a callback is provided, then constraint-violating alerts won't be printed and *CallbackPredicate* is called instead. If a violation is detected, that first argument will be bound to the list of violating instances of the constraint, so the callback could take a corrective action. For instance, for our conflict-of-interests constraint, we could define the following corrective action:

```
resolve_COI(?Violators,?File,?Line) :-
    deleteall{?Y| COI(?Empl,?Comp) \in ?Violators, ?Y=consults(?Empl,?Comp)},
    (
        write('Violation of constraint in file '), write(File),
        write(' on line '), write(Line), write(' was detected: '),
        writeln(?Violators).
        write('Corrective action was deletion of '),
        write(consults(?Person,?Company)),
        writeln(' for all violated instances of the constraint')
    )@\io.
```

which deletes the facts that cause the violation and then prints a message about what was done to correct the violation. (In databases, this is called a "compensating" transaction.) To tell $\mathcal{E}$RGO that this corrective action is to be taken, the constraint must be activated using the 2-argument version of `+constraint{...}` instead of the 1-argument version used before:

```
?- +constraint{COI(?,?), resolve_COI(?,\@F,\@L}.
```

We see that the first argument (the variable ?) is left unbound because it will be bound by the system to the list of violating instances of the constraint; arguments 2 and 3 are instantiated with the quazi-constants `\@F` and `\@L`. To remind (from Section 7.6), the first quazi-constant is replaced by the compiler with the name of the file in which the above `+constraint{...}` statement appears and `\@L` is replaced with the appropriate line number. This is how the callback in our example gets its file and line arguments.

Finally, in the `-constraint{...}` statement, the second argument is checked for syntax, but is ignored. Therefore, it makes little sense to provide a second argument here—regardless of whether the constraint being deleted does or does not use a callback.

Performance considerations. One should keep in mind that an active constraint injects a query (to find all violating instances) after each top-level query, which cannot but affect the performance. Therefore, one should strive to keep the number of active constraints to a minimum and to design the constraints to be lightweight. It would be an understatement to say that, if a constraint check goes into an infinite loop, the resulting problem would be very hard to figure out.

37 Alerts

An *alert* statement is a request to monitor a predicate or a frame (including isa- and subclass-literals) for certain predefined conditions. Note: the monitored predicate cannot be a complex subgoal, nor can it be a conjunction or disjunction, not even a `\naf` of something. It *can* be a `\neg` of a predicate or a frame, however. If the status of a predicate (or frame) monitored by an alert changes as a result of an update transaction, the alert is raised. At present, $\mathcal{E}$RGO supports these types of alerts:

- `conflict` alerts
- `truth` alerts

A `conflict` alert is raised if a monitored predicate or frame has a conflict *after* a transaction and that conflict did not exist *before* the transaction.

For a true literal, `foo`, defined via *non-defeasible* rules (see Section 38), this means that `\neg foo` can also be derived and therefore the information about `foo` is contradictory. If foo is *defeasible*, a conflict means that there is an *opposing* literal `bar` (i.e., `\opposes(foo,bar)` is true) and both `foo` and `bar` are derivable if defeasibility is disregarded. Note that both `foo` and `bar` could be false due to being defeated and yet they maybe in conflict (and this is likely the reason why they were defeated).

`Truth` alerts are simpler. A truth alert for a literal `foo` is raised after a transaction if `foo` was false before the transaction and became true after.

Conflict alerts. To illustrate the concept of a conflict, suppose `employee(Acme,John,2016)` means that John works for Acme and `\neg employee(Acme,Bob,2017)` means that Bob was fired by Acme. If neither `employee(foo,bar)` nor `\neg employee(foo,bar)` is true, it means that `foo`

neither hired nor fired `bar`. The management might want to be alerted if the same employee was hired and fired in the same year. Note that this is not a constraint, as hiring and firing of the same employee in the same year is *not* an integrity violation.

To tell $\mathcal{E}$RGO that we would like to be notified about the aforesaid events, we can use the alert primitive:

```
?-  +alert{conflict,employee(?,?,?)}.
```

Then, if `employee(Acme,John,2016)` already exists and we execute the transaction `insert{\neg employee(Acme,John,2016)}` then this alert will be raised:

```
?-  insert{\neg employee(Acme,John,2016)}.

*** Conflict alerts raised after transaction insert{\neg employee(Acme,John,2016)}:
        employee(Acme,John,2016)
    These alerts were activated on line 3 in file mykb.ergo
```

Alerts with callbacks. Similarly to constraints, one can tell $\mathcal{E}$RGO to call a HiLog predicate when an alert is raised. Again as in the case of a constraint, the callback predicate must be at least one argument, and that argument must be unbound. It will be bound by the system to the list of alerts discovered during the transaction. For instance, we could define

```
employee_alert_callback(?Alerts,?Module) :-
        insert{employee_alerts(?Alerts)},
        fmt_write('Employee alerts in module %s\n', arg(?Module))@\io,
        fmt_write('The alerts %S are logged.\n', arg(?Alerts)).
```

Then we could activate the alert using the 3-argument version of the `+alert{...}` primitive:

```
?-  +alert{conflict,employee(?,?,?), employee_alert_callback(?,\@)}.
```

Recall that `\@` above is a quasi-constant denoting the current module.

Complex alerts. Lest the reader thinks that only simple alerts can be specified using the above mechanism, here is a much more involved example. Suppose we want to be notified if an employee was fired not only in the same year, but within less that 4 years. To do so, we cannot simply use the `employee` predicate as before, but we can define what we need via rules:

```
longtime_or_current_employee(?Company,?Empl) :- employee(?Company,?Empl,?).
\neg longtime_or_current_employee(?Company,?Empl) :-
        employee(?Company,?Empl,?Year),
        \neg employee(?Company,?Empl,?Year2),
        ?Year2 >= ?Year,
        ?Year2 - ?Year < 4.
?-  +alert{conflict,longtime_or_current_employee(?,?)}.
```

Now, a conflict alert will be raised whenever an employee is fired within 4 years from hiring. With a little bit of thinking, arbitrarily complex alerts can be specified with the help of rules using the above idea.

Truth alerts. Truth alerts are defined similarly to conflict alerts except that `truth` is used as the first argument instead of `conflict`:

```
+alert{truth,longtime_or_current_employee(?,?)}.
```

If some instances of `longtime_or_current_employee(?Company,?Empl)` become true as a result of a transaction, a warning similar to conflict updates will appear. For instance,

```
*** Truth alerts raised after transaction insert{\neg employee(Acme,John,2016)}:
        longtime_or_current_employee(Acme,John)
        longtime_or_current_employee(GroupLTD,Bob)
    These alerts were activated on line 3 in file mykb.ergo
```

Callbacks and deactivation of truth alerts are specified analogously to conflict alerts.

Note that alerts of type `truth` have many similarities with integrity constraints described in Section 36. Indeed, instead of

```
?-  +constraint{COI(?,?)}.
```

in the examples in that section we could as well request

```
?-  +alert{truth,COI(?,?)}.
```

and the results would be similar. Indeed, in both cases, true instances of the query `COI(?Company,?Person)` would be reported after each transaction. There are, however, several significant differences:

- An alert is *not* conceptualized as an integrity constraint, so transactional updates will *not* be rolled back, if the monitored predicate becomes true.

- An alert is raised only for the instances of the monitored predicate that became true as a result of the preceding transaction. Instances that were true before the transaction are not reported. In contrast, constraint violations are always reported—new and old—after each update transaction.
- Alerts are more expensive computationally than constraints because they monitor the state of the knowledge base before and after each transaction. A constraint checks only the after-states.

Finally, we would like to remind something that important even if obvious: each alert results in a query (to find prior conflicts or truths) prior to evaluation of *any* top-level query. It also results in an execution of *another* query right after the previous top-level query, if the latter was a state-changing transaction. Therefore, one should keep the number of active alerts to absolute necessity and one should strive to keep the monitored predicates rather lightweight.

38 Defeasible Reasoning

Defeasible reasoning is a form of non-monotonic logical reasoning where some rule instances can be *defeated* by other rule instances. The defeated rules do not need to be satisfied by the intended model of the knowledge base. $\mathcal{E}$RGO supports the form of defeasible reasoning known as *Logic Programs with Defaults and Argumentation Theories*, as described in [15].

The basic mechanisms by which the user can control defeasible reasoning include the notions of rule *overriding* and literal *opposition*. We will not describe the semantics of defeasible reasoning formally, but instead will give an informal overview.

The basic idea is that the user specifies overriding and opposition using the appropriate predicates and then the *associated argumentation theory* decides which ground instances of which rules are defeated (and thus do not need to be satisfied by the intended model).

To this end, any rule can be *tagged* by adding a primitive @*tagname* to the left of the rule. Tags do not have to be unique and should not be confused with rule Ids. However, if no rule tag is explicitly given, the Id of the rule is used.

A rule instance ρ_1, tagged i_1, *opposes* another rule instance, ρ_2 tagged i_2, if and only if `\opposes(`i_1`,`$head(\rho_1)$`,`i_2`,`$head(\rho_2)$`)` is true, where $head(\rho_i)$ denotes the head-literal of the corresponding rule. Literals for the form `lit` and `\neg lit` always oppose each other so this fact does not need to be explicitly specified using `\opposes`. A rule instance ρ_1, tagged i_1, *overrides* another rule instance ρ_2, tagged i_2, if and only if `\overrides(`i_1`,`$head(\rho_1)$`,`i_2`,`$head(\rho_2)$`)` is true. (For convenience, $\mathcal{E}$RGO also defines 2-argument versions of these predicates, but this is unimportant for the present discussion.) We say that two rule instances are in *conflict* if they oppose each other

and their bodies are true in the intended model of the knowledge base. A more detailed description of rule conflict and overriding appears in Section 38.2.

A rule can be defeated in several different ways. It can be

- *refuted*
- *rebutted*
- *disqualified*.

Different argumentation theories may assign different meanings to these concepts, but roughly they mean the following. A rule instance ρ_1, tagged i_1, *refutes* another rule instance ρ_2, tagged i_2, if and only if these rules are in conflict and the former rule instance overrides the latter. The rule instance ρ_1, tagged i_1, *rebuts* the rule ρ_2 with tag i_2 if and only if the two instances are in conflict and neither instance is refuted by some other rule instance.

Rule disqualification is not so commonly agreed upon a notion as the other two. A rule is *disqualified*, if it is *canceled*, i.e., if it matches a cancellation literal that is true and not defeated. More details on cancellation appear in Section 38.3. More details on cancellation appear in Section 38.3. In addition, some argumentation theories disqualify rules in other cases as well. The default argumentation theory in $\mathcal{E}$RGO postulates that a rule instance ρ_1, tagged i_1, is disqualified if

- If it transitively defeats itself, i.e., there is a sequence of tagged rule instances (i_1, ρ_1), (i_2, ρ_2), ..., (i_{n-1}, ρ_{n-1}), (i_n, ρ_n), such that $i_n = i_1$, $\rho_n = \rho_1$, and at each step (i_i, ρ_i) either refutes or rebuts (i_{i+1}, ρ_{i+1}).

These notions are very useful for debugging the knowledge base (finding out why certain inferences were or were not made), and we will discuss the corresponding debugging primitives in Section 38.8.

It is important to keep in mind that both `\overrides` and `\opposes` are user-level predicates, which are defined by the user as a set of facts (most commonly) and (less commonly) rules. The system uses these predicates to determine which rules are defeated but the user normally does not query these predicates explicitly except for debugging. These predicates are typically queried by the system to get answers to questions like "is there a rule that overrides (or opposes to) a given rule?" (e.g., `\overrides(?X,r123)`) or "check that this rule is not overwritten (or is opposed to) by some other rule" (e.g., `\naf \overrides(?X,r345)`). In other words, `\overrides` and `\opposes` are likely to be called with some of the arguments unbound. Because of that, one must be careful with what is in the body of the rules that define these predicates. For instance, in the following rule for `\overrides`, one parametrized rule-tag (as we shall see, rule tags can be terms with variables) is said to override another based on the values of the parameters, which are assumed to be numbers:

```
\overrides(rule1(?h),rule2(?j)) :- ?h >?j.
```

The problem here is that > is a mode-sensitive predicate, which expects both of its arguments to be bound to an integer. However, as discussed above, `\overrides` might be called with one of the arguments, say the first, unbound. In this case, `?h` will be unbound at the time `?h >?j` is called and a runtime error will ensue.

38.1 Specifying an Argumentation Theory to Use

$\mathcal{E}$RGO supports a generalized form of defeasible reasoning. The user can enable defeasible reasoning on a per-module basis, and different theories of defeasibility can be used in different modules. The type of defeasible reasoning to be used depends on the chosen *argumentation theory*, which defines the *arguments* (as in "arguing") that the reasoner has to use in order to decide what inferences are to be defeated. Syntactically, defeasible reasoning is requested by placing one of the following directives at the top of the appropriate module (before any rules are given):

```
:- use_argumentation_theory.
:- use_argumentation_theory{Module}.
```

The first instruction directs $\mathcal{E}$RGO to use the default argumentation theory module, `\gcl`, which stands for Generalized Courteous Logic (or GCL). One can use a different theory of defeasible reasoning by implementing an appropriate argumentation theory and loading its file into some module `foo`. A file that actually uses this argumentation theory should have the directive

```
:- use_argumentation_theory{foo}.
```

at the top. Thus, different modules of the same $\mathcal{E}$RGO knowledge base can use different theories of defeasible reasoning. For instance, if some argumentation theory is implemented in file `myargth.ergo`, then it can be declared in a $\mathcal{E}$RGO module using

```
:- use_argumentation_theory{foo}.
?- [myargth>>foo].
```

Defeasible theories must use certain API, which will be described in a later version of this manual. Meanwhile, one can construct such theories by analogy with GCL—see `AT/flrgclp.ergo`.

Note that argumentation theories affect only defeasible rules and the defeasibility status of a rule can be changed from `strict` to `defeasible` and vice versa—see Section 34.

38.2 Rule Overriding and Conflicts

Rule overriding can be specified via one of the following two forms:

```
\overrides(RuleLab1,RuleLab2).
\overrides(RuleLab1,AtomForm1,RuleLab2,AtomForm2).
```

The first form of the `\overrides` statement says that the rule with tag *RuleLab1* overrides the rule with the rule tag *RuleLab2* regardless of what the heads of those rules are. The second form of `\overrides` says the following. Let ρ_1 be a rule with the tag *RuleLab1* and head *H1*, and ρ_2 be a rule with the tag *RuleLab2* and head *H2*. Assume that the variables in ρ_1, ρ_2, and the `\overrides` rule are standardized apart (i.e., are not shared, which can be always achieved by renaming). Then for any substitution θ such that $\theta(H1) = \theta(AtomForm1) = \theta(H2) = \theta(AtomForm2)$, the rule-instance $\theta(\rho_1)$ overrides the rule-instance $\theta(\rho_2)$.

The `\opposes` predicate specifies which literals in the rule heads should be considered to be in conflict, if derived simultaneously. As with the `\overrides` predicate, `\opposes` has two forms:

```
\opposes(AtomForm1,AtomForm2).
\opposes(RuleLab1,AtomForm1,RuleLab2,AtomForm2).
```

The first form of the `\opposes` predicate says that the base formulas *AtomForm1* and *AtomForm2* contradict each other. More precisely, for any variable substitution θ, the knowledge base must not infer $\theta(AtomForm1)$ if $\theta(AtomForm2)$ is also inferred, and vice versa. The 4-argument version of `\opposes` is more restrictive: it says that, for any substitution θ, the knowledge base must not infer $\theta(AtomForm1)$ by means of a rule with tag *RuleLab1*, if $\theta(AtomForm2)$ is inferred by means of a rule with tag *RuleLab2*, and vice versa. Note that by "inference" here we mean inference with respect to the argumentation theory in use—not just with respect to the usual first-order logic. For instance, `a :- b` and `b` do not necessarily imply `a` because the rule `a :- b` may be defeated by a different rule.

Both `\opposes` and `\overrides` can be defined via user rules as well as facts, and these rules and facts can even be added and deleted or disabled/enabled at run-time using the statements `insert{...}/delete{...}`, `insertrule{...}/deleterule{...}`, `enable{...}/disable{...}`. However, excessive use of these facilities is not recommended because the knowledge base might get hairy and hard to understand. This is especially true in the case of the `\opposes` predicate. Although it is advised to be conservative with these predicates lest the meaning of a knowledge base becomes obscure, there are no syntactic restrictions on the use of `\overrides` and `\opposes`: they can appear in rule bodies, heads, in aggregate functions, and so on.

It should be kept in mind that for any non-transactional literal L (base HiLog formula, frame, ISA, or subclass), L and `\neg` L always oppose each other, and there is no need to state this

explicitly: it is part of the underlying theory of defeasible reasoning. Transactional literals do not participate in defeasible reasoning. They cannot be negated using `\neg`, but the compiler does not check if they appear as arguments to `\opposes`. If they do, this information is ignored.

38.3 Cancellation of Rules

Sometimes it is useful to turn off—or *cancel*—a rule instance, if certain conditions are satisfied. This can be specified using the special predicate `\cancel`. The default argumentation theory in $\mathcal{E}$RGO understands two versions of that predicate:

```
\cancel(tag).
\cancel(tag,head).
```

The first rule cancels all rule instances whose tag unifies with *tag* and the second version cancels only those rules whose tag unifies with *tag* and whose head unifies with *head*.

Cancellation rules can also be tagged, overridden, or canceled by other cancellation rules. To illustrate, consider the following example:

```
t(aa).
t(bb).
@{L1} tt1(?X) :- t(?X).
@{L2} tt2(?X) :- t(?X).
@{L3} tt3(?X) :- t(?X).

@{c1} \cancel(?,tt1(bb)).
@{c2} \cancel(L2).
@{c3} \cancel(?,tt3(aa)).
@{c4} \cancel(?,tt3(bb)).
@{c5} \cancel(c4).
```

Here `tt1` is true only of `aa` and `tt3` only of `bb`. `tt2` is false for both `aa` and `bb`. For instance, `tt1(bb)` is false because of the cancellation rule `c1`. Similarly, `tt3(a)` is false because it is canceled by the rule `c3`. More interesting, however, is the reason for `tt3(bb)` being true. Note that rule tagged with `c4` cancels the derivation of `tt3(bb)`. However, rule `c5` cancels rule `c4`, so `tt3(bb)` stays put.

Observe the use of the 1-argument and 2-argument cancellation predicate. The first argument is always a term that matches rule tags for the rules to be canceled. The second argument, if present, matches the heads of the rules to be canceled. For instance, `\cancel(L2)` cancels only

the rule tagged `L2`. If more than one rule matches `L2` (this is possible, but is not the case in our example), then all such rules are canceled. Two-argument cancellation predicates cancel only the rules that match both arguments. In our case, however, the first argument is a variable, as in `\cancel(?,tt3(aa))`, which means that all the rules matching the head are canceled, regardless of the tag.[17]

Here is a more complicated example:

```
device(printer).          abused(Bob,printer).       pardoned(printer,Bob).
device(scanner).          abused(Bob,scanner).       pardoned(scanner,Bill).
device(fax).              abused(Bill,scanner).
                          abused(Bill,printer).
                          abused(Mary,fax).
person(Bob), person(Bill), person(Mary).
@{id1}                authorized(?Persn,?Device) :- device(?Device), person(?Persn).
@{id2(?Dev,?Persn)}   \cancel(id1,authorized(?Persn,?Dev)) :- abused(?Persn,?Dev).
@{id3}                \cancel(id2(?Device,?Persn)) :- pardoned(?Device,?Persn).
```

The most interesting feature here is the tag `id2` parametrized with variables and the rule tagged `id3`, which uses that parametrized tag in the head. Note also that here the cancellation rules are conditional. The effect is that rule `id3` cancels the instances of the cancellation rule tagged `id2(printer,Bob)` and `id2(scanner,Bill)`, which has the effect of authorizing Bob to use the printer and Bill to use the scanner, despite the reported abuses. The reader can verify that the net effect of all the cancellations and counter-cancellations is that Bill and Bob can use the fax, Bob and Mary are authorized to use the printer, and Bill and Mary can use the scanner.

Another interesting situation arises when cancellation rules are defeated not due to overriding, but due to a conflict with some other cancellation rule. For instance,

```
@{r}   P.
@{c1}  \cancel(r).
@{c2}  \neg \cancel(r).
```

Here rule `c1` cancels rule `r`, but `c1` conflicts with `c2`, since the two rules contradict each other. As a result, `P` stays true.

Finally, we should note that, if a rule that is being canceled overrides the canceling rule, than the first-mentioned rule stands and the cancellation rule is defeated instead. For example, in

```
@{L1} foo(1).
```

[17] Again, in our case, only one rule head matches the literal `tt3(a)`, but in general there can be several.

```
@{L2} foo(2).
@{L3} \cancel(?,foo(?)).
\overrides(L1,foo(1),?,\cancel(L1,foo(1))).
```

the instance `\cancel(L1,foo(1))` is defeated and `foo(1)` remains true. However, `foo(2)` gets canceled.

One might think that it makes little sense to specify the rule being canceled as having higher priority than the canceling rule and that it is simply a case of ill-design. However, there can be good reasons to design rules in this way. Suppose the following information is given at two different sites:

```
Site 1:
   @{L11} foo(1).
   @{L12} bar.

Site 2:
   @{L21} \cancel(?,foo(?)).
   @{L22} foo(2).
```

Site 1 might want to merge the rules from Site 2, but it is unwilling to let the `\cancel(?,foo(?))` statement tagged `L21` to have effect on the rules of Site 1. That is, for the merged set of rules, it is ok to let the `L22` statement to be canceled (because it was supposed to be canceled at Site 2), but it is not ok for the newly merged rules to cancel `L21`. To achieve this effect, we can add the following statements to the above:

```
origin(L11,site1).
origin(L12,site1).
origin(L21,site2).
origin(L22,site2).
\overrides(?lab1,?, ?lab2,?) :- origin(?lab1,site1), origin(?lab2,site2).
```

The first four facts simply tell us where each rule came from. The last rule says that the rules that came from `site1` take precedence over the rules that came from `site2`. Note that we are in the same situation as described earlier: the fact `L21` cancels both `L11` and `L22`, but `L11` has higher priority. As a result, `L22` remains canceled and is false, but `L11` is not canceled and remains true.

38.4 Changing the Default Defeasibility Status

In Section 34, we discussed the default defeasibility policy and the fact that defeasibility can be changed using the `makestrict`/`makedefeasible` primitives. This policy says that rules that have

no explicit tags and no explicit defeasibility descriptor `@@{defeasible}` are considered strict. To change this default policy, $\mathcal{E}$RGO provides two compiler directives:

```
:- default_is_defeasible_rules.
:- default_is_defeasible.       // short form
:- default_is_strict_rules.
:- default_is_strict.           // short form
```

The `\default_is_defeasible_rules` directive changes the default so that untagged rules become defeasible. The directive `\default_is_strict_rules` changes the default back to strict. These directives can appear any number of times in the file, changing the treatment of the untagged rules to the desired default. (Of course, such frequent switching is not advisable.)

Sometimes it is useful to be able to query the rule base based on rule tags and heads. Such queries can be issued using the `clause{...}` primitive. However, if you do not need to query the rule bodies, a faster fay is to use the `tag` primitive. For instance, if we have the rules

```
:- use_argumentation_theory.
abc(?X) :- cde(?X).
@{p} foo :- bar.
@@{defeasible} abc2(?X) :- cde2(?X).
@@{strict} foo2 :- bar2.
```

then the query

```
?- tag{?X,?Y@main}.
```

will return the following answers:

```
?X = 2
?Y = ${abc(?_h8660)@main}

?X = 4
?Y = ${abc2(?_h8623)@main}

?X = 5
?Y = ${foo2@main}

?X = p
?Y = ${foo@main}
```

Note that the query `tag{?X,?Y}` (without the module specification for the head, will likely return many answers, as there can be many rules in different loaded modules.

38.5 Supported Argumentation Theories

At present, $\mathcal{E}$RGO supports several different argumentation theories: the cautious, the original, and the strong courteous logics plus also a logic with general exclusion constraints. The $\mathcal{E}$RGO library of argumentation theories also contains a number of experimental packages that are not described here.

38.5.1 The Cautious, Original, and Strong Courteous Argumentation Theories

These argumentation theories are very similar to each other; the differences lie on the edges. All these theories use the `\opposes` and `\overrides` predicates as well as the notions of rebuttal, refutation, and cancellation as the means for determining which rule instances are to be defeated. The difference is that, in the original courteous theory, a rule, R, is defeated if another rule defeats, rebuts, or cancels R. In the cautious theory, that other (defeating) rule must not itself be defeated in order to defeat R. The cautious theory does not also let that other rule be involved in a circular defeating relationships, i.e., there can be no sequence of tagged rule instances (i_1, ρ_1), (i_2, ρ_2), ..., (i_{n-1}, ρ_{n-1}), (i_n, ρ_n), such that $i_n = i_1$, $\rho_n = \rho_1$, and at each step (i_i, ρ_i) either refutes or rebuts (i_{i+1}, ρ_{i+1}).

The strong Courteous argumentation theory is in-between the cautious and the original Courteous theories. It is very similar to the original theory except that it does not allow the defeating rule, ρ_1, to be rebutted or refuted by the rule ρ_2 that is being defeated by ρ_1.

The cautious argumentation theory is the default. It can be invoked by placing

```
:- use_argumentation_theory.
```

at the top of a $\mathcal{E}$RGO module where defeasible reasoning is to be used. This loads the new courteous argumentation theory into the module `\gcl`. For the original courteous argumentation theory, place the following at the top:

```
:- use_argumentation_theory{ogcl}.
?- [ogclp>>ogcl].
```

The original argumentation theory is then available in the module `ogcl`. (Of course, the user can choose a different module name).

The strong argumentation theory can be requested as follows (where, again, the choice of the module is up to the user):

```
:- use_argumentation_theory{sgcl}.
?- [sgclp>>sgcl].
```

38.5.2 Courteous Logic with Exclusion Constraints

This argumentation theory can be invoked by placing the following at the top of an appropriate module:

```
:- use_argumentation_theory{gcle}.
?- [gclpe>>gcle].
```

Again, the user can choose to load this argumentation theory into a differently named module.

The main difference between this argumentation theory and the two previous ones is that it allows more than two rule heads to oppose each other. The syntax is

```
id:\Exclusion[\opposers->{opposer1, ..., opposerN}].
```

This kind of statements is called an *exclusion constraint* and it means that `opposer1`, ..., `opposerN` cannot be all true at the same time. The opposers `opposer1`, ..., `opposerN` must be all reified. The term `id` is the identifier of the exclusion constraint. The usual `\opposes` statements are also understood; they are treated as binary exclusion constraints. As with `\opposes`, the above exclusion constraints can have variables and they can also be defined by rules.

In case all opposers in an inclusion constraint are true, overriding determines which of them are defeated. Intuitively, the defeated opposers are those that do not "beat" (rebut) any other opposer in the same exclusion constraint.

38.6 Defeasible Rules Must Be Purely Logical

It must be kept in mind that all rules involved in defeasible reasoning must be *purely logical*. This includes both the tagged rules of the core knowledge base (whether they are explicitly mentioned in the overrides/opposes statements or not) as well as the rules that define the predicates `\overrides` and `\opposes`. If such a rule depends on a strict rule then the latter must also be purely logical. "Purely logical" here means that the bodies of such rules *cannot* use non-logical or dynamic features, such as:

- I/O statements
- Insert/delete statements
- The cut (!), the predicates `ground/1`, `var/1`, `nonvar/1`, and similar. These predicates can be used only if the alternatives (when the predicates fail) cause an abort and issue an error message for the user. For instance, the following predicate

  ```
  %check_state(?s) :- ground(?s), !.
  %check_state(?s) :- abort(['Nonground state found: ', ?s,
                              '. Might cause infinite recursion.'])@\sys.
  ```

  can be used in the body of a defeasible rule (or of the rules defining `\overrides` and `\opposes`) because the alternative to being ground here aborts the inference process. This is because `ground/1` is a logically clean predicate: if its argument is not ground, the predicate `%check_state(?s)` aborts.

- *Modal* predicates that require bound arguments, which include:
  - Comparison operators >, <, etc., unless it is assured that both sides of the comparison are ground during the inference.
  - Inequality operators, such as `!=`, `!==`, `\=`, `\==`, unless it is certain that both arguments get bound during the evaluation.
  - The evaluation operator *left* `\is` *right*, unless *right* is bound to an appropriate evaluable expression.

  In all these cases, a general remedy in case of a runtime error that complains about unbound arguments is to identify the appropriate domain for the variables that must be bound and bind the variables in question before the modal predicate is invoked. For instance, suppose we wish to say that argument 2 of the predicate `price` must not have two different values for the same item (in argument 1):

  ```
  \opposes(price(?x,?p1),price(?x,?p2)) :- ?p1 \= ?p2.
  ```

  Unfortunately, such a definition will almost certainly cause a runtime error because one of the `?p1` or `?p2` will end up unbound in the course of the reasoning performed inside the argumentation theory. However, we do know that these variables must be bound to prices and thus we could write

  ```
  \opposes(price(?x,?p1),price(?x,?p2)) :-
        price(?x,?p1),
        price(?x,?p2),
        ?p1 \= ?p2.
  ```

This will likely bind ?p1 and ?p2 before the inequality \ = is used.[18]

On the other hand,

```
\opposes(price(?x,?p1),price(?x,?p2)) :- ?p1 != ?p2.
```

is likely to work fine because != is delayed until both arguments become bound. Generally, != and !== are slower than but safer and more declarative than \= and \== because their evaluation is delayed when necessary and they are not that dependent on their position in the rule body. The comparison operators <, >, etc., as well as the \is operator are also generally safe, *if* their arguments get bound eventually.

38.7 Some Useful and Common Definitions

One of the most common uses of defeasible reasoning is specifying functional dependencies among the attributes of a predicate. We saw one example of this in Section 38.6 where we wanted to say that each product has at most one price. If prices were specified via a frame feature, i.e., ?product[price->?value], then we could write

```
\opposes(?x[price->?p1],?x[price->?p2]) :- ?p1 != ?p2.
```

and if we had two facts like

```
@{amazon} milk[price->3].
@{tweeter} milk[price->3.2].
```

then both will be defeated (will be false) unless one statement overrides the other.

Observe that this specification of the functional property of the price-attribute is quite different from specifying a signature like

```
Product[|price{0..1} => \decimal|].
```

In the first case, as we said, both milk[price->3] and milk[price->3.2] would be *false*. In the second case, we have a constraint and both facts remain *true*. However, when the constraint is *checked*, these two facts will be pronounced as constraint *violators*. This second approach is similar

[18] We say "likely" because price(?x,?p) itself may be defined by rules (rather than by a collection of facts) and these rules might involve comparisons, arithmetic operations, etc., and thus may require that it be called with some arguments (e.g., ?x) bound. However, it cannot be guaranteed that, in the above rule, price(?x,...) will always be called with ?x bound.

