Maude as a Metalanguage ¹

M. Clavel a, F. Durán b, S. Eker b, P. Lincoln b, N. Martí-Oliet c, J. Meseguer b, and J. F. Quesada d

a Department of Philosophy, University of Navarre, Spain
b SRI International, Menlo Park, CA 94025, USA
c Facultad de Ciencias Matemáticas, Universidad Complutense, Madrid, Spain
d CICA (Centro de Informática Científica de Andalucía), Seville, Spain

Abstract

One of the key goals of rewriting logic from its beginning has been to provide a semantic and logical framework in which many models of computation and languages can be naturally represented. There is by now very extensive evidence supporting the claim that rewriting logic is indeed a very flexible and simple logical and semantic framework. From a language design point of view the obvious question to ask is: how can a rewriting logic language best support logical and semantic framework applications, so that it becomes a metalanguage in which a wide variety of logics and languages can be both semantically defined, and implemented? Our answer is: by being reflective. This paper discusses our latest language design and implementation work on Maude as a reflective metalanguage in which entire environments—including syntax definition, parsing, pretty printing, execution, and input/output—can be defined for a language or logic \( L \) of choice.

1 Introduction

This paper is an overview of our latest language design and implementation work on Maude. The goals of language design and implementation are at the service of the uses that we envision for Maude, and in turn, these should help furthering the overall goals of the rewriting logic research program.

We are particularly interested in supporting uses of rewriting logic as a logical and semantic framework, in which many different logics, models of computation, and languages can be represented, can be given a precise semantics, and can be executed. Since rewriting logic is used as what we might

¹ Supported by DARPA through Rome Laboratories Contract F30602-97-C-0312, by Office of Naval Research Contract N00014-96-C-0114, and by National Science Foundation Grant CCR-9633363.
call a \textit{metalogic}, it seems very natural to conceive a rewriting logic language as a \textit{metalanguage}, in which we can create executable \textit{environments} for many other languages and logics. Such metalanguage capabilities are among our highest priorities in designing Maude.

Rewriting logic is itself reflective [12,7]. By efficiently supporting this reflection in Maude through its predefined module \texttt{META-LEVEL}, we in fact obtain a very useful \textit{gateway} to the metatheory of rewriting logic, enabling Maude to be used as a metalanguage. The essential reason for the appropriateness of this solution is that the desired representation maps and environments manipulate metalevel entities such as program or specification modules in the object language of interest and in Maude. Reflection then gives us the desired disciplined access to such a metalevel within the logic, and underlies very generic techniques for syntax definition, parsing, pretty printing, and input/output that allow us to build language environments in Maude. We explain all this in more detail—and how it is supported by the Maude language and implementation—in the rest of the paper.

It is useful to distinguish a CoreMaude sublanguage of \textit{flat} functional and system modules—corresponding, respectively, to equational and rewriting logic specifications that do not have any submodules. This core language is efficiently supported by Maude's rewriting engine, as it is discussed—along with other aspects of the Maude system—in Section 2. The reflective capabilities are then supported by the module \texttt{META-LEVEL}. Section 3 explains how this module provides the cornerstone for making Maude highly extensible and flexible. For extending Maude, and, more generally, for creating an environment for any other language \( \mathcal{L} \) using Maude, we need generic syntax definition, meta-parsing, and meta.pretty printing capabilities that can deal with expressions in any language, including languages like Maude whose modules have user-definable syntax. And we need a general facility for input/output that can be customized for each language of interest. Section 4 explains how all this can be done in Maude thanks to its reflective design. We conclude the paper with some remarks about related work and future plans.

\section{The CoreMaude Interpreter}

The Maude system is built around the CoreMaude interpreter, which is implemented in C++ and consists of two parts, each of which is made up of a number of components.

