Tutorial Proposal

ACM Symposium on Applied Computing – SAC 2006 -

Title:

"SQL/XML: Closing the Gap between the Relational World and XML

Abstract:

Relational database technology has a long tradition in providing methods to persistently store and efficiently query large and highly structured data sets. The advent of XML as a state-of-the-art (meta-) format to exchange information between businesses however, challenges the relational world to provide means for extended database functionality in order to support an efficient handling of XML documents. Although, there is still a huge battle going on discussing the optimal way of implementing XML-structured data sets by either shredding documents into multiple relational tables or by providing native database support, the question of how to query these data seems to be solved. SQL/XML (http://www.sqlx.org) provides an extension of SQL – the core relational database language – to query and generate XML-style data sets. Within this tutorial, we will present the main ideas of the two worlds (relational and XML), compare the two concepts, discuss the proposed mapping on the structural level (data types, …) and show in great detail the integration of XQuery – the upcoming standard for querying XML data sets - functionality into the SQL language. Since the tutorial is addressing one of the most interesting issues in exploiting the de-facto-standard of modeling and processing flexible data structures in the context of the well-established relational context, it demonstrates an extremely high degree of relevance and addresses a broad spectrum of potential attendees.

The proposed tutorial consists of four major parts. Within the first part, the core concepts of XML and XML-related standards and technologies will be discussed in order to provide a common understanding for the auditorium. This introductory part will range from the requirements to the structure of XML documents via a presentation of XML-Schema related issues (type concept, constraints).
The second part of the tutorial compares the XML/XQuery data model with the relational data model and starts the in-depth discussion of SQL/XML from a structural point of view. Therefore, the main constructs of the XQuery data model like sequences, different node types and node properties are reviewed. Thereafter, this part comprises a detailed discussion of mapping data structures from the relational world into the XML world and vice versa.
The third part finally gives a broad overview of XQuery capabilities. Within this part, the main concepts of XQuery expressions are explained in great detail. A presentation of the different ways of generating XML elements is followed by a discussion of the main expressions of XQuery: path and FLWOR expressions. This third part of the tutorial

concludes with a short presentation of the capabilities in writing user-defined functions in XQuery.

The fourth part of the tutorial finally focuses on the SQL/XML approach from an operational point of view and exploits the methods learned in the preceding sections (XQuery functionality in the third part and the structural mapping explained in the second part). The SQL/XML description starts with the introduction of the new relational base type XML und proceeds with the description of the SQL/XML functions like XMLGEN, XMLELEMENT, etc.

The tutorial and the description of closing the gap between XML and relational database systems closes with a discussion of the current state of XML support in commercial systems. In this section the tutorial will briefly sketch the SQL/XML support level within the three big players on the market, i.e., Microsoft SQL Server 2005, Oracle 10g, IBM DB2 8.2.

In summary, the tutorial provides a deep understanding of the problems and the currently proposed solution SQL/XML to close the gap between the XML world and the classical relational row-based world. The mapping will be discussed from a structural (mapping data types) as well as from an operational (SQL/XML functions) point of view. Finally, the tutorial provides insight into the current state of SQL/XML support in major commercial systems.

Presenters:

Prof. Dr.-Ing. Wolfgang Lehner earned his master and PhD. degree from University of Erlangen-Nuernberg, in 1995 and 1998 respectively. After spending a year at the IBM Almaden Research Center, San Jose (CA), he was temporarily assigned to the professorship of databases at the University of Halle. Since 10/2002, Wolfgang Lehner is full professor and head of the database technology group at the Dresden University of Technology. From 8/2004 to 9/2004, he was visiting scientist at Microsoft Research in Redmond (WA).

His major research areas are data management issues related to complex structured data sets. Up to now, Wolfgang Lehner published four text books on multidimensional database systems (1999), subscription systems (2001), database technology for data warehouse systems (2003), and - most recently together with Harald Schöning - the XQuery database language (2004). Additionally, he published more than 70 reviewed research papers and is editor of journals and four other books. His major research interests are in the field of data analysis and querying of XML-based database systems.

Dr.-Ing. Harald Schöning has studied computer science at the University of Kaiserlautern, and has acquired a PhD in database systems. He worked in a research project that designed and implemented an advanced complex-object database. In 1993, he joined Software AG, starting in the development of Adabas, a high-end database system. When Software AG decided to start the XML database system Tamino, Harald was appointed architect for this system, and has followed standardization of XQuery from its beginning. Recently, Harald Schöning's working area was extended to cover information integration issues.

Harald Schöning has given several lectures at various German universities, and regularly teaches "XML and databases" for the DIA (Deutsche Informatik Akademie). He has written a book on "XML and databases", and another one on XQuery (together with Wolfgang Lehner).

Preliminary Content for a 3h Tutorial:

- XML core concepts (15min)
    - XML documents (wellformedness, validity)
    - XML schema (type concept, constraints, …)
- SQL/XML: Relationship of XQuery and the relational data model (45min)
    - Sequences and atomic types
    - node types and node properties
    - comparison with relational data model
    - mapping data types (namespace XMLSQL)
- XQuery expressions (60min)
    - element generation (computed elements, …)
    - path expressions
    - flwor expressions
    - user-defined functions in XQuery
- SQL/XML: Functional perspective (45min)
    - base data type: XML (storage, comparison, …)
    - SQL/XML functions (XMLGEN, XMLELEMENT, …)
- Current status of XML support in commercial systems (15min)
    - querying native XML database systems (Tamino system)
    - integration of the XQuery language within relational database systems
        - Microsoft SQL Server 2005
        - Oracle 10g
        - IBM DB2 8.2