Portions Copyright © 2013–2017 Coherent Knowledge

to the treatment of integrity constraints in databases and is very common. Nevertheless, the first approach might be better in situations where information is merged from different sources and some losses (like the price of milk) are acceptable. The second approach is also better if, for example, we trust one source (e.g., `amazon`) more than the other. In such a case, we can state that, in case of a conflict, everything that comes from the more trusted source overrides what comes from the other source.

The predicate `price` with two arguments and a functional dependency of the second argument on the first can be generalized in various ways. Suppose we have a predicate of the form

```
project(Company,ProjectName,Budget,Location)
```

where the budget argument functionally depends on the company and the project (but the same project may have different locations and so location is not functionally dependent on anything). With the defeasibility approach, we can specify this using the following rule for `\opposes`:

```
\opposes(project(?c,?p,?b1,?), project(?c,?p,?b2,?)) :- ?b1 != ?b2.
```

We could even generalize this to *all* predicates of the same kind. To do so, we could introduce a class, say `functional_4arg_pred([1,2],3)`, and write

```
\opposes(?Pred(?arg1,?arg2,?arg3a,?), ?Pred(?arg1,?arg2,?arg3b,?)) :-
        ?Pred : functional_4arg_pred([1,2],3),
        ?arg3a != ?arg3b.
```

Then, if we also have the fact `project:functional_4arg_pred([1,2],3)` then the previous `\opposes`-rule is implied as would be the case for any other predicate in that class.

The last rule can be generalized for any number of arguments and any functional dependency among the arguments with the help of $\mathcal{E}$RGO meta-facilities, like `=..` (Section 16.3). However, such a general rule is considerably harder to write. But something almost as general is easier to write using frames:

```
\opposes(?arg1[?Pred(?arg2)->?arg3a], ?arg1[?Pred(?arg2)->?arg3b]) :-
        ?Prop : functional_property,
        ?arg3a != ?arg3b.
```

Here we assume that `?arg2` is a list of the relevant arguments on which `?arg3` depends functionally.

38.8 Debugging Defeasible Knowledge Bases

To help the user debug defeasible knowledge base, argumentation theories provide a special API, which can be used to find out why certain inferences were or were not made. The API consists of several methods, which take status-objects of the form `status(ruleTag,ruleHead)` and returns information such as why a certain rule was defeated, which rules are defeated by the given one, etc. The list of methods follows:

- `status(?T,?H)[howDefeated -> ?Reason].`
  Here `?T` is a rule Id and `?H` is a rule head. The rule head cannot be a variable. There is no need to reify the rule head: $\mathcal{E}$RGO understands that a rule-head literal is expected and will compile it accordingly.

  If the corresponding rule is not defeated, the query fails. Otherwise, `?Reason` is the result of the query. It can take three different forms:

  - `refutedBy(ruleTag,ruleHead)`: In this case, all rule tag/head pairs that refute the rule(s) represented by the `?T/?H` pair (i.e., whose tag unifies with `?T` and head with `?H`) will be returned.
  - `rebuttedBy(ruleTag,ruleHead)`: All the rule tag/head pairs that rebut the rule(s) represented by the `?T/?H` pair will be returned.
  - `disqualified`: This is returned if the rule is disqualified. In the default `\gcl` theory, a rule is disqualified if it is canceled, overridden by a strict rule, or if it transitively refutes/rebuts itself. In this case, an auxiliary method, `howDisqualified`, can provide additional information, as described next.
  - `canceled`: Other argumentation theories disqualify a rule only if it is canceled. These theories return `canceled` in this case instead of `disqualified`.
  - `beatenByStrictRule(?ruleHead)`: This means that a strict rule with the head `?ruleHead` opposes the rule with tag `?T` and head `?H`.

- `status(?T,?H)[howDisqualified->defeatCycle(?Defeater,?Defeated)].`
  If the rules represented by the `?T/?H` pair are disqualified (in the default argumentation theory), this method returns the set of terms of the form `defeats(?Defeater,?Defeated)`. Here both `?Defeater` and `?Defeated` are tag/head pairs that are defeated by the `?T/?H` pair and, in addition, `?Defeater` defeats `?Defeated`. By following these pairs one should be able to discover a cycle of defeats starting and ending with `?T/?H`, which constitutes a self-defeating cycle.

- `status(?T,?H)[howDisqualified->canceled].`
  This query is true if all the rules whose tag unifies with `?T` and head with `?H` are canceled.

- `status(?T,?H)[howDisqualified->beatenByStrictRule(?SRH)]`.
  This literal is true if all the defeasible rules with tag `?T` and head `?H` have an opposing strict rule. In that case, `?SRH` is bound to the head of that opposing rule.

- `status(?T,?H)[info->?Info]`.
  This method provides all kinds of details about the behavior of the rule with head `?H` and tag `?T`. The information returned includes `candidate` (if the corresponding rule is a candidate), `conflictsWith(?Head)`, `competes(?Exclusion,?Head)`, `refutes(?Head)`, and `rebuts(?Head)`.

All these methods are provided by all argumentation theories and are available in their respective modules (`\gcl` for the default argumentation theory; for other argumentation theories, these methods are available in the modules in which these argumentation theories are loaded). However, the information returned by these methods differs from one argumentation theory to another. For instance, for the default, `sgclp`, and `ogclp` theories, the method `howDefeated` may return `refutedBy(...)` and `rebuttedBy(...)`, but in case of `gclpe` this method might return `notBeaterFor(ExclusionConstraintId)` as well as `canceled` and `beatenByStrictRule(?SRH)`. Here are some of the examples of these queries:

```
?- status(?T,configuration(0,block4,square7))[info->conflictsWith(?X)]@\gcl.
?- status(?T,configuration(nxt(nxt(0)),block4,square3))[howDefeated->?X]@\gcl.
?- status(?T,configuration(0,block4,square7))[info->competes(?Exclusion,?O)]@gcle.
```

39 Primitive Data Types

$\mathcal{E}$RGO supports the following built-in data types: `\boolean`, `\long`, `\integer`, `\double`, `\decimal`, `\string`, `\symbol`, `\charlist`, `\iri` (international resource identifier; generalization of URLs), `\time`, `\date`, `\dateTime`, and `\duration`.

Following the now accepted practice on the Semantic Web, $\mathcal{E}$RGO denotes the constants that belong to a particular primitive data type using the idiom `"`*literal*`"^^`*type*. The *literal* part represents the lexical form of the constant and the *type* part is the type name. For instance, `"2004-12-24"^^\date`, `"2004-12-24T15:33:44"^^\dateTime`.

A type name must be a Prolog atom. Some data types, like `time`, `dateTime`, etc., are exact analogues of the corresponding XML Schema types. In this case, their names will be denoted using symbols that have the form of a URL. For instance, `'http://www.w3.org/2001/XMLSchema#time'`. However, for convenience, all type names will have one or more $\mathcal{E}$RGO-specific abbreviated forms, such as `\time` or `\t`. These abbreviated forms are case-insensitive. So, `\time` and `\TiMe` are assumed to be equivalent. In addition, when the type names have the form of an IRI, the compact

prefix representation is supported (see Section 39.2 below). For instance, if `xsd` is a prefix name for `'http://www.w3.org/2001/XMLSchema#'` then the constant `"12:33:55"^^'http://www.w3.org/2001/XMLSchema#time'` can be written as `"12:33:55"^^xsd#time'`. Taking into account the various abbreviations for this data type, we can also write it as `"12:33:55"^^\time` or even `"12:33:55"^^\t`.

Variables can be also *typed*, i.e., restricted to be bound only to objects of a particular primitive data type. The notation is `?`*`variablename`*`^^`*`typename`*. For instance, the variable `?X^^\time` can be bound only to constants that have the primitive type `\time`. This mechanism is more general and allows bounding of variables to arbitrary classes, not just data types; it has already been discussed in Section 10.

The methods that are applicable to each particular primitive type vary from type to type. However, certain methods are more or less common:

- `toSymbol`, which applies to a data type constant and returns its printable representation (a Prolog atom). For instance, if ?Y is bound to `"12:44:23"^^\time` then `?Y[toSymbol->'12:44:23']@\basetype` will be true.
- `toType`(parameters), which applies to any class corresponding to a primitive data type (for instance, `\time`). Most types will have two versions of this method. One will apply to arguments that represent the components of a data type. For instance, `\time[toType(12,23,45) -> "12:23:45"^^\time]@\basetype`. The other will apply to the general constant symbol (≡ Prolog atom) representation of the data type. For instance, `\time[toType('12:23:45')->"12:23:45"^^\ti`
- `isTypeOf`(constant), which applies to every data type class (e.g., `\time`) and determines whether constant has the given primitive type (`\time` in this example).
- `equals`(constant), which tells when the given datatype constant equals some other term.
- `lessThan`(constant), which tells when one constant is less than some other term. For integers, floats, time, dates, durations, and strings, this method corresponds to the natural order on these types. For other types, this method returns false.
- `typeName`, which tells the type name (and thus also class) of the given data type.

All these methods are available in the $\mathcal{E}$RGO system module `\basetype`.

In addition, each primitive data type has a built-in class associated with it. For instance, the primitive data type `\integer` has an associated class named `\integer` and the data type `\dateTime` has an associated class with the same name.

Note: Since built-in classes have infinite extensions, $\mathcal{E}$RGO allows only ground membership tests with respect to these classes. Non-ground tests are permitted, but evaluate to **false** or `true` depending on the situation. For instance, the following query

```
?- f(?X):\symbol.
```

returns the answer "No." On the other hand,

```
?- ?X:\integer, ?X=1.    // true
?- ?X:\integer, ?X=abc.  // false
```

If at the top level `?X` in `?X:\integer` remains an unbound variable, the expression should evaluate to `undefined` and leave `?X` unbound. (Currently, though, such an expression evaluates to `true`). □

The following subsections describe each data type separately.

39.1 $\mathcal{E}$RGO Symbols

Before describing the actual data types, we remind that in Section 7.1 we introduced alphanumeric constants, such as `abc12`, and sequences of characters enclosed in single quotes, such as `'aaa 2*)@'`, and called them *general constant symbols* or *Prolog atoms*. These are not the only constants in $\mathcal{E}$RGO. In the following subsections we will introduce typed literals that represent time, date, and more.

The general constant symbols mentioned above are partitioned into three subcategories: *strings* (class `\string`) , *IRIs* (class `\iri`), and *abstract symbols* (class `\symbol`).

Basically, abstract symbols are all the Prolog atoms that do not serve as internal representations for the data types of the form `"..."^^\iri` or `"..."^^\strings`.[19]

The class `\symbol` has only the following method apart from the already mentioned `toSymbol`, `lessThan`, etc.:

- `\symbol[concat(`*List*`) => \object]`

Note that that the above is *not* a class-level signature, but an object level one. That is, it applies to the object `\symbol` itself and not to the individual objects that belong to the class `\symbol`. For instance,

```
ergo>  \symbol[concat([abc,cde,fgh])->?X]@\basetype.

?X = abccdefgh
```

[19] For efficiency, IRIs and strings are encoded as atoms that are prefixed with a character sequence that is unlikely to be engaged by the user. IRIs are thus atoms prefixed with the character `i` and the backspace character; strings are atoms that are prefixed with the character `s` and the backspace symbol.

The *List* argument in the `concat` method can be a list of anything, not necessarily of other symbols. For instance,

```
ergo>  \symbol[concat(["11:11:11"^^\time,cde,fgh])->?X]@\basetype.

?X = '11:11:11cdefgh'
```

39.2 The \iri Data Type

The canonical representation of constants of type IRI (international resource identifiers, a generalization of URLs, uniform resource locators) is `"`*some iri*`"^^\iri`, where *literal* must have a lexical form corresponding to IRIs on the World Wide Web. IRIs have the shorthand notation `\"`*some iri*`"`, as mentioned before. The full IRI name of this type is `'http://www.w3.org/2007/rif#iri'`.

IRIs can come in the usual full form or in an abbreviated form known as the *curi* form (for *compact IRI*).

A *compact form* of an IRI (curi) consists of a prefix and a local-name as follows: *PREFIXNAME#LOCALNAME*. Here *PREFIXNAME* is an alphanumeric identifier that must be defined as a shortcut for an IRI elsewhere (see below). *LOCALNAME* can be a string, an alphanumeric identifier, or a quoted atom. (If *LOCALNAME* contains non-alphanumeric symbols, it must be enclosed in double quotes as in `"ab%20"`.) A compact IRI is treated as a macro that expands into a full IRI by concatenating the expansion of *PREFIXNAME* with LOCALNAME.

The prefix of a compact IRI must be defined in one of the following ways:

```
:- iriprefix{PREFIXNAME = PREFIXIRI}.
:- irilocalprefix{PREFIXNAME = PREFIXIRI}.
?- iriprefix{PREFIXNAME = PREFIXIRI}.
```

Here `PREFIXIRI` can be an alphanumeric identifier or a quoted atom. Prefixes can also be defined at run time using a query:

```
?- iriprefix{PREFIXNAME = PREFIXIRI}.
```

Such a prefix becomes defined only after the command is executed. If a prefix is used before it is defined, an error will result. For example,

```
:- iriprefix{w3c = 'http://www.w3c.org/', AAAWEB = 'http://www.AAA.com/'}.
```

Defines two prefixes, which can be used in subsequent commands like this:

```
?- ?X = w3c#a.
```

This will bind `?X` to `\"http://www.w3c.org/a"`. Likewise,

```
?- ?Y = AAAWEB#"ab%20"
```

binds `?Y` to `\"http://www.AAA.com/ab%20"`.

Scope of IRI prefixes: module locality. IRI prefix definitions are local to the module where they are defined. If we define the following in module `foo`:

```
:- iriprefix{W3='http://w3.org/'.
C[a->\"http://w3.org/abc"].
```

and then load the following file into module `main`

```
r(?X):-?X[a->W3#abc]@foo.
s(?X):-?X[a->W3#cde].
```

then `W3` will have an expansions for calls to `foo`, but not to the current module. Thus, the answer to

```
?- r(?X).
```

will be `C` but

```
?- s(?X).
```

will get an error saying that the prefix W3 is not defined. If the same prefix is defined both in the file and in a module being referenced by a query then the prefix definition in the current file takes precedence. For instance, if in the above example the module main had another declaration for the IRI prefix `W3` then `?X[a->W3#abc]@foo` will use that definition rather than the one in module `foo`.[20]

The prefix macro-expansion works also for transactional predicate and method names. For instance,

[20] The rationale here is the theory of the "least surprise."

```
:- iriprefix{W3 = 'http://w3.org/'}.
C[%W3#aaa(b)].
%W3#r(?Y)(?X):-?X[%W3#abc(?Y)]@foo.
```

Note, however, that transactional (%-prefixed) symbols can occur only as the names of predicates and methods (possibly higher-order predicates and methods).

For convenience, some IRI prefixes are predefined:

```
xsd   'http://www.w3.org/2001/XMLSchema#'
rdf   'http://www.w3.org/1999/02/22-rdf-syntax-ns#'
rdfs  'http://www.w3.org/2000/01/rdf-schema#'
owl   'http://www.w3.org/2002/07/owl#'
rif   'http://www.w3.org/2007/rif#'
swrlb 'http://www.w3.org/2003/11/swrlb#').
dc    'http://purl.org/dc/elements/1.1/').
```

However, one can always override these built-in definitions using either a compile time directive `iriprefix` or a runtime query `iriprefix`.

$\mathcal{E}$RGO also provides the necessary primitives to enable querying the available prefix definitions at run time: `prefix{?prefix,?expansion}` and `prefix{?prefix,?expansion}@module`. When `@module` is not specified, the current module is assumed. When `module` is given, only the prefixes defined for the given module are returned. The aforesaid predefined prefixes are considered to be defined for every module. For example:

```
:- iriprefix{foo = 'bar.com'}.
?- prefix{foo,?exp}@?M.

?exp = bar.com
?M = main


?- prefix{?p,?exp}@main.

?p = foo
?exp = bar.com

?p = owl
?exp = http://www.w3.org/2002/07/owl#
```

```
?p = rdf
?exp = http://www.w3.org/1999/02/22-rdf-syntax-ns#

?p = rdfs
?exp = http://www.w3.org/2000/01/rdf-schema#

?p = rif
?exp = http://www.w3.org/2007/rif#

?p = xsd
?exp = http://www.w3.org/2001/XMLSchema#
```

Scope of IRI prefixes: file locality. The `iriprefix` directives are *global* within the module. This means that these directives take effect not only in the file in which they appear, but also in all files *added* to the same module afterwards. Prefixes defined in the files loaded/added to module `main` also become known in the $\mathcal{E}$RGO shell. However, if an *added* file has an `iriprefix` directive for a previously defined prefix then the new definition overrides the old one (and is inherited by all subsequently added files). Note: *loading* a file into a module zaps all the prefix definitions known to the module and replaces them with whatever is defined in that file.

Sometimes, however, it is preferable to have prefix definitions that are truly local to a file and are not inherited by subsequently added files. This can be done using the `irilocalprefix` directive. For instance, suppose file A was loaded into module `Mod` and has these prefix definitions:

```
:- iriprefix{foo='http://123'}.
:- iriprefix{moo='http://abc'}.
:- iriprefix{doo='http://456'}.
:- irilocalprefix{bar='http://cde'}.
```

Suppose also that file B has these definitions and was added to `Mod` later:

```
:- iriprefix{foo='http://789'}.
:- irilocalprefix{moo='http://fgh'}.
```

Finally, suppose file C was added later to the same module `Mod`. Then file A has prefixes `foo`, `moo`, `doo`, and `bar`. File B does not have prefix `bar` defined because it is local to file A. Moreover, prefixes `foo` and `moo` are overridden and have a different meaning than they have in file A. Prefix `doo` has the same meaning, however.

In file C, the situation is as follows. Prefixes `foo`, `moo`, `doo` are inherited. Prefix `doo` has the same meaning as in files A and B. However, prefix `foo` was overridden in file B, so it is file B's

expansion that is inherited by file C. As to prefix `moo`, it is inherited from file A and has the same meaning. Note that even though `moo` was overridden by file B, this overriding was local and did not propagate to file C.

Note that

```
:- iriprefix{foo=bar}.
```

is equivalent to the pair

```
:- irilocalprefix{foo=bar}.
?- iriprefix{foo=bar}.
```

The `irilocalprefix` directive takes case of compiling prefix `foo` within the given file and the executable `iriprefix` directive propagates the prefix to all subsequently added files (and to the $\mathcal{E}$RGO shell, if applicable).

Class \iri. All constants of the primitive type IRI are members of the built-in class `\iri`.

The IRI data type supports the following methods, which are available in the $\mathcal{E}$RGO module `\basetype` (or, in the abbreviated form, `\btp`). They are described here by their signatures.

Class methods:

- `\iri[toType(\symbol) => \iri]`
- `\iri[=> isTypeOf(\object)]`

Component methods:

- `\iri[|scheme => \symbol|]`
- `\iri[|user => \symbol|]`
- `\iri[|host => \symbol|]`
- `\iri[|port => \symbol|]`
- `\iri[|path => \symbol|]`
- `\iri[|query => \symbol|]`

- `\iri[|fragment => \symbol|]`

Note that the exact meaning of the above components depends on the IRI scheme. For `http`, `ftp`, `file`, etc., the meaning of the first five components is clear. The query is an optional part of the IRI that follows the ?-sign, and fragment is the last part that follows #. Some components might be optional for some IRI schemes. For instance, for the `urn` and `file` schemata, only the path component is defined. For the `mailto` scheme, port, path, query, and fragment are not defined. If a scheme is not recognized, then the part of the IRI that follows the scheme goes into the `path` component unparsed.

Other methods:

- `\iri[|toSymbol => \symbol|]`
- `\iri[|=> equals(\object)|]`
- `\iri[|typeName => \symbol|]`

Examples:

- `\"http://foo.bar.com/abc"`
- `"http://foo.bar.com/abc"^^\iri`
- `?- \iri[toType('http://foo.bar.com/abc') -> "http://foo.bar.com/abc"^^\iri]@\basetype`
- `?- \"http://foo.bar.com/abc"^^\iri[host -> 'foo.bar.com']@\btp`

39.3 The Primitive Type \dateTime

This data type corresponds to the XML Schema `dateTime` type. The constants of this data type have the form `"ZYYYY-MM-DDTHH:MM:SS.sZHH:MM"^^\dateTime`. The symbols `-`, `:`, `T`, and `.` are part of the syntax. The leftmost `Z` is an optional sign (`-`). The part that starts with the second `Z` is optional and represents the time zone (the second `Z` is a sign, which can be either `+` or `-`; note that the first `Z` can be only the minus sign or nothing). The part that starts with `T` is also optional; it represents the time of the specified day. The part of the time component of the form `.s` represents fractions of the second. Here `s` can be any positive integer.

The constants of this primitive type all belong to the class `\dateTime`. The name of this type has the following synonyms: `\dt`, `'http://www.w3.org/2001/XMLSchema#dateTime'`.

The following methods are available in the $\mathcal{E}$RGO system module `\basetype`; they are described by their signatures below.

Class methods:

- `\dateTime[toType(\integer,\integer,\integer,\integer,\integer,\integer,\decimal,\integer,\integer,\integer) => \dateTime]`
  The meaning of the arguments is as follows (in that order): date sign (1 or -1), year, month, day, hour, minute, second, zone sign (1 or -1), zone hour, zone minute. All arguments, except date sign and zone sign, are assumed to be positive integers; date sign and zone sign can be either 1 or -1.
- `\dateTime[toType(\symbol) => \dateTime]`
- `\dateTime[=> isTypeOf(\object)]`
  Tells if object belongs to the primitive type `\dateTime`.
- `\dateTime[now => \dateTime]`
  Current local date+time.
- `\dateTime[now(utc) => \dateTime]`
  Current UTC date+time.
- `\dateTime[now(\decimal) => \dateTime]`
  Current UTC date+time adjusted for the time zone given by the argument. The decimal argument has the form `SHH.MM`, where `S` is the plus sign (or empty) or the minus sign. If more than two `MM` digits are provided, the rest are cut off. If `HH>24` then the query fails.

Component methods:

- `\dateTime[|dateSign => \integer|]`
- `\dateTime[|year => \integer|]`
- `\dateTime[|month => \integer|]`
- `\dateTime[|day => \integer|]`
- `\dateTime[|hour => \integer|]`
- `\dateTime[|minute => \integer|]`
- `\dateTime[|second => \integer|]`

- `\dateTime[|zoneSign => \integer|]`
- `\dateTime[|zoneHour => \integer|]`
- `\dateTime[|zoneMinute => \integer|]`
- `\dateTime[|date => \date|]`
- `\dateTime[|time => \time|]`

Other methods:

- `\dateTime[|toSymbol => \symbol|]`
- `\dateTime[|=> equals(\object)|]`
- `\dateTime[|=> lessThan(\object)|]`
- `\dateTime[|typeName => \symbol|]`
- `\dateTime[|add(\duration) => \dateTime|]`
- `\dateTime[|minus(\dateTime) => \duration|]`

Examples:

- `"2001-11-23T12:33:55.123-02:30"^^\dateTime`
- `"2001-11-23T12:33:55.123-02:30"^^'http://www.w3.org/2001/XMLSchema#dateTime'`
- `"2001-11-23"^^\dateTime`
- `"-0237-11-23T12:33:55"^^\dateTime`
  Note that this date refers to year 238 BCE.
- `?- "2001-11-23"^^\dateTime[day -> 23]@\basetype.`
- `?- "2001-11-23"^^\dateTime[toSymbol -> '2001-11-23T00:00:00+00:00']@\basetype.`
- `?- "2001-11-23T18:33:44-02:30"^^\dateTime[add("-P22Y2M10DT1H2M3S"^^\duration) -> "1979-09-13T17:31:41-02:30"^^\dateTime]@\btp.`
- `?- "2011-12-22+2:19"^^\dt[minus("2019-11-29T2:30:30-2:09"^^\dt)-> "-P0007Y11M07DT02H30M30S"^^\duration]@\btp`

39.4 The Primitive Type \date

This type corresponds to the XML Schema `date` type. Constants of this type have the form `"ZYYYY-MM-DDSHH:MM"^^\date`. The symbols `-` and `:` are part of the syntax. The symbol `S` represents the timezone sign (`+` or `-`). The timezone part (beginning with `S`) is optional. The leftmost `Z` is the optional sign (`-`). Note that unlike `\dateTime`, which represents a single time point, `\date` represents *duration* of a single day.

All constants of this type belong to the built-in class `\date`. The type name `\date` has the following synonyms: `\d`, `'http://www.w3.org/2001/XMLSchema#date'`.

The following methods are defined for this type and are available through the system module `\basetype`.

Class methods:

- `\date[toType(\integer,\integer,\integer,\integer,\integer,\integer,\integer) => \date]`
  The meaning of the arguments is as follows (in that order): date sign (1 or -1), year, month, day, zone sign (1 or -1), zone hour, zone minute. All arguments, except date sign and zone sign, are assumed to be positive integers; date sign and zone sign can be either 1 or -1.
- `\date[toType(\symbol) => \date]`
- `\date[=> isTypeOf(\object)]`
  Tells if object belongs to the primitive type `\date`.
- `\date[now => \date]`
  Current local date.
- `\date[now(utc) => \date]`
  Current UTC date.
- `\date[now(\decimal) => \date]`
  Current UTC date adjusted for the time zone given by the argument. The decimal argument has the form `SHH.MM`, where `S` is the plus sign (or empty) or the minus sign. If more than two `MM` digits are provided, the rest are cut off. If `HH>24` then the query fails.

Component methods:

- `\date[|dateSign => \integer|]`

- `\date[|year => \integer|]`
- `\date[|month => \integer|]`
- `\date[|day => \integer|]`
- `\date[|zoneSign => \integer|]`
- `\date[|zoneHour => \integer|]`
- `\date[|zoneMinute => \integer|]`

Other methods:

- `\date[|toSymbol => \symbol|]`
- `\date[|=> equals(\object)|]`
- `\date[|=> lessThan(\object)|]`
- `\date[|typeName => \symbol|]`
- `\date[|add(\duration) => \date|]`
- `\date[|minus(\date) => \duration|]`
- `\date[|toDateTime(\integer,\integer,\decimal) => \dateTime|]`
  The arguments are hours, minutes, and seconds (with possible milliseconds).

Examples:

- `"2001-11-23-2:30"^^\date`
- `"2001-11-23"^^\date`
- `"-237-11-23"^^\date`
  Note that this date refers to year 238 BCE.
- `?- "2001-11-23"^^\date[day -> 23]@\basetype.`
- `?- "2001-11-23"^^\date[toSymbol -> '2001-11-23+00:00']@\basetype.`
- `?- "2011-12-22"^^\date[toDateTime(11,15,7.6) ->`
  `"2011-12-22T11:15:07.6"^^\dateTime]@\basetype.`

- `?- "2011-12-22+2:09"^^\d[minus("2019-10-29+3:19"^^\d)->`
  `"-P0007Y10M07DT00H00M00S"^^\du]@\btp.`

- `?- "2001-11-23-02:30"^^\date[add("-P2Y2M10DT"^^\duration) ->`
  `"1999-09-13-02:30"^^\date]@\basetype.`
  Note that when adding a duration to a date, the time-part of the duration constant must be empty.

39.5 The Primitive Type \time

This primitive type corresponds to the XML Schema `time` data type. Constants of this type have the form `"HH:MM:SS.sZHH:MM"^^\time`. The symbols `:` and `"."` are part of the syntax. The part `.s` is optional. It represents fractions of a second. Here `s` can be any positive integer. The sign `Z` represents the sign of the timezone (`+` or `-`). The following `HH` represents time zone hours and `MM` time zone minutes. The time zone part is optional.

The name of this type has the following alternative versions: `\t` and `'http://www.w3.org/2001/XMLSchema#time'`. All constants of this type are also assumed to be members of the built-in class `\time`.

The following methods are available for the class `\time` and are provided by the module `\basetype`. Their signatures are given below.

Class methods:

- `\time[toType(\integer,\integer,\decimal,\integer,\integer,\integer) => \time]`
  The arguments represent hour, minute, second, time zone sign, time zone hour, and time zone minute.

- `\time[toType(\symbol) => \time]`

- `\time[=> isTypeOf(\object)]`
  Tells if object belongs to the primitive type `\time`

- `\time[now => \time]`
  Current local time.

- `\time[now(utc) => \time]`
  Current UTC time.

- `\time[now(\decimal) => \time]`
  Current UTC time adjusted for the time zone given by the argument. The decimal argument has the form `SHH.MM`, where `S` is the plus sign (or empty) or the minus sign. If more than two `MM` digits are provided, the rest are cut off. If `HH>24` then the query fails.

Component methods:

- `\time[|hour => \integer|]`
- `\time[|minute => \integer|]`
- `\time[|second => \integer|]`
- `\time[|zoneSign => \integer|]`
- `\time[|zoneHour => \integer|]`
- `\time[|zoneMinute => \integer|]`
  The arguments are years, months, and days.

Other methods:

- `\time[|toSymbol => \symbol|]`
- `\time[|=> equals(\object)|]`
- `\time[|=> lessThan(\object)|]`
- `\time[|typeName => \symbol|]`
- `\time[|add(\duration) => \time|]`
- `\time[|minus(\time) => \duration|]`
- `\time[|toDateTime(\integer,\integer,\integer) => \dateTime|]`

Examples:

- `"11:24:22"^^\time`
- `"11:24:22"^^'http://www.w3.org/2001/XMLSchema#time'`
- `?- \time[toType(12,44,55) -> "12:44:55"^^\time]@\basetype.`

- `?- "12:44:55"^^\time[minute -> 44]@\basetype.`
- `?- "12:44:55"^^\time[toSymbol -> '12:44:55']@\basetype.`
- `?- "20:12:22"^^\time[toDateTime(2011,12,22) -> "2011-12-22T20:12:22"^^\dateTime]@\basetype.`
- `?- "11:22:33+2:22"^^\t[minus("12:23:44+2:09"^^\t)-> "-P0000Y00M00DT01H01M11S"^^\du]@\btp.`
- `?- "12:44:55"^^\time[add("PT2M3S"^^\duration) -> "12:46:58"^^\time]@\btp.`
  Note that when adding a duration to a time, the date-part of the duration constant must *not* be present.

39.6 The Primitive Type \duration

The primitive type duration corresponds to the XML Schema `duration` data type. The constants that belong to this type have the form `"sPnYnMnDTnHnMdS"^^\duration`. Here `s` is optional sign `-`, `P` indicates that this is a duration data type, and `Y`, `M`, `D`, `H`, `M`, `S` denote year, month, day, hour, minutes, and seconds. `T` separates date from time. The symbols `P`, `Y`, `M`, `D`, `H`, `M`, and `S` are part of the syntax. The symbol `n` stands for any positive integer (for instance, the number of hours can be more than 12 and the number of minutes and seconds can exceed 60) and `d` stands for a decimal number. The part that starts with `T` is optional and any element in the date and the time parts can be omitted.