\subsection{The Rewrite Engine}

The rewrite engine is an improved version of that described in [11]. The overall design is highly modular and does not contain any Maude-specific code. Two key components are the "core" module and the "interface" module. The "core" module contains classes for objects which are not specific to a particular
theory, such as equations, rules, sorts, and connected components of sorts. The “interface” module contains abstract base classes for objects that may have a different representation in different equational theories, such as symbols, term nodes, dag nodes, and matching automata. New equational theories can be “plugged in” by deriving from the classes in the “interface” module. To date, all combinations of associativity, commutativity, left and right identity, and idempotence have been implemented apart from those that contain both associativity and idempotence. New built-in symbols with special rewriting (equation or rule) semantics may be easily added.

The engine is designed to provide the look and feel of an interpreter with source level tracing and the possibility of adding user interrupts and debugging facilities. These goals prevent a number of optimizations that one would normally implement in a compiler, such as transforming the user’s term rewriting system, or keeping pending evaluations on a stack and only building reduced terms. The actual implementation is a semi-compiler where the term rewriting system is compiled to a system of tables and automata which is then interpreted.

2.2 The Mixfix Frontend

The mixfix frontend consists of a bison/flex based parser for Maude’s surface syntax, a grammar generator (which generates the context free grammar for the mixfix parts of Maude over the user’s signature), a context free parser, the built-in functions for the metalevel, together with a considerable amount of “glue” code which holds everything together. Many of the C++ classes in this part of the system are derived from those in the rewrite engine.

2.3 The SCP Parsing Algorithm

The current version of Maude implements a parser based on the SCP parsing algorithm [31]. Basically, SCP is a bidirectional, bottom-up and event-driven parser for unrestricted context-free grammars. From a mathematical point of view, we have proved that SCP is sound and complete for the set of CFGs. From a computational perspective, SCP avoids overparsing [30], a common problem of CF parsers like Earley, chart or GLR. The use of multi-virtual trees [29] at the level of representation and the relations of coverage, partial derivability and adjacency as top-down predictions over the basic bottom-up strategy, obtain a high level of efficiency without diminishing the generality of the algorithm.

2.4 Other Aspects of the System

The current Maude system exhibits many other important properties. It supports a metalevel with efficient rewriting and representation-shifting operations. Heavy use is made of the rewrite engine’s extensible design in the
implementation of many aspects of the metalevel. The current mixfix front-
end supports flat functional and system modules, i.e., there is little or no
support for importation, renaming, parameterization or object-oriented mod-
ules. Typical performance with the current version is 800K-840K free-theory
rewrites per second and 27K-111K AC-theory rewrites per second on standard
hardware (300MHz P6). The figure for AC rewriting is highly dependent on
the complexity of the AC patterns (AC matching is NP-complete) and the
size of the AC subjects. The above results were obtained using fairly simple
linear and non-linear patterns and large (hundreds of nested AC operators)
subjects.

3 The Module META-LEVEL

Informally, a reflective logic is a logic in which important aspects of its metathe-
ory can be represented at the object level in a consistent way, so that the
object-level representation correctly simulates the relevant metatheoretic as-
pects. In other words, a logic which can be faithfully interpreted in itself.
Rewriting logic is a reflective logic [12,7] in the precise sense that there is
a finitely presented universal rewrite theory $\mathcal{U}$ that can simulate all other
finitely presented theories, including itself. Maude’s language design and im-
plementation make systematic use of this fact to give the user a gateway to the
metatheory of rewriting logic in a clear and principled way. This gateway is
the predefined functional module META-LEVEL, that provides the user with key
functionality in the universal theory for rewriting logic. In other words, the
module META-LEVEL can be seen as a (partial) specification of the metatheory
of rewriting logic. In particular, in the module META-LEVEL:

- Maude terms can be reified as elements of a data type Term of terms;
- Maude modules can be reified as terms in a data type Module of modules;
- the processes of reducing a term to normal form in a functional module
  and of finding whether such a normal form has a given sort are reified by a
  function meta-reduce;
- the process of applying a rule of a system module to a subject term is reified
  by a function meta-apply;
- the process of rewriting a term in a system module using Maude’s default
  interpreter is reified by a function meta-rewrite; and
- parsing and pretty printing of a term in a signature, as well as key sort
  operations such as comparing sorts in the subsort ordering of a signature
  are also reified by corresponding metalevel functions.

