XPath, XSLT, XQuery: Formal Approach

Author: Pavel Hloušek

This thesis is devoted to formal properties of query languages for XML - XPath, XQuery and XSLT. Particularly, their formal semantics and expressivity are studied. It is important to remind that the development of these languages was done intensively during last 5 year by various research groups, the languages changed mainly in their semantics. For example, the semantics of XPath 1.0 and XPath 2.0 distinguish each other. One of the thesis's goal was also to deal with the semantics of all the languages one unified way.

After a necessary background in Chapter 2 summarizing various parts of XML technology related to the thesis, the author introduces in Chapter 3 a notation, or more precisely, a formal (denotational) system used throughout the thesis. The approach is inspired by Wadler's work to semantics of XPath 1.0.

In Chapter 4, the XPath language 2.0 is described. The author also proves one of the main results of the thesis: XPath 2.0 is capable of sorting sequences. The second result shows that quantified expressions are redundant in the language. As a sufficiently representative subset of the language its core is used for the proof. The core has the same expressive power as the full language. Chapter 5 is focused on the XQuery 1.0 language. Remind that XPath 2.0 is a proper subset of XQuery 1.0. The third result is proved here: every XQuery expression can be rewritten to an equal XQuery expression without an ORDER BY clause.

Chaper 6 concerns the XSLT 2.0 language. The author specifies and proves his own Core XSLT language (the minimality of this core is not considered) and describes its semantics. The contribution of this part is significant since (i) no formal semantics of XSLT has been proposed till now; (ii) the approach used here is unified w.r.t. the XPath and XQuery.

The thesis is written on a very high formal level. Moreover, it is readable since all fundamental steps are accompanied by motivation remarks and examples. The result about sorting has been submitted to the XIME-P 2005 workshop organized in cooperation with the prestige ACM SIGMOD/PODS Conference. The accepted paper was presented there and it is published in the workshop proceedings.

The author has proved that he is able to do creative scientific work. The thesis contains new results partially published at a reviewed international conference. I recommend the thesis to be accepted for a presentation aimed at delivering the PhD degree.

Prague 31. 1. 2006