The constants of this data type all belong to the class `\duration`.

The type name has the following synonyms: `'http://www.w3.org/2001/XMLSchema#duration'`, `\du`.

The following classes are available in module `\basetype`. Their signatures are shown below.

Class methods:

- `\duration[toType(\integer,\integer,\integer,\integer,\integer, => \duration]`
  The meaning of the arguments (in that order) is: year, month, day, hour, minute, second.
- `\duration[toType(\symbol) => \duration]`
- `\duration[=> isTypeOf(\object)]`
  Tells if an object belongs to the primitive type `\duration`.

Component methods:

- `\duration[|year => \integer|]`
- `\duration[|month => \integer|]`
- `\duration[|day => \integer|]`
- `\duration[|hour => \integer|]`
- `\duration[|minute => \integer|]`
- `\duration[|second => \integer|]`

Other methods:

- `\duration[|toSymbol => \symbol|]`
- `\duration[|=> equals(\object)|]`
- `\duration[|=> lessThan(\object)|]`
- `\duration[|typeName => \symbol|]`
- `\duration[|add(\duration) => \duration|]`

Examples:

- `"P5Y5M10DT11H24M22S"^^\duration`
- `?- "-P2Y05M10DT11H24M22S"^^\duration[minute -> 24]@\basetype.`

39.7 The Primitive Type \boolean

This corresponds to the XML Schema `boolean` type. Constants of this type have the form `"true"^^\boolean "false"^^\boolean` or the shorter form `\true`, `\false`. A synonym for the `\boolean` type name is `'http://www.w3.org/2001/XMLSchema#boolean'`.

All constants in this type belong to the built-in class `\boolean`. The following methods are available in module `\basetype`.

Class methods:

- `\boolean[toSymbol => \symbol]`
- `\boolean[=> isTypeOf(\object)]`

Other methods:

- `\boolean[|toSymbol => \symbol|]`
- `\boolean[|=> equals(\object)|]`
- `\boolean[|=> lessThan(\object)|]`
  Note: `\false[lessThan(\true)]`.
- `\boolean[|typeName => \symbol|]`
- `\boolean[|rawValue => \symbol|]`
  Extract the content value from the `\boolean` data type. For instance,
  `?- "true"^^\boolean[rawValue->?X]@\basetype`.
  `?X = true`

39.8 The Primitive Type \double

This corresponds to the XML Schema type `double`. The constants in this type all belong to the class `\double` and have the form `"value"^^\double`, where `value` is a floating point number that uses the regular decimal point representation with an optional exponent. Doubles have a short form where the `"..."^^\double` wrapper is removed.

In $\mathcal{E}$RGO, the `\float` and `\double` type designators are interchangeable, and the constants of these data types are treated as regular floating point numbers. This means that, for example, `"1.2"^^\double`, `"1.2"^^\float`, and `1.2` represent the same number.

This type name has a synonym `'http://www.w3.org/2001/XMLSchema#double'`. The following methods are available for type `\double` in module `\basetype`.

Class methods:

- `\double[toType(\decimal) => \double]`
  Converts decimals to doubles. Error, if overflow.

- `\double[toType(\long) => \double]`
  Converts long integers to doubles.
- `\double[toType(\string) => \double]`
  `\double[toType(\symbol) => \double]`
  Converts strings and symbols into doubles, if the textual representation of these values is a number.
- `\double[=> isTypeOf(\object)]`

Instance methods:

- `\double[|floor => \integer|]`
- `\double[|ceiling => \integer|]`
- `\double[|round => \integer|]`

Other methods:

- `\double[|toSymbol => \symbol|]`
- `\double[|=> equals(\object)|]`
- `\double[|=> lessThan(\object)|]`
- `\double[|typeName => \symbol|]`
- `\double[|rawValue => \double|]`
  Extract the number part of the `\double` data type.

Examples: `"2.50"^^\double`, `2.50`, `25E-1` — different forms of `\double`.
`?- "3.54"\double[round->?X]@\basetype.` // answer: `?X = 4`,
`?- 5.51[floor->?X]@\basetype.` //answer: `?X = 5`,
`?- \double[toType(51)->?X]@\basetype.` // `?X = 51.0000` — long-to-double conversion.

39.9 The Primitive Type \long

This data type corresponds to XML Schema's long integers. The constants in this data type belong to class `\long` and have the form `"value"^^\long`, where value is an integer in its regular representation in the decimal system. A shorter form without the `"..."^^\long` wrapper is also allowed. This type name has a synonym: `'http://www.w3.org/2001/XMLSchema#long'`.

Class methods:

- `\long[toType(\symbol) => \long]`
  `\long[toType(\symbol) => \long]`
  Converts strings to long integers, if the string represents an integer in textual form. If it does not then this method fails.
- `\long[toType(\integer) => \long]`
  Converts long integers to arbitrary big integers.
- `\long[=> isTypeOf(\object)]`

Other methods:

- `\long[|toSymbol => \symbol|]`
- `\long[|=> equals(\object)|]`
- `\long[|=> lessThan(\object)|]`
- `\long[|typeName => \symbol|]`
- `\long[|rawValue => \long|]`
  Extract the number part of the `\long` data type.

Examples: `123`, `55`, `"55"^^\long`.

39.10 The Primitive Types \decimal, \integer, and \short

At present, $\mathcal{E}$RGO does not implement the `\decimal` and the `\integer` types, which correspond to XML Schema arbitrary precision types decimal and integer. Instead, `\decimal` is a synonym for `\double`, while `\short` and `\integer` for `\long`, which are described in previous subsections. As usual, there are corresponding classes `\integer` and `\decimal`.

39.11 The Primitive Type \string

This corresponds to the XML Schema type string. The constants in this class belong to type `\string` and the type name has the synonym `http://www.w3.org/2001/XMLSchema#string`. The values of this class have the form `"value"^^\string`. Alphanumeric strings that start with a letter

do not need to be quoted. In the full representation (with the `"..."^^\string` wrapper), the double quote symbol and the backslash must be escaped with a backslash.

The following methods are available in module `\basetype`:

Class methods:

- `\string[=> isTypeOf(\object)]`
- `\string[toType(\object) => \string]`

Note that the method `toType` in class `\string` can be used to serialize any term as a string. For instance,

```
ergo>  \string[toType(abc(cde))->?val]@\basetype.

?val = "abc(cde)"^^\string
```

Instance methods:

- `\string[|=> contains(\string)|]`
- `\string[|concat(\string) => \string|]`
- `\string[|reverse => \string|]`
- `\string[|length => \integer|]`
- `\string[|toUpper => \string|]`
- `\string[|toLower => \string|]`
- `\string[|=> startsWith(\string)|]`
- `\string[|=> endsWith(\string)|]`
- `\string[|substring(\integer,\integer) => \string|]`
  Returns a substring of the object string, where the starting and the ending position of the substring are given by the arguments of the method. -1 in argument 2 means the end of the string.

Other methods:

- `\string[|=> equals(\object)|]`
- `\string[|=> lessThan(\object)|]`
- `\string[|typeName => \symbol|]`

Examples:

- `"abc"^^\string`
- "a `string\n"^^\string`
- `"a\tstring\b"^^\string`
- `"string with a 'quoted' substring"^^\string`
- `?- "abc"^^\string[concat("bbb"^^\string)->?X]@\basetype.`
  `?X = "abcbbb"^^\string`

Note that internally the string `"abc"^^\string` and the atom `'abc'` are different. To extract the actual pure Prolog atom (stripped from the internal stuff), use the method `rawValue`. For instance, `?- "abc"^^\string[rawValue->?X]@\basetype`.

39.12 The Primitive Type \list

This is the usual Prolog list type. The members of this type have the form `"[elt1, ..., eltn]"^^\list` (short form `[elt1, ..., eltn]`) and belong to class `\list`.

The following methods are available from the standard module `\basetype`:

Class methods:

- `\list[=> isTypeOf(\object)]`
- `\list[toType(\list) => \list]`

Other methods:

- `\list[|=> contains(\list)|]`
  Tells if a list object contains the method's argument as a sublist.
  Since checking list containment is a very common operation, this method has a special shortcut `\in`:

  ```
  ?- [b,?A] \subset [b,c,a,d].
  ?A = c
  ?A = a
  ?A = d
  ```

- `\list[|=> member(\object)|]`
  The method's argument and the list-object may not be fully ground. In this case, the method succeeds if the argument to the method unifies with a member of the list.
  Since checking list membership is a very common operation, this method has a special shortcut `\in`:

  ```
  ?- a \in [b,c,a,d].
  Yes
  ```

- `\list[|select(\object) => \list|]`
  Find an member in list that unifies with the object-argument and return the list with the selected member removed. For instance,

  ```
  ?- "[a,b(1),c,b(2)]"^^\list[select(b(?X))->?R]@\btp.

  ?X = 1
  ?R = [a, c, b(2)]

  ?X = 2
  ?R = [a, b(1), c]
  ```

- `\list[|delete(\object) => \list|]`
  Delete all occurrences of the object in list. Selection is made using `==`, not unification. For instance,

  ```
  ?- [a,b(1),c,b(?X),b(1)][delete(b(?X))->?R]@\btp.

  ?X = ?_h4800
  ?R = [a, b(1), c, b(1)]
  ```

Note that `b(?X)` was deleted, but not `b(1)`. On the other hand,

```
?- "[a,b(1),c,b(1)]"^^\list[delete(b(1))->?R]@\btp.

?R = [a, c]
```

i.e., all occurrences of `b(1)` are deleted.

- `\list[|append(\list) => \list|]`
  Appends one list to another and returns the resulting list.
- `\list[append(\list) => \list]`
  Here the signature is attached directly to `\list` as an object. This means that the `append` method applies directly to class `\list`. In this case, it takes a *list of lists* and returns the list that is the result of appending the lists found in that argument list of lists. For instance

  ```
  ?- \list[append(["[a,b]"^^\list,[c,d],"[e,f]"^^\list])->?R]@\btp.

  ?R = [a, b, c, d, e, f]
  ```
- `\list[|ith(\integer) => \object|]`
  Given a list, returns the object in the ith position. If the position is a variable, returns the position in the list at which the result-object is found. If both the position and the object are variables, enumerates all elements in the list and their position number.
- `\list[|length => \long|]`
  Computes the length of the list.
- `\list[|reverse => \list|]`
- `\list[|sort => \list|]`
- `\list[|=> startsWith(\list)|]`
- `\list[|=> endsWith(\list)|]`
- `\list[|=> subset(\list)|]`
  True if the list object contains the argument list.

Other methods:

- `\list[|toSymbol => \symbol|]`
- `\list[|=> equals(\object)|]`
- `\list[|typeName => \symbol|]`

Examples:

- `[a,b,c]`
- `[a,b|?X]`
- `[a,b,c|[d,e]]`
- `"[a,b,c]"^^\list`
- `"[a,b|?X]"^^\list`
- `"[a,b,c|[d,e]]"^^\list`

As in Prolog, the part of a list term that follows the bar `|` represents the tail of the list.

39.13 Character Lists

$\mathcal{E}$RGO character lists, *charlists*, are represented as `"..."^^\charlist`. Since character lists are ... lists, they can also be represented using the list notation. For instance, `[102,111,111]` is the same as `"foo"^^\charlist`. The main reason for the existence of the `\charlist` data type is that writing `"foo"^^\charlist` is a lot easier than consulting the ASCII table to find the numeric code for each character in order to write `[102,111,111]`. In addition, *expert* users can simply write `"foo"`, but this syntax is disabled by default because novice users tend to not understand this data structure and misuse it in various ways.

Do not confuse character lists with symbols: symbols are *not* lists and have a completely different representation and one should never use charlists in place of symbols. Character lists are useful in situations when it is necessary to parse the contents of a sequence of characters.

Escape sequences and Unicode are recognized inside $\mathcal{E}$RGO charlists similarly to $\mathcal{E}$RGO symbols. However, inside a charlist, a single quote character does not need to be escaped. A double quote character, however, needs to be escaped by another double quote, e.g., `"""foo"""`, or by a backslash.

Instance methods. All methods applicable to the `\list` datatype are also applicable to charlists. In addition, some methods applicable to the datatype `\string` also apply to character lists:

- `\charlist[|substring(\integer,\integer) => \charlist|]`
- `\charlist[|toUpper => \charlist|]`
- `\charlist[|toLower => \charlist|]`
- `\charlist[|concat(\charlist) => \charlist|]`

Class methods. The usual class-level methods likewise apply to charlists:

- `\charlist[=> isTypeOf(\object)]`
- `\charlist[toType(\charlist) => \charlist]`

39.14 Special Classes for Callable Literals

In addition to the above, $\mathcal{E}$RGO provides the following builtin meta-classes:

- `\modular` — this is a class for atomic formulas whose truth value depends on the module. This includes F-logic molecules, HiLog predicates, and Prolog predicates declared as `:- prolog` or `:- table`. For instance, `${a[b->c]}:\modular` is true.
- `\callable` — this class includes all atomic formulas that can possibly have a truth value in $\mathcal{E}$RGO. It includes `\modular` formulas as well as various `@\prolog` formulas and builtin primitives like `isinteger{...}`.
  `\callable` does exclude terms that, by their semantics, are not supposed to be truth-valued. These include HiLog terms (as opposed to HiLog predicates), datatype constants, and builtin class names. For instance, `${p(b,?X)}:\callable` and `isinteger{...}:\callable` are true, but `p(b,?X):\callable` and `"abc"^^foobar:\callable` are false. The former is false because `p(b,?X)` in a HiLog term, not a predicate, and the latter is false because `"abc"^^foobar` is a datatype constant.

In both cases, if the class member being tested is a variable then the result is `\undefined` unless this is a typed variable of the matching type. For instance, `?X:\modular` is `\undefined`, but `?X^^\callable:\callable` is `\true`.

39.15 User-defined Types

$\mathcal{E}$RGO also supports *user-defined types*. A user-defined type can be any atom, say `foo`, that is not reserved for the builtin types, i.e., is not prefixed with a "\". A literal of a user-defined type, such as `foo`, has the form "*some string*"`^^foo`, i.e., it has the same form as the built-in data types. However, $\mathcal{E}$RGO does not prescribe the contents of the data type and, at present, there is no hook to let the user plug in a personal parser to sort out which literals belongs to the data type and which do not.

Typed variables, introduced in Section 10, also work with user-defined types. Since `foo` may denote a class, a user-defined type, or both, such a variable, `?X^^foo` binds to *both* the members of the class `foo` and the literals of type `foo`. For example,

```
{a,b}:foo.

?- a=?X^^foo, b=?Y^^foo.      // true
?- "abc"^^foo=?X^^foo.        // true
?- d=?X^^foo.                 // false: d is not in class foo
?- "abc"^^moo=?X^^foo.        // false: type mismatch
?- "abc"^^foo=?X^^(foo;moo).  // false: don't use foo both as a class & type
?- insert{"abc"^^foo:foo}, "abc"^^foo=?X^^(foo;moo).  // true
```

However, it is not recommended to use the same symbol both as a class and as a user-defined type at the same time in the same module. For example, in the next-to-last example above, `foo` is used as a class in the expression `?X^^(foo;moo)`, so it will not unify with the literal of type `foo` unless `"abc"^^foo` is a member of the class `foo`. That last possibility is illustrated by the last line in the above example.

39.16 Regular Expression Matching

.

Regular expression matching is important in many applications. $\mathcal{E}$RGO provides convenient facilities for such operations based on the well-known PCRE package,[21] which supports Perl-style extended regular expressions, which are documented at this RL: http://perldoc.perl.org/perlre.html#Regular- The Web contains many sites that include various tutorials for this type of expressions.

On Linux and Mac, in order to use the pattern matching facility, PCRE must be installed before $\mathcal{E}$RGO is installed. On Windows, PCRE gets installed together with $\mathcal{E}$RGO, so the below instructions can be skipped.

To install PCRE on Linux, use the system's package manager and search for `pcre`. The names of the packages differ from distribution to distribution. On Debian, Ubuntu, and Mint, two packages need to be installed: `libpcre` and `libpcre-dev`; on Fedora, the package is `pcre-devel`; on the Mac, use Homebrew to install the `pcre` package.

If, on Linux or Mac, $\mathcal{E}$RGO was installed *first* and PCRE is installed afterwards, $\mathcal{E}$RGO suite must be reconfigured as follows: change to the top directory of the $\mathcal{E}$RGO suite (the one that has `Ergo` and `XSB` as subdirectory) and then execute

```
sh ./Ergo/ergosuite_config.sh
```

Follow the prompts.

[21] http://pcre.org/

Pattern matching commands. Like other string and symbol operations, $\mathcal{E}$RGO provides the pattern matching facility through its `\basetype` (or `\btp`) module. Three API calls are supported:

- *?Symbol* `[matchOne(`*?Pattern*`) -> `*?Result*`]`: find the first match.
  Here *?Symbol* must be a Prolog atom (which includes the data types `\url`, `\string`, or `\symbol`); it is a string to be matched against the *?Pattern* regular expression—also a Prolog atom that represents a Perl regular expression. **Important**: Perl regular expressions include various keywords that start with a backslash, e.g., `\d`. These backslashes must be *doubled* in *?Pattern*, i.e., `\\d`.

  ?Result is a HiLog term of the form `match(`*?Match*`,`*?Prematch*`,`*?Postmatch*`,`*?Submatches*`)`, that represents the *first* match. Here the first argument, *?Match*, is the substring of *?Symbol* that matches the pattern; *?Prematch* is the substring of *?Match* that precedes *?Match*; *?Postmatch* is the substring that follows *?Match*; and *?Submatches* is a list of submatches. These submatches are substrings of the pattern that match the *subpatterns* of *?Pattern* that are enclosed in parentheses (see the examples below).

- *?Symbol* `[matchAll(`*?Pattern*`) -> `*?Result*`]`: find all matches.
  Here *?Symbol* and *?Pattern* are as before and *?Result* is a *list* of the terms of the form `match(`*?Match*`,`*?Prematch*`,`*?Postmatch*`,`*?Submatches*`)` described above.

- *?Symbol* `[substitute(`*?Pattern*`,`*?Substitution*`) -> ?Result]`: substitute a string for all matching substrings.
  Here *?Symbol* and *?Pattern* are as before and *?Substitution* is a Prolog atom. *?Result* is *?Symbol* with each matching substring replaced with *?Substitution*.

Examples:

```
ergo> 'Hello12345-6789 NYwalk'[matchOne('(\\d{5}-\\d{4})\\ [A-Z]{2}')->?R]@\btp.
?R = match('12345-6789 NY',Hello,walk,['12345-6789'])

ergo> 'a@b.com@c.net@d.edu'[matchAll('[a-z]+@[a-z]+\.(com|net|edu)')->?Result]@\btp.
?Result = [match('a@b.com',,'@c.net@d.edu',[com]),
           match('com@c.net','a@b.','@d.edu',[net]),
           match('om@c.net','a@b.c','@d.edu',[net]),
           match('m@c.net','a@b.co','@d.edu',[net]),
           match('net@d.edu','a@b.com@c.',,[edu]),
           match('et@d.edu','a@b.com@c.n',,[edu]),
           match('t@d.edu','a@b.com@c.ne',,[edu])]

ergo> 'This is a Mississippi issue'[substitute(is,was)->?Result]@\btp.
?Result = 'Thwas was a Mwasswassippi wassue'
```

40 Cardinality Constraints

The earlier versions of F-logic made a distinction between functional and set-valued attributes and methods. The former were allowed to have only one value for any particular object and the latter could have any. In $\mathcal{E}$RGO, this dichotomy was replaced with the much more general mechanism of cardinality constraints. These constraints can be specified in signature expressions, which we have earlier used only to define types of attributes and methods. The extended syntax is as follows:

```
Cl[Meth{LowerBound..UpperBound}=>Cl2]
Cl[|Meth{LowerBound..UpperBound}=>Cl2|]
```

The first signature applies to object `Cl` and to its method `Meth`. The second expression is a class-level statement, so it applies to all members of `C1` (now viewed as a class) and to all subclasses of `C1`.

The lower and upper bounds in cardinality constraints can be non-negative integers, variables, or the symbol `*` (which denotes infinity). Variables can occur in signatures in rule bodies, which is useful especially when one wants to query the bounds of the cardinality constraints.

For example,

```
?- c1[m{2..?X}=>c2].
```

means that the method `m` of class `c1` must have at least 2 at most 3 values. Similarly,

```
c1[m{2..*}=>c2].
```

means that `m` has at least 2 values; there is no upper bound.

We can query the specified cardinality constraints by putting variables in the appropriate places. For instance, consider the following knowledge base loaded into module `foo`:

```
C[|m{3..*}=>B|].
C[m{?x..1}=>B] :- ?x=0.

v:C.
C2::C.
v2:C2.

C[m->{1,2}].
v[m->2].
C2[|m->{1,2,3}|].
```

Portions Copyright © 2013–2017 Coherent Knowledge

The query

```
?- ?C[?M{?L..?H}=>?]@foo.
```

will yield three solutions:

```
?C = C
?M = m
?L = 0
?H = 1

?C = v
?M = m
?L = 3
?H = *

?C = v2
?M = m
?L = 3
?H = *
```

Note that the objects `v` and `v2` are in the answer to the query because they inherited the cardinality constraint from the first clause, `C[|m3..*=>B|]`.

On the other hand, the query

```
?- ?C[|?M{?L..?H}=>?|]@foo.
```

has two solutions:

```
?C = C
?M = m
?L = 3
?H = *

?C = C2
?M = m
?L = 3
?H = *
```

Class `C` is in the result because the constraint is specified explicitly and `C2` is in the result because it inherited the constraint from `C`.

41 Exception Handling

$\mathcal{E}$RGO supports the common catch/throw paradigm through the primitives `catch{?Goal, ?Error, ?Handler}` and `throw{?Error}`. Here `?Goal` can be any $\mathcal{E}$RGO query, `?Error` is a HiLog (or Prolog) term, and `?Handler` is a $\mathcal{E}$RGO query that will be called if an exception that unifies with `?Error` is thrown during the execution of `?Goal`. For instance,

```
%someQuery(?Y) :- ?Y[value->?X], ?X > 0, %doSomethingUseful(?X).
%someQuery(?Y) :- ?Y[value->?X], ?X =< 0, throw{myError('?X non-positive', ?X)}.

?- %p(?Y), catch{%someQuery(?Y), myError(?Reason,?X), %handleException(?Reason,?X)}.

%handleException(?Reason,?X) :-
        format('~w: ?X=~w~n',[?Reason,?X])@\prolog(format), \false.
```

The `catch` construct first calls the query `%someQuery/1`. If `?X` is positive then nothing special happens, the query executes normally, and `catch{...}` has no effect. However, if `?X` turns out to be non-positive then the query throws an exception `myError('?X non-positive', ?X)`, where `?X` is bound to the non-positive value that was deemed by the logic of the program to be an exceptional situation. The term thrown as an exception is then unified with the term `myError(?Reason,?X)` that was specified in `catch{...}`. If the two terms do not unify (*e.g.*, if the error specified in `catch` was something like `myError(foo,?X)`) then the exception is propagated upwards and if the user does not explicitly catch it, the exception will eventually be caught by the $\mathcal{E}$RGO command loop. In the above example, however, the thrown term and the exception specified in `catch` unify and thus `%handleException/2` is called with `?Reason` and `?X` bound by this unification.

The queries `?Goal` and `?Handler` in the `catch{...}` primitive can be frames, not just predicates. However, `?Error` — both in `catch` and in `throw` — must be HiLog or Prolog terms. No frame literals are allowed inside these terms unless they are reified. That is, `myError('problem found', a[b->c])` will result in a parser error, but an exception of the form `myError('problem found', ${a[b->c]})` is correct because the frame is reified.

Some exceptions are thrown by $\mathcal{E}$RGO itself, and applications might want to catch them:

- `'_$ergo_undefined'(?MethodSpec,?ErrMsg)` — thrown when undefinedness checking is in effect (see Section 43.1) and an attempt is made to execute an undefined method or predicate. The first argument in the thrown exception is a specification of the undefined predicate or the method that caused the exception. The second argument is the error message.
- `'_$ergo_abort'` or `'_$ergo_abort'(?Message)` — thrown when $\mathcal{E}$RGO encounters other kinds of errors. This exception comes in two flavors: with an error message and without.

A rule can also throw this exception when immediate exit to the top level is required. The safest way to do so is by calling `abort(?Message)@\sys`, as explained in Section 46.3.

These exceptions are defined by $\mathcal{E}$RGO under the symbolic names `ERGO_UNDEFINED_EXCEPTION` and `ERGO_ABORT`. When a user application needs to catch these errors we recommend that the applicable files include `flora_exceptions.flh` and use the above symbolic names. For instance,

```
#include "flora_exceptions.flh"
?- ..., catch{myQuery(?Y),
               ERGO_ABORT(ERGO_UNDEFINED_EXCEPTION(?MethSpec,?Message),?_),
               myHandler(?MethSpec)}.
?- ..., catch{yourQuery(?Y),ERGO_ABORT(?Message,?_),yourHandler(?Message)}.
```

The `catch{...}` primitive can also catch exceptions thrown by the underlying Prolog system. For this to happen you need to know the format of the exceptions thrown by Prolog (which can be found in the manual). These exceptions have the form

`error(`*errortype*`(`*arguments*`),context(`*Message*, *Backtrace*`))`

However, $\mathcal{E}$RGO aims to intercept all Prolog exceptions and contextualize them in the appropriate $\mathcal{E}$RGO terms and any non-caught Prolog exception should be treated as an omission to be fixed in the next release.

42 The Compile-time Preprocessor

$\mathcal{E}$RGO supports a C-style preprocessor, which is invoked during the compilation. The most important commands are

```
#define variable value
#define macro(arg_1,...,arg_n) expression
#ifdef variable
#ifndef variable
#else
#endif
#include "file"
```

There are many features that go beyond the C preprocessor such as the tests

```
#if exists("file")
#if !exists("file")
```

and many others. For example, it is possible to enable macro substitution inside quotes. An advanced user is referred to the XSB manual where this preprocessor, called gpp, is described in an appendix to Manual 1.

43 Debugging User Knowledge bases

$\mathcal{E}$RGO comes with an interactive, Prolog-style debugger, which is described in Appendix B. The compiler makes many useful checks, such as the occurrence of singleton variables, which is often an error (see Section 7.1). When a problem is deemed serious enough, errors are reported.

The **most important rule** in debugging $\mathcal{E}$RGO knowledge bases is: **never ignore any kind of warnings** issued by system. This golden rule actually applies to programming in *any* language. In addition, it is possible to tell $\mathcal{E}$RGO to perform various run-time checks, as described below.

43.1 Checking for Undefined Methods and Predicates

$\mathcal{E}$RGO has support for checking for the invocation of undefined methods and predicates at run time. This feature can be of great help because a trivial typo can cause a method/predicate call to fail, sending the user on a wild goose chase after a hard-to-find bug. It should be noted, however, that enabling these checks can slow the runtime by up to 2 times (typically about 50% though), so we recommend that this be done during debugging only.

To enable runtime checks for undefined invocations, $\mathcal{E}$RGO provides two methods, which can be called at any time during execution (and thus enable and disable the checks dynamically):

```
?- Method[mustDefine(?Flag)]@\sys.
?- Method[mustDefine(?Flag(?Module))]@\sys.
```

The argument `?Flag` can be `on`, `off`, or it can be a variable. The argument `?Module` must be a valid loaded $\mathcal{E}$RGO module name or it can be a variable. When the flag argument is `on`, the first method turns on the checks for undefinedness in all modules. The second method does it in a specific module. When the flag argument is off, the above methods turn the undefinedness checks off globally or in a specific module, respectively.

When either `?Flag` or `?Module` (or both) is a variable, the above methods do not change the way undefined calls are treated. Instead, they query the state of the system. For instance, in

```
?- Method[mustDefine(?Flag)]@\sys.
?- Method[mustDefine(?Flag(foo))]@\sys.
?- Method[mustDefine(on(?Module))]@\sys.
```

the first query binds `?Flag` to `on` or `off` depending on whether the checks are turned on or off globally. The second query reports on the state of the undefinedness checks in $\mathcal{E}$RGO module `foo`, while the third query tells in which modules these checks are turned on.

In addition to turning on/off the checks for undefinedness on the per-module basis, $\mathcal{E}$RGO provides a way to turn off such checks for individual predicates and methods:

```
?- Method[mustDefine(off,Predicate/Method-spec)]@\sys.
```

For example,

```
?- Method[mustDefine(off,?(?)@foo)]@\sys.
```

specifies that all undefinedness errors of predicates that unify with `?(?)@foo` are ignored, provided that `foo` is a loaded module. Note that the module must *always* be specified. For instance, to ignore undefinedness checking in the current module, use

```
?- Method[mustDefine(off,?(?)@ \@)]@\sys.
```

The use of the current module symbol `\@` is essential in this example. Omitting it is probably not what you want because the module specification `\sys` propagates inward and so the above statement (without the `\@`) would turn off undefinedness checks in module `\sys` instead of the current module.

One can also turn undefinedness checks off in *all* modules by putting a variable in the module position:

```
?- Method[mustDefine(off,?(?)@ ?Mod)]@\sys.
```

However, this must *not* be an anonymous variable like `?`, `?_`, or a don't care variable like `?_Something`. If one uses an anonymous or a don't-care variable then undefinedness checks will be ignored only in some randomly picked module.

A pair of parentheses is needed when multiple predicates/methods are listed in one call.

```
?- Method[mustDefine(off,(?:class@foo, ?[%?]@ \@))]@\sys.
```

The undefinedness exception in $\mathcal{E}$RGO can be caught using $\mathcal{E}$RGO's `catch{...}` built-in. For instance, suppose `FOO` is a predicate or a frame whose execution might trigger the undefinedness exception. Then we can catch this exception as follows:

```
#include "flora_exceptions.flh"

  ..., catch{FOO, \FLUNDEFEXCEPTION(?Call,?ErrorMessage), handler(?Call)}, ...
```

Here `ERGO_UNDEFINED_EXCEPTION` is the exception name defined in the $\mathcal{E}$RGO system file `flora_exceptions.flh`, which must be included as shown. The predicate `handler/1` is user-defined (can be a frame as well), which will be called when an undefinedness exception occurs. The variable `?Call` will be bound to an internal representation of the method or predicate call that caused the exception. For instance, if we define

```
  handler(?_) :- !.
```

then the undefinedness exception that occurs while executing *FOO* will be ignored and the call to *FOO* will succeed.

Undefinedness checks and meta-programming. We should note one subtle interaction between these checks and meta-programming. Suppose the user knowledge base does not have any class membership facts and the undefinedness checks are turned on. Then the meta-query

```
    ?- a:?X.
```

would cause the following error:

++Error[Ergo]: Undefined class ?:? in user module main

Likewise, if the knowledge base does not have any method definitions, the query `?- ?X[?Y->?Z].` would cause an error. This might not be what one expects because the application in question might be *exploring* the schema or the available data, and the intention in the above cases might be to fail rather than to get an error.

