A Three Layer Meta Model for Specifying Business Domain Semantics with XML

Wolfgang Schuetzelhofer
IBM Austria
Obere Donaustrasse 95
A-1020 Vienna, Austria
wschuet@attglobal.net

Karl M. Goeschka
Vienna University of Technology
Gusshausstrasse 27-29/384
A-1040 Vienna, Austria
goeschka@acm.org

ABSTRACT

As XML is gathering more and more importance in the field of data interchange in distributed business to business (B2B) applications, it is increasingly important to provide a formal definition of XML-structures together with a well defined way to map business domain semantics to these structures. An XML-algebra, similar to the relational algebra, is required for the formal definition of operations and transformations and to prove the correctness and completeness of design methods. To develop an XML-algebra, we propose a sound mathematical foundation, modeling XML-structures as typed directed graphs based on set theory. Together with a formal method to apply domain semantics to directed graphs we present a three layer meta model to address the separation of structure and content, and we introduce extensible type hierarchies on nodes and links. This allows to model and validate business domain semantics on different levels of abstraction.

Keywords

XML, business domain semantics, set theory, formal software engineering, graph theory

1. PROBLEM STATEMENT AND GOAL

The importance of modeling business domain semantics, thus focusing on the user's perception of data, in contrast to physical data representation, has been highly recognized for many years. The field of research on semantic data models has grown rapidly over the last years [2].

With XML as a standardized and broadly used data description language, this area is even gathering more importance. As more and more XML schemas are introduced for different domains, it is increasingly important to provide a formal definition of XML-structures together with a well defined way of mapping domain semantics to these structures. As standards for XML-based B2B infrastructures (ebXML, UDDI) are evolving rapidly, new requirements for a sound mathematical foundation arise.

To ensure robustness, XML-data which serve as input for business applications, have to be proved for validity with respect to schemas. Algebraic specifications, based on algebraic equations, are a formal means to define validity.

XML-based data exchange between different systems and querying of XML-data require the formal definition of operations, mappings and transformations of XML-structures, which have to be proved for equivalence and soundness. Transformations are also needed as building blocks for design methods in order to develop XML-based data models for different business domains. On one hand this requires a formal method to map business domain semantics to XML-structures.

On the other hand it requires an appropriate algebra, similar to the relational algebra for RDBMS, in order to allow formal specification and verification to formally prove the correctness and completeness of single transformations and the whole design method respectively. Therefore, as a first step, a sound mathematical foundation has to be build.

2. KEY CONCEPTS

2.1 Modeling XML-structures as directed graphs

XML models documents and data structures as a composition of elements. The basic structure of an XML-document is given by the hierarchically nested elements, thus it is a natural approach to model an XML-document as a tree-graph. This is the approach of the World Wide Web Consortium (W3C) in its specification of the Document Object Model called DOM.

Several linking mechanisms such as ID/IDREF or the XML Linking Language extend the basic tree structure of XML-documents to express more complex relationships between XML-elements. In order to provide a seamless integration of different linking mechanisms, our concept is to model XML-structures as directed graphs rather than as tree-graphs.

We use the term *link* to abstractly express a connection between two nodes of a directed graph, that is a path through the graph (in contrast to other publications, where *links* are used as synonyms for *edges* only). Moreover, we allow to recursively compose a link out of other links, which makes our approach of link composition more powerful and serves as an important means of abstraction.

2.2 Types as a means to map business domain concepts

Applying business domain concepts to XML-structures is a key issue of this work. Hypertext systems, which in fact are the origin of XML, may profitably be viewed as semantic nets [1]. A semantic net is a directed graph in which concepts are represented as nodes and relations between concepts are represented as links. Using this approach, we introduce types on nodes and links of a directed graph and we appropriately map business domain concepts and relationships between business domain concepts to node and link types.

2.3 Type hierarchies as a means of abstraction

The process of modeling business domains requires the possibility to view and describe domain models at different levels of abstraction. To provide this ability, we introduce hierarchies on node types and link types.

