

XML Semantics

Ferenc Havasi and Miklós Kálmán

These days one of the most popular standards in use for storing structured information is XML. More and more applications are able to export data in an XML format, more databases are stored in XML, and XML processing techniques are becoming more generic. If this trend continues, XML will eventually be present in almost every part of the informatics sphere. Because of this the new research results related to XML should prove important the future.

The main idea behind our paper is based on a connection between XML documents and attribute grammars. The analogy makes it possible to apply techniques of attribute grammar (semantics rules) to the XML environment. The first notion of including semantics to XML was published in [1], but here we shall introduce a new approach.

We create a format (XML based) which makes it possible for us to define a real XML attribute via semantics rules. The new set of semantics rules then become an organic part of XML documents and do not violate the original XML specification.

This method consists of two major modules, the reduce and the complete module, each having a significant role in the completion and reduction of the designated XML file. The reduce module removes the specific attributes which can be calculated via the semantic rules stored in the SRML file. The complete module recreates the original XML file using the reduced XML and the semantic XML file (SRML). The rules specified will only be applied to the reduction phase when their original attribute values and the calculated values are equal. The SRML file format keeps the original DTD of the XML, since all attributes and nodes are mentioned which need to be IMPLIED.

The method was implemented using the JAVA language thus making it platform independent. It was successfully tested on various CPPML files, each varying in size and complexity. During these tests the number of attributes, thus the size was decreased considerably in case of reduction. The running time of the method is not significant, therefore a file size of 11MB can be reduced in a matter of minutes (approx. 2.30) achieving an average of 64%.

The future of the method includes the dynamic creation of the semantic file using Artificial Intelligence machine learning techniques. This will enable the clarification of the relationship between attributes towards the user, aside from reducing the size of the appointed XML file.