The operations meta-reduce, meta-apply, and meta-rewrite, the pars-
ing and pretty printing operations, and the sort operations, are all functions
equationally definable in the module META-LEVEL. However, for efficiency and
convenience, they are built-in functions in Maude. The Maude implementa-
tion produces the exact same behavior as if \textit{META-LEVEL} had been defined as a normal functional Maude module, that could itself be represented as a term of sort \texttt{Module} in \textit{META-LEVEL}. In this way, a “reflective tower” with an arbitrary number of levels of reflection is supported. A detailed explanation of the signature and operations of the current version of \textit{META-LEVEL} is given in [8].

Notice that, since the Maude system is essentially a particular implementation of the metatheory of rewriting logic, the module \textit{META-LEVEL} can also be used as a gateway to the Maude system itself, so that the user can effectively re-design Maude (in Maude) to fit his/her particular computational needs. Following this line of thought:

- Clavel and Meseguer [13,7] have shown how user-definable \textit{internal strategy languages}, that typically extend the module \textit{META-LEVEL}, can be used to change the (default) operational semantics of Maude for system modules;
- Durán and Meseguer [15] have explained how Maude’s parameterized programming style in the Clear and OBJ tradition [5,17] can be internalized in Maude in an extension of \textit{META-LEVEL} and enriched with new modes of parameterization, new methods of program composition, and new ways of defining views of other program modules; and
- Maude’s (default) syntax for programs can be re-defined in an extension of \textit{META-LEVEL}, so that programs written in a new user-defined syntax will be understood by Maude’s parser (see Sections 4.1 and 4.2).

This potentiality of the module \textit{META-LEVEL} as a building block that users can employ to build customized versions of Maude has inspired us to make it the cornerstone of the Maude system design. The part of the Maude system implemented in C++ (the CoreMaude interpreter presented in Section 2) basically consists of the rewrite engine and a parser for flat modules. The rest of the system is fully specified in Maude as an extension of the module \textit{META-LEVEL}. We plan to further optimize the module \textit{META-LEVEL} in order to make metalevel computations even more efficient, and to extend it with new built-in functions to increase its usability and range of applications.

4 Maude as a Metalanguage

One of the key goals of rewriting logic from its beginning has been to provide a \textit{semantic framework} in which many models of computation—particularly concurrent and distributed ones—and languages can be naturally represented. Because of the intrinsic duality between logic and computation that rewriting logic supports, the very same reasons making rewriting logic a suitable semantic framework, make it also an attractive \textit{logical framework} [20] to represent many different logics. Thanks to the sustained efforts of many researchers, particularly in the ELAN, Pisa, Stanford, and Maude teams, there is by now very extensive evidence supporting the claim that rewriting logic is indeed a
very flexible and simple semantic framework \cite{21,23,24,4}, and logical framework \cite{20,18,34,19,3,32,6,14,7,10}. Moreover, object-oriented design languages, architectural description languages, and languages for distributed components also have a natural semantics in rewriting logic \cite{35,25,33,27,28}.

What is common to all these applications is that the models of computation, logics, or languages are represented in rewriting logic by mappings of the form

$$\Phi : \mathcal{L} \longrightarrow RL\text{ogic}.$$ 

From a language design point of view the obvious question to ask is: how can a rewriting logic language best support such representation maps, so that it becomes a metalanguage in which a very wide variety of programming, specification, and design languages, and of computational and logical systems can be both semantically defined, and implemented in it?

Our answer is: by being reflective. Specifically, we can define an abstract data type $\text{Module}_\mathcal{L}$ representing modules in the logic or language $\mathcal{L}$, and internalize a representation map $\Phi$ as an equationally defined function

$$\overline{\Phi} : \text{Module}_\mathcal{L} \longrightarrow \text{Module},$$

where $\text{Module}$ is the data type representing finitely presentable rewrite theories in the language—in Maude this data type is part of the META-LEVEL module (see Section 3). In fact, thanks to the general metarelut of Bergstra and Tucker \cite{2}, any computable representation map $\Phi$ can be specified in this way by a finite number of Church-Rosser and terminating equations. Using this metalevel gateway, we can then execute in Maude the rewrite theory $\overline{\Phi}(M)$ associated to a module $M$ in $\mathcal{L}$. This has been done, for example, for linear logic in \cite{7}, for structured Maude modules in \cite{15}, and could be done for a very wide range of other languages and logics using the same method.

