# A Logic-Based Approach to XML Data Integration with "lazy materialization"

Wolfgang May

Institut fur Informatik,
Universitat Gottingen, Germany
may@informatik.uni-goettingen.de

XPathLog is a Datalog-like extension of XPath for querying, manipulating and integrating XML data. The querying part extends XPath with binding variables to XML nodes that are "traversed" when evaluating an XPath expression. In contrast to other approaches, the XPath syntax and semantics is also used for a declarative specification how the database should be updated: when used in rule heads, XPath filters are interpreted as specifications of elements and properties which should be added to the database. Special operations for data integration include additional cross links (subelement relationship and attributes) between fragments of the original database, declaring synonym relationships between notions from different sources, and XML element fusion.

As a data manipulation and integration language, XPathLog is originally based on a graph-based, edge-labeled model, called XTreeGraph. The XTreeGraph extends the basic XML data model by modeling multiple overlapping trees, and thus allows for restructuring existing XML trees into a densely connected graph database. XML result trees are then defined as XML tree views by projections from this database. The LoPiX implementation follows a "warehouse approach" where the integrated XTreeGraph is materialized.

In the present talk, a "lazy materialization" strategy is proposed that does not materialize the complete internal database, and that is based on the original XML tree model combined with XLink: Data items that are not (yet) changed are integrated into the internal database as references, represented by XLinks. The approach uses our recent proposal for a logical, transparent data model for XLinked data. Only if a referenced data item is actually modified in course of the integration process, it is (partially) loaded into the database; unchanged fragments of it still being represented in a lazy way by XLinks.