UML-based Frameworks for Engineering of Interoperable Information Systems

Marie-Noëlle Terrasse, George Becker, Marinette Savonnet, and Eric Leclercq
Laboratoire LE2I, Université de Bourgogne
B.P. 47870, 21078 Dijon Cedex, France
E-mail: {terrasse,becker,savonnet,leclercq}@khali.u-bourgogne.fr

March 9, 2003

Abstract

We propose a four-layer metamodeling architecture in the context of UML metamodeling. This allows modelers to use a three-step modeling process: 1) giving an informal description of the universe of the discourse (in terms of modeling paradigms); 2) defining a corresponding UML dialect (in terms of metamodels); 3) describing a model of an information system by using the chosen dialect. By using particular properties of our metamodeling architecture, we can define formal and semantical operations on metamodels, e.g. integration of metamodels and measure of semantical distance between them. Such operations are then used as a core of two innovative frameworks. The first framework allows to combine formal and informal approaches for both engineering and interoperability of information systems. The second framework uses metamodeling and reengineering in order to reuse know-how that has been established by researchers in interoperability of information systems in the Semantic Web.

1 Introduction

The role of metamodels in information system engineering has been studied for a long time. Finally, a consensus emerged (e.g., with UML and OMG) in which:

- metamodels form the core of metamodeling architectures by implementing two abstraction mechanisms: abstraction by projection (i.e., separation and combination of concerns [2, 3, 18, 26]), and abstraction by conceptualization (e.g., the Model Driven Architecture approach emphasizes abstraction by conceptualization by using metamodels in a manner analogous to the use of models in the classical modeling).

- metamodels define languages (or dialects [21]) for model descriptions. Recent examples of such languages can be found in UML extensions for synchronization [11] and Architecture Description Languages [25].
• metamodels describe semantics of application domains: metamodels serve as domain ontologies.

UML-based metamodeling architectures use the UML metamodel to support two abstraction mechanisms. First, the UML metamodel is used to define a modeling language which satisfies specific needs of a given application domain. Second, the UML metamodel is used to define diagrams necessary for description of an application belonging to a given domain.

In such metamodeling architectures, information system engineering has been conceived as a two-step process: first, modelers choose a metamodel; and second, modelers instantiate their metamodel into a model. In practice, directly searching for a convenient metamodel could be really difficult. It would be necessary to know all available extensions of a candidate metamodel (in order to decide whether an existing metamodel can be used or whether a new metamodel needs to be defined). It would also be necessary to fully understand the manner in which existing metamodels have been built (in order to be able to extend them properly). One possibility to avoid this difficulty is to reduce metamodel proliferation by providing modelers with standardized UML’s metamodel extensions for specific application domains: this is achieved by OMG’s profiles [23, 22]. Another possibility is to organize metamodels into a hierarchy [5, 7]: every extension of a metamodel is an heir of the metamodel being extended.

When studying modelers’ practice, we notice that modelers do not work directly with metamodel descriptions: information is provided to them in the form of “semi-formal” descriptions (which can be ambiguous but tend to be more readable). We call such a description a modeling paradigm. The modeling process— in practice— is a three-step process:

• defining a modeling paradigm (by identifying a modeling paradigm which is reasonably close to the needed paradigm, and then building a variant of such a modeling paradigm),

• instantiating the chosen modeling paradigm into a metamodel,

• instantiating the obtained metamodel into a model.

For example, Price’s proposal [24] of a UML extension for time models is defined by reference to Kakoudakis’s time models [12], and Robbins’s proposal for Architecture Description Languages is defined by reference to Medvidovic & al.’s proposal [15, 16].

Due to the above, metamodels and models do not provide comprehensive information about the actual modeling process: all the initial work (in defining modeling paradigms) is lost. Our objective is to integrate such information into a metamodeling architecture which could then evolve into a comprehensive framework for information system engineering.

In the rest of this paper, we first provide an overview of our metamodeling architecture which encompasses both metamodels and modeling paradigms (Section
2). We then present two example frameworks based upon our metamodeling architecture: a framework for engineering and interoperability of information systems (Section 3), and a framework for the Semantic Web (Section 4). The conclusion (Section 5) discusses our ongoing work.

2 Overview of our metamodeling architecture

In order to help modelers to obtain the maximum benefits from metamodels, we introduce modeling paradigms into our metamodeling architecture. These modeling paradigms describe—in terms of concepts that are interrelated by constraints—the semantics that modelers assign to the real world. Our hypothesis is that there is nearly always, in practice, a one-to-one correspondence between modeling paradigms and metamodels. This hypothesis leads to the metamodeling architecture described in the following paragraphs and depicted in Figure 1.