In practice, however, we would like to be able not only to represent and transform the modules of a language $\mathcal{L}$ as terms of a data type $\text{Module}_\mathcal{L}$ within Maude. We would like to use Maude to generate a whole environment for $\mathcal{L}$, including a facility for defining and modifying the language's syntax, an input/output facility, a parser, a pretty printer, and an execution environment for it. Furthermore, we would like to be able to do this for languages that—like Maude itself, and many other formal specification languages—have modules with user-definable syntax, so that expressions in those modules cannot be parsed with a fixed syntax for the language, but need to be parsed with the particular syntax introduced in the module.

4.1 Syntax Definition

In order to generate in Maude a whole environment for a language $\mathcal{L}$, the first thing we need to do is to define the syntax for $\mathcal{L}$-modules. This can be done by extending the module META-LEVEL with a data type $\text{Module}_\mathcal{L}$ for $\mathcal{L}$-modules, and with other auxiliary data types for commands and other constructs. This
can be easily and naturally done using the mixfix frontend, and the built-in
data types Token (any identifier) and Bubble (any string of identifiers). The
intuition behind these types is that they correspond to pieces of a module
in a language that can only be parsed once the grammar introduced by the
signature of the module is available.

The idea is that for a language that allows modules with user-definable
syntax—as it is the case of Maude—it is natural to see its syntax as a combined
syntax, at two different levels: what we may call the top level syntax of the
language, and the user-definable syntax introduced in each module. The data
types Token and Bubble allow us to reflect this duality of levels in the syntax
definition.

To illustrate these concepts, suppose that we want to define the syntax of
Maude in Maude. Consider the following Maude module:

\begin{verbatim}
fm\text{od }\text{NAT}3 \text{ is}
sort Nat3 .
op 0 : \rightarrow Nat3 .
op s_\_ : Nat3 \rightarrow Nat3 .
eq \left[ \begin{array}{ccc} s & s & s \\ s & s & 0 \end{array} \right] = \left[ \begin{array}{c} 0 \end{array} \right] .
endfm
\end{verbatim}

Notice that the strings of characters inside the boxes are not part of the
top level syntax of Maude. In fact, they can only be parsed with the grammar
associated to the signature of the module NAT3. A correct and useful (in terms
of defining afterwards a meaningful parser) definition of the syntax of Maude
in Maude must reflect this duality of syntax levels. We show below parts of
the module MAUDE that defines the syntax for functional modules in Maude:

\begin{verbatim}
fm\text{od }\text{MAUDE is}
sort PreModule PreCommand .
subsort Decl < DeclList .
op eq_=\_ : Bubble Bubble \rightarrow Decl .
op f\text{mod}_\text{is}_\text{endfm} : \text{Token DeclList} \rightarrow \text{PreModule} .
op \text{red in}_\text{\_}\_\_ : \text{Token Bubble} \rightarrow \text{PreCommand} .
\end{verbatim}

Notice how we explicitly declare operators that correspond to the top level
syntax of Maude, and represent as terms of sort Bubble those pieces of the
module that can only be parsed with the user-defined syntax; for example,
the left and righthand sides of an equation.

Then, the functional module NAT3 above can be parsed as a term of sort
PreModule in MAUDE. The name of this sort reflects the fact that not all terms
of PreModule actually represent Maude modules. In particular, for a term of
sort PreModule to represent a Maude module all the Bubbles must be correctly
parsed in the user-defined syntax. We will come back to this important point
in the next section. Finally, notice that we have also defined the syntax for
commands like reduce in Maude. We will see in Section 4.4 how the user can
define the semantics of these commands.