One way of circumventing this problem is to insert some "weird" facts into the knowledge base and special-case them. For instance, one could add the following facts to silence the above errors:

```
    ads_asd_fsffdfd : ads_asd_fsffdfd.
    ads_asd_fsffdfd[ads_asd_fsffdfd -> ads_asd_fsffdfd].
```

The user can then arrange the things so that anything that contains `ads_asd_fsffdfd` would be discarded.

Another way to circumvent the problem is to turn the undefinedness checks off temporarily. For instance, suppose the query `?- ?X:a` causes an unintended undefinedness error in module `foo`. Then we can avoid the problem by posing the following query instead:

```
?- Method[mustDefine(off(foo))]@\sys,
   ?X:a,
   Method[mustDefine(on(foo))]@\sys.
```

A more selective way to circumvent this problem is to turn off undefinedness checking just for the offending classes. For instance,

```
?- Method[mustDefine(off,?:a@ \@)]@\sys.
```

The fourth way is to deal with the exception is to use $\mathcal{E}$RGO's `catch{...}` built-in (note the curly braces):

```
#include "flora_exception.flh"

?- catch{?X:a, \FLUNDEFEXCEPTION(?,?)@\prolog, true}.
```

Undefinedness checks and update operators. Although undefinedness checking can be turned on and off at will, it cannot always capture all cases correctly. Namely, if an insert or delete statement is executed while undefinedness checking is off, the corresponding methods will not be properly captured and spurious undefinedness errors might result. For instance, if

```
?- insert{a[meth->b]}, delete{a[meth->b]}.
?- Method[mustDefine(on)]@\sys.
```

are executed then the query `?- a[meth->b]` will cause the undefinedness error. However,

```
?- insert{a[meth->b]}, delete{a[meth->b]}.
?- Method[mustDefine(on)]@\sys.
?- a[meth->b].
```

will not flag the method `meth` as undefined.

43.2 Type Checking

Although $\mathcal{E}$RGO allows specification of object types through signatures, type correctness is not checked automatically. A future version of $\mathcal{E}$RGO might support some form of run-time type checking. Nevertheless, run-time type checking is possible even now, although you should not expect any speed here and this should be done during debugging only.

Run-time type checking is possible because F-logic naturally supports powerful meta-programming, although currently the knowledge engineer has to do some work to make type checking happen. For instance, one can write simple queries to check the types of methods that might look suspicious. Here is one way to construct such a type-checking query:

```
type_error(?O,?M,?R,?D) :-
       %% Values that violate typing
       ?O[?M->?R], ?O[?M=>?D], \naf ?R:?D
       ;
       %% Defined methods that do not have type information
       ?O[?M->?R], \naf ?O[?M=>?_D].
?- type_error(?Obj,?Meth,?Result,?Class).
```

Here, we define what it means to violate type checking using the usual F-logic semantics. The corresponding predicate can then be queried. A "no" answer means that the corresponding attribute *does not* violate the typing rules.

In this way, one can easily construct special purpose type checkers. This feature is particularly important when dealing with *semistructured* data. (Semistructured data has object-like structure but normally does not need to conform to any type; or if it does, the type would normally cover only certain portions of the object structure.) In this situation, one might want to limit type checking only to certain methods and classes, because other parts of the data might not be expected to have regular structure.

Note that in a multi-module knowledge base, the module information should be added to the various parts of the above type-checker. It is reasonable to assume that the schema information and the definition for the same object resides in the same module (a well-designed knowledge base is likely to satisfy this requirement). In this case, a type-checker that take the module information into account can be written as follows:

```
type_error(?O,?M,?R,?D) :-
       %% Values that violate typing
       (?O[?M->?R], ?O[?M=>?D])@?Mod1, \naf ?R:?D@?Mod1
       ;
       %% Defined methods that do not have type information
       (?O[?M->?R], \naf ?O[?M=>?_D])@?Mod1.
?- type_error(?Obj,?Meth,?Result,?Class).
```

We should note that type-checking queries in $\mathcal{E}$RGO are likely to work only for "pure" queries, i.e., ones that do not involve built-ins like arithmetic expressions. Built-ins pose a problem because they typically expect certain variable binding patterns when these built-ins are called. This assumption may not hold when one asks queries as general as `type_error`.

To facilitate all these checks, $\mathcal{E}$RGO provides a method, `check`, in class `Type` of module `\typecheck` (which can be abbreviated to `\tpck`). Its syntax is:

```
?- Type[check(?Atom,?Result)]@\typecheck.
?- Type[check(?Atom,?Result)]@\tpck.
```

The `?Atom` variable must be bound to a base frame, as described below. `?Result` gets bound to the evidence of type violation (one or two atoms that violate the typing constraint).

- If `?Atom` is of the form `?[?Meth->?]@?Mod` then all type constraints for `?Meth` are checked in module `?Mod`. Missing types (semistructured data) are flagged. If `?Mod` is an unbound variable, then the constraints are checked in all modules. `?Meth` can also be a variable. In this case all non-transactional methods will be checked.
- If `?Atom` is of the form `?[?Meth=>?]@?Mod` then the type constraints for `?Meth` are checked in module `?Mod` but missing types (semistructured data) are ignored. As before, `?Mod` and `?Meth` can be unbound variables.
- If `?Atom` is of the form `?[|?Meth->?|]@?Mod` then only the consistency between `->` and `=>` is checked and only for frames that are statements about classes as a whole, i.e., the frame formulas of the form `?[|?Meth->?|]@?Mod` and `?[|?Meth=>?|]@?Mod`. The `obj[...]`-style frames are ignored. Missing types (semistructured data) are flagged.
- If `?Atom` is of the form `?[|?Meth=>?|]@?Mod` then again only the consistency between `?[|?Meth->?|]@?Mod` and `?[|?Meth=>?|]@?Mod` is checked, but missing types are <u>not</u> flagged.

For example, if our knowledge base consists of:

```
a[b->c].
a[b=>d].
c:d.
```

then the query will fail, as the typing is correct:

```
?- Type[check(?[?Meth->?],?Result)]@\typecheck.
```

But if, in addition, we had

```
a[b->e].
a[foo->e].
```

then the above query would yield multiple evidences of type inconsistency:

```
?Result = [(${a[b -> e]}, ${a[b => d]}), ${a[foo -> e]}]
```

The first item in the list (the pair inside parentheses) means that the frame `a[b -> e]` violates the type constraint specified by the signature `a[b => d]`. The second item means that the frame `a[foo -> e]` does not have a corresponding signature. On the other hand,

```
?- Type[check(?[?Meth=>?],?Result)]@\typecheck.
```

will yield only the first evidence because `a[foo->e]` does not violate any typing constraint for semistructured data.

If the object position in the first argument of `check` is bound then this object is treated as a class and only the objects in that class will be type-checked. For instance, if we also had

```
q[foo->bar].
q:qq.
```

in our knowledge base then the query

```
?- Type[check(qq[?Meth->?],?Result)]@\typecheck.
```

will return one evidence of type inconsistency:

```
?Result = [${q[foo -> bar]}]
```

because `q` is the only object in class `qq` that has type violations.

An easy way to remember which type of constraint represents what kind of type checking is to think that `=>` represents typing and, therefore the `=>`-style constraints mean that only the methods that have typing information will be type-checked. The `->`-style constraints, on the other hand, mean that all methods will be checked—whether they have signatures or not. Similarly, `...[|...|]`-style constraints indicate that only information about classes as a whole will be type-checked, while information specified explicitly for individual objects will not be. In contrast, `...[...]`-style constraints indicate that all type information will be verified.

43.3 Checking Cardinality of Methods

$\mathcal{E}$RGO does not automatically enforce the cardinality constraint specified in method signatures. However, the `type` system module in $\mathcal{E}$RGO provides methods for checking cardinality constraints for methods that have such constraints declared in their signatures.

In practice as well as in theory things are more complicated, however. First, it is theoretically impossible to have a terminating query that will flag a violation of a cardinality constraint if and only if one exists.

In practice, the constraint checking methods in the `type` system library may trigger run-time errors if there are rules that use non-logical features or certain built-ins in their bodies. Therefore, in practice, the user should do constraint-checking methods only for purely logical methods. Cardinality constraints declared for methods that are defined with the help of non-logical features should be used for documentation only.

The above problems aside, in $\mathcal{E}$RGO it is easy to verify that a particular method satisfies a cardinality constraint. For instance, if method `foo` is declared as

```
someclass[|foo {2..3}=> sometype|].
```

then to check that the cardinality constraint is not violated, one can ask the following query:

```
?- Cardinality[check(?Obj[foo =>?])]@\typecheck.
```

If no violations are found, the above query will *fail*. If there *are* violations of this constraint then `?Obj` will get bound to the objects for which the violation was detected. For instance, consider the following knowledge base:

```
cl[|foo {2..3}=> int|].
c::cl.

o1:c.
o2:c.
o3:c.

o1[foo->{1,2,3,4}].
o3[foo->{3,4}].

c[|foo -> 2|].
cl[|foo -> {3,4,5}|].
```

Then the query

```
?- Cardinality[check(?O[foo=>?])]@\typecheck.
```

will return `?O = o1` and `?O = o2` because `o1` has a method `foo` with four values while at most 3 are allowed according to the signature. The object `o2` is returned because `foo` has no values for

that object, while at least 2 are required. The object `o3` is not returned because it does not violate the constraint. Similarly, the query

```
?- Cardinality[check(?O[|foo=>?|])]@\typecheck.
```

will return `?O = c` because the method `foo` has only 1 value for that class, while at least two are required by the signature. The class `c1` is not returned because it does not violate the constraint.

In general, the allowed forms of the method `check` in class `Cardinality` are as follows. The argument is always an atomic signature frame (no need to specify reification ${...}). The method type of the signature can be only `=>`, but the frames can have the `...[...]`-style or `...[|...|]`-style. The former checks the cardinality constraints of object methods, while the latter checks cardinality constraints only for default values of the methods.

- `Cardinality[check(?Object[?Method => ?])]@\typecheck`
  Checks cardinality constraints for `?Method` of type `=>` in the current module. That is, whether there are instances of the literal `?Object[?Method -> ?Val]` that violate a cardinality constraint imposed by some signature of the form `?Object[?Method{?Low..?High}=>?Type]` (which may be a derived signature).

- `Cardinality[check(?Obj[|?Method =>?|]@?Module)]@\typecheck`
  Checks cardinality constraints for the default values of `?Method` in module `?Module`. If `?Module` is unbound and a cardinality constraint violation is detected in some module then `?Module` is bound to that module. That is, it is a check for whether there are instances of the literal `?Object[|?Method -> ?Val|]` that violate a cardinality constraint imposed by some signature of the form `?Object[|?Method{?Low..?High}=>?Type|]` (which may be a derived signature).

- `Cardinality[check(?Obj[?Method {?LoBound..?HiBound} => ?]@?Mod)]@\typecheck`
  Like the previous query, but the variables `?LoBound` and `?HiBound`, which must be unbound variables, can be used to indicate which bounds are violated. If the lower bound is violated, then `?LoBound` will be bound to the violated lower bound; otherwise, it is bound to `ok`. If the higher bound is violated, then `?HiBound` is bound to the violated higher bound; otherwise it is bound to `ok`.

  If `?Mod` is unbound then it will be bound to the module(s) in which the cardinality constraint is violated.

  For instance, for the above knowledge base, the query

  ```
  ?- Cardinality[check(?O[|foo {?Low..?High} => ?|]@?Module)]@\typecheck.
  ```

will bind `?O` to `c`, `?Mod` to `main`, `?Low` to 2, and `?High` to `ok`. Indeed, only the lower bound of the cardinality constraint `c[|foo {2..3}=> int|]` (which was inherited from `cl`) is violated by the class `c`.

```
?- Cardinality[check(?O[foo {?Low..?High} => ?])]@\typecheck.
```

will return the following results:

```
?O = o1
?Low = ok
?High = 3

?O = o2
?Low = 2
?High = ok
```

43.4 Logical Assertions that Depend on Transactional and Non-logical Features

On page 145 we mentioned the potential problems when tabled predicates or frames depend on updates. A similar problem arises when such statements depend on non-logical features, such as `var(...)` or on statements that have side effects, such as I/O operations (e.g., `write('foo bar')@\prolog`). Since tabled statements in $\mathcal{E}$RGO are considered purely logical, one cannot assume that the evaluation happens in the same way as in Prolog. For instance, consider the following knowledge base:

```
?O[bar] :- ?O:foo.
?O:foo :- writeln('executed')@\prolog.
?- abc[bar].
```

Despite what one might expect, the above query will cause "executed" to be printed twice — once when `abc[bar]` will be proved for the first time and once when the system will attempt some other way of proving `abc[bar]`. (The system may not realize that the second proof is not necessary.) In general, transactional and side-effectfull statements might be executed even if the attempt to prove the statement in the rule head ultimately fails.

$\mathcal{E}$RGO issues warnings when it finds that a tabled predicate depends on non-logical or side-effectfull statements, but it does not warn about all Prolog predicates of this kind. Therefore, caution needs to be exercised in specifying purely logical statements and warnings should not be ignored. If you are certain that a particular suspicious dependency is harmless, use the `ignore_depchk` directive to suppress the warning.

43.5 Explaining Answers

When Ergo runs as an Ergo Studio application, the user can request explanations for the answers either by running a query using the *Query Tool* or on command line, through the *Studio Listener*. In the first case, explanations can be requested by double-clicking on any particular answer (or by choosing an appropriate menu item—Why? or Why not?). In the latter case this can be done by running the query

```
why{ Goal }.
```

in the Studio Listener. In both cases, a window pops up, which explains how answers were obtained or why they were *not* obtained. Statements marked in green in the explanation window are proved to be true, those displayed in the blue color are false. Red means inconsistency (both p and `\neg` p can be proved), and gray means undefined (neither p nor `\naf` p can be proven). Finally, brown color denotes facts that should have been true, but were defeated.

Explanations can be combined with the ErgoText feature, described in Section 47, to generate textual explanations in English instead of the more technical $\mathcal{E}$RGO syntax. This feature is described in Section 47.7 and is called *text generation*. The way to do this is to create Enlish sentence templates for the subgoals that one desires to appear in explanations as English sentences. See Section 47.7 for more details.

43.6 Examining Tables

Sometimes it is useful to be able to examine the tables that XSB has generated while answering queries. To this end, $\mathcal{E}$RGO provides the following library predicates:

```
\tabledump(File,AtomicGoal)
\tabledump(File,AtomicGoal,Option)
```

The `File` argument specifies the file in which to place the results. The results are in the $\mathcal{E}$RGO format, as explained below. If `File` is `userout` then the results go to the standard output. `AtomicGoal` is a HiLog predicate or an atomic frame (e.g., `?[?->foo]`). It specifies the subgoals for which tables are being requested. `Option` is the option selected. Currently three options are supported: `summary` (minimalist output that summarizes the overall statistics of tables), `subgoals` (more details about individual subgoals). The third option, `answers`, will output full details, including the information about each called subgoal and all answers to all subgoals.

The first (binary) form of `\tabledump` above is equivalent to `\tabledump(File,AtomicGoal,summary)`.

In all three cases, `\tabledump` generates information about the tables for subgoals that are subsumed by `AtomicGoal`. If `AtomicGoal` is a variable, information is displayed about all tables.

When the `summary` option is used, the information is displayed in the following format:

```
AtomicGoal[total_subgoals->..., total_subgoal_answers->...].
```

For instance,

```
?- \tabledump(userout,?).

${?A(?B)}[total_subgoals->1, total_subgoal_answers->0].
${?A(?B,?C)}[total_subgoals->3, total_subgoal_answers->2].

?- \tabledump(userout,p(2,?)).

${p(2,?A)}[total_subgoals->1, total_subgoal_answers->1].
```

If the third option is used (`answers`), then in addition to the output produced for the `summary` option the system will show the individual subgoals subsumed by `AtomicGoal` and the answers to each:

```
AtomicGoal[
   total_subgoals->1,
   subgoal_details->{Subgoal1[total_answers->..., answer_list->[answer1,...],
                     Subgoal2[total_answers->..., answer_list->[answer2,...],
                     ...}].
```

For instance, the following might be produced for the two earlier requests, if the `details` option is specified :

```
?- \tabledump(userout,?,answers).

${?A(?B)}[total_subgoals->1, total_subgoal_answers->0].
${?A(?B)}[total_subgoals->1,
          subgoal_details->{${q(?_h34)}[total_answers->0, answer_list->[]]}].
${?A(?B,?C)}[total_subgoals->3, total_subgoal_answers->2].
${?A(?B,?C)}[total_subgoals->3,
             subgoal_details->{${p(2,b)}[total_answers->1, answer_list->[${p(2,b)}]],
                               ${p(1,a)}[total_answers->1, answer_list->[${p(1,a)}]],
                               ${p(?_h65,?_h67)}[total_answers->0, answer_list->[]]}].

?- \tabledump(userout,p(2,?),answers).
```

```
${p(2,?A)}[total_subgoals->1, total_subgoal_answers->1].
${p(2,?A)}[total_subgoals->1,
           subgoal_details->{${p(2,b)}[total_answers->1, answer_list->[${p(2,b)}]]}].
```

If the second option is used (*subgoals*), then the output is similar to the third option, but the `answer_list` part is not shown. In this case, the output is slightly smaller. However, it should be kept in mind that for large knowledge bases with large numbers of answers table dumps can be huge (hundreds of megabytes) and it can take considerable time to dump these tables. In this case only the first option (and maybe the second, if you must) is recommended.

43.7 Examining Incomplete Tables

Sometimes it becomes necessary to examine incomplete tables (i.e., tables to subgoals that have not yet been completely evaluated) in the middle of execution or upon exception.

To get a dump of all incomplete tables in the middle of the computation, one has to insert the predicate `\dump_incomplete_tables` in an appropriate place in user's rule bodies. For instance,

```
r(0,?):- !, \dump_incomplete_tables(temp).
r(3,?A):- r(5,?A).
r(?N,?A):- ?N1 is ?N - 1, r(?N1,?A).
?- r(5,foo(a)).
```

will put the following into the file `temp`:

```
${r(5,foo(a))}[scc->1].
${r(4,foo(a))}[scc->1].
${r(3,foo(a))}[scc->1].
${r(2,foo(a))}[scc->2].
${r(1,foo(a))}[scc->3].
${r(0,foo(a))}[scc->4].
```

It says that there are four strongly connected components of subgoals, numbered 1, 2, 3, and 4. All of these subgoals are still waiting to be fully computed, but at the moment their truth values are still unknown.

More often, though, one might need to examine incomplete tables after an exception, if the user suspects that the exception has something to do with tabled subgoals. Such a table dump looks

exactly like the dump produced by `\dump_incomplete_tables/1` but it is requested differently. First, one must execute the query

```
?- \set_dump_incomplete_tables_on_exception.
```

Then, after an exception took place, the user should execute the query

```
?- \dump_incomplete_tables_after_exception(filename).
```

For instance,

```
?- \set_dump_incomplete_tables_on_exception.
q(0,?):- !, abort@\sys.
q(3,?A):- q(5,?A).
q(?N,?A):- ?N1 is ?N - 1,q(?N1,?A).
?- q(5,foo(a)).
?- \dump_incomplete_tables_after_exception(temp).
```

The dump of the tables that were incomplete at the time of the abort will be in the file `temp` and will have the same structure as before:

```
${q(5,foo(a))}[scc->1].
${q(4,foo(a))}[scc->1].
${q(3,foo(a))}[scc->1].
${q(2,foo(a))}[scc->2].
${q(1,foo(a))}[scc->3].
${q(0,foo(a))}[scc->4].
```

$\mathcal{E}$RGO, in addition, has a much more useful primitive, `showgoals{...}`, which can show the information about incomplete computations without the need to modify the program rules and in a more focused way.

43.8 Non-termination Analysis

It is a fact of life that an $\mathcal{E}$RGO query might not terminate. In fact, due to the expressive power of $\mathcal{E}$RGO, it is *undecidable* whether a query will terminate or not. However, $\mathcal{E}$RGO comes with a powerful analysis tool called *Terminyzer* (for non-*Termin*ation Anal*yzer*), which can greatly help in determining whether a given query might not terminate.

Terminyzer examines the trace of the query run and tries to detect symptoms of non-termination. Since the problem is fundamentally undecidable, it may not find any problems even if they exist and it may report a non-terminating loop falsely. In some cases, the analysis might turn out to be too time consuming and thus not feasible. Details of the approach are given in [10, 11]. It is guaranteed that Terminyzer will find a problematic loop *given enough time*. But the time we have is limited and it is pointless to wait beyond 10 or 15 minutes. When Terminyzer reports a loop, however, it means that the query will indeed not terminate or it might terminate after a very long period of time and the reported loop may be a computational bottleneck. It is worth to examine the reported loop and see if you can spot any inefficiency or redundancy.

Non-termination can happen due to two reasons:

- the execution might be invoking some subgoals repeatedly, with larger and larger arguments; or
- some subgoal has an infinite number of answers.

Here is a typical example of a non-terminating query due to the first reason:

```
p(a).
q(b).
p(?X) :- q(f1(?X)).
q(?X) :- p(f2(?X)).
?- p(?X).
```

Clearly, the two rules on lines 3 and 4 repeatedly call each other with increasingly growing arguments f1(...), f1(f2(...)), f1(f2(f1(...))) etc., and Terminyzer will dutifully report the following:

```
*** Report: subgoals that form a possibly infinite call loop ***
    p(f2(f1(f2(f1(?A)))))@main
        in rule on line 4, file test1.ergo, module main
    q(f1(f2(f1(f2(f1(?A))))))@main
        in rule on line 3, file test1.ergo, module main
```

The user can then examine the reported subgoals in the rules that occur at the specified lines and will easily see the problem. In general, several subgoals in several rules can be involved, and the larger the cycle the harder it is to understand what is going on. However, this is certainly much easier than manually finding and examining a sometimes astronomical number of possible loops in a knowledge base that has a few dozens to thousands of of rules.

If the cause of non-termination is sue to repeated calls to the same subgoals with increasingly large arguments, a simple remedy may sometimes help; it is called *call abstraction* and can be

requested using the `setruntime` directive described in Section 43.12.2. Call abstraction is a technique that the underlying XSB inference engine can use to prevent calls with increasingly deeply nested arguments by replacing deeply nested subterms with variables. In our example, subgoal abstraction at term size 5 will take the call such as

```
p(f1(f2(f1(f2(f1(?X))))))
```

and transform it into an equivalent call

```
p(f1(f2(f1(?Y)))), ?Y = f2(f1(?X))
```

The effect is that the new call has no more than 5 symbols and the first cause of non-termination is impossible. In our case, issuing the command

```
?- setruntime{goalsize(5,abstract)}.
```

prior to running the query `p(?X)` would terminate the query with the answer

```
?X =  a
```

Call abstraction is no a panacea, however. In some cases, it may transform a non-terminating program with an infinite loop of subgoal calls into a non-terminating program in which some subgoal has an infinite number of answers. That is, the first cause of non-termination will morph into the second. To see if call abstraction might help your query, check the box "Abstract large subgoals?" when you start Terminyzer (which is in the sequel).

If the cause of non-termination is an infinite number of answers to some subgoal or a set of subgoals, Terminyzer will try to identify such subgoals. To illustrate, consider the following knowledge base and a query:

```
p(a).
q(b).
p(f1(?X)) :- q(?X).
q(f2(?X)) :- p(?X).
?- p(?X).
```

This is similar to the previous query, but the function symbols are in the rule heads. It is easy to see that the query has the answers `p(f1(a))`, `p(f1(f2(f1(a))))`, `p(f1(f2(f1(f2(f1(a))))))`, etc., ad infinity. In this case, Terminyzer will report

```
*** Report: subgoals that form a possibly infinite answer-producing pattern ***
    q(?A)@main
        in rule on line 3, file test2.ergo, module main
    p(?A)@main
        in rule on line 4, file test2.ergo, module main
```

What this means is that answers to the subgoal `q(?A)` that feeds into the rule on line 3 might cause production of new answers for the head `p(...)`. But the next line in the report says that production of new answers for `p(...)` might cause production of new answers for `q(...)` because `p(...)` feeds its answers into the rule on line 4, which generates answers for its head `q(...)`.

In the rare case when you find that the knowledge base is correct and the infinite set of answers is inevitable, the computation will not terminate and one will not get any answers. However, not everything is lost. The engine also supports *answer abstraction* — a technique similar to call abstraction. Since we cannot cheat the nature, something will have to give, and what we have to give is precision. Executing both of these commands (where the sizes are chosen for the sake of an example):

```
?- setruntime{goalsize(7,abstract)},
   setruntime{answersize(3,abstract)}.
```

will give the following answers:

```
?X = a
?X = f1(b)
?X = f1(f2(a))
?X = f1(f2(f1(b)))
?X = f1(f2(f1(?(?))))  - undefined
```

what this means is that we got four certain answers and the instances of the fifth answer are *undefined*. This is because some of them are true answers and some are not.

Answer abstraction together with call abstraction is guaranteed to terminate any query that has the following properties:

- uses only tabled (non-transactional predicates and methods)
- does not use arithmetic operators
- does not perform updates of any kind.

To see how Terminyzer works in practice, start $\mathcal{E}$RGO and load the knowledge base. Then type

```
terminyzer{}.
```

at the command prompt or select "*Use Terminyzer*" from the *Debug* menu, if using the Studio.

If the monitor has graphical capabilities, a window will pop up with suggested values of various "tripwires." A tripwire is a condition that will cause XSB to pause if certain events happen. Terminyzer uses the following tripwires:

- *timer*: the query computation will pause if a timeout occurs
- *goal size limit*: if the computation generates a subgoal of the size exceeding the specified number, the computation will pause; the Terminyzer window also lets the user check a box to request that call abstraction should be used.
- *answer size limit*: the computation will pause if a subgoal answer gets generated that exceeds the given limit.
- *limit on the number of currently unfinished subgoals*: the computation reaches the state where the number of unfinished subgoals (that it is still trying to evaluate) exceeds the given limit.

When the computation pauses, various messages will be printed for your perusal and Terminyzer will start its analysis. When it is done, a window will pop up telling you the results. It may find the problem on the first try and you should then try to find the problem in the knowledge base based on Terminyzer's report. It is also possible that no problems were found. In that case, you might want to ask Terminyzer to try harder.

When Terminyzer finishes its analysis, it asks the user whether to continue the analysis further or to stop. If you would like to continue, respond appropriately and the analysis will continue after you resume the query. (The query is not resumed by itself, giving the user an opportunity to change mind or to examine the reported statistics and to use other tools.) When the query is resumed, the computation will pause again if any of the tripwires is tripped. Note that if the pause happened due to the times then the timer is reset to the same value as before. If the pause was due to any other reason, say, reaching the goal size limit, that limit is automatically increased by 20%. (Any other tripwire limit is not affected unless the computation pauses due to that tripwire.)

Note that with each new pause the time it takes to analyze the computation increases, since Terminyzer now tries to look at a bigger chunk of the computation trace. It does try to give the user an idea about how long the wait might be, but it is only a very rough figure.

Finally, if you decide to stop Terminyzer, this does not stop the query. You can still resume it, use other analysis tools, or you can abort it.

Portions Copyright © 2013–2017 Coherent Knowledge

43.9 Runtime Inspection of Computation

$\mathcal{E}$RGO allows the user to stop any computation by hitting `Ctrl-C` in the terminal or by clicking the `Pause` button in the $\mathcal{E}$RGO listener. During the pause, one can ask certain queries about the state of the computation, but not arbitrary queries. Specifically, no tabled (non-transactional) queries are allowed and insert/delete operations are not permitted either. The most useful queries that one *can* ask are described below.

The `showgoals{...}` command to display information about the goals still being computed. The information shown is the number of answers for each goal computed so far and the number of calls to each goal. The output is sorted first by the number of answers and then by the number of calls. There are two types of the `showgoals` command:

- `showgoals{}`: this shows all incomplete subgoals that have over 1000 calls or over 50 answers.
- `showgoals{`*CallCutoff,AnwerCutoff*`}`: This will show only the subgoals that have more than *CallCutoff* calls or over *AnswerCutoff* answers.

$\mathcal{E}$RGO also provides a companion primitive, `showtables{...}`, which is usually more convenient than the table dump because it provides a more focused output and is easier to use. This primitive is intended not for the pause in computation but for analysis to be performed after a query of interest is finished. It then displays subgoals that took part in the computation of the last (and previous queries) using the same format as `showgoals{...}`. There are two types of the `showtables` command:

- `showtables{}`: this shows all incomplete subgoals that have over 1000 calls or over 50 answers.
- `showtables{`*CallCutoff,AnwerCutoff*`}`: This will show only the subgoals that have over *CallCutoff* calls or more than *AnswerCutoff* answers.

The `showgoals{...}` primitive is typically used to investigate reasons for *runaway* computation, i.e., non-termination of queries as well as diagnose performance problems. It is designed to be used during pauses in query computation, before the query finishes. The `showtables{...}` primitive is designed to be used after the query finishes and thus is useful only for analyzing the performance aspect of runaway computations. The call and answer cutoffs must be chosen judiciously. In both of the above runaway cases, the goals of interest are those that either have many calls to them (in the thousands, but depends on the case at hand) or have many answers (in the tens or hundreds, as appropriate). Examining the information provided by $\mathcal{E}$RGO during pauses and after query completion can provide a number of insights:

- *Compare the number of active goals and the number of active recursive components,* which is shown when the computation is paused. The active goals/components are the ones that are currently being computed and therefore are of interest during pauses in the computation.

  - Check the *ratio* of the *number of active recursive components to the number of active goals.* This ration is normally ≤ 1, but if it is very small (say, under 3% and keeps falling during subsequent pauses), then it is an indication that the rules are poorly structured. Every subgoal depends on many other subgoals and there likely are too many calls that do not contribute to the query results, but just waste the computational resources. Consider restructuring the rules so that the recursive components form some kind of a hierarchy. The use of $\mathcal{E}$RGO modules is strongly recommended in this case. Also, try to mix HiLog and frame-based representation.
  - If the aforesaid ratio stays approximately constant and *both* the number of active recursive components and the number of active goals climb steadily, it is an indication of an infinite recursion where subgoals are called with increasingly deeply nested terms (e.g., `p(f(a))`, `p(f(f(a)))`, `p(f(f(f(a))))`, etc.). The aforesaid ratio in such cases is typically non-negligible (> 0.3).

- *Check the memory usage*, which is shown during each pause. If memory usage keeps climbing steadily, it might indicate a runaway computation or a possible scalability problem.

- *Watch the cpu time* shown during each pause. It can be helpful as a way of tracking progress. This is cpu time since the time $\mathcal{E}$RGO started; it includes all the pauses.

- *Check the number of derived facts* shown at each pause. If this number does not grow and $\mathcal{E}$RGO keeps going on and on, it is an indication of some kind of a runaway computation.