2.1 A poset of modeling paradigms

Modeling paradigms are described informally yet they are organized in a poset\(^1\). Modeling paradigm descriptions possibly mix several different languages: the English language, logics, the set theory, the Z notation, etc. Modeling paradigms may use a various number of concepts, each of them being described with more or less precision. For example, a modeling paradigm may use a unique concept of \textit{class}, while another modeling paradigm may use different concepts such as \textit{interface}, \textit{abstract class}, and \textit{implementation class}. Modeling paradigms may have more or less precise constraints such as \textit{any object must belong to a class} or \textit{any object must belong to one and only one class}. We define a partial order for modeling paradigms by using a subsumption relation between them.

We denote by \textit{gmp} the generic modeling paradigm which corresponds to the standard UML semantics. We restrict ourselves to the set of modeling paradigms that are subsumed by \textit{gmp}: we denote this set by \textit{Restrict\textsubscript{MP}}. We denote by \(\leq\) the partial order among modeling paradigms.

In Figure 1 we depict some modeling paradigms of \textit{Restrict\textsubscript{MP}}: \textit{mp7} with a concept of spatiality is subsumed by \textit{mp5}. \textit{mp5} itself has a constraint of time model uniqueness and is subsumed by \textit{mp3} supporting several time models. There is no subsumption between \textit{mp3} and \textit{mp4} (\textit{mp4} supports the C2 Architecture Description Language).

2.2 A mirroring inheritance hierarchy of metamodels

Our objective is to build the metamodel layer of our architecture as a mirror of the poset of modeling paradigms: the generic modeling paradigm \textit{gmp} is instantiated

\(^1\)We use Gratzer’s definition of posets [8]: posets (also called partially ordered sets) are “sets equipped with partial ordering relations”.

3
into the UML metamodel itself (which we denote by $mm_{UML}$), and all other modeling paradigms are then instantiated into specializations of the UML metamodel (by using UML’s extension mechanisms: constraints, tag-values, and stereotypes). Furthermore, we require that each metamodel instantiates some modeling paradigm, and that each inheritance link between metamodels instantiates a subsumption link between modeling paradigms.

For example, in Figure 1 the general modeling paradigm $gmp$ is instantiated into the UML metamodel $mm_{UML}$ while $mp_4$ is instantiated into a metamodel $mm_4$. Large grey arrows represent instantiations of modeling paradigms into metamodels.

2.3 An example using ADL (Architecture Description Language)

Medvidovic & al. [15] describe the C2-style Architecture Description Language in English2: “connectors transmit messages between components, while components maintain state, perform operations, and exchange messages with other components via two interfaces (named "top" and "bottom"). ... Inter-component messages are either requests for a component to perform an operation or notifications that a given component has performed an operation or changed state”. Let us call by $mp_4$ the modeling paradigm described above. As depicted in Figure 1, the description of $mp_4$ includes concepts (e.g., connector, component, interface, message) and constraints (e.g., “components may not directly exchange messages; they may only do so via connectors”).

Robbins & al. [25] propose an extension of UML’s metamodel for C2-ADL. Their extension is an instantiation of the Medvidovic’s modeling paradigm in terms of a metamodel which we denote by $mm_4$. They define \texttt{C2 - interface} as a stereotype of the UML interface with a tagged value (top, bottom). \texttt{C2 - operation} (which can be either a request for operation or a notification message) is defined as a stereotype of the UML operation with a constraint forbidding any return value. The tagged values request, notification are used to distinguish requests from notifications. The set of elementary UML constructs of $mm_4$ is depicted in Figure 1. The set of UML constraints of $mm_4$ contains instantiations of constraints of modeling paradigm $mp_4$.

2.4 Semantical integration of metamodels

Our objective is to define a metamodel which can achieve a good semantical approximation of concepts shared between two given metamodels. The basis of our strategy is to use our mirroring structure for building a convenient common semantical ancestor from the two given metamodels, i.e., a common ancestor which is semantically close to these metamodels. Different configurations (i.e., relative positions of two metamodels to be integrated in the inheritance hierarchy) can be

\footnote{A formal description of C2 in language Z is given in [16].}
encountered. This leads to more or less complex solutions: relaxation (directly choosing a common ancestor in the inheritance hierarchy), formal integration (of consistent ancestors), instantiation (of a semantically close modeling paradigm chosen from the poset of modeling paradigms). See [28] for details. Due to our construction process, we can guarantee nominal quality of the proposed semantically integrated metamodel, as well as nominal quality of our instantiation function. However, semantical quality of such an integrated metamodel is not yet guaranteed. We believe that semantical quality of semantically integrated metamodels depends on the quality of our mirroring structure (and, ultimately, on the quality of the poset of modeling paradigms). We propose to delegate the responsibilities of ensuring the semantical quality of the poset of modeling paradigms to domain experts. In addition to the experts’ participation in the overall quality of the poset of modeling paradigms, we intend to measure semantical distance between metamodels in order to evaluate quality of semantically integrated metamodels.

