

Maintaining XML Data Integrity in Programs - An Abstract Datatype Approach

Patrick Michel

Technische Universität Kaiserslautern

XML technology can be supported in high level programming languages in many different ways. APIs like DOM and SAX offer support for data in the generic XML format. Data binding approaches like JAXB and language extensions like XJ offer support for the combination of XML with XML schemata. Utilizing the type system of languages to guarantee validity of XML data is one way to improve language support.

There is almost no support, however, for the combination of XML with more powerful schema languages like Schematron. Such languages allow the definition of arbitrary boolean expressions on the shape and more important the content of XML data. Tools supporting Schematron are able to check all these integrity constraints for any given document, but offer no support for the correct manipulation of such documents. Schematron also does not define any method of manipulation, but allows to define valid sets of documents only.

We propose to view XML documents with such complex constraints as abstract datatype, with an interface of schema-specific procedures. Such an interface integrates well with object-oriented languages like Java and encapsulates the alien aspects of tree-shaped data. By automatically guarding these procedures with minimal preconditions, the lasting correctness and validity of an XML document can be guaranteed.

The procedures are implemented in a language accessible to Java programmers and domain experts, incorporating abstractions of the XML domain. It is powerful enough to define basic atomic manipulations that can leave valid documents in a valid state. An automated analysis is able to derive minimal preconditions that check if a procedure is applicable to a valid document. Failure of these preconditions can be dealt with by the Java programmer, for example, by changing parameters or calling other procedures.

Together with typical data binding approaches, which allow the type correct and convenient navigation and reading on documents, our approach offers comprehensive support for XML data with complex integrity constraints within languages like Java.