2.4 A three-layer meta model

The consequent separation of type and instance, with the goal to separate structure and content, leads to a three-layer meta model approach: Firstly, we differentiate between the *type-level* and the *instance-level*. Now in order to define the type hierarchies, node types and link types themselves are typed nodes of a typed graph at the type-level. Consequently we introduce the *meta-type-level*, so that node types and link types, as typed nodes of the *type-level*, are *instances of meta-types*.

The three layer meta model of a typed graph, including a formal method to map business domains, in terms of domain concepts and business knowledge (-rules), is called **D**omain **G**raph **M**odel (**DGM**), see [5].

2.5 Expressing business domain knowledge through structural constraints

Business domain concepts and relationships between business domain concepts are mapped to node types and link types. Our concept to express business domain knowledge and rules is to define structural constraints on these types (e.g. link cardinalities, restrictions on link sequences).

We say that a graph type defines the structure of a business domain, and consequently structural constraints have to be specified on types and hence to be defined at the *type-level*. Our approach is to model structural constraints as nodes and links of a directed graph at the *type-level*.

3. RELATED WORK

A lot of similar approaches stresses the need for our work, the most important papers: The formal hypertext model described in Tochtermann and Dittrich [3] provides some formally defined structural concepts, lacking mechanisms to define more powerful structural and relational constraints. Wang and Rada [1] have developed a semantic data model based on the concept of a semantic net, introducing organizational and relational link types on a DAG, but in contrast to our work do not provide an extensible type system. Abiteboul and Hull [4] in their approach on formalizing a semantic model recognize the importance of types which are used to model object structures, but a concept of link type hierarchies, as elaborated in our paper, is missing.

The W3C's working activity on the XML Schema Language [6] introduces a formal XML-based language to define structural constraints on classes of XML-documents in order to apply domain semantics to these documents. This approach has achieved a high acceptance, as it is targeted towards practical implementations, but the specification does not provide a set theory based mathematical formalism.

4. CONTRIBUTION AND FUTURE WORK

The results of our work, seen to contribute to the field of formal software engineering are:

- A three-layer meta model, which strictly separates structure and content and which has proven to be a powerful means of abstraction to model complex business domains.
- A hierarchically extensible type system, including the definition of link type hierarchies.
- A formal method to map business domain concepts to XML structures.
- The formal definition of validity, which allows the development of algebraic specifications for business domains.
- XML has been defined as a specific business domain, which allows arbitrary business domains to be expressed as XML-structure.
- A sound mathematical foundation, using set- and graph theory, considered to serve as a basis for algebraic specifications and for an XML-algebra.

The set theory based DGM, together with the formal specification of validity, build a sound mathematical basis to develop an XML-algebra allowing to define operations on XML-structures as future work, such as inserting, deleting or updating nodes and links. This mathematical basis also provides for formal specifications and proofs of correctness, consistency, and completeness of design methods by proving the equivalence and soundness of transformations. New and extended design methods, which can be formally specified and verified, are seen to be a profitable output of future work based on this paper.

5. REFERENCES

1. W. Wang, R. Rada, “Structured Hypertext with Domain Semantics”; ACM TIS, 16,4, 1998, pp. 372-412

2. J.L. Schnase, J.J. Leggett, D.L. Hicks, R.L. Szabo, “Semantic data modeling of hypermedia associations”, ACM Trans. Inf. Syst. 11, 1, Jan. 1993, pp. 27-50

3. K. Tochtermann, G. Dittrich, “Towards a family of formal models for hypermerdia”, HIM’95 Proceedings, 1995, pp. 77–91

4. S. Abitoul, R. Hull, “IFO: a formal semantic database model”, ACM Trans. Database Syst. 12, 4, 1987

5. Schützelhofer, W.; Göschka, K.: “A Set Theory Based Approach on Applying Domain Semantics to XML-Structures.” Proceedings of HICSS-35, Jan 7-10, IEEE 2002, p.120

6. W3C XML Schema, *XML Schema Specification*, http://www.w3c.org/XML/schema, 2001