2.5 Semantical distance between modeling paradigms

In order to be able to evaluate semantical quality, we need to measure semantical distances between metamodels. We define a measure of semantic distance

Figure 1: Our mirroring structure of meta-metamodel and metamodel layers
as a weighted sum of elementary distances between corresponding elements of metamodels (i.e., corresponding constructs and corresponding constraints). The principal question in such an approach is how to identify corresponding pairs of elements. In the rest of this section, we outline a generic method for determining pairs of such elements. Elementary distances between corresponding elements, as well as weights used for combining these elementary distances, are not discussed in the general case: they need to be fine-tuned in the context of a specific application domain. We propose to delegate the responsibilities of defining these elementary distances and weights to domain experts. For a formal presentation of our method please refer to [27].

Let us consider a simplified version of Robbins & al.’s metamodel (metamodel \(mm_4\) in Figure 1). Metamodel \(mm_4\) is a specialization of \(mm_{UML}\), and \(mp_4\) is subsumed by \(gmp\). For evaluation of semantical distance between \(mm_4\) and \(mm_{UML}\) we introduce the following sets:

- At the meta-metamodel level (\(mp_4\) and \(gmp\)), the set of common concepts:
  \[Com(mp_4, gmp) = \mathcal{E}l^3(gmp)\]
- At the meta-metamodel level (\(mp_4\) and \(gmp\)), the set of specialized concepts:
  \[Spe(mp_4, gmp) = \{interface, message\}\]
- At the meta-metamodel level (\(mp_4\) and \(gmp\)), the set of new concepts:
  \[New(mp_4, gmp) = \{connector, component\}\]
- At the metamodel level (\(mm_4\) and \(mm_{UML}\)), the pairs of initial UML constructs and specialized constructs:
  \[Spe_{UML}(mm_4, mm_{UML}) = \{(\llangle C2 - operation \rrangle, operation), (\llangle C2 - interface \rrangle, interface), (\llangle C2 - component \rrangle, class), (\llangle C2 - connector \rrangle, class)\}\]

We note that two new concepts of \(mp_4\) are instantiated as specializations of the UML class: their instantiations belong to \(Spe_{UML}(mm_4, mm_{UML})\). Thus, \(New_{UML}(mm_4, mm_{UML}) = \emptyset\).

An example evaluation of the mismatch between UML constructs and the above stereotypes is given in the following paragraph (see Figure 2). For such an evaluation, we use a measure of elementary distances with only three possible values (low, middle and high) from which a more precise measure can be derived.

- Specialized constructs belonging to \(Spe_{UML}\) introduce:
Table 1: Stereotypes and Evaluation Levels

<table>
<thead>
<tr>
<th>Stereotype</th>
<th>UML construct</th>
<th>Extension mechanism</th>
<th>$S_{\text{speUML}}$</th>
<th>$A_{\text{ddUML}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2-operation</td>
<td>operation</td>
<td>tagged value and constraint</td>
<td>low</td>
<td>low middle high</td>
</tr>
<tr>
<td>C2-interface</td>
<td>interface</td>
<td>tagged value and constraint</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>C2-component</td>
<td>class</td>
<td>constraint</td>
<td>•</td>
<td>• •</td>
</tr>
<tr>
<td>C2-connector</td>
<td>class</td>
<td>constraint</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

Figure 2: The first step in measurement of mismatch between the UML metamodel and Robbins & al.’s metamodel

- A stereotype $\langle C2 - \text{operation} \rangle$ of UML operations with a tagged value (request, notification) and a constraint $\langle C2 - \text{operation} \rangle$ has no return value. The tagged value is evaluated as low and the constraint is evaluated as middle.
- A stereotype $\langle C2 - \text{interface} \rangle$ of UML interfaces with a tagged value (top, bottom) and a constraint $\langle C2 - \text{interface} \rangle$ operations must have stereotype $\langle C2 - \text{operation} \rangle$. Both the tagged value and the constraint are evaluated as low. The constraint does not belong to $S_{\text{speUML}}$ (since it is a constraint between two stereotypes) but rather to the set of additional UML constraints $A_{\text{ddUML}}$.
- A stereotype $\langle C2 - \text{component} \rangle$ of class with constraints:
  - $\langle C2 - \text{component} \rangle$ do not contain features: evaluated as high,
  - $\langle C2 - \text{component} \rangle$ implements exactly two interfaces (with stereotype $\langle C2 - \text{interface} \rangle$) top and bottom: evaluated as middle and belonging to $A_{\text{ddUML}}$,
  - requests travel upward only and notifications travel downward only: both evaluated as low,
  - each $\langle C2 - \text{component} \rangle$ has exactly one instance in the running system: evaluated as high.
- A stereotype $\langle C2 - \text{connector} \rangle$ of class with a constraint: $\langle C2 - \text{connectors} \rangle$ are limited to binary associations: evaluated as low.