- *Use* `showgoals` *during the computation pauses.* As mentioned earlier, you will likely be interested in goals with many calls or many answers. Keep increasing the cutoffs during subsequent pauses. Examining the active goals can indicate two things:

  - Weird, unexpected stuff showing up as active goals might provide a hint about a possible bug in the user's rules.
  - Some goals have unusually high numbers or calls and/or answers. This may mean two things:

    * Non-terminating computation. The goals in question might be involved in a non-terminating recursive loop. *Terminyzer* (Section 43.8) may supply additional information if this is the case, including the information about the rules that are involved in such a loop.
    * Gross inefficiency. The computation may be terminating but, for some reason, the goals in question get called unusually often or they produce too many results. The latter might also indicate a bug.

- *Use* `showtables` *after the computation.* This is used after and if the query finishes. In that case, the query obviously terminates, but one might be still dissatisfied with the performance.

The `showtables` primitive supplies the same kind of information as `showgoals`, but since the former is called after the query is done, it supplies complete information about the subgoals that took part in the computation. This can give too kinds of hints: a possible bug or a possible inefficiency. As before, a bug may be lurking if unexpected subgoals show up in the output of `showtables`. Inefficiency should be suspected if some subgoals have unusually high number of calls and/or answers.

The table dump described in Section 43.6 provides much additional information, but examining it manually is usually overwhelming. The table dump feature is designed for expert users who know how to write Prolog and/or $\mathcal{E}$RGO programs to analyze such dumps automatically and check the various hypotheses about the computational behavior of queries.

Another very useful informational primitive that is allowed during pauses is `peek{`*query*`}`, where *query* is an atomic subgoal (tabled, non-transactional). It examines the current state of the computation of *query* and reports the answers that have been computed so far. This primitive is described in more detail in Section 22.3.

43.10 Continuous Runtime Monitoring

Sometimes, especially when queries take a long time, it might be useful to see the various statistics pertaining the queries. To request this service, the user can execute the following command in the $\mathcal{E}$RGO shell:

```
ergo>  setmonitor{Secs,Type}.
```

The effect will be that all subsequent queries will be monitored and every *Secs* seconds certain statistics will be printed out to the standard output. Here *Type* is the type of the monitor. The `heartbeat` monitor just shows the elapsed time. The `performance` monitor shows time, memory, and other key statistics. The `extended` monitor shows additional statistics of interest.

Secs must be a non-negative integer. If *Secs* = 0, monitoring is turned off. Note: it is not guaranteed that the statistics will be refreshed every *Secs* seconds, as certain intermediate computations are not interruptable and may take several seconds. In such cases, the next statistics report would be output right after such a non-interruptable computation ends.

If the user instead types

```
ergo>  setmonitor{}.
```

then the parameters can be input via a dialog window, which will pop up.

If $\mathcal{E}$RGO is run under the studio or in a graphical desktop environment out of a command line window, the output from the monitor will go into a separate pop-up window. Otherwise, it will be redirected to the standard output (which might interfere with the output of the query *if* the query has write-statements of its own; this *does not* interfere with the display of the query results however).

Some of the statistics shown by the `performance` and `heartbeat` monitors are described in Section 43.9 on runtime inspection of computation and they thus provide indication of the state of the system as described there.

43.11 Tracing Tabled Calls via Forest Logging

While the regular Call-Redo-Exit-Fail logging is useful in many cases, it is extremely slow and generates large amounts of output. A query that runs mere 10 seconds can take hours to execute under tracing and it may generate hundreds of megabytes worth of trace output. Clearly, it is hard to use the regular tracing facility under such conditions. On top of this, even though the regular trace is capable of producing queriable output, such trace does not supply parent-child relationships between calls, and, due to this, automatic analysis of such trace is very hard.

Fortunately, there is an alternative: *forest logging*. Forest logging is a kind of tracing that keeps track only of tabled predicate calls. It is very fast (time overhead is less than 80% compared to orders of magnitude for the regular trace), it generates drastically smaller trace logs, and it preserves the parent-child relationship between the calls. The drawback is that logforest traces track tabled calls only, but in $\mathcal{E}$RGO this is not a serious problem since most calls that are of interest to the user are tabled (except for transactional predicates and methods).

$\mathcal{E}$RGO's forest logging is implemented as a presentation layer on top of XSB's forest logging. Its format and other details are described next.

Forest logging. To start forest logging, the user must issue the command `\logforest` at the $\mathcal{E}$RGO prompt or include the query `?- \logforest` where appropriate. In the latter case, logging will start after this subgoal gets executed. To stop forest logging, issue the command `\nologforest`.

The entries in the log represent the following actions that occur during tabled evaluation of queries and every entry has an Id which is a non-negative integer.

- *A call to a tabled subgoal.* When a call to a tabled subgoal $\mathtt{S}_1$ is made from a derivation tree for $\mathtt{S}_2$, a frame literal is recorded in the following format:

  ```
  call(Id)[goal->S1, stage->Stage, parent->S2].
  ```

where `Id` is the generated Id of the call and `Stage` is

- `new` if S_1 is a new subgoal.
- `comp` if S_1 is an old *completed* subgoal.
- `incmp` if S_1 is an old *incomplete* subgoal (i.e., it has not been fully evaluated yet).

If S_1 is the first tabled subgoal in an evaluation, S_2 is represented by the atom `null`. If the call is negative, a similar fact of the form `negative_call(Id)[goal->S1, stage->Stage, parent->S2]` is logged.

- *Derivation of a new answer.* When a new unconditional answer `A` is derived for subgoal `S` and added to the table, the following fact is logged:

```
answer(Id)[goal->S, answer->A].
```

  As before, `Id` is the identity number generated for this particular action.

  When a new conditional answer `A` is derived for subgoal `S` and the delayed literals are `D`, a log of the form `conditional_answer(Id)[goal->S, answer->A, delayed_literals->D]` is recorded.

- *Return of an answer to a consuming subgoal.* When an answer A is computed and returned to a consumer subgoal `S` in a derivation tree for `ST` and the table for `S` is incomplete, the following fact is recorded:

```
answer_to_consumer(Id)[goal->S, answer->A, consumer->ST].
```

  If `A` is conditional, this entry `delayed_answer_to_consumer(Id)[goal->S, answer->A, consumer->ST]` is recorded.

- *Delaying a negative literal.* When a selected negative literal `N` of a node `S` is delayed due to its involvement in a loop through negation, and `S` is in a derivation tree for S_T, a fact of this form is logged.

```
delay(Id)[delayed_literal->N, parent->ST].
```

- *Subgoal completion.* When a set $\mathcal{S}$ of subgoals is completely evaluated, for each $S \in \mathcal{S}$ a fact of the following format is logged for each `S`:

```
completed(Id)[goal->S, sccnum->SCCNum].
```

  Here `SCCNum` is the identifier generated for the set of subgoals $\mathcal{S}$. If `S` is completed early, `SCCNum` is the atom `ec`.

- *Table abolishes.* There are three occasions where tables are abolished.
  - When a tabled subgoal `S` is abolished, a fact of the following form is logged:

    ```
    table_abolished(Id)[type->subg, goal->S].
    ```

  - When all tables for a predicate p/n are abolished, a fact of the following form is logged:

    ```
    table_abolished(Id)[type->pred, goal->[p/n]].
    ```

  - When all tables are abolished, the following fact is logged:

    ```
    table_abolished(Id)[type->all].
    ```

- *Recording of errors.* If an error is thrown and the execution is in a derivation tree for subgoal `S`, forest logging records the following fact:

  ```
  error(Id)[goal->S].
  ```

By default, logs are sent to the current output stream. However, it is usually more convenient to dump the logs to a file using the following command: `\logforest(File)`. For instance, executing

```
?- \logforest('foobar.ergo').
```

will direct $\mathcal{E}$RGO to dump the entire forest log into the file `foobar.ergo`. In case the user also wants to examine the original XSB's forest logging trace, the following command can be executed: `\logforest(FloraTraceFile,XSBTraceFile)`. For instance, executing

```
?- \logforest('foobar.ergo', 'foobar.P').
```

will dump the trace in the above $\mathcal{E}$RGO format into the file `foobar.ergo` and keep the original XSB's forest log in the file `foobar.P`.

We may want to skip certain types of log entries in some circumstances such as reducing log file sizes. $\mathcal{E}$RGO provides `\logforest(HideOptions)` and `\logforest(File, HideOptions)`, where `HideOptions` is a list of log types to be skipped. The elements of `HideOptions` can be one of the following: `call, negative_call, delayed_call, answer, conditional_answer`, `answer_to_consumer`, `delayed_answer_to_consumer, completed, table_abolished`, and `error`. For instance,

```
?- \logforest([table_abolished, error]).
?- \logforest('foobar.ergo', [table_abolished, error]).
```

will not record logs for table abolish and error actions.

Note that `\logforest(FloraTraceFile,XSBTraceFile)` and `\logforest(File,HideOptions)` will not be mixed since $\mathcal{E}$RGO checks whether their second argument is a list or not to tell which command is executed. Similarly, `\logforest(HideOptions)` and `\logforest(File)` will not be mixed.

It is important to keep in mind, however, that *no* output is produced (to the file or output stream) unless the user issues the command `\nologforest`. In other words, forest logging can be obtained only *after* the evaluation is finished. If you expect the query to throw an error, it is a good idea to use the $\mathcal{E}$RGO `catch...` primitive. If the query does not terminate, you should also wrap the query with the `timed_call/3` XSB predicate.[22]

Low-level forest logging. Sometimes it is necessary to look into the guts of forest logging without converting it into the $\mathcal{E}$RGO format. Typically this is needed for low-level debugging of $\mathcal{E}$RGO itself. The command for this is `\logforestlow`; it directs $\mathcal{E}$RGO to display forest logs directly in the XSB format. There is also a version of this command for saving the log in a file:

```
?- \logforestlow('foobar.ergo').
```

As before, one should remember to issue the command `\nologforest` in order to flush the log to the output. $\mathcal{E}$RGO also provides `\logforestlow(HideOptions)` and `\logforestlow(File, HideOptions)` to skip certain types of low-level log entries.

43.12 Controlling Subgoal and Answer Size, Timeouts, Unification Mode

Sometimes it is useful to be able to control the term-size of the subgoals that can be generated during evaluation and the size of the answers returned. The former is useful if the user knowledge base has recursive rules in which the size of the body literals is greater than the size of the head. The latter is useful when a query has an infinite number of answers. In both cases, limiting the size can terminate a run-away computation. Timeouts are useful when it is desirable to stop the computation if it does not finish within a preset amount of time. To control these features, $\mathcal{E}$RGO provides the builtin primitive that can appear in any query or rule body:

```
?- setruntime{Opt1,Opt2,...}
```

Several options can be used in the same `setruntime` command. These options are described below.

[22] `timed_call/3` is described in the XSB Manual, Part 1. Do not forget to reify the query argument to `timed_call/3`.

43.12.1 Timeouts

The following `setruntime{...}` options can be used to control various types of timeouts:

```
timeout(Spec)
timeout(Spec,Spec)
timeout(0)
```

where *Spec* is either `max(Time,Handler)` or `repeating(Time,Handler)`. Here *Time* is a positive integer that specifies the number of seconds after which queries should be interrupted. The first form will interrupt the queries once (so this form is used to specify timeouts) and the second will interrupt queries periodically, after each *Time* seconds. Either `max`, or `repeating`, or both can be specified (via the 1-argument and 2-argument form of `timeout`). In either case, *Handler* will be called at each interrupt. The last form above removes all timeout restrictions. Note that each subsequent `setruntime` timeout-setting command will override the previous one.

Here are some examples:

```
?- setruntime{timeout(max(4,fail))}.
?- setruntime{timeout(max(100,abort))}.
?- setruntime{timeout(repeating(4,Handler(?)),max(100,abort))}.
?- setruntime{timeout(0)}.
```

Note that once any of these commands is issued, it applies to all subsequent queries. The last command resets the timeout to infinity.

Handler in the `max` and `repeating` specifications must have one of the following forms:

- A predefined error handler: `ignore` (ignore the timer interrupt), `abort` (abort the current goal), `fail` (make the current goal fail); or
- A Prolog predicate defined in some Prolog module; or
- A predicate declared using $\mathcal{E}$RGO's `:- prolog{...}` directive.
- A predefined error handler `pause`, which is available only **in** $\mathcal{E}$RGO. In that case, the execution pauses and the user is given the opportunity to inspect the state of the system and then either continue or abort the computation.[23] The `pause` interrupt handler is perhaps the most useful when the system runs in interactive mode. For instance,

[23] The ability to pause a computation exists only in $\mathcal{E}$RGO and this is why the pause interrupt handler is available only there.

```
?- setruntime{timeout(repeating(20,pause))}.
```

If *Handler* is a Prolog predicate that exists in the default Prolog's `usermod` module then the handler should look like `foobar(?)` or `foobar` (i.e., one or zero arguments) . The argument (if present) must be an unbound variable, which will be bound at runtime to the goal being interrupted. A Prolog handler can be also specified as `foobar(?)@\prolog` or `foobar(?)@\prolog(`*mod*`)`. The latter is especially useful if `foobar/1` is in the Prolog's module *mod* rather than the `usermod` module. In either of these cases, `foobar/1` must be defined in a Prolog program that $\mathcal{E}$RGO has already loaded. In $\mathcal{E}$RGO, information primitives such as `showgoals{}`, can also be used as handlers.

If the handler is an $\mathcal{E}$RGO predicate declared as `:- prolog` then it must be specified as `foobar(?)@`*mod* (note: not `\prolog(`*mod*`)`, but *mod*), where *mod* is an $\mathcal{E}$RGO module (or `\@`, if the current module is desired (see Section 15.3).

For instance, the following fragment in $\mathcal{E}$RGO sets the system for periodic timer interrupts:

```
:- prolog{periodic_handler/1}.
periodic_handler(?Goal) :- writeln(interrupted_goal=?Goal)@\prolog.
?- setruntime{timeout(repeating(2,periodic_handler(?)@\@))}.
```

In this example, the $\mathcal{E}$RGO predicate `periodic_handler/1` will be called every two seconds.

Alternatively, one could define `periodic_handler/1` as a Prolog predicate in a `.P` file, e.g., `foo.P`, and then set periodic interrupts as follows:

```
?- ['foo.P'].
?- setruntime{timeout(repeating(2,periodic_handler(?)@\prolog))}.
```

Care must be taken to not obliterate periodic handlers during loading. For instance, in either of the above cases `periodic_handler/1` may get erased. In case of a Prolog program (the second example), the danger is not as great: `periodic_handler/1` may get replaced only if another Prolog file gets loaded and happens to have the same predicate. A Prolog predicate may also be explicitly abolished by the user, but this is rare. In case of periodic handlers defined in $\mathcal{E}$RGO modules, however, the problem occur more frequently. For instance, the first example may be loaded into some module, `foo`, and then, due to an oversight, something else may get loaded (as opposed to "added") into the same module. In that case, `periodic_handler/1` will be obliterated and the periodic timer interrupt specification will become orphaned, resulting in a runtime error. If this happens, errors will be issued whenever the interrupt handler is called.

43.12.2 Subgoal Size Control

The following options provide subgoal size control:

```
goalsize(TermSize,abort)
goalsize(TermSize,abstract)
goalsize(TermSize,pause)          // Ergo only
```

These options control the maximum size of the subgoals called during the evaluation. *TermSize* specifies the max size of the terms returned as answers. This takes into account both nesting of all function symbols except lists (lists are considered to be of size 1) and the arity (width) of the terms.

The first form of the above options will abort the query if the specified limit is reached. The second form will *perform call abstraction* and replace the deeply nested subterms with new variables. The `pause`-action will pause the computation and allow the user to examine the state of the computation before deciding what to do next. Examples:

```
?- setruntime{goalsize(100,abort)}.
?- setruntime{goalsize(100,abstract)}.
```

The above options are incompatible, so each subsequent option overrides the previous one.

43.12.3 Answer Size Control

The following options enable answer size control:

```
answersize(TermSize,abort)
answersize(TermSize,abstract)
answersize(TermSize,pause)          // Ergo only
```

The meaning of the `abort` options is the same as for the goal size. The meaning of the `abstract` options is that the size of terms is limited to the specified numbers. If higher size answer-objects are generated, their truth value is set to *undefined*. This is called *answer abstraction*. For instance, if the computation generates the answers `p(f(a))`, `p(f(f(a)))`, etc., then answer abstraction at size 5 will generate the answers up to `p(f(f(f(a))))` and the last answer, `p(f(f(f(f(?)))))`, will be *undefined*. This means that some of the instances of that last answer might be true and some false.

Examples:

```
?- setruntime{answersize(100,abort)}.
?- setruntime{answersize(100,abstract)}.
```

43.12.4 Controlling the Number of Active Goals

The following options enable one to control the number of active (i.e., not completely evaluated) tabled subgoals during query evaluation, which is useful for termination and performance analysis.

```
activegoals(TermSize,abort)
activegoals(TermSize,pause)          // Ergo only
```

The meaning of the `abort` and `pause` actions is the same as in the case of subgoal and answer size controls.

43.12.5 Memory Usage Limit

The following option is available:

```
memory(memory-limit-in-GBs)
```

For instance,

```
?- setruntime{memory(12)}.
```

will set the limit to 12 GBs. If the $\mathcal{E}$RGO process exceeds this amount, the computation will be aborted.

$\mathcal{E}$RGO (but not $\mathcal{F}$LORA-*2*) also supports the following:

```
memory(memory-limit-in-GBs,Action)
```

where *Action* is either `abort` or `pause`. The `abort` option works exactly like `memory(Size)`, while if `pause` is given then the computation will pause instead of aborting. The user can then ask informational queries about the system state and decide what to do. For instance, one may decide whether to abort or to increase the memory limit and continue.

43.12.6 Unification Mode

The default unification mode in $\mathcal{E}$RGO is unsound, in general. That is, non-unifiable terms might be unified, as explained below. The following runtime options can be used to change the unification mode:

```
unification(fast)
unification(pedantic)
```

For instance,

```
?- setruntime{unification(fast)}.
?- setruntime{unification(pedantic)}.
```

These commands controls the experimental feature of unification mode. The default unification mode is `fast`. This is a logically unsound mode: it does not do the occurs-check and so the unification `?X = f(?X)` will succeed even though the two terms are not unifiable in the classical sense. Under the `pedantic` mode, such a unification would fail.

Unifying in the fast mode can lead to hard-to-find errors, although $\mathcal{E}$RGO makes attempts to limit the damage and the probability of serious problems occurring due to this mode of unification is low.

44 Considerations for Improving Performance of Queries

Left-to-right processing. The first rule in improving the performance of $\mathcal{E}$RGO queries is to remember that query evaluation proceeds from left to right. Therefore it is generally advisable to place subgoals with smaller answer sets as close to the left of the rule body as possible. And, as with databases, Cartesian products should be avoided at all costs.

Nested frames and path expressions. The $\mathcal{E}$RGO compiler makes decisions about where to place the various parts of complex frames, and the knowledge engineer can affect this placement by writing frames in various ways. For instance,

```
?- ..., ?X[attr1 -> ?Y, attr2 -> ?Y], ...
```

is translated as

```
?- ..., ?X[attr1 -> ?Y], ?X[attr2 -> ?Y], ...
```

so the first attribute will be computed first. If the second attribute has a smaller answer set, the attributes in the frame should be written in the opposite order. The other consideration has to do with literals that have nested frames in them. For instance, the following query

```
?- ..., ?X[attr1->?Y[attr2->?Z]], f(?P.attr3), ...
```

is translated as

```
?- ..., ?X[attr1->?Y], ?Y[attr2->?Z], f[?P->?_newvar], f(?_newvar), ...
```

i.e., the nested literals follow their hosts in the translation. Thus, writing terms in this way is considered a hint to the compiler, which indicates that bindings are propagated from `?X[attr1->?Y]` to `?Y[attr2->?Z]`, etc. If, on the other hand, `?Y[attr2->?Z]` has only one solution then, perhaps, writing `?Y[attr2->?Z], ?X[attr1->?Y]` might produce more efficient code.

Similar considerations apply to `f(?P.attr3)`, but here `?P.attr3` is computed first and the result is passed to the predicate `f/1`. Note that frames and truth-valued path expressions are not allowed as nested arguments in predicates and functions. That is, `f(?Y[attr2->?Z])` or `f(?Y.attr2[])` would cause compiler to issue errors. However, *pure* path expressions, which have no truth values, can be nested inside predicates and functions. For instance, `f(?Y.attr2)` or even `f(?Y[foo->bar].attr2)` are acceptable except inside `insert...` and `delete...` primitives.

As with nested frame literals, the $\mathcal{E}$RGO compiler assumes that path expressions represent a hint that bindings are propagated left-to-right. In other words, in `?X.?Y.?Z`, `?X` will be bound first. Based on this, the oids of the objects `?X.?Y` are computed, and then the attribute `?Z` is applied. In other words, the translation will be `?X[?Y->?Newvar1], ?Newvar1[?Z->?Newvar2]`.

Unfortunately, unlike databases, statistical information is not available to the $\mathcal{E}$RGO compiler and only a few heuristics (such as variable binding analysis, which the compiler does not perform) can be used to optimize such queries. If the order chosen by the compiler is not right, the user may unnest the literals and place them in the right order in the rule body.

Open calls vs. bound calls. In Prolog it is much more efficient (space- and time-wise) to make one unbound call than multiple bound ones. For instance, suppose we have a class, `cl`, that has hundreds of members, and consider the following query:

```
?- ?X:cl[attr->?Y].
```

Here, Prolog would first evaluate the open call $?X : cl$ and then for each answer `x` for `?X` it will evaluate `x[attr->?Y]`. If the cost of computing `x[attr->?Y]` is higher than the cost of $x : cl$ and the number of answers to `?X[attr->?Y]` is not significantly higher than the number of answers to `?X:cl`, then the following query might be evaluated much faster:

```
?- ?X[attr->?Y], ?X:cl.
```

In this query, a single call `?X[attr->?Y]` is evaluated first and then `x:cl` is computed for each answer for `?X`. Since, as we remarked, the cost of this call can be much smaller than the combined cost

of multiple calls to `x[attr->?Y]` for different `x`. If the number of bindings for `?X` in `?X[attr->?Y]` that are not members of class `cl` is small, the second query might take significantly less space and time.

The delay quantifiers `wish` and `must`, described in Section 33, can also be helpful in ensuring that certain variables are bound.

45 Compiler Directives

45.1 Executable vs. Compile-time Directives

Like a Prolog compiler, the $\mathcal{E}$RGO compiler can take compiler directives. As in Prolog, these directives can be *executable* or *compile-time*, and this distinction is very important. Executable directives are treated as queries and they begin with `?-`. Compile-time directives begin with `:-`.

Executable directives are mostly used to control how the $\mathcal{E}$RGO shell interprets the expressions that the user types in. These directives have no effect during the compilation of the knowledge base. Instead, when they are executed as queries they affect the shell. In contrast, compile-time directives affect the compilation of the files they occur in. Also, if a module is loaded into the main module in the shell, then all compile time directives in that module are executed in the shell as well, so there is no need to explicitly execute these directives in the shell. $\mathcal{E}$RGO requires that all compile-time directives appear at the top of the file prior to the first appearance of a rule or a fact, because such directives have effect only after they are found and processed.

To better understand the issue, consider the following simple example (say, in the file `test.ergo`):

```
:- op{400,xfx,fff}.
a fff b.
?- ?X fff ?Y.
```

If one loads this example, it will execute correctly and return the bindings `a` and `b` for `?X` and `?Y`, respectively. If you execute the same query `?X fff ?Y` in the $\mathcal{E}$RGO shell, the result will still be correct because $\mathcal{E}$RGO made sure that the directive `op{400,xfx,fff}` in `test.ergo` was executed in the shell as well. On the other hand, if the example were

```
?- op{400,xfx,fff}.
a fff b.
?- ?X fff ?Y.
```

then `fff` would be known to the shell, but, unfortunately, we will not get that far to find out: The compiler will issue an error, since `fff` will not be known as an operator during the compilation.

Summary of directives. The following is a summary of all supported compiler directives:

`:- setsemantics{`*`Option1, Option2, ...`*`}.`

Sets the semantic options in the current user module. The currently allowed options are:

`equality=none` (default), `equality=basic`,
`inheritance=none`, `inheritance=flogic` (default), `inheritance=monotonic`,
`tabling=reactive` (default), `tabling=passive`,
`tabling=variant` (default), `tabling=subsumptive`,
`custom=none` (default), `custom=`*`filename`*.

The form `option=val1+val2+val3` is also supported. For instance, `tabling=variant+reactive`.

These options are explained in detail in Section 21.

`?- setsemantics{`*`Option1, Option2, ...`*`}.`

This is an executable version of the `setsemantics` directive. The following options can be used only with the executable version of `setsemantics`:

`subclassing=strict` (default), `subclassing=nonstrict`,
`class_expressions=on`, `class_expressions=none` (default),

`?- setsemantics{`*`Option1, Option2, ...`*`}@`*`module`*`.`

Same as above, except that the semantics is set for the specified user module.

`:- index{`*`Arity-Argument`*`}.`

Says that all tabled HiLog predicates of arity `Arity` should be indexed on argument number `Argument` (the count starts at 1). This directive should appear at the beginning of a module to have any effect. Normally predicates in $\mathcal{E}$RGO are indexed on predicate name only. The above directive changes this so that indexing is done on the given argument number instead.

Note that the `index` directive is not very useful for predicates that mostly contain facts, because these are trie-indexed anyway (regardless of what you say). Thus, this instruction is useful only for predicates with partially instantiated arguments that appear in the rule heads.

This is an executable version of the `index` directive. The module of the predicates can be specified.

`:- index{`*`%Arity-Argument`*`}.`

The `index` directive for non-tabled HiLog predicates.

`?- (index{`*`%Arity-Argument`*`)[@`*`module`*`]}.`

The executable `index` for non-tabled HiLog predicates.

`:- op{`*precedence*`,`*type*`,`*operator*`}.`
Defines *operator* as a $\mathcal{E}$RGO operator with the given precedence and type. The *type* is the same as in Prolog operators, i.e., `fx`, `xf`, `xfy`, etc. Note that the `op` directive is confined to the module in which it is executed or defined. For instance, if `example.ergo` has a call `(a foo b)@bar`, the symbol `foo` is declared as an operator in the knowledge base loaded to module `bar`, but not in `example.ergo`, then a syntax error will result, because `example.ergo` does not know about the operator declaration for `foo`.

`:- op{`*precedence*`,`*type*`,[`*operator*`, ..., `*operator*`]}.`
Same as above, except that this directive defines a list of operators with the same precedence and type.

`?- op{`*precedence*`,`*type*`,`*operator*`}@module.`
Same as above, except that a module is also given. However, unless the module is `main`, this directive acts as a no-op.

45.2 Miscellaneous Compiler Options

Sometimes it is desirable to pass miscellaneous compiler options to the $\mathcal{E}$RGO compiler. To do this, $\mathcal{E}$RGO provides the directive `compiler_options`. It takes one argument — a list of options that is understood by the underlying Prolog compiler. At present, the following options are supported:

- `production=on` – compile the file without various debugging features such as the rule Ids embedded in the heads and bodies of rules.
- `production=off` – compile the file in the development mode. In this mode, rule Id information is added to the heads and the bodies of the rules, which facilitates tracing and profiling of $\mathcal{E}$RGO queries. This mode is the default. This compiler option was introduced to make it possible to override production mode when it is requested from the $\mathcal{E}$RGO shell via the primitive `production{on}` (see Section 50.2).
- `expert=on` – allow advanced syntax. Do *not* turn this option on if you are not a very experienced $\mathcal{E}$RGO user: the expert mode was introduced to prevent misuse by novice users. See Section 53 to learn about the features that are available only in the expert mode.
- `expert=off` – do not allow advanced syntax. Note that, like `production`, the expert mode is a compile-time option. In the $\mathcal{E}$RGO shell, use `expert{on}` and `expert{off}`.
- `prolog=[Opt1,Opt2,...,OptN]` – pass the list of Prolog compiler options `[Opt1,Opt2,...,OptN]` to the underlying Prolog compiler. The `prolog` options are not checked by the $\mathcal{E}$RGO compiler and are simply passed to the underlying Prolog engine. If any of the given options is

invalid, *runtime* warnings or errors will result. Also, unlike `expert` and `production` options, the `prolog` option cannot be issued at run time via the ℰRGO shell.

Example:

```
:- compiler_options{production=on,expert=on,prolog=[spec_repr]}.
...
?- expert{off}.
```

will cause the module that contains this directive to be compiled without the special debugging features that slow the running code down. This will also pass the `spec_repr` option to the underlying Prolog compiler so it will perform the specialization optimization.[24]

46 ℰRGO System Modules

ℰRGO provides a number of useful libraries that the user can use. These libraries are statically preloaded into modules that are accessible through the special `@\modname` syntax, and they are called *system modules*. We describe the functionality of these modules below. Some of these modules also have longer synonyms. These synonyms are mentioned below, if they exist.

46.1 Input and Output

This library simplifies access to the most common Prolog I/O facilities. This library is preloaded into the system module `\io` and can be accessed using the `@\io` syntax.

The purpose of the I/O library is not to replace the standard I/O predicates with ℰRGO methods, but rather to relieve the user from the need to do explicit conversion of arguments between the HiLog representation of terms used in ℰRGO and the standard Prolog representation of the underlying engine.[25] The methods and predicates accessible through the `\io` library are listed below. Note that some operations are defined as transactional methods and others as predicates. This is because we use the object-oriented representation only where it makes sense — we avoid introducing additional classes and objects that require more typing just for the sake of keeping the syntax object-oriented.

[24] Although the `spec_repr` optimization option is the default in XSB, ℰRGO turns it off because, we believe, it is buggy and causes some ℰRGO queries to go wrong.

[25] See Section 16 for a discussion of the problems associated with this representation mismatch.

Standard I/O interface. This interface to input-output uses the concept of default streams that can be opened with `see` (for input) and `tell` (for output). Subsequent `see`/`tell` commands push the current streams to the respective input and output stacks and open new default streams. Subsequent reads and writes operate with those default streams. The commands `seen` and `told` close the current default streams and pop up the appropriate streams from the appropriate stacks, making them the default streams. The `seeing` and `telling` commands obtain references to the current default streams. If these streams are pushed on the stack and are no longer the default ones, those references can still be used to write to or read from the streams that are no longer default. This could be useful, for example, if one wants to copy one file to another. The following is a list of commands in this interface followed by some examples.

