Chapter VIII

XML Semantics

Yasser Kotb
Ain Shams University, Egypt

Katsuhiko Gondow
Tokyo Institute of Technology, Japan

Takuya Katayama
Japan Advanced Institute of Science and Technology, Japan

ABSTRACT

In this chapter, we consider the problem of checking the consistency of the semantics associated with extensible markup language (XML) documents. We propose a novel technique to add semantics to XML documents by attaching semantic information to the XML element tag attributes. We call this technique XML semantics. It is based on the same concept as attribute grammars (AGs), attaching and checking static semantics of programming languages through their attributes. Furthermore, we show how the attribute dependencies in this approach can be expressed in the SLXS language. The SLXS language is a new description language based on XML; we have designed it to describe the semantic dependencies of XML documents. By this approach we remain compliant with the XML core technology. Finally, we give a practical example to illustrate the power of our approach: we have successfully applied this approach to check the semantic consistency of the several holy books that are mentioned on the Religion 2 Website.

INTRODUCTION

The Extensible Markup Language, abbreviated XML (Bray, Paoli & Sperberg-McQueen, 2000), describes a class of data objects called XML documents and partially describes the behaviors of computer programs that process them. It is a useful way of describing declarative, structured documents. XML is a meta-language for describing
markup languages. In other words, XML provides a facility to define tags and the structural relationships between them. However, XML lacks semantics in its construction (Psaila & Crespi-Reghizzi, 1999). XML specifies neither semantics nor a tag set. Since there is no predefined tag set, there cannot be any preconceived semantics. Meanwhile, document type definitions (DTDs) and other XML schemas (Thompson et al., 2001), which try to make XML documents more reliable and consistent, still lack the essential ability to describe semantics in XML documents.

Declarative structures, like XML, separate their semantic and syntax definitions, and have their own local descriptions that result in high readability and high maintainability. Therefore, the aim of this chapter is to propose a novel technique to add semantics to XML documents and to show how we can check the consistency of these semantics. This way is intended to use attribute grammars (AGs) (Knuth, 1968) in the area of Web content mining. By adding semantic information to XML attributes it is possible to describe the page’s semantic dependencies, evaluate and check page consistency, and improve the automatic understanding of the page content. We focus primarily, but not only, on the design of a proper method to add semantics to XML documents by associating semantics with the element tag attributes. By extracting such a semantic description, we are able to notify document writers of semantic errors in XML documents. This method takes the advantage of attribute grammars. We use the positive characteristics of AGs in the sense that they provide a clear description by the functional computation of attributes. Of course we can add some semantics to XML documents by writing Java or Perl programs, but these programs are likely to be very ad-hoc, in the sense that:

- They tend to be large, including non-essential details, which results in low readability and low maintainability.
- They are located outside the data schema (DTD); syntax and semantics are loosely coupled, which implies that semantic checking cannot be forced upon XML users.

We propose a novel technique to add semantics to XML documents by attaching the semantic information to the XML element tag attributes (Kotb, Gondow & Katayama, 2002a). We called this XML semantics. This approach is based on the AGs concept of attaching and checking the static semantics of programming languages through their attributes.

Furthermore, in order to specify the attribute element tag dependencies in the XML documents, and to show how to compute automatically their values as functions of other attributes, we propose an XML-based specification language called SLXS (Specification Language for XML Semantics) (Kotb, Gondow & Katayama, 2002b). Any SLXS specification document must follow the predefined syntactic structure of an SLXS DTD. The complete list of SLXS DTD rules will be given and discussed later. SLXS allows us to describe the semantic consistency in a functional, declarative, and local manner. It uses the element tags only, without using any attributes attached to the element tags, which guarantees more readable and maintainable XML specification documents. Mainly, SLXS is used to automatically generate the necessary code for the attribute evaluation procedures and/or the functions that check the consistency of the semantics associated with an XML document. This code is called the generated code.

As a practical example to illustrate the real power of our approach, we have successfully applied it to check the semantic consistency of the several holy books that
Web Page Design
www.igi-global.com/chapter/web-page-design/31188?camid=4v1a