4.2 Parsing and Pretty Printing

A built-in function meta-parse is declared in META-LEVEL with syntax

\[
\text{op meta-parse : Module Bubble} \rightarrow \text{Term}.
\]

The function \text{meta-parse}(M, i_1 \ldots i_n) checks whether the string of iden-
tifiers \(i_1 \ldots i_n\) is a well-formed term in the module represented by \(M\); if this is
the case, \text{meta-parse} returns the representation of \(i_1 \ldots i_n\) as a term of sort
\text{Term} in META-LEVEL; otherwise, it returns error*.

This built-in function can be very useful to specify in Maude an efficient
parser for a language \(\mathcal{L}\) that allows modules with user-definable syntax. For
example, a parser for Premodules in Maude essentially consists of a function
\text{parse-premodule} with syntax

\[
\text{op parse-premodule : PreModule} \rightarrow \text{Module?}.
\]

Given a PreModule \(M\), parse-premodule generates, first, a Module \(M'\) that
represents a module in Maude with the exact same signature as \(M\) but without
equations; then, parse-premodule attempts to transform the PreModule \(M\)
into a Module \(M''\) by reducing each Bubble \(b\) in \(M\) to a Term \(t\), where \(t\) is the
successful result of \text{meta-parse}(M', b). In case of failure, an error term in the
supersort Module? of Module is returned.

The inverse function to \text{meta-parse} is also included as a built-in function
in the module META-LEVEL, with syntax

\[
\text{op meta-pretty-print : Module Term} \rightarrow \text{Bubble}.
\]

It takes as arguments the representation of a module \(M\) and the representation
of a term \(t\). It returns a string of identifiers produced by pretty printing \(t\) in
the syntax given by \(M\).

4.3 Module Algebra and Execution

We are interested in using Maude to build environments for languages \(\mathcal{L}\) such
as formal specification languages and modular programming languages whose
modules can have user-definable syntax. In addition, such modules can be
highly structured and parameterized. That is, there can be a rich collection
of module composition operations endowing \(\mathcal{L}\) with a module algebra. In such
cases we typically have two data types of modules, a data type \text{Module}_\mathcal{L}\ of
flat or unstructured modules, and a more general data type \text{StrModule}_\mathcal{L}\ of
structured modules.

The point is that both data types and all the module algebra operations
for \(\mathcal{L}\) can be defined within Maude as an extension of the module META-LEVEL.
In this way, the environment that we can build for \(\mathcal{L}\) using Maude can also
support all the module composition operations of \(\mathcal{L}\). Among those module
operations a common and important one is \textit{flattening}, that is, the process of passing from a structured module to its unstructured flat form. This can be understood as a function

\[
\text{StrModule}_\mathcal{L} \xrightarrow{(\_)} \text{Module}_\mathcal{L}.
\]

Since modularity constructs can change from language to language, it may be simpler to represent \( \mathcal{L} \) in rewriting logic by representing only its flat modules, that is, by a function \( \Phi : \text{Module}_\mathcal{L} \rightarrow \text{Module} \) which makes the language \( \mathcal{L} \) executable on top of Maude. But using the function \( (\_)^{\mathcal{L}} \) we can also make structured modules in \( \mathcal{L} \) executable by means of the function composition

\[
\text{StrModule}_\mathcal{L} \xrightarrow{(\_)^{\mathcal{L}}} \text{Module}_\mathcal{L} \xrightarrow{\Phi} \text{Module}.
\]

The first most obvious language \( \mathcal{L} \) to which we can apply these ideas is Maude itself. The CoreMaude sublanguage has flat modules represented by the data type \texttt{Module}. But general Maude modules can be structured and parameterized, and can contain very complex \textit{module expressions} that instantiate and rename several, possibly parameterized, modules; and all this can also happen for object-oriented modules. A module algebra for Maude written in Maude is presented in [15]. In particular, such an algebra contains as well a flattening function making structured Maude modules executable in the Maude engine.