3 A Framework for Engineering and Interoperability of Information Systems

The increasing interest in UML made it possible to develop a large number of various modeling environments, both in the academia and in the industry. These environments provide modelers with numerous tools (for design, validation, and implementation of information systems). The evolution of many Analysis & Design environments seems to inherently proceed towards the integration of formal methods with programming tools. Analogously, building common knowledge (called

---

Footnote: In order to make our example sufficiently short, we do not use all the constraints given in [25].
a basis of agreement) that is shared by interoperable systems can be carried out—by formal operations— at the metamodel level and then instantiated at the model and instance levels. We therefore seek to provide modelers with integrated environments for engineering of interoperable information systems. Such an integrated environment can be used as outlined in Figure 3 (where processes are depicted by numbered dark-grey thick arrows). A core component is used for elaboration of modeling paradigms, metamodels, models and their documentation. A specification toolbox encompasses various tools for formal operations on models. By using such tools, modelers can produce a reliable specification of an information system from the description constructed in the core component (Arrow 1). A programming toolbox is then used for generation and testing of executable code based upon the specification of an information system (Arrow 2). A formal interoperation toolbox produces an abstract basis of agreement—in terms of a metamodel—from a set of specifications of interoperating information systems (Arrow 3). An agreement toolbox is used for generation of actual bases of agreement from abstract bases of agreement (Arrow 4). In the following paragraphs, we describe these components in more details (except for the formal interoperation toolbox which is discussed in Section 2).

The core component is dedicated to elaboration of metamodels, models and their documentation. This core component is supported by our mirroring structure. It allows modelers to choose a convenient initial modeling paradigm or to define their own modeling paradigm when necessary, to instantiate it into their metamodel, and finally to create their own model. For example, Clark & al.’s [4] have proposed an environment for metamodel construction. Such an environment includes formal tools for model validation.
The **programming toolbox** is used for generation and testing of executable code. Most of the commercial products for UML-based modeling include toolboxes that provide modelers with tools for generation of code (e.g., C++ or Java) from a model of the system. In many cases, code generation is carried out from an unvalidated model: an error in the model is thus propagated into the code. By producing code from a complete and consistent specification, and by using tests produced by our specification toolbox, it is possible to guarantee the quality of the resulting code.

For example, Norstrom & al. [20] have proposed a UML-based environment, called Model Integrated Computing (MIC), for generation of programs. MIC encompasses a partial validation of the model. They have further extended MIC into a Generic Modeling Environment [14].

The **formal interoperoperation toolbox** produces an abstract basis of agreement from a set of specifications of interoperating information systems. Such an abstract basis of agreement is expressed in terms of metamodels. The formal interoperoperation toolbox is based upon our semantical integration of metamodels: abstract bases of agreement are produced by semantical integration of the metamodels of information systems involved. For a given group of information systems, several abstract bases of agreement can be proposed (depending on the strategy used for building instantiated metamodels). By using our semantical distance between metamodels, we can help users to evaluate the proposed bases of agreement and to choose the most suitable one.

The role of domain experts is particularly relevant for this toolbox: they have to validate the part of the poset of modeling paradigms that describes their application domain and they have to determine the needed parameters (weights and elementary distances) for the semantical distance measure.

The **agreement toolbox** is used for generation of actual bases of agreement from abstract bases of agreement. We use abstract bases of agreement in the same way Gruber [9] uses ontologies of representation (i.e., ontologies which capture representation primitives but do not define their content). Defining the content of an actual basis of agreement is a process similar to the instantiation of a metamodel into a model. The main difference is that the information to be modeled is—in general—much more difficult to identify.

### 4 A Framework for the Semantic Web

With the worldwide development of web-enabled information systems, the need for efficient data access integrators (such as portals) and comprehensive interoperating services is urgent. The idea of a Semantic Web, as “a web talking to machines” has been widely accepted. According to Euzenat [6], manipulating only unstructured information has become an obsolete option: information models are needed (e.g., RDF, UML, XML) in order to balance “formalized knowledge” against “actual
informal contents”. In any case, even by using the pre-existing information models, it appears difficult to build the Semantic Web from scratch without reusing the know-how that has been established and refined for interoperability [17].