- `see(?Filename), ?Filename[see]` — open `?Filename` and make it the current input stream. The file can live remotely at some URL. In that case, `?Filename` must be instantiated to the form `url(...)`.
- `seeing(?Stream)` — binds `?Stream` to the current input stream.
- `seen` — closes the current input stream.
- `tell(?Filename), ?Filename[tell]` — opens `?Filename` as the current output stream. The file can live remotely at some URL. In that case, `?Filename` must be instantiated to the form `url(...)`. In that case, ERGO will attempt to use the HTTP POST request to store the file remotely.
- `telling(?Stream)` — binds `?Stream` to the current output stream.
- `told` — closes the current output stream.
- `write(?Obj)` — writes `?Obj` to the current output stream.
- `writeln(?Obj)` — same as above, except that the newline character is output after `?Obj`.
- `write(?Obj,?Options)`, `writeln(?Obj,?Options)` — like `write(?Obj)` and `writeln(?Obj)` but takes the `?Options` argument, which is a list. At present, the only options in that list that are supported are `oid` and `goal`. The difference shows only when printing reified statements: with the `oid` option, `?Obj` is printed as an object, while with `goal` it is printed as a goal (without the reification symbol). For instance,

```
?- write(${p(1)},[oid])@\io.
${p(1)}
?- write(${p(1)},[goal])@\io.
p(1)
```

If the argument is not a reified formula then the two options give the same result. The 1-argument versions `write(?Obj)`, `writeln(?Obj)` are equivalent to `write(?Obj,[oid])`, `writeln(?Obj,[oid])`, respectively.

- `nl` — writes the newline character to the current output stream.
- `read(?Result)` — binds `?Result` to the next term in the current input stream.
- `read_canonical(?Term)` — read input and bind `?Term` to the next term in the input. The term *must* be in canonical Prolog form, or else an error will result. This method is much faster than the usual `read` operation, but it is not as versatile, as it assumes that input is in canonical form.
- `write_canonical(?Term)` — write `?Term` to standard output in canonical Prolog form.
- `fmt_write(?Format,?Term)` — C-style formatted output to the standard output. See the XSB manual, volume 2, for the description of all the details. Here we just mention that the format is an atom `'....'` whose structure is like in C (some formats might not be supported) with the addition of the format specified `%S`, which can take any term. `?Term` must have the format $args(arg_1, ..., arg_n)$ (i.e., all the arguments to be printed must be grouped under a singe functor; the name of the functor is immaterial).
- `fmt_write(?Format,?Term,?Options)` — like `fmt_write(?Format,?Term)`, but take options whose meaning is the same as in case of `write`/`writeln` above.
- `fmt_write_string(?String,?Format,?Obj)` — same as above, but binds `?String` to the result. See the XSB manual for the details.
- `fmt_write_string(?String,?Format,?Obj,?Options)` — same as above, but takes options whose meaning is the same as in the case of `write`/`writeln` above.
- `fmt_read(?Format,?Result,?Status)` — C-style formatted read from standard input. See the XSB manual.
- `readline(?Type,?String)` — read the current input and bind `?String` to the next line. `?Type` is either the word `atom` or `charlist`. The former means that `?String` is to be bound to a Prolog atom and the latter binds it to a list of characters.

Stream-based I/O. Stream-based I/O is like the standard I/O interface except that it does not use the default streams that can be pushed to or popped from the stack of streams. Instead, there is a notion of standard input and output ports, which are usually associated with and interactive window, and user-defined ports. Standard input and output ports always exist, while user-defined

ports are created and destroyed when files are open and closed. This is somewhat similar to stream-based I/O, but nothing gets pushed on stacks. Instead, all read and write operations must use the appropriate ports explicitly.

- `?Filename[open(?Mode,?Stream)]` — opens `?Filename` with mode `?Mode` (which can be `read`, `write`, `append`, `write_binary`, or `append_binary`) and binds `?Stream` to the file handle. The file can live remotely at some URL. In that case, `?Filename` must be instantiated to the form `url(...)`.

  The modes `write_binary` and `append_binary` are used in Windows only. In Unix-based systems (Linux, Mac, etc.), these modes coincide with `write` and `append`, respectively.

  One way to think of `?Filename[open(read,?Stream)]` is that it is like `see(?Filename)` followed by `seeing(?Stream)`. Similarly, `?Filename[open(write,?Stream)]` is like `tell(?Filename)` followed by `telling(?Stream)`. But `open` provides one more option, `append`, which does not exist in stream-based I/O, plus the binary modes for Windows.

- `?Stream[close]` — closes `?Stream`. It is similar to `told` but it can close any stream, not just the default one.

- `?Stream[read(?Result)]` — bind `?Result` to the next term in the previously open input `?Stream`.

- `?Stream[write(?Obj)]`, `?Stream[writeln(?Obj)]` — like `write(?Obj)`, `writeln(?Obj)` but write to a previously open output `?Stream` instead of the current output stream.

- `?Stream[write(?Obj,?Option)]`, `?Stream[writeln(?Obj,?Option)]` — like `write(?Obj,?Option)`, `writeln(?Obj,?Option)` but write to a previously open output `?Stream` instead of the current output stream.

- `?Stream[fmt_write(?Format,?O)]` — same as `fmt_write(?Format,?O)`, but uses `?Stream` for the output.

- `?Stream[fmt_write(?Format,?O,?Options)]` — same as `fmt_write(?Format,?Term,?Options)` but uses a previously open stream.

- `?Stream[fmt_read(?Format,?Result,?Status)]` — same, but use `?Stream` for input.

- `?Stream[write_canonical(?Term)]` — same, but use `?Stream` for output.

- `?Stream[read_canonical(?Term)]` — same, but use `?Stream` for input.

- `?Stream[readline(?Type,?String)]` — same, but use `?Stream` for input.

Portions Copyright © 2013–2017 Coherent Knowledge

Here are some examples. To read a file, one must open it first. If, say, `foo.txt` has `foo.p(a).` then the following will result:

```
?- (see('foo.txt'), read(?X), read(?Y), seen)@\io.

?X = foo
?Y = p(a)
```

The following illustrates writing to a file.

```
?- (tell('foo.txt'), telling(?Stream1), tell('bar.txt'),
    ?Stream1[writeln(abc)], writeln(cde), told, ?Stream1[close])@\io.
```

In this case, the files `foo.txt` and `bar.txt` will be created (or erased, if they exist) and `abd` will be written out to the first file and `cde` to the second. In more detail, the first `tell` opens `foo.txt` and the next `telling` command obtains a reference to the stream associated with that file. The next `tell` opens `bar.txt`. The default stream now becomes associated with that file, but we still have a reference `?Stream1` to the first stream, so we can write to it, which is done by the next command. After that, `cde` gets written to the default stream and `told` closes that stream. The last command closes (the non-default) `?Stream1` explicitly.

To read from or write to the window through which the user interacts with a running session of ℰRGO, one does not need to open (or `see`/`tell`) any files: when ℰRGO starts, the default input and output streams are connected to that window. For instance,

```
?- write(foobar)@\io.
```

will display `foobar` to the user unless the current default was changed by an earlier `tell` command. If the user needs to work with several files at once, he must keep track of the open streams by binding them to variables, as in the above example.

Debugging output. The above output methods and predicates make every effort to present the output in the valid ℰRGO syntax. However, sometimes—usually for debugging—one would like to see the internal representation of certain ℰRGO constructs. Using `write(...)@\prolog` is the ultimate way to see the internals, while `setdisplaymode{...}` and `setdisplayschema{...}` can be seen as an intermediate device. There are three display modes: `default`, `answer`, and `debug`. At present, the only difference is that in the `debug` mode the Skolems are written out using their internal representation while in the default mode the output looks like a ℰRGO term. In the `answer` mode the Skolems are shown both in the user-readable as well as the internal representation. For instance,

```
? setdisplaymode{debug}.
?- insert{p(\#)}, p(?_X), writeln(?_X)@\io.
_$_$_ergo'skolem2|_2

? setdisplaymode{default}.
?- insert{p(\#)}, p(?_X), writeln(?_X)@\io.
\#
```

The primitive `displaymode{?X}` can be used to find out the current mode in effect.

The primitive `setdisplayschema{...}` can be used to affect whether the answers and terms are printed using the ℰRGO's representation or the internal Prolog representation. For instance,

```
?- setdisplayschema{raw}.
?- ergo> ?X=p(?p).

?X = flapply(p,_h9195)
```

while in the default schema, `flora`, the answer will be shown as

```
ergo> ?X=p(?p).

?X = p(?_h9195)
```

The display schema also affects the `write(...)@\io` predicates.

The primitive `displayschema{?X}` tells which schema is currently in effect.

Common file operations. The `\io` module also provides a class `File`, which has methods for the most common file operations. These include:

- `File[exists(?F)].` True if file `?F` exists.
- `File[readable(?F)].` True if file `?F` is readable.
- `File[writable(?F)].` True if the file is writable.
- `File[executable(?F)].` True if the file is executable.
- `File[modtime(?F,?T)].`
  `File[modtime(?F)->?T].` Binds `?T` to the last modification time of `?F`.

- `File[mkdir(?Dir)].` Makes a directory named after the value of `?Dir`.
- `File[rmdir(?Dir)].` Removes the directory `?Dir`.
- `File[chdir(?Dir)].` Changes the current directory to `?Dir`.
- `File[cwd(?Dir)].`
  `File[cwd->?Dir].` Binds `?Dir` to the current working directory in the shell.
- `File[link(?F,?Dest)].` Creates a link named after `?F` to the existing file `?Dest`.
- `File[unlink(?F)].` Removes the link `?F`.
- `File[remove(?F)].` Removes the file `?F`.
- `File[tmpfilename(?F)].` Binds `?F` to a temporary file with a completely new name.
- `File[isabsolute(?F)].` True if `?F` is an absolute path name.
- `File[rename(?F,?To)].` Renames file `?F` to file `?To`.
- `File[basename(?F,?Base)].` Binds `?Base` to the base name of file path `?F`. For instance, `?- File[basename('/a/b/cde',?Base)].` would bind `?Base` to `cde`.
- `File[extension(?F,?Ext)].` Binds `?Ext` to the extension of the file ?F. For instance, `?- File[extension('/a/b/cde.exe',?Ext)]` would bind `?Ext` to `exe`.
- `File[dirname(?F,?Dir)].` Binds `?Dir` to the directory name of file `?F`.
- `File[expand(?F,?Expanded)].` Expands the file `?F` by attaching the directory name (if the file is not absolute) and binds `?Expanded` to that expansion.
- `File[newerthan(?F,?F2)].` True if `?F` is a newer file than `?F2`.
- `File[copy(?F,?To)].` Copies the contents of the file ?F to `?To`.

46.2 Storage Control

ℰRGO keeps the facts that are part of the knowledge base or those that are inserted at runtime in special data structures called *storage tries*. The system module `\db`, accessible through the module reference `@\db`, provides primitives for controlling this storage. This module also has a longer synonym `\storage`.

- `%commit` — commits all changes made by transactional updates. If this statement is executed in the middle of an update transaction, changes made by transactional updates prior to this will be committed and will not be undone even if a subsequent subgoal fails.
- `%commit(?Module)` — commits all changes made by transactional updates to facts in the user module `?Module`. Transactional updates to other modules are unaffected.
- `%purgedb(?Module)` — deletes all facts previously inserted into the storage associated with module `?Module`.

46.3 System Control

The system module `\sys` provides primitives that affect the global behavior of the system. It is accessible though the system module reference `@\sys` (or through its synonym `\system`).

- `Libpath[add(?Path)]` — adds `?Path` to the library search path. This works similarly to the `PATH` environment variable of Unix and Windows in that when the compiler or the loader are trying to locate a file specified by its name only (without directory) then they examine the files stored in the directories on the library search path.
  Using `Libpath[add(a(?Path))]` will move the directory to the front of the library search path (deleting any other occurrences of that directory on that search path).
  Using `Libpath[add(push(?Path))]` will put the directory at the front of the library search path. The other occurrences of that directory on the search path stay put.
- `Libpath[remove(?Path)]` — removes one `?Path` from the library search path.
- `Libpath[removeall(?Path)]` — removes all occurrences of `?Path` from the library search path.
- `Libpath[query(?Path)]` — queries the library search path. If `?Path` is bound, checks if the specified directory is on the library search path. Otherwise, binds (through backtracking) `?Path` to each directory on the library search path.
- `Tables[abolish]` — discards all tabled data in Prolog.

This module also provides the following amenities:

- `Method[mustDefine(?Mode)]` — affects the system behavior when stumbling upon an undefined predicate or method. This method is described separately, in Section 43.1.

Portions Copyright © 2013–2017 Coherent Knowledge

- `abort(?Message)` — prints `?Message` on the standard error stream and terminates the current execution. Message can also be in the form `(?M1, ?M2, ..., ?Mn)`. In this case, all the component strings are concatenated before printing them out.

  User aborts can be caught as follows:

  ```
  ?- catch{?Goal, ERGO_ABORT(ERGO_USER_ABORT(?Message),?_), ?Handler}
  ```

  In order to be able to use the predefined constants `ERGO_ABORT` and `ERGO_USER_ABORT` the file must contain the include statement

  ```
  #include "flora_exceptions"
  ```

- `warning(?Message)` — prints a warning header, then message, `?Message`, and continues. Output goes to the standard error stream. `?Message` can be of the form `(?M1, ?M2, ..., ?Mn)`.

- `message(?Message)` — Like `warning/1`, but does not print the warning header. `?Message` can be of the form `(?M1, ?M2, ..., ?Mn)`.

46.4 Type and Cardinality Checking

This system module of ERGO provides methods for testing type and cardinality constraints of the methods defined in the ERGO knowledge base. The module defines the method `check` in classes `Type` and `Cardinality` of module `\typecheck` (or, abbreviated, `\tpck`). This method is described in Section 43.3.

46.5 Data Types

The system module `\basetype` of ERGO provides methods for accessing the components of data types such as `\dateTime`, `\iri`, and so on. Data types are described in Section 39.

46.6 Reading and Compiling Input Terms

Sometimes it may be necessary for an application to read and compile ERGO statements from an input source. To this end, the `\parse` system library provides the following predicates and methods.

- `?- read(?Code,?Stat)@\parse.`

Read the next term from the standard input and compile it. The resulting term is bound to the variable `?Code`. The term can also be a reified formula and even a reified rule. Such a formula/rule can be used in a query or inserted into the knowledge base as appropriate.

`?Status` is bound to the status code returned by the predicate and has the form `[OutcomeFlag, EOF_flag|ErrorList]`, where:

OutcomeFlag = null/ok/error
null - a blank line was read, no code generated (?Code = null)
ok - good code was generated, no errors
error - parsing/compilation errors
EOF_flag = eof/not_eof
not_eof - end-of-file has not been reached
eof - if it has been reached.
ErrorList - if OutcomeFlag=null/ok, then this list would be empty.
if OutcomeFlag=error, then this would be a list of the form `[error(N1,N2,Message), ...]`, where N1, N2 encode the line and character numbers, which are largely irrelevant in this context. Message is an error message. Error messages are displayed.

Example:

```
?- read(?Code,?Stat)@\parse.
f(a).   <-----user input

?Code = f(a)
?Stat = [ok, not_eof]

?- read(?Code,?Stat)@\parse.
${a[b->c]@foo}.   <-----user input

?Code = ${a[b -> c]@foo}
?Stat = [ok, not_eof]
```

- `?- readAll(?Code,?Stat)@\parse.`

Used for reading terms one-by-one and returning answers interactively. The meaning of the arguments is the same. Under one-at-a-time solution (`\one`), will wait for input, return compiled code, then wait for input again, if the user types ";". If the user types `RET` then this predicate succeeds

and exits. Under all-solutions semantics (`\all`), will wait for inputs and process them, but will not return answers unless the file is closed (e.g., Ctl-D at standard input).

- `?- ?Source[readAll(?CodeList)]@\parse.`

This collects all answers from a source, which can be either a file or a string. If the source is a string, it should be specified as `string(Str)`, where `Str` is either an atom or a variable that is bound to an atom. If the source is a file, then it should be specified as `file(FileName)`, where `FileName` is an atom that specifies a file name (or a variable bound to it). `?CodeList` contains the result. It is bound to a list of the form `[code(TermCode1,Status1), code(TermCode2,Status2), ...]`, where `TermCode` is the compiled code of a term in the source, and `?Status` is the status of the compilation for this term. It has the form `[OutcomeFlag, EOF_flag|ErrorList]`, as explained before.

46.7 Displaying ℰRGO Terms and Goals

The system library loaded into the module `\show` can be used to obtain printable representation of ℰRGO terms and goals. This is needed when one needs to show those terms to the user in a form that the user can recognize (rather than in the internal form into which the terms are compiled).

The available methods are:

- `?Term[show->?Result]` — `?Result` is bound to an atom that represents the printable view of the term. For instance,

```
?- ${d(c,k,?M)}[show->?P]@\show.

?P = '${d(c,k,?_h0)@main}'

?- [f(a,?X),b,${d(c,k,?M)}][show->?Result]@\show.

?P = '[f(a,?_h0), b, ${d(c,k,?_h1)@main}]'
```

- `?Term[show(?Option)->?Result]` — like the above but also takes an option argument, which can be `goal` or `oid`. `?Term[show->?Result]` is the same as `?Term[show(oid)->?Result]`. The `goal` option affects the display of reified statements only. In that case, these statements are shown as goals (without the reification symbol), while with `oid` they are shown as objects. For instance,

```
?- ${p(a,b)}[show(oid) -> ?R]@\show.
```

```
?R = '${p(a,b)@main}'

?- ${p(a,b)}[show(goal) -> ?R]@\show.
?R = 'p(a,b)'
```

For HiLog terms, which are not reified, the two forms give the same result. For instance,

```
?- p(a,b)[show(oid) -> ?R]@\show.
?R = 'p(a,b)'
```

- `?List[splice(?Separator)->?Result]` — `?Result` is bound to an atom that represents the printable view of the sequence of elements in `?List` with `?Separator` (an atom) inserted in-between every pair of consecutive elements. For instance,

```
?- [f(a,?X),b,${d(c,k,?M)}][splice(' ')->?P]@\show.

?P = 'f(a,?_h0) b ${d(c,k,?_h1)@main}'

?- [f(a,?X),b,${d(c,k,?M)}][splice('| |')->?P]@\show.

?P = 'f(a,?_h0)| |b| |${d(c,k,?_h1)@main}'
```

47 ErgoText

ExgoText is a high-level facility intended to facilitate knowledge entry via English-like sentences and is primarily targeting engineers who are not highly skilled $\mathcal{E}$RGO users. However, skilled $\mathcal{E}$RGO users may also find ErgoText very convenient in many situations.

ErgoText documents typically have two parts: the actual knowledge base that contains template English sentences (with the usual extension `.ergo`) and another file, with the extension `.ergotxt`, which contains *template definitions*. Template definitions are typically created by a highly skilled $\mathcal{E}$RGO user; they provide mappings from English templates to $\mathcal{E}$RGO statements. The interesting feature of ErgoText is that template English can be freely mixed with $\mathcal{E}$RGO statements in the same document or even in the same rule. When an English sentence is found, template definitions are automatically looked up. If no matching template exists for the sentence in question, a compile-time error is issued. Otherwise, the $\mathcal{E}$RGO template definition is used instead of the sentence.

47.1 ErgoText Knowledge Bases

In this section, we assume that we have a file, `mykb.ergo`, which contains $\mathcal{E}$RGO statements intermixed with ErgoText sentences. Before such sentences can be used, the template definition file (say, `mytemplates.ergotxt`)[26] must be loaded either by executing the *runtime* query

```
?- ergotext{mytemplates}.
```

before compiling `mykb.ergo` or it must be declared at compile time by including the directive

```
:- ergotext{mytemplates}.
```

(note: `:-`, **not** `?-`) at the top of `mykb.ergo`. Note that `mytemplates.ergotxt` must be *locatable,* i.e., $\mathcal{E}$RGO must be able to find it. If the above `ergotext` query or directive appears inside a file, such as `mykb.ergo`, then `mytemplates.ergotxt` must be a path *relative to the folder* where `mykb.ergo` resides. It can also be an absolute path name, but this is **strongly discouraged** because it will make your knowledge base non-portable between machines. Thus, if `mykb.ergo` and `mytemplates.ergotxt` are in the same directory then simply `ergotext{mytemplates}` will do. If, say, all templates are in the `templates_dir` subfolder with respect to the file `mykb.ergo` that contains the directive, then `ergotext{templates_dir/mytemplates}` would work (the forward slash will work both for Unix *and* Windows).

If, on the other hand, `ergotext{mytemplates}` is posed as a query *on command line* in the $\mathcal{E}$RGO shell *then* `mytemplates.ergotxt` would typically be an absolute path name. It can also be a path name relative to the *current folder*, but in case of $\mathcal{E}$RGO Studio it is not always obvious what the current folder is.

Invoking `ergotext{?X}` as a query with an unbound variable argument will bind the argument to the currently active template file or will fail if none is active.

Note that <u>at most one</u> ErgoText template is *active* at any given moment. The active template is determined by either the most recent query of the form `?- ergotext{...}` or (if set within a file) by the nearest previous directive `:- ergotext{...}` within the same (or `#include'd`) file. This implies that a file can have several declared templates, but all the statements in-between the consecutive `:- ergotext` declarations will be interpreted by exactly one template. For instance, in

```
:- ergotexttemplate1.
statement1
:- ergotexttemplate2.
statement2
```

[26] It is perfectly fine to use the same base name for the template file, i.e., `mykb.ergotxt`.

```
statement3
:- ergotexttemplate3.
statement4
statement5
```

the statements 2 and 3 will be interpreted using `template2` and statements 4 and 5 via `template3`. This also means that all the ErgoText sentences that appear in statements 1 to 3 must be defined in `template2.ergotxt`. If one is defined, say, in `template3` or in `template1` but not in `template2` then an error will be issued. It should be said, however, that using more than one template per $\mathcal{E}$RGO module is a bad ides. It is a sure sign of poor design and **should be avoided**.

An *ErgoText sentence* has the form

```
\(word1 word2 ... wordN\)
```

where each word is either an English word, a HiLog term, including variables, or even an ISA or frame formula. Special symbols like parentheses, commas, semicolons, +, -, =, |, etc., are also allowed. Note that English words are a special case of HiLog terms. In fact, words do not need to be in English or, for that matter, use Latin characters, but then they have to be enclosed in single quotes, like any atom that contains non-alphanumeric characters. This also applied to text that contains single or double quotes. For instance,

```
\(My parent's house\)
\(My parent\'s house\)
```

will be rejected by the parser, but all of the ones below

```
\(My 'parent''s' house\)
\(My 'parent\'s' house\)
\(My parent''s house\)
```

are fine. (Note that in the last case `parent"s` actually consists of *three* separate tokens: `parent`, `"`, and `s` and separating them with spaces will yield the same three tokens, `parent " s`, and the same template.)

The following is a sample $\mathcal{E}$RGO knowledge base that consists mostly of ErgoText sentences, some of which (S3,S5) are intermixed with the normal $\mathcal{E}$RGO syntax.

```
\(Tom is not only a Student, but also a (salaried) Employee\).      // S1
\(If ?someone has ?something and paid for it then s/he likes it\).  // S2
\(?X cares for ?Y\), ?Y[inGoodShape] :-                             // S3
```

```
                \(?X has ?Y\), ?Y:ValuableThing.
\(insert rule: if ?X has more than 5M then he is rich\).          // S4
?- delete{\(Tom is an Employee\)}, ?X:Employee.                    // S5
?- deleterule{\(?someone with at least 5M is rich\)}.              // S6
```

Each sentence must match one and only one template (matching templates for the above are given below). All details of the syntax of the corresponding template must be followed to the letter: commas, parentheses (cf. S1), letter case, etc. Only variables can be replaced with HiLog terms when templates are used. For instance, as will be seen shortly, in S1 we replaced the variables in the template (T1) with the terms `Tom`, `Student`, and `Employee`, respectively. The sentences S1, S2, and S4 also illustrate the use of special features, like parentheses, commas, slashes, etc., which might be desirable for readability and other purposes.

Needless to say that the template text must be clear, unambiguous, and easily understood in order for ErgoText to be useful; achieving these goals is the job of the template designer. Typically, templates are designed for a particular domain (e.g., legal documents) and thus must involve both a domain expert to supply the template text and semantics and a skilled $\mathcal{E}$RGO user to provide the $\mathcal{E}$RGO translation.

To summarize, ErgoText sentences are English sentences that can be mixed with $\mathcal{E}$RGO syntax, such as terms, variables, etc. These sentences must be bound to $\mathcal{E}$RGO statements using template definitions (see below). These definitions are kept in separate template files (with the extension `.ergotxt`) and these files must be either declared or loaded (with the `ergotext{...}` directive) before compiling the $\mathcal{E}$RGO knowledge bases that contain the ErgoText sentences in question. These sentences can also be used in the $\mathcal{E}$RGO shell, provided that the corresponding template file is loaded beforehand—again using an appropriate `ergotext{...}` directive.

47.2 ErgoText Template Definitions

ErgoText template definitions are supposed to be stored in files with the extension `.ergotxt`. Each template has the form

`template(`*Context*, *Template*, *Ergo translation*`)`

The `Context` argument is one of the following reserved keywords (they are reserved only in the template definitions, not in general):

- `head`, `body`, `headbody`
- `insert`, `delete`, `insdel`

- `dynrule`

Why is context needed? The problem is that the same $\mathcal{E}$RGO formula may need to be compiled differently depending on where it occurs: rule head, rule body, in an insert or delete statement, etc. Some syntax may be allowed, say, in the rule body and insert/delete statements (e.g., `@module`) but not in rule heads; rule Ids and tags can occur only in facts and rules, but not in rule bodies; delete statements allow syntax that is not allowed in insert statements, and so on. Thus, the `head` context means that the corresponding template is allowed in rule heads (and facts), but not in rule bodies (e.g., S3 above). The `body` context means the template is suitable only for rule bodies (also S3). Since often the same template might be allowable in both contexts, a combined `headbody` context is provided for convenience (e.g., for S1 above). The `insert` and `delete` contexts are provided for the templates that are meant to be used inside the `insert{...}` and `delete{...}` statements, respectively (cf., S5). Some sentences may be suitable for both of these update commands and also for rule heads and bodies. An all-encompassing context `insdel`, which subsumes all four of the aforesaid contexts, is provided for this kind of sentences, for convenience. Finally, `dynrule` is a context for sentences that are meant to appear inside `insertrule{...}` and `deleterule{...}` statements (as in S6 above).

Apart from the above contexts, there are, of course, rules (outside of `insertrule{...}` and `deleterule{...}`), like S2 above, queries as in S4, and latent queries (not shown). The context for these is "anything else" and any keyword other than the above reserved ones can be used for them. It is a good practice to use `rule`, `fact`, `query`, and `latentquery` in these cases, for documentation purposes (so that you will be able to recognize and easily understand the purpose of the corresponding templates). Internally, these latter contexts are represented by the keyword `toplevel`, and this is what is shown by various debugging primitives described in Section 47.5.

Here are examples of template definitions that are suitable for the above knowledge base (S1-S6) to be placed in `mytemplates.ergotxt`:

```
template(headbody,   // headbody is reserved context                  (T1)
         \(?X is not only a ?Y, but also a (salaried) ?Z\),
         (?X:{?Y,?Z})).
template(rule,       // rule is not a reserved context keyword        (T2)
         \(If ?P has ?Thing and paid for it then s/he likes it\),
         (?P[likes->?Thing] :- ?P[{has,paidfor}->?Thing])).
template(head,      // reserved context, used for documentation       (T3)
         \(?X cares for ?Y\), ?X[caresfor->?Y]).
template(body,      // reserved context keyword                       (T4)
         \(?X has ?Y\), ?X[has->?Y]).
template(query,     // not reserved: query is used for documentation  (T5)
         \(insert rule: if ?X has more than 5M then he is rich\),
```

```
        (?- insertrule{?X:rich :- ?X[networth->?W], ?W>50000000})).
template(insdel,    // reserved for insert{...}/delete{...}          (T6)
        \(?X is an ?Y\), ?X:?Y).
template(dynrule,   // reserved                                      (T7)
        \(?someone with at least 5M is rich\),
        (?someone:rich :- ?someone[networth->?W], ?W>50000000))
```

The reader may have noticed that the $\mathcal{E}$RGO side of the above definitions (argument 3) is sometimes enclosed in parentheses, as in (T1), (T2), (T5), and (T7). This is needed in cases when the definitions contain infix operators, like `?-`, `:-`, `<==`, and the like, which bind its arguments weaker than the comma (or when the definitions contain commas, colons, as in (T1), or other operators, like semicolons). In these cases, the templates might not parse without the aforesaid parentheses.

Note that template definitions can be arbitrarily complex as long as they fit the declared context. For instance,

```
template(fact,
         \(Every state has a senator\),
         (forall(?State)^exist(?Sen)^senator(?State,?Sen) <== ?State:State))
```

In case one wants to create a template for rules or facts with descriptors, such as rule Ids, tags, and the like, keep in mind that rule descriptor constructs like `@{...}` and `@!{...}` are not permitted in ErgoText sentences. Instead, one should just use English and whatever symbols are allowed to stick in the variables for rule Ids and tags. For instance, a template definition that defines a rule with descriptors could look like this:

```
template(rule,
        \(Here's a rule with Id=?I and Tag=?T: If it rains, it pours\),
        (@!{?I} @{?T} pours(?X) :- rains(?X))).
```

One can then construct a concrete $\mathcal{E}$RGO rule via the following ErgoText sentence:

```
\(Here is a rule with Id=badweather & Tag = watchout: If it rains, it pours\).
```

47.3 ErgoText Queries and Command Line

So far we have been talking about ErgoText statements that appear in files. What if we want to ask an ErgoText query using the $\mathcal{E}$RGO shell, i.e., on command line (in a terminal, the Studio Listener, or the Studio Query tool)? Just as with file-based ErgoText, one must tell the command line processor which templates to use. There are two ways to do that.

The most common and convenient way to import a template file into the $\mathcal{E}$RGO shell is to load or add a file that contain the directive `:- ergotext{...}`. The template specified in that directive will be then imported into the module to which the file was loaded or added. For instance, if a file, say `example.ergo` is loaded into the module `main` and contains a directive of the form `:- ergotext{'some/place/templ'}` then one can issue a query like

```
?- \(my ?H template\).
```

in the $\mathcal{E}$RGO shell (assuming such a template exists in `some/place/templ.ergotxt`). If `example.ergo` is loaded into the module `foo` then

```
?- \(my ?H template\)@foo.
```

can be used as a command line query.

The other method is to load a template file explicitly. This can be used if the first method cannot be used for some reason. In Section 47.1, we already mentioned that ErgoText templates can be loaded at runtime by executing the query

```
?- ergotext{mytemplates}.
```

where `mytemplates.ergotxt` is a suitable template file. To load a template into a module other than `main`, a two-argument directive can be used. For instance,

```
?- ergotext{mytemplates,foobar}.
```

47.4 ErgoText and $\mathcal{E}$RGO Modules

ErgoText understands $\mathcal{E}$RGO modules, so the expressions like

```
p(?H) :- \(my ?H template\)@foo.
```

are understood subject to the following:

1. A compiler directive

   ```
   :- ergotext{mytemplate,foo}.
   ```

   must appear at the top of the file containing a sentence like `\(my ?H template\)@foo` above and the template file `mytemplate.ergotxt` must exist and be locatable, i.e., it must either be an absolute path name (not recommended) or a name relative to the directory of the containing file.