4.4 Input/Output

Using object-oriented concepts, we can specify in Maude a general input/output facility provided by a module \texttt{LOOP} that extends the module \texttt{META-LEVEL} into a generic read-eval-print loop. This facility can then be specialized for each language \( \mathcal{L} \). There is a class \texttt{I/O} of input/output objects acting on behalf of users, with two attributes: an input and an output buffer storing \texttt{Bubbles}. There is also a class \texttt{System} with an input buffer of sort \texttt{Input}, which is defined as a supersort of \texttt{PreModule} and \texttt{PreCommand}, an output buffer of sort \texttt{Bubble}, and an attribute of sort \texttt{Database}. The sort \texttt{Database} is left completely unspecified, so that, depending on the language \( \mathcal{L} \) of interest it can be specialized accordingly. For example, for Maude it will contain the current database of modules already entered in the system.

subsorts \texttt{PreCommand} \texttt{PreModule} < \texttt{Input} .

class \texttt{I/O} | input : \texttt{Bubble}, output : \texttt{Bubble} .

class \texttt{System} | db : \texttt{Database}, input : \texttt{Input}, output : \texttt{Bubble} .

The module \texttt{LOOP} provides a \textit{persistent} pair of objects: one of class \texttt{I/O} called \texttt{user}, and one of class \texttt{System} called \texttt{system} that interact with each other by exchanging data in and out between their buffers.

For each particular language both the sort \texttt{Database} for the \texttt{system} object and the additional rewrite rules defining the system behavior are specified according to the specific details of the language in question. We illustrate
below the case of Maude. Processing of a PreModule once it has been entered into the system is done by the rule

\[
\text{var DB : Database . var PM: PreModule .}
\]
\[
\text{var I : Identifier . var B : Bubble .}
\]
\[
\text{rl [premodule] :}
\]
\[
\text{< system : System | db : DB, input : PM >}
\]
\[
\Rightarrow < \text{system : System | db : processPreModule(PM, DB),}
\]
\[
\text{input : empty} > .
\]

The function processPreModule attempts to parse the PreModule using the function parse-premodule, and, if it succeeds, it introduces the resulting Module into the database.

Then, for user-defined commands as \text{red in} \_\_\_ we can define rules of the form

\[
\text{rl [reduce] :}
\]
\[
\text{< system : System | db : DB, input : red in I : B .} >
\]
\[
\Rightarrow < \text{system : System | input : empty,}
\]
\[
\text{output : meta-pretty-print(getModule(I, DB),}
\]
\[
\text{meta-reduce(getModule(I, DB),
\]
\[
\text{meta-parse(getModule(I, DB), B)))} > .
\]

where \text{getModule} is a function that extracts from the database the flat module whose name is given in its first argument.

5 Concluding Remarks

We have explained our current design and implementation work to support logical and semantic framework applications of Maude as a rewriting logic metalanguage. This work should be seen as a further step in the context of previous efforts to use algebraic languages to give executable semantic definitions of other languages, including OBJ [16], ASF+SDF [1], action semantics [26], and ELAN [18]. What seems to be new is the systematic use of reflection to achieve more powerful metalanguage functionalities, as well as the capacity to deal with languages that can be extended by user-definable syntax. In particular, when these capabilities are applied to Maude itself, they allow new very flexible ways of extending the language and its module operations, and of changing its syntax.

Much more work remains ahead. The first, most immediate goal is the upcoming release of Maude in the next few months, that will include the metalanguage capabilities described here. This will enable many new logical and semantic framework applications to be developed, such as execution environments for logics, theorem provers, architecture description languages, specification languages, and natural language processing systems. After the release of Maude we will probably concentrate on further extensions of the lan-
guage such as built-in unification modulo different equational theories—that will permit narrowing computations—, built-in objects and foreign interface modules, and Mobile Maude, a distributed and mobile extension. We will also work on further developing Maude’s theorem proving environment [9].

Acknowledgement

We thank Peter Mosses for his remarks on a draft of this paper. The work of Maude is part of a larger international collaboration on the rewriting logic research program, including important contributions to the tool-building effort by the ELAN and CafeOBJ groups, and many theoretical developments and applications [24]. We are particularly grateful to those researchers with whom we have worked most closely on Maude and rewriting logic ideas, including our fellow members at the Logic and Specification Group at SRI, as well as Adel Bouhoula, Roberto Bruni, Jean-Pierre Jouannaud, Ugo Montanari, and Carolyn Talcott.

References


12