We propose a framework for development of the Semantic Web. Since web sites used in searches are not necessarily described at a conceptual level (and even when they are described at this level, it may be done so using different models), we propose to systematically build structural descriptions of web-enabled systems by using reengineering techniques. The main advantage of such a systematic reengineering is the possibility to provide homogeneous descriptions of web-enabled information systems: any web-system can be described using the same information model (i.e. UML) and the same description strategy [29]. Our metamodeling architecture can be used as a basis for a new approach to the Semantic Web, in which: 1) each domain is described top-down by using a domain library which contains both metamodels and models, 2) web data-model of the involved information systems are produced bottom-up by using systematic reengineering under the control of a domain description, 3) narrowly focused domain descriptions can be constructed separately and then integrated for cross-domain applications.

The framework we propose encompasses the following processes which are depicted by numbered arrows in Figure 4:

- The attachment process (arrows 1, 2) establishes a dependence between a web-enabled information system and a metamodel which belongs to a library of domain descriptions. Requiring existence of such a referring metamodel for any web-enabled system may appear rather restrictive. Yet due to the
development of UML as a standard tool for high-level system descriptions\(^4\) this requirement is reasonable (or will soon become reasonable). Domain libraries also cover the model and instance levels of descriptions. They are organized as UML packages. Each package is a narrowly targeted description of a specific domain. These packages can be mutually linked in order to produce cross-domain descriptions. The aim of our ongoing work is to precisely define the attachment component of our framework.

- The \textit{reengineering process} (arrows 3, 4) is a reversal of Koch’s hypermedia UML-based modeling methods [10, 13]. This reengineering process collects a representative sample of navigation paths and expresses them as a web data-model of the corresponding information system (in terms of well-balanced [19] Class and Collaboration Diagrams):
  
  - First, we navigate the web site in order to extract a graph of inter-object links. Such a graph is used as a representative sample of the whole navigation space. We propose this graph as a Collaboration Diagram for the model to be built.
  
  - Second, by using our domain integrator, we produce a “minimal” Class Diagram for the model being built.
  
  - Third, we measure the adequacy of the Collaboration Diagram in combination with the Class Diagram. As long as its adequacy is insufficient, we keep modifying the Class Diagram and repeating the above steps.

- The \textit{integration process} (arrow 5) produces an integrated metamodel from which the cross-domain description can be built. The integration component of our framework has been presented in Section 2.

\section{Conclusion}

Our metamodeling architecture takes advantage of knowledge of modelers’ behaviors, abstract approaches to information system engineering, and formal methods. Modeler’s behavior is represented by a poset structure of modeling paradigms at the meta-metamodel level. Abstraction is implemented by metamodels which form the core of our architecture. Cohesion of the two uppermost layers of our architecture is guaranteed by our mirroring structure of the poset of modeling paradigms and the inheritance hierarchy of metamodels. Therefore, we may “induce” properties of modeling paradigms from properties that have been formally defined on metamodels (e.g., ambiguity, consistency).

\(^4\)See, for example, OMG’s profiles [23], Cook’s prefaces [5], and UML-based ontologies [1].
We further define formal operations on metamodels. We also extend these “purely” formal operations into semantical ones: semantical integration of metamodels and a measure of semantical distance between metamodels. These semantical operations form the basis of two frameworks which apply metamodeling to information system interoperability and the Semantic Web. We continue our work on several technical questions related to this project:

- Providing modelers with a convenient method for using our mirroring structure in the form of a library of domain descriptions (i.e., for describing modeling paradigms and instantiating corresponding metamodels).

- Providing domain experts with a method for adapting our generic measure of semantical distance between metamodels to the experts’ application domains.

- Defining generic rules for quality measurement of the sub-poset of modeling paradigms corresponding to closely related application domains.

- Defining what is the required functionality of agreement toolboxes (for generation of actual bases of agreement from abstract bases of agreement) in our framework for integrated engineering and interoperability of information systems.

- Defining an attachment process of our framework for the Semantic Web. The main anticipated problem is how to determine limitations and pre-requirements of such a process.

Finally, we believe that the extensive use of our proposal would open a “political issue”, namely the need for a new organization of domain modeling. Modelers would be responsible for “local semantics” (i.e., for describing their own application domain as a variation of an existing domain description). Domain experts would be responsible for the global semantics (i.e., for validating semantical dependencies between domain descriptions).

References


14