2. The module specification must be a constant; a variable or a quasi-constant like \@ will cause the compiler to issue an error. That is, something like the following

```
?- ?Mod = foo, \(my ?H template\)@?Mod.
?- \(my ?H template\)@\@.
```

are not allowed.

47.5 Debugging ErgoText Templates

$\mathcal{E}$RGO makes many syntactic checks to verify the templates are syntactically correct and otherwise make sense. It also checks that there is no *ambiguity* among templates; that is, that templates with *overlapping* contexts have disjoint template sentences.

Two contexts overlap if their sentences can be used in the same syntactic position in $\mathcal{E}$RGO statements. Clearly, identical contexts (like `body/body` or `head/head`) overlap. In addition, `headbody` is a shorthand for a set of contexts `body`, `head`, and `toplevel`. The context `head` actually also includes `toplevel`, while `insdel` includes `body`, `head`, `insert`, `delete`, and `toplevel`. So, `head` overlaps with `insdel` and `toplevel`, while `insdel` overlaps with `body`, `head`, and others. Two template sentences are *disjoint* if they cannot match the same ErgoText sentence; in other words, if they do not unify.

Besides the checks, $\mathcal{E}$RGO provides several primitives to help the template designer to debug the templates before releasing them to the end user. The most typical problem is when an ErgoText sentence *seems* to match a template, but it does not and the compiler issues an error claiming that no template definition matches the sentence. To better see the problem, consider the following template:

```
template(body,\(Is ?X true?\), ?X).
```

and a query

```
?- \(Is p:q true?\).
```

Intuitively, one might think that this should be translated into the query `?- p:q`, but it does not. The reason is that $\mathcal{E}$RGO's syntax is rich with infix operators, like `:`, `::`, `->`, comma, etc. These operators have precedence values that indicate which operator binds tighter. The magic of the operator grammars combines the above text into regular terms of the form `functor(argument_list)`. Since the details of that algorithm are complicated and one cannot and does not want to consult the table of precedence values, it is sometimes hard to see how a sentence, like the above, will

be parsed. The bottom line here is that the above query will *not* match the template—to much surprise and chagrin of the template developer and the user. To help debug such cases, $\mathcal{E}$RGO provides a parser API call `show_ergotext_sentence_as_term/1`, which can be used as follows:

```
?- show_ergotext_sentence_as_term(\(Is p:q true?\))@\prolog(flrparser).
ErgoText sentence parsed form:
    [Is, p : (q, (true, ?))]
```

This means that the above ErgoText query gets parsed as a list of two terms the second of which is `p:(q,(true,?))`. Asking the same question about the template sentence, we get:

```
?- show_ergotext_sentence_as_term(\(Is ?X true?\))@\prolog(flrparser).
ErgoText sentence parsed form:
    [Is, ?X, true, ?]
```

which is quite different. In such a situation, one can force the parser do the needful by placing commas in strategic places. For instance,

```
?- show_ergotext_sentence_as_term(\(Is, p:q, true?\))@\prolog(flrparser).
ErgoText sentence parsed form:
    [Is, p : q, true, ?]
```

yields the same parse as the sentence in the template, and one can verify that the above ErgoText sentence will now match the template.

To list all the active templates (i.e., templates that are currently affecting the parser), $\mathcal{E}$RGO's parser provides `ergo_show_active_templates/0`. The following example shows a use of that call along with some of the results it produces.

```
?- ergo_show_active_templates@\prolog(flrparser).
Active ErgoText templates:

  Template context:    body
  Template sentence:   ?A is not only a ?B but also a (salaried) ?C
  Template definition: ?A : {?B, ?C}

  Template context:    head
  Template sentence:   ?A cares for ?B
  Template definition: ?A[caresfor->{?B}]
```

```
Template context:    body
Template sentence:   ?A has ?B
Template definition: ?A[has->{?B}]

Template context:    delete
Template sentence:   ?A is an ?B
Template definition: ?A : ?B
```

If one wants to see templates that are *not* currently loaded (active), the API call `ergo_show_templates/1` will do it:

```
?- ergo_show_templates(template3)@\prolog(flrparser).
```

The output is similar to `ergo_show_active_templates/0`.

Given an ErgoText sentence, one might want to see the templates that match it. This is done via the calls `ergo_show_matching_templates/2` and `ergo_show_active_matching_templates/2`. The difference between these calls is that the latter matches against the active templates that are in effect in the *module* specified in the second argument, while the former takes the template file argument explicitly. For instance,

```
?- ergo_show_matching_templates(\(Is,p:q,true?\),template3)@\plg(flrparser).
Context:    body
Definition: p : q
```

or

```
?- ergo_show_active_matching_templates(\(Is,p:q,true?\),main)@\plg(flrparser).
```

The latter produces similar output but it contains the matches against the templates currently active in module `main`. The context part in the above listing is either one of the reserved keywords `body`, `head`, `insert`, `delete`, or `dynrule`, or it is `toplevel`. The latter corresponds to all other contexts, i.e., rule, query, latent query, or fact (facts can match either `head` context or `toplevel`).

47.6 ErgoText and Error Reporting

ErgoText templates present a number of challenges when it comes to error reporting, and the template designer must be aware of the possible pitfalls. One such pitfall has to do with templates that obscure so much of the underlying $\mathcal{E}$RGO formulas that it is hard for the end user to understand the reasons for various errors and warnings. To illustrate, consider the following buggy template:

```
template(body,\(test\),(\neg insert{foobar})).
```

Here explicit negation is applied to the insert operator, which is an error. The problem is that the template compiler uses a shallow parser, which does not catch all the errors and leaves further error detection to a later stage. As a result, the above template would be compiled without any errors or warnings. The error ("illegal use of explicit negation") will be reported only when the template will be actually used at compile or run time:

```
:- ergotext{mytemplates}.
?- \(test\).
```

Although the error will point to the line number where the above ErgoText sentence was used, the template designer (who failed to debug his templates) — and certainly the end user — will find it difficult to understand what is going on because the template sentence provides no hint regarding the underlying $\mathcal{E}$RGO statement.

For another example, consider the template

```
template(head,\(test\),p(b)).
```

and the knowledge base

```
:- ergotext{mytemplates}.
f(p(a)).
\(test\).
```

Although this is not a mistake, it is suspicious: the symbol `p/1` is used in two different ways here—once explicitly, as a HiLog function symbol, and once implicitly, as a predicate symbol (in the template). $\mathcal{E}$RGO will issue a warning about the ambiguous use of `p/1` (see Section 24), but the end user will likely be at a loss as to why.

The lesson of these examples is that ErgoText templates must be designed in a way that the template sentences will have enough information to help the end user resolve the possible issues with errors and warnings reported by $\mathcal{E}$RGO.

47.7 ErgoText and Text Generation

Text generation refers to the ability to generate English textual description of $\mathcal{E}$RGO facts, queries, and rules. Currently text generation is used for generating textual explanations to query answers—see Section 43.5.

Text generation is intimately tied to ErgoText in that a textual explanation is generated whenever a subgoal is chosen as an explanation and that subgoal matches the definitional part in an ErgoText template. In that case, the English sentence in the template will be shown as an explanation instead of the subgoal. Only the templates of type `body` (and, by implication, `headbody` and `insdel`) will be used for text generation. Thus, if you want to *exclude* a certain template from being used for text generation, use a context other than the above three. For instance, `rulehead`, `rulebody`, or some other easily recognizable keyword.

48 Unicode and Character Encodings

If all your applications deal only with ASCII characters, like the ones found on an English keyboard, you can safely skip this section. Otherwise, if your are dealing with alphabets other than English (even if it is just one of the European languages based on Latin alphabet) then read on.

48.1 What Is a Character Encoding?

Data (including programs, databases, and knowledge bases) is represented as sequences of characters, where each character is encoded as a sequence of bits. All these bits are useless, however, unless one knows what character each particular subsequence represents. The mapping between bit sequences and characters is known as a *code table*. The best-known coding table is *ASCII*, and it assigns a number between 0 and 127 to various characters found on English keyboards, including digits, lower- and uppercase letters, and punctuation.

Unfortunately, ASCII encodes too few characters and various groups took initiative to define bigger code tables, typically appropriating the characters in the 128 – 255 range for various national alphabets. One of the best-known such tables is *Latin-1,* which encodes all special accented letters in Latin-based alphabets. But these extra 128 symbols are too few to accommodate other alphabets, like Russian, Greek, Hebrew, etc. As a result, many more coding tables sprang up—all appropriating the same characters from the range 128 – 255 for the different alphabets, and this is where many problems originate. Data could now come in any of the dozens of different encodings and, unless one knows which encoding was used, that data is all but useless. On top of all this, Far-Eastern languages need so many characters for writing that they require different kinds of coding tables all together.

To bring some order into this business, *Unicode Consortium* was created and tasked to come up with *one* coding table that is capable of encoding all existing alphabets and common characters, including many mathematical characters, currency symbols, and then some. This eventually led to the Unicode standard known as UTF-8, which is a coding table that represents all known characters using sequences of one to four bytes. Thus, ASCII characters are represented using one

byte, Latin-1 (and much more) using two bytes, etc. This picture is complicated by the fact that Unicode Consortium did not come to UTF-8 right away, but through a series of less successful standards, like UTF-16 and UTF-32, which are still around and make things more complicated.

48.2 Character Encodings in $\mathcal{E}$RGO

With all this multitude of encodings, $\mathcal{E}$RGO supports only the three most common ones: UTF-8, CP1252, and Latin-1. Since these three (and, in fact, all coding systems) subsume and agree on the ASCII character set, the latter is, of course, also supported.

CP1252 is very similar to Latin-1 in its intent and their encodings agree everywhere except for about two dozen of symbols in the 128 – 159 range. The only reason why CP1252 is supported by $\mathcal{E}$RGO is that it happens to be the default on Windows and much of the data produced on that platform still tends to be encoded using the CP1252 coding table.

48.3 Specifying Encodings in $\mathcal{E}$RGO

To specify the encoding to use in processing programs or data, $\mathcal{E}$RGO provides one compiler directive and one executable directive as follows:

```
:- encoding{enc_name}.  // compiler directive
?- encoding{enc_name}.  // executable directive
```

where *enc_name* can be one of the following: `utf8`, `latin1`, `cp1252`.

First, if a knowledge base or a data file contains only ASCII characters, the encoding is immaterial and nothing needs to be specified. Otherwise, if one has a ruleset or a set of $\mathcal{E}$RGO facts that contain non-ASCII symbols, then the *compiler directive* `encoding` should be used at the top of the corresponding file. If a data file that contains non-ASCII characters needs to be read in or written out, the *executable directive* must be executed just before the file is opened (say, using the `see`, `tell`, or `open` operations).

Note that, on Linux and Mac, the default encoding is UTF-8, but on Windows it is CP1252. Therefore, forgetting to place or execute the `encoding` directive in a non-ASCII situation may result in the same knowledge base being compiled differently or the same file being ingested or written out differently under Linux and Windows.

We also note that the *scope* of the compiler directive `encoding` is the file in which the directive occurs. After the file is compiled and loaded, the encoding returns to what it was before. On the other hand, the effect of the *executable* directive persists until the next `encoding` directive is executed.

49 Notes on Style and Common Pitfalls

Knowledge engineering in $\mathcal{E}$RGO is similar to programming in Prolog, but is more declarative. For one thing, frame literals are always tabled, so the knowledge engineer does not need to worry about tabling the right predicates. Second, there is no need to worry that a predicate must be declared as dynamic in order to be updatable. Third — and most important — the facts specified in the knowledge base are stored in special data structures so that their order does not matter and duplicates are eliminated automatically.

49.1 Facts are Unordered

The fact that $\mathcal{E}$RGO does not assume any particular order for facts has a far-reaching impact on the knowledge engineering style and is one of the pitfalls that an engineer should avoid. In Prolog, it is a common practice to put the catch-all facts at the end of a program block in order to capture subgoals that do not match the rest of the program clauses. For instance,

```
p(f(?X)) :- ...
p(g(?X)) :- ...
%% If all else fails, simply succeed.
p(?_).
```

This will *not* work in $\mathcal{E}$RGO, because `p(?_)` will be treated as a database fact, which is placed in no particular order with respect to the program. If one wants the same effect in $\mathcal{E}$RGO, represent the catch-all facts as rules:

```
p(f(?X)) :- ...
p(g(?X)) :- ...
%% If all else fails, simply succeed. Use rule notation for p/1.
p(?_) :- \true.
```

49.2 Testing for Class Membership

In imperative programming, users specify objects' properties together with the statements about the class membership of those objects. The same is true in $\mathcal{E}$RGO. For instance, we would specify an object `John` as follows, which is conceptually similar to, say, Java:

```
John : person
[ name->'John Doe',
```

```
  address->'123 Main St.',
  hobby->{chess, hiking}
].
```

However, in $\mathcal{E}$RGO attributes can also be specified using rules. For instance, we can say that (in our particular enterprise) an employee works in the same building where the employee's department is located:

```
?X[building->?B] :- ?X:employee[department->?_[building->?B]].
```

Our experience in teaching F-logic-based knowledge engineering to users indicates that initially there is a tendency to confuse premises with consequents when it comes to class membership. So, a common mistake is to write the above as

```
?X:employee, ?X[building->?B] :- ?X[department->?[building->?B]].
```

A minute of reflection should convince the reader that this is incorrect, since the above rule is equivalent to two statements:

```
?X[building->?B] :- ?X[department->?_[building->?B]].
?X:employee :- ?X[department->?_[building->?B]].
```

It is the second statement, which is problematic. Certainly, we did not intend to say that any object with a `department` attribute pointing to an object with a `building` attribute is an employee!

It is interesting to note that such a confusion between premises and consequences is common only when it comes to class membership. Therefore, the user should carefully check the validity of placing class membership formulas in rule heads.

49.3 Composite Frames in Rule Heads

Another common mistake is the inappropriate use of complex frames in rule heads. When using a complex frame, such as `a[b->c, d->e]`, one must always keep in mind that its meaning is `a[b->c] and a[d->e]` whether the frame occurs in a rule head or in its body. Therefore, if `a[b->c, d->e]` occurs in the head of a rule like

```
a[b->c, d->e] :- body.
```

then the rule can be broken in two using the usual logical tautology $((\mathtt{X} \wedge \mathtt{Y}) \leftarrow \mathtt{Z}) \equiv (\mathtt{X} \leftarrow \mathtt{Z}) \wedge (\mathtt{Y} \leftarrow \mathtt{Z})$:

```
a[b->c] :- body.
a[d->e] :- body.
```

Forgetting this tautology sometimes causes logical mistakes. For instance, suppose `flight` is a binary relation that represents direct flights between cities. Then a rule like this

```
flightobj[from->?F, to->?T] :- flight(?F,?T).
```

is likely to be a mistake if the user simply wanted to convert the relational representation into an object-oriented one. Indeed, in the head, `flightobj` is a *single* object and therefore both `from` and `to` will get multiple values and it will not be possible to find out (by querying that object) which cities have direct flights between them. The easiest way to see this is through the use of the aforesaid tautology:

```
flightobj[from->?F] :- flight(?F,?T).
flightobj[to->?T] :- flight(?F,?T).
```

Therefore, if the `flight` relation has the following facts

```
flight(NewYork,Boston).
flight(Seattle,Toronto).
```

then the following frames will be derived (where the last two are unintended):

```
flightobj[from->NewYork, to->Boston].
flightobj[from->Seattle, to->Toronto].
flightobj[from->NewYork, to->Toronto].
flightobj[from->Seattle, to->Boston].
```

To rectify this problem one must realize that each tuple in the `flight` relation must correspond to a separate object in a rule head. The error in the above is that *all* tuples in `flight` correspond to the same object `flightobj`. There are two general ways to achieve our goal. Both try to make sure that a new object is used in the head for each `flight`-tuple.

The first method is to use a new function symbol, say `f`, to construct the oids in the rule head:

```
f(?F,?T):flight, f(?F,?T)[from->?F, to->?T] :- flight(?F,?T).
```

As an added bonus, we also created a class, `flight`, and made the flight objects into the members of that class. While it solves the problem, this approach might not always be acceptable, since the oid essentially explicitly encodes all the information in the tuple.

An alternative approach is to use the `skolem{...}` primitive from Section 13. Here we are using the fact that each time `flight(?F,?T)` is satisfied, `skolem{?X}` generates a new value for `?X`.

```
%convert_rel_to_oo :-
     flight(?F,?T), skolem{?O}, insert{?O:flight[from->?F,to->?T]}.
```

This approach is not as declarative as the first one, but it saves the user from the need to figure out how exactly the oids in the rule head should be constructed.

50 Miscellaneous Features

50.1 Suppression of Banners

When $\mathcal{E}$RGO initializes itself, it generates quite a bit of chatter, which is suppressed by default. The user who needs this information (e.g., the developer), can force the chatter to appear by starting $\mathcal{E}$RGO with

```
runergo --devel
```

In normal operation, $\mathcal{E}$RGO issues a prompt after every query or command. However, sometimes it might be necessary to suppress the prompt. For instance, when $\mathcal{E}$RGO interacts with other programs (*e.g.*, with a GUI) then sending the prompt to the other program just complicates things, as the receiving program needs to remember to ignore the prompt. To avoid this complication, the invocation flag `-noprompt` is provided. Thus,

```
runergo --noprompt
```

will print no chatter, not even the prompt, on startup and will be just waiting for user input. When the input occurs, $\mathcal{E}$RGO will evaluate the query and return the result. After this, it will return to wait for input without issuing any prompts.

50.2 Production and Development Compilation Modes

By default, $\mathcal{E}$RGO compiler compiles everything in the development mode. However, before deploying an application, it is desirable to recompile it in the production mode to gain significant performance benefits. One way to do this is to put the directive

```
:- compiler_options{production=on}.
```

in each file. However, this is often inconvenient, takes time, and might be error-prone. An alternative method is to execute the following command at the prompt:

```
?- production{on}.
```

Executing

```
?- production{off}.
```

puts $\mathcal{E}$RGO back into the development mode.

Note that the mode is reset to the default development mode after compiling any file, so `production{on}` must be re-executed each time when compilation in that mode is desired.

Also note that the explicit compiler directives

```
:- compiler_options{production=on}.
:- compiler_options{production=off}.
```

placed at the top of a file override any prior `production{on}/production{off}` commands.

51 Useful XSB Predicates Without a Counterpart in $\mathcal{E}$RGO

This section contains a list of useful predicates that are available in XSB, have no direct counterpart in $\mathcal{E}$RGO, and for which such counterparts are not being planned.

51.1 Time-related Predicates

- `cputime(?X)@\prolog` — returns the CPU time in seconds (floating point number; including fractions of seconds) used by the $\mathcal{E}$RGO process so far. This excludes the idle time.
- `walltime(?X)@\prolog` — returns the total time in (including fractions of seconds) seconds that elapsed since $\mathcal{E}$RGO started, including the idle time.
- `epoch_seconds(?X)@\prolog(machine)` — time in seconds (no fractions) that has passed since Thursday, 1 January 1970.
- `epoch_milliseconds(?Secs,?Fraction)@\prolog(machine)` — time that has passed since Thursday, 1 January 1970. Both `?Secs` and `?Fraction` are integers, where `?Secs` is the number of seconds and `?Fraction` is the number of additional milliseconds since the above date.

51.2 Hashing

- `term_hash(?Term, ?Size, ?Value)@\prolog(machine)` — binds `?Value` to the hash number of `?Term` in the range from 0 to `?Size-1`.
- `crypto_hash(?Type, ?Input, ?Result)@\prolog(machine)` — produce a cryptographic hash of the input and bind `?Result` to it. Input must be either an atom or have the form of `file(filename)@\prolog`. `?Type` specifies the type of the hash function to use. Currently, only `md5` (for MD5 hash) and `sha1` (for SHA1) are supported.

51.3 Input/Output

- `write(?X)@\prolog` and `writeln(?X)@\prolog` — these are similar to the corresponding predicates in the ℰRGO module `\io`, but they print out the internal form of the term bound to `?X`. For atoms, numbers, and variables, there is no substantial difference with `\io`, but for more complex terms there is. These predicates are mostly useful for debugging ℰRGO itself and for bug reporting.

51.4 Meta-programming

ℰRGO has a high-level predicate `=..`, which lets one inspect the internal structure of most ℰRGO's constructs in high-level terms of that system. However, if more sophisticated, low-level parsing of terms is needed, the following Prolog predicates can be used:

- `(?Left =.. ?Right)@\prolog` — decomposes the `?Left` term into a list of the form `[functor, arg1,...,argN]`. The result is bound to `?Right`.
- `functor(?Term,?Functor,?Arity)@\prolog` — binds `?Functor` to the function symbol used in `?Term` and `?Arity` to the arity of `?Term`.
- `arg(?ArgNum,?Term,?Arg)@\prolog` — binds `?Arg` to the `?ArgNum`'s argument of `?Term`.

52 Bugs in Prolog and ℰRGO: How to Report

The ℰRGO system includes a compiler and runtime libraries, but for execution it relies on Prolog. Thus, some bugs that you might encounter are the fault of ℰRGO, while others are Prolog bugs. For instance, a memory violation that occurs during the execution is in all likelihood an internal Prolog bug. (ℰRGO is a stress test — all bugs come to the surface.)

An incorrect result during the execution can be equally blamed on Prolog or on $\mathcal{E}$RGO— it requires a close look at the knowledge base. A compiler or a runtime error issued for a perfectly valid $\mathcal{E}$RGO specification is probably a bug in $\mathcal{E}$RGO.

Bugs that are the fault of the underlying Prolog engine are particularly hard to fix, because $\mathcal{E}$RGO knowledge bases are translated into mangled, unreadable to humans Prolog code. To make things worse, this code might contain calls to $\mathcal{E}$RGO system libraries.

To simplify bug reporting, $\mathcal{E}$RGO provides a utility that makes the compiled Prolog program more readable. The `dump{...}` primitive can be used to strip the macros from the code, making it much easier to understand. If you issue the following command

```
?- dump{foo}.
```

then `foo.ergo` will be compiled without the macros and dump the result in the file `foo_dump.P`. This file is pretty-printed to make it easier to read. Similarly,

```
?- dump{foo,bar}
```

will compile `foo.ergo` for module `bar` and will dump the result to the file `foo_dump.P`.

Unfortunately, this more readable version of the translated $\mathcal{E}$RGO specification might still not be executable on its own because it might contain calls to $\mathcal{E}$RGO libraries or other modules. The set of guidelines, below, can help cope with these problems.

Reporting $\mathcal{E}$RGO-related Prolog bugs. If you find a Prolog bug triggered by $\mathcal{E}$RGO, here is a set of guidelines that can simplify the job of the XSB developers and increase the chances that the bug will be fixed promptly:

1. Reduce the size of your $\mathcal{E}$RGO knowledge base as much as possible, while still being able to reproduce the bug.
2. Eliminate all calls to the system modules that use the `@\lib` syntax. (Prolog modules that are accessible through the `@\prolog(modname)` syntax are OK, but the more you can eliminate the better.)
3. If the knowledge base has several user modules, try to put them into one file and use just one module.
4. Use `dump{...}` to strip $\mathcal{E}$RGO macros from the output of the $\mathcal{E}$RGO compiler.

5. See if the resulting program runs under XSB (without the $\mathcal{E}$RGO shell). If it does not, it means that the program contains calls to $\mathcal{E}$RGO runtime libraries. Try to eliminate such calls.

   One common library call is used to collect all query answers in a list and then print them out. You can get rid of this library call by finding the predicate `fllibprogramans/2` in the compiled `.P` program and removing it while preserving the subgoal (the first argument) and renaming the variables (as indicated by the second argument). Make sure the resulting program is still syntactically correct!

   Other calls that are often no longer needed in the dumped code are those that load $\mathcal{E}$RGO runtime libraries (which we are trying to eliminate!). These calls have the form

   ```
   ?- flora_load_library(...).
   ```

   If there are other calls to $\mathcal{E}$RGO runtime libraries, try to delete them, but make sure that the bug is still reproducible.

6. If the program still does not run because of the hard-to-get-rid-of calls to $\mathcal{E}$RGO runtime libraries, then see if it runs after you execute the command

   ```
   ?- bootstrap_ergo.
   ```

   in the Prolog shell. If the program runs after this (and reproduces the bug) — it is better than nothing. If it does not, then something went wrong during the above process: start anew.

7. Try to reduce the size of the resulting program as much as possible.

8. Tell the XSB developers how to reproduce the bug. Make sure you include all the steps (including such gory details as whether it is necessary to call `bootstrap_ergo/0`).

Finally, remember to include the details of your OS and other relevant information. Some bugs might be architecture-dependent.

Reporting $\mathcal{E}$RGO bugs. If you believe that the bug is in the $\mathcal{E}$RGO system rather than in the underlying Prolog engine, the algorithm is much simpler:

1. Reduce the size of the program as much as possible by deleting unrelated rules and squeezing a multi-module program into just one file.

2. Remove all the calls to system modules, unless such a call is the essence of the bug.

3. Tell $\mathcal{E}$RGO developers how to reproduce the bug.

The current version contains the following known bugs, which are due to the fact that certain features are yet to be implemented:

1. Certain queries may cause the following XSB error message:

   ++Error[XSB]: [Compiler] '!' after table dependent symbol

   or something like that. This is due to certain limitations in the implementation of tabled predicates in the XSB system. This problem might be eliminated in a future release of XSB.

2. Inheritance of transactional methods is not supported: `a[|%p(?X)|]`.

53 The Expert Mode

Skip this section unless you are an experienced $\mathcal{E}$RGO user who has good understanding of the syntactic, semantic, and computational aspects of $\mathcal{E}$RGO.

$\mathcal{E}$RGO has plethora of syntactic constructs, which may be daunting for a novice and certain constructs are especially prone to be misused by such users to detrimental effect. To fence off these features from a novice, $\mathcal{E}$RGO's parser works in the novice mode by default. Only very experienced users should work in the expert mode—see Section 45.2 for how to do that.

The following syntactic constructs are available only in the expert mode:

- *Shortcut for charlists.* This shortcut permits to write `"abc"` instead of `"abc"^^\charlist`. Charlists are commonly used for parsing, but novice users tend to misuse them for the same purpose as atoms. Because charlists are much more expensive in terms of memory, compilation time, and various runtime operations, using them in place of atoms is a very bad idea.

- *Embedded ISA-literals in rule heads.* An embedded ISA-literal is one that appears as an argument to a predicate, function symbol, or is part of a composite frame. For instance, `p(a:b)`, `p(a,f(c::d))`, `a:b[c->d:f]`. $\mathcal{E}$RGO allows these to appear in rule bodies and in facts, but their appearance in rule heads (that have non-empty body) is restricted to the expert mode.

  One reason why embedded ISA-literals in rule heads are not advisable to novice users is because they tend to confuse them with typed variables. Note that

  ```
  p(?X:foo) :- ...
  p(?X^^foo) :- ...
  ```

are very different things and using the first form is frequently a mistake. An embedded ISA-literal, `p(?X:foo)`, represents the conjunction `p(?X), ?X:foo`, so the first rule above is equivalent to

```
p(?X), ?X:foo :- ...
```

That is, `?X:foo` is *derived*, not checked. Inexperienced users tend to incorrectly assume the latter. On the other hand, the second rule, one that uses typed variables is equivalent to

```
p(?X) :- ?X:foo, ...
```

but using `?X^^foo` can be much more efficient, since the actual tests for class membership are done only when the variable gets bound.

- *Expanded scope of the operators* `->` *and* `=>`. The symbols `->` and `=>` are infix operators when they appear in the context of a frame. In other contexts, using them as infix operators will cause an error, as mistyping round parentheses for square brackets is a common mistake. Advanced users, however, may wish to use these operators for other purposes as well, e.g., to simulate predicates with named arguments, such as `p(foo->1,bar->2)`. However, this syntax, is allowed only in the expert mode.

- *The* `<==>` *and* `<~~>` *double implications.* These double-implications are blocked off in the novice mode because inexperienced users tend make logical mistakes by using "iff" when only "if" is called for.

Appendices

A A BNF-style Grammar for $\mathcal{E}$RGO

This BNF is an approximation of the operator-based, context sensitive syntax used in $\mathcal{E}$RGO. Not all features mentioned in the preceding sections (especially directives and a number of non-logical commands) are listed in this BNF.

```
%% To avoid confusion between some language elements and meta-syntax
%% (e.g., parentheses and brackets are part of BNF and also of the language
%% being described), we enclose some symbols in single quotes to make it
%% clear that they are part of the language syntax, not of the grammar.
%% However, in ERGO these symbols can be used with or without the quotes.

Statement := (Rule | Query | LatentQuery | Directive) '.'

Rule := (RuleDescriptor)? Head (':-' Body)? '.'

Query := '?-' Body '.'

LatentQuery := RuleDescriptor '!-' Body '.'

Directive := ':-' ExportDirective | OperatorDirective | SetSemanticsDirective
             | IgnoreDependencyCheckDirective | ImportModuleDirective
                     | PrefixDirective | CompilerDirective | IndexDirective

RuleDescriptor := '{' RuleTag '}'
                  | '{' BooleanRuleDescriptor '}'
                  | '@!{' RuleId ( '[' DescrBody ']' )? '}'

RuleTag := Term
RuleId := Term
BooleanRuleDescriptor := Term
DescrBody := DescrBodyElement (',' DescrBodyElement)*
DescrBodyElement := Term | Term '->' Term

%% Heads in ERGO (not FLORA-2) can also have ==>, <==, <==>, \or, and quantifiers
Head := ('\neg')? HeadLiteral
```

```
Head := Head (',' | '\and') Head

HeadLiteral := BinaryRelationship | ObjectSpecification | Term

Body := BodyLiteral
Body := BodyConjunct | BodyDisjunct | BodyNegative | ControlFlowStatement
%% 'exists' can be used instead of 'exist'.
%% Body-parentheses are optional, if Body is BodyLiteral
Body := 'forall(' VarList ')' '^' '(' Body ')' | 'exist(' VarList ')' '^' '(' Body ')'
Body := Body '@' ModuleName
Body := BodyConstraint

ModuleName := Atom | 'Atom()' | Atom '(' Atom ')' | ThisModuleName

BodyConjunct := Body (',' | '\and') Body
BodyDisjunct := Body (';' | '\or') Body
BodyNegative := (('\naf' | '\neg | '\+') Body)
BodyConstraint := '{' CLPR-style constraint '}'

ControlFlowStatement  := IfThenElse | UnlessDo
                         | WhileDo | WhileLoop
                         | DoUntil | LoopUntil
IfThenElse := '\if' Body '\then' Body  ('\else' Body)?
               | Body '<~~' Body | Body '~~>' Body | Body '<~~>' Body
               | Body '<==' Body | Body '==>' Body | Body '<==>' Body
UnlessDo   := '\unless' Body '\do' Body
WhileDo    := '\while' Body '\do' Body
WhileLoop  := '\while' Body '\loop' Body
DoUntil    := '\do' Body '\until' Body
LoopUntil  := '\loop' Body '\until' Body

BodyLiteral := BinaryRelationship | ObjectSpecification | Term
               | DBUpdate | RuleUpdate | Refresh
               | NewSkolemOp | Builtin | Loading | Compiling
               | CatchExpr | ThrowExpr | TruthTest

Builtin := ArithmeticComparison | Unification | MetaUnification | ...

Loading := '[' LoadingCommand (',' LoadingCommand)* ']'
           | 'load{' LoadingCommand (',' LoadingCommand)* '}'
```

```
LoadingCommand := Filename ('>>' Atom)

Compiling := 'compile{' Filename '}'

BinaryRelationship := PathExpression ':' PathExpression
BinaryRelationship := PathExpression '::' PathExpression

ObjectSpecification := PathExpression '[' SpecBody ']'

SpecBody := ('\naf')? MethodSpecification
SpecBody := ('\neg')? ExplicitlyNegatableMethodSpecification
SpecBody := SpecBody ',' SpecBody
SpecBody := SpecBody ';' SpecBody

MethodSpecification :=  ('%')? Term
MethodSpecification :=  PathExpression
                        (ValueReferenceConnective | SigReferenceConnective)
                        PathExpression

ValueReferenceConnective :=  '->' | '+>>' | '->->' | '-->>'

SigReferenceConnective := ('{' (Integer|Variable) ':' (Integer|Variable) '}')? ('=>' )

ExplicitlyNegatableMethodSpecification := Term
ExplicitlyNegatableMethodSpecification :=
            PathExpression ExplicitlyNegatableReferenceConnective PathExpression

ExplicitlyNegatableReferenceConnective := '->' | SigReferenceConnective

PathExpression := Atom | Number | String | Iri | Variable | SpecialOidToken
PathExpression := Term | List | ReifiedFormula
PathExpression := PathExpression  PathExpressionConnective PathExpression
PathExpression := BinaryRelationship
PathExpression := ObjectSpecification
PathExpression := Aggregate

Iri := SQname | FullIri

SQname := Identifier '#' String
```

Portions Copyright © 2013–2017 Coherent Knowledge

```
FullIri := 'Śtring

PathExpressionConnective :=  '.' | '!'

SpecialOidToken :=  AnonymousSkolem | NumberedSkolem | ThisModuleName

ReifiedFormula := '${' (Body | '(' Rule ')') '}'

%% No quotes are allowed in the following quasi-constants!
%% No space allowed between \# and Integer
%% AnonymousSkolem & NumberedSkolem can occur only in rule head
%% or in reified formulas
AnonymousSkolem   := '\#'
%% No space between \# and Integer
NumberedSkolem    := '\#'Integer
ThisModuleName := '\@'

List := '[' PathExpression (',' PathExpression)* ('|' PathExpression)? ']'

Term := Functor '(' Arguments ')'

Term := '%' Functor '(' Arguments ')'

Functor := PathExpression

Arguments := PathExpression (',' PathExpression)*

Aggregate := AggregateOperator '{' TargetVariable (GroupingVariables)? '|' Body '}'
AggregateOperator := 'max' | 'min' | 'avg' | 'sum' | 'setof' | 'bagof'
%% Note: only one TargetVariable is permitted.
%% It must be a variable, not a term. If you need to aggregate over terms,
%% as for example, in setof/bagof, use the following idiom:
%%      S = setof{ V | ... , V=Term }
TargetVariable := Variable
GroupingVariables := '[' VarList ']'

Variable := '?' ([_a-zA-Z][_a-zA-Z0-9]*)?
VarList := Variable, (',' Variable)*

DBUpdate := DBOp '{' UpdateList ('|' Body)? '}'
```

```
DBOp := 'insert' | 'insertall' | 'delete' | 'deleteall' | 'erase' | 'eraseall'
%% In ℰRGO, UpdateList can also contain stealth literals
UpdateList := HeadLiteral ('@' Atom)?
UpdateList := UpdateList (',' | 'and') UpdateList
Refresh := 'refresh{' UpdateList '}'

RuleUpdate := RuleOp '{' RuleList '}'
RuleOp := 'insertrule' | 'insertrule_a' | 'insertrule_z' |
               'deleterule' | 'deleterule_a' | 'deleterule_z'
RuleList := Rule | '(' Rule ')'  ( (',' | 'and')  '(' Rule ')' )*

NewSkolemOp := 'skolem{' Variable '}'

CatchExpr := 'catch{' Body, Term, Body '}'
ThrowExpr := 'throw{' Term '}'
TruthTest := 'true{' Body '}' | 'undefined{' Body '}' | 'false{' Body '}'
               | 'truthvalue{' Variable '}'
```

B The ℰRGO Debugger

The ℰRGO debugger is implemented as a presentation layer on top of the Prolog debugger, so familiarity with the latter is highly recommended (XSB Manual, Part I). Here we sketch only a few basics.

The debugger has two facilities: *tracing* and *spying*. Tracing allows the user to watch the execution step by step, and spying allows one to tell ℰRGO that it must pause when execution reaches certain predicates or object methods. The user can trace the execution from then on. At present, only the tracing facility has been implemented in ℰRGO.

Tracing. To start tracing, you must issue the command `\trace` at the ℰRGO prompt. It is also possible to put the subgoal `\trace` in the middle of a program. In that case, tracing will start after this subgoal gets executed. This is useful when you know where exactly you want to start tracing the program. To stop tracing, type `\notrace`.

During tracing, the user is normally prompted at the four parts of subgoal execution: `Call` (when a subgoal is first called), `Exit` (when the call exits), `Redo` (when the subgoal is tried with a different binding on backtracking), and `Fail` (when a subgoal fails). At each of the prompts, the user can issue a number of commands. The most common ones are listed below. See the XSB manual for more.

- `carriage return (creep)`: to go to the next step
- `s (skip)`: execute this subgoal non-interactively; prompt again when the call exits (or fails)
- `S (verbose skip)`: like `s`, but also show the trace generated by this execution
- `l (leap)`: stop tracing and execute the remainder of the program

The behavior of the debugger is controlled by the predicate `debug_ctl`. For instance, executing `debug_ctl(profile, on)` at the $\mathcal{E}$RGO prompt tells XSB to measure the CPU time it takes to execute each call. This is useful for tuning your knowledge base for performance. Other useful controls are: `debug_ctl(prompt, off)`, which causes the trace to be generated without user intervention; and `debug_ctl(redirect, foobar)`, which redirects debugger output to the file named `foobar`. The latter feature is usually useful only in conjunction with the aforesaid prompt-off mode. See the XSB manual for additional information on debugger control.

$\mathcal{E}$RGO provides a convenient shortcut that captures some of the most common uses of the aforesaid `debug_ctl` interface. Executing

```
?- \trace('foobar.txt').
```

will switch $\mathcal{E}$RGO to non-interactive trace mode and the entire trace will be dumped to file `foobar.txt`. Note that you have to execute `\notrace` or exit Prolog in order for the entire file to be flushed to disk.

Another useful form of non-interactive tracing is to dump the trace into a file in the form of $\mathcal{E}$RGO facts, so that the file could later be loaded and queried. This is accomplished with the following call:

```
?- \trace('foobar.txt',log).
```

The second argument denotes the option to be passed to the trace facility. Currently the only available option is `log`. The form of the facts is as follows:

```
flora_tracelog(CallId,CallNumber,PortType,CurrentCall,Time)
```

Here `CallId` is an identifier generated when the engine encounters a new top-level call. This identifier remains the same for all subgoals called while tracing that top-level call. `CallNumber` is the call number that the underlying generates 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'`. `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.

Low-level tracing. The $\mathcal{E}$RGO debugger also supports low-level tracing via the shell command `\tracelow`. With normal tracing, the debugger converts low-level subgoals to $\mathcal{E}$RGO syntax and are thus meaningful to the programmer. With low-level tracing, the debugger displays the actual Prolog subgoals (of the compiled .P program) that are being executed. This facility is useful for debugging $\mathcal{E}$RGO runtime libraries.

As with `\trace`, $\mathcal{E}$RGO provides a convenient shortcut that allows the entire execution trace to be dumped into a file:

```
?- \tracelow('foobar.txt').
```

As with the `\trace/1` call, there is a `\tracelow/2` version:

```
?- \tracelow('foobar.txt',log).
```

which dumps the trace in the form of queriable facts. However, in this case the facts are in the low-level Prolog form, not $\mathcal{E}$RGO form.

References

[1] A.J. Bonner and M. Kifer. An overview of transaction logic. *Theoretical Computer Science*, 133:205–265, October 1994.

[2] A.J. Bonner and M. Kifer. A logic for programming database transactions. In J. Chomicki and G. Saake, editors, *Logics for Databases and Information Systems*, chapter 5, pages 117–166. Kluwer Academic Publishers, March 1998.

[3] W. Chen and M. Kifer. Sorted HiLog: Sorts in higher-order logic programming. In *Int'l Conference on Database Theory*, number 893 in Lecture Notes in Computer Science, January 1995.

[4] W. Chen, M. Kifer, and D.S. Warren. HiLog: A foundation for higher-order logic programming. *Journal of Logic Programming*, 15(3):187–230, February 1993.

[5] K. Clark. Negation as failure. *Logic and Databases*, pages 293–322, 1978.

[6] J. Frohn, G. Lausen, and H. Uphoff. Access to objects by path expressions and rules. In *VLDB*, pages 273–284, 1994.

[7] M. Gelfond and V. Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Computing*, 9(3/4):365–386, 1991.

[8] M. Kifer, W. Kim, and Y. Sagiv. Querying object-oriented databases. In *Proceedings of the ACM SIGMOD International Conference on the Management of Data*, pages 393–402, New York, June 1992. ACM.

[9] M. Kifer, G. Lausen, and J. Wu. Logical foundations of object-oriented and frame-based languages. *Journal of the ACM*, 42:741–843, July 1995.

[10] Senlin Liang and Michael Kifer. A practical analysis of non-termination in large logic programs. *Theory and Practice of Logic Programming*, 13:705–719, September 2013.

[11] Senlin Liang and Michael Kifer. Terminyzer: An automatic non-termination analyzer for large logic programs. In *Practical Aspects of Declarative Languages (PADL)*, January 2013.

[12] J. W. Lloyd. *Foundations of Logic Programming*. Springer-Verlag, 1984.

[13] A. Van Gelder. The alternating fixpoint of logic programs with negation. In *ACM Principles of Database Systems*, pages 1–10, New York, 1989. ACM.

[14] A. Van Gelder, K.A. Ross, and J.S. Schlipf. The well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):620–650, 1991.

Portions Copyright © 2013–2017 Coherent Knowledge

[15] H. Wan, B. Grosof, M. Kifer, P. Fodor, and S. Liang. Logic programming with defaults and argumentation theories. In *International Conference on Logic Programming*, July 2009.

[16] G. Yang and M. Kifer. Well-founded optimism: Inheritance in frame-based knowledge bases. In *Intl. Conference on Ontologies, DataBases, and Applications of Semantics for Large Scale Information Systems (ODBASE)*, October 2002.

[17] G. Yang and M. Kifer. Inheritance and rules in object-oriented semantic web languages. In *Rules and Rule Markup Languages for the Semantic Web (RuleML03)*, number 2876 in Springer Verlag, October 2003.

[18] G. Yang and M. Kifer. Inheritance in rule-based frame systems: Semantics and inference. *Journal on Data Semantics*, 2800:69–97, 2003.

[19] G. Yang and M. Kifer. Reasoning about anonymous resources and meta statements on the Semantic Web. *Journal on Data Semantics, LNCS 2800*, 1:69–98, September 2003.

Portions Copyright © 2013–2017 Coherent Knowledge

Index

++ operator, 24
− − operator, 24
&& operator, 24
! ∼
    meta-disunification operator, 72
|| operator, 24
<~~>, 309
<==>, 309
∼
    meta-unification operator, 69, 71, 117, 147
'\\compile"/1 predicate, 55
'\\flimport' operator, 55
'\\load'/1 predicate, 55
:-
    in UDF definitions, 125
:=
    in UDF definitions, 124
@!{...} descriptor query, 185
@\plgall, 79, 138
@\prologall, 79, 138
[+file], 51
#, 213
${...} reification operator, 72
runergoscript, 2
ERGO_ABORT, 241
bootstrap_ergocommand, 54
ERGO_RC_FILE environment variable, 2
ERGO_UNDEFINED_EXCEPTION, 241
\+, 79
\naf, 79
\?C quasi-variable, 28
\?F quasi-variable, 28
\?L quasi-variable, 28
\@ quasi-constant, 28
\@! quasi-constant, 28
\@!, current rule Id quasi-constant, 182
\@F quasi-constant, 28
\@L quasi-constant, 28
\# quasi-constant, 28
\#1, \#2, etc., quasi-constant, 28
\abolishtables, 145
\callable built-in class, 235
\charlist, 234
\dump_incomplete_tables/1, 254
\false, 10
\gcl module, 198
\if
    in UDF definitions, 124
\in list containment operator, 232
\in list membership operator, 232
\isloaded/1 predicate, 56
\modular built-in class, 235
\nontabled_module directive, 172
\notrace, 7, 314
\opposes predicate, 199
\overrides predicate, 199
\repeat option in do-loops, 168
\set_dump_incomplete_tables_on_exception/0, 255
\skolem class, 40
\table directive, 176
\tabledump, 252
\trace, 7, 314
\trace/1, 315
\trace/2, 315
\tracelow, 315
\true, 10
\undefined, 10
\until, 169
<~~, 85, 167
<~~>, 85
~~>, 85, 167
~~>vs. ==> vs. :-, 88
<==>

omniformity, 90
<==
omniformity, 90
==>
omniformity, 90
\@!, current rule Id quasi-constant, 181
\dump_incomplete_tables_after_exception/1, 255
\@, 47
+>>, 166
-noprompt, 303
->->, 165
[file], 5, 48
<==>, 85
<==, 85, 89
==>, 85, 89
=.. meta-decomposition operator, 75
\?C quasi-variable, 48

active ErgoText template, 288
activegoals option, 270
add{...}, 51
add{...} primitive, 5
aggregate operator
avg, 161
avgdistinct, 161
bagof, 161
count, 161
countdistinct, 161
max, 161
min, 161
setof, 161
sum, 161
sumdistinct, 161
aggregates
set-valued methods, 165
aggregation
aggregate operator, 161
grouping, 161
variable, 161
alert, 186, 193
of type conflict, 193
of type truth, 193, 195
with callback, 194
alert{...} primitive, 194
ambiguity
among ErgoText templates, 294
anonymous variable, 6
answer abstraction, 258
answer size, 266
answersize option, 269
arithmetic expression, 21
inline, 25
arithmetic operator, 21
atomic formula, 15
attach method, 236
attachNew method, 236
attribute
class-level, 94
object-level, 94

base formula, 10
in F-logic, 12
base part of predicate, 135
body
ErgoText context, 290
boolean method
class-level, 35
builtin
delayable, 42
bulk delete, 139
bulk insert, 137

call abstraction, 256, 269
callback
in alert, 194
in integrity constraint, 191
caller{...} primitive, 48
cardinality constraint, 248
catch{...}, 243, 245

Portions Copyright © 2013–2017 Coherent Knowledge

catch{...} primitive, 240
character encoding, 298
    CP1252, 299
    Latin1, 298, 299
    UTF-8, 298, 299
character list, 234
chatter{...}, 7
check method
    class Cardinality, module \tpck, 246
    class Cardinality, module \typecheck, 246
    class Type, module \typecheck, 249
class, 14
    \boolean, 226
    \date, 221
    \dateTime, 218
    \decimal, 229
    \double, 227
    \duration, 225
    \integer, 229
    \iri, 217
    \list, 231
    \long, 228
    \string, 229
    \symbol, 212
    \time, 223
    expression, 114
    instance, 14
    subclass, 14
class_expressions
    setsemantics option, 109, 113, 274
clause{...}, 185
clause{...} primitive, 153
cloneterm{...} primitive, 41
code inheritance, 106
comment, 18
compact IRI, 213
compile{...} primitive, 48
compiler directive, 273
    compile-time, 273
    equality, 110
    executable, 273
    export, 61
    index{...}, 274
    op, 19, 275
    setsemantics{...}, 274
compiler_options directive, 275
constant
    Skolem, 36
      numbered, 37
constant symbol, 11
    general, 11
constoff, 134
constraint
    automatically maintained, 188
    integrity, 188
    type, 96
Constraint solving, 170
constraint{...}
    adding, 189
    deleting, 189
constraint{...} primitive, 189
counter{...} primitive, 149
CP1252 character encoding, 299
curi, 213
curi prefix, 213
current directory, 66
current folder, 66
current rule Id quasi-constant, \@!, 181, 182
custom
    setsemantics option, 108, 274
cut, 123
    across tables, 123
    in ERGO, 123

data frame, 13
datatype
    \boolean, 210, 226
    \date, 210, 221

\dateTime, 218
\decimal, 229
\double, 210, 227
\duration, 210, 225
\float, 210
\integer, 210, 229
\iri, 210, 213
\list, 231
\long, 210, 228
\string, 229
\symbol, 212
\time, 223
debugger, 314
debugging, 242
decimal, 12
default negation, 79
default property, 95
defeasibility-capable document, 182
defeasible
descriptor property, 181, 186
defeasible rule, 202
defsensor directive, 173
delay quantifier, 176
delayable builtin, 42
delayable subgoal, 177
delete
bulk, 139
ErgoText context, 290
delete{...} primitive, 139
deleteall{...} primitive, 139
deleterule, 151
deleterule_a, 151
deleterule_z, 151
deleterule{...} primitive, 150
dependency checking
ignore_depchk, 146
warnings{compiler}, 146
derived part of predicate, 135
descriptor
see *rule descriptor*, 180
directive, *see* compiler directive
\nontabled_module, 172
defsensor, 173
encoding{...}, 299
importmodule, 63
prolog, 171
table, 171
use_rule_structure_db, 156
usesensor, 173
useudf, 130
directory
current, 66
disable{...} primitive, 182
displaymode{...} primitive, 281
displayschema{...} primitive, 281
do-until, 167
don't care variable, 6
dump{...} primitive, 306
dynamic module, 111
dynamic rule, 150
dynrule
ErgoText context, 290

empty frame, 14
enable{...} primitive, 182
encapsulation, 60
encoding{...} directive, 299
environment variable
ERGO_RC_FILE, 2
equality, 110, 131
setsemantics option, 108, 274
equality maintenance level, 109
equality with user defined functions, 124
Ergo Studio, 252
Ergo Suite, 252
ergo_show_active_templates/0, 295
ErgoText, 288
ErgoText context
body, 290
delete, 290

dynrule, 290, 291
head, 290
headbody, 290
insdel, 290
insert, 290
ErgoText sentence, 289
ErgoText template, 290
active, 288
ambiguity, 294
debugging, 294
disjoint sentences, 294
overlapping contexts, 294
ergotext{...} primitive, 288, 293
escape sequences, 16
expert compiler option, 275
expert{...} primitive, 275
explicit negation, 83
export compiler directive, 61
expression
arithmetic, 21

F-logic frame formula, 15
false{...} primitive, 82
feedback{...}, 7
file
descriptor property, 185, 186
file name
absolute, 66
relative, 66
floating number, 17
floating point number, 12
flora_query/4 predicate, 56
folder
current, 66
forest logging, 263
formula
atomic, 15
base, 10
frame, 15
data, 13
empty, 14
logic expressions, 21
object value, 34
signature, 13
truth value, 34
frame literal, 15
free variable, 87
function
in UDF definitions, 124

GCL, 198
general constant symbol, 212
Generalized Courteous Logic, 198
goalsize option, 268

head
ErgoText context, 290
head-formula, 89
headbody
ErgoText context, 290
here{...} primitive, 68
HiLog, 68
translation, 68
unification, 69
HiLog to Prolog conversion, 78

I/O
standard interface, 276
stream-based, 276
Id-term, 12
if
in if-then-else statements, 166
ignore_depchk, 146, 251
immediate execution of delayed subgoals, 178
immediate execution operator, 178
importmodule directive, 63
index compiler directive, 274
inheritance
behavioral, 93
monotonic, 93

non-monotonic, 93
of code, 106
of value, 106
setsemantics option, 108, 274
structural, 93
initialization file, 2
inline arithmetic expression, 25
insdel
ErgoText context, 290
insert
bulk, 137
ErgoText context, 290
insert{...} primitive, 136
insertall{...} primitive, 136
insertrule, 150
insertrule_a, 150
insertrule_z, 150
insertrule{...} primitive, 150
integer, 12, 17
integrity constraint, 186, 188
with callback, 191
IRI prefix scope, 214
irilocalprefix directive, 213, 216
iriprefix directive, 213
isatom{Arg,Mode}, 43
isatom{Arg}, 41
isatomic{Arg,Mode}, 43
isatomic{Arg}, 41
isbasefact{...} primitive, 155
ischarlist{Arg,Mode}, 43
ischarlist{Arg}, 41
iscompound{Arg,Mode}, 43
iscompound{Arg}, 41
isdecimal{Arg,Mode}, 43
isdecimal{Arg}, 41
isdefeasible{...}, 184
isdisabled{...} primitive, 182
isenabled{...} primitive, 182
isfloat{Arg,Mode}, 43
isfloat{Arg}, 41
isground{Arg}, 41
isinteger{Arg,Mode}, 43
isinteger{Arg}, 41
isiri{Arg,Mode}, 43
isiri{Arg}, 41
islist{Arg,Mode}, 43
islist{Arg}, 41
isloaded{FileAbsName,Module,
FileLocalName,Mode} primitive, 50
isloaded{FileAbsName,Module,Mode} primitive,
50
isloaded{Module} primitive, 49
isnonground{Arg}, 41
isnonvar{Arg}, 41
isnumber{Arg,Mode}, 43
isnumber{Arg}, 41
isskolem{...} primitive, 40
isstrict{...}, 184
isstring{Arg,Mode}, 43
isstring{Arg}, 41
issymbol{Arg,Mode}, 43
issymbol{Arg}, 41
isvar{Arg,Mode}, 43
isvar{Arg}, 41

latent query, 186
insertion and deletion, 187
meta-querying, 188
Latin1 character encoding, 298, 299
list containment
\subset, 232
contains, 232
list membership
\in, 232
member, 232
literal
isa, 14
subclass, 14
load{...} primitive, 48
loading files, 48

Portions Copyright © 2013–2017 Coherent Knowledge

local name, 213
logical expressions, 20
logical operator, 21
loop-until, 169

makedefeasible{...}, 184
makestrict{...}, 184
matchAll method, 236
matchOne method, 236
meta-data
    in inserted rules, 151
    in reified rules, 74
    in rules, 180
meta-decomposition operator =.., 75
meta-disunification operator !∼, 72
meta-programming, 70
meta-unification operator ∼, 69, 71, 117, 147
method, 13
    boolean, 35
      class-level, 35
    self, 30
    transactional, 119
    value-returning, 13
Method class in module \system, 242
module, 44
    \prolog, 52
    \prolog(modulename), 52
    \prologall, 53
    \prologall(modulename), 53
    \@, 47
    \modulename, 59
    contents, 44
    descriptor property, 185, 186
    isloaded{Module}, 49
    name, 44
    Prolog, 45, 52
    rules for, 46
    system, 45, 59, 276
    user, 45
      compilation of, 48
      reference to, 45
module \basetype, 217
module \btp, 217
mustDefine/1 in class Method, module \system, 242
mustDefine/2 in class Method, module \system, 242

negation
    default, 79
    explicit, 83
    Prolog-style, 79
    well-founded semantics for, 79
newmodule/1, 150
non-logical operator, 21
non-tabled predicate
    in $\mathcal{E}$RGO, 116
non-tabled predicates
    importing into Prolog, 55
non-termination, 255
non-transactional update, 136
    delete, 138
    deleteall, 138
    erase, 138
    eraseall, 138
    insert, 136
    insertall, 136
nonstrict
    subclassing semantics, 112
number, 17
numbered Skolem constant, 37

object
    base part of, 135
    derived part of, 135
object constructor, 11
object identifier, 12
oid, 12
    generated at run time, 37

omniformity of <==, ==>, and <===>, 90
op compiler directive, 275
operator
    ++, 24
    −−, 24
    &&, 24
    ||, 24
    arithmetic, 21
    logical, 21
    non-logical, 21
operators, 18
    precedence level, 19
    type, 19

p2h{...}, 78
p2h{...} primitive, 54
passive tabling, 143
    interaction with reactive, 143
path expression, 29
peek{...} primitive, 120
peeking into query results, 120
persistence, 65
predicate
    base part of, 135
    derived part of, 135
prefix{...} primitive, 215
primitive
    add{...}, 5
    alert{...}, 194
    caller{...}, 48
    catch{...}, 240
    clause{...}, 153
    cloneterm{...}, 41
    compile{...}, 48
    constraint{...}, 189
    counter{...}, 149
    delete{...}, 139
    deleteall{...}, 139
    deleterule{...}, 150
    disable{...}, 182
    displaymode{...}, 281
    displayschema{...}, 281
    dump{...}, 306
    enable{...}, 182
    false{...}, 82
    here{...}, 68
    insert{...}, 136
    insertall{...}, 136
    insertrule{...}, 150
    isbasefact{...}, 155
    isdisabled{...}, 182
    isenabled{...}, 182
    isloaded{...}, 49
    isskolem{...}, 40
    load{...}, 48
    p2h{...}, 54
    prefix{...}, 215
    production{...}, 304
    query{...}, 187
    semantics{...}, 115
    setdisplaymode{...}, 280
    setdisplayschema{...}, 281
    setruntime{...}, 266
    showgoals{...}, 255
    skolem{...}, 38
    tag{...}, 203
    tdelete{...}, t_delete{...}, 141
    tdeleteall{...}, t_deleteall{...}, 141
    tdisable{...}, 182
    tenable{...}, 182
    terase{...}, t_erase{...}, 141
    teraseall{...}, t_eraseall{...}, 141
    throw{...}, 240
    true{...}, 82
    truthvalue{...}, 83
    undefined{...}, 82
    variables{...}, 41
primitive expert{...}, 275
production compiler option, 275
production{...} primitive, 304

Prolog atom, 11, 212
prolog directive, 171
Prolog module, 45, 52
Prolog to HiLog conversion, 78
property
    default, 95

quantifier
    all, 85
    delay, 176
    each, 85
    exist, 85, 89
    exists, 85, 89
    forall, 85, 89
    some, 85
quasi-constant, 10, 28
quasi-variable, 10, 28
query
    latent, 186
    latent; insertion and deletion, 187
    latent; meta-querying, 188
Query Tool, 252
query{...} primitive, 187

reactive tabling, 142
    interaction with passive, 143
refresh{...}, 144
regular expression, 236
reification operator ${...}, 72
relative file name, 66
rule
    defeasible, 202
    dynamic, 150
    static, 150
    strict, 202
rule deletion, 150
rule descriptor, 180
rule Id, 180, 182
    in inserted rules, 151
    in reified rules, 74
rule insertion, 150
rule meta-data, 180
rule structure database, 156
runtime Skolem symbol, 37

scratchpad code, 50
semantics{...} primitive, 115
sensor, 173
set-valued methods
    aggregation, 165
setdisplaymode{...} primitive, 280
setdisplayschema{...} primitive, 281
setmonitor..., 262
setruntime{...} primitive, 266
setsemantics
    custom=filename, 115
    directive, 108
    equality=basic, 110
    equality=none, 110
    inheritance=flogic, 112
    inheritance=monotonic, 112
    inheritance=none, 112
    subclassing=nonstrict, 112
    subclassing=strict, 112
    tabling=passive, 143
setsemantics{...} compiler directive, 274
setwarnings{...}, 7
show
    method in module \show, 286
show_ergotext_sentence_as_term/1, 295
showgoals{...} primitive, 255, 260
showtables{...} primitive, 260
signature
    in F-logic, 13
Skolem
    constant, 36
    function, 36
    generated at run time, 37
    symbol, 36
    term, 36

Portions Copyright © 2013–2017 Coherent Knowledge

Skolem constant, 36
skolem{...} primitive, 38
skolem..., 37
splice
    method in module \show, 286
spying, 314
static rule, 150
stealth update, 145, 148
stealth update and UDFs, 149
stealth{...}, 148
strict
    descriptor property, 181, 186
    subclassing semantics, 112
strict rule, 202
Studio Listener, 252
subclass, 14
subclassing
    setsemantics option, 109, 274
subclassing semantics
    nonstrict, 112
    strict, 112
subgoal
    controlled by delay quantifier, 177
    delayable, 177
subgoal size, 266
symbol, 16, 133
symbol_context, 133, 134
system module, 45, 59, 276

t_insert, 191
table directive, 171
tabling, 116
    interaction of passive and reactive, 143
    passive, 143
    reactive, 142
    setsemantics option, 108, 274
tag
    descriptor property, 180, 186
tag{...} primitive, 203
tdelete{...}, t_delete{...} primitives, 141
tdeleteall{...}, t_deleteall{...} primitives, 141
tdisable{...} primitive, 182
tenable{...} primitive, 182
terase{...}, t_erase{...} primitives, 141
teraseall{...}, t_eraseall{...} primitives, 141
Terminyzer, 255
throw{...} primitive, 240
timeout, 266
timeout option, 267
timer
    max, 267
    repeating, 267
timer interrupt handler
    abort, 267
    fail, 267
    ignore, 267
    pause, 267
tinsert, 191
tracing, 314
transactional update, 141
    t_delete, tdelete, 141
    t_deleteall, tdeleteall, 141
    t_erase, terase, 141
    t_eraseall, teraseall, 141
    t_insert, tinsert, 141
    t_insertall, tinsertall, 141
tripwire, 259
true{...} primitive, 82
truthvalue{...} primitive, 83
type
    descriptor property, 186, 188
    user-defined, 235
type checking, 245
type constraint, 13, 96
typed variable, 32, 235

UDF
    :=, 124

\if, 124
and stealth update, 149
user-defined function, 124
undefined{...} primitive, 82
Unicode, 299
Unicode character, 16, 234
unification option, 270
unless-do, 167
until, 167
update, 135
non-transactional, 136
stealth, 145, 148
transactional, 141
updates
and tabling, 142
URL, 210
use_argumentation_theory directive, 198
use_rule_structure_db directive, 156
usefunction
implicit directive, 130
user defined functions, 124
user module, 45
user-defined type, 235
usesensor directive, 173
useudf directive, 130
UTF-8 character encoding, 298, 299

value inheritance, 106
variable, 11
aggregation, 161
anonymous, 6
don't care, 6
free, 87
implicitly quantified, 87
typed, 32, 235
variables{...} primitive, 41

warnings{...}, 8
warnoff, 134
well-founded semantics, 102
well-founded semantics for negation, 79
while-do, 167
while-loop, 169
why{...}, 252