XML Query Processing

Daniela Florescu
BEA Systems
danielaf@bea.com

Donald Kossmann
University of Heidelberg
kossmann@informatik.uni-heidelberg.de

XQuery is starting to gain significant traction as a language for querying and transforming XML data. It is used in a variety of different products. Examples to date include XML database systems, XML document repositories, XML data integation, workflow systems, and publish and subscribe systems. In addition, XPath of which XQuery is a superset is used in various products such as Web browsers.

Although the W3C XQuery specification has not yet attained Recommendation status, and the definition of the language has not entirely stabilized, a number of alternative proposals to implement and optimize XQuery have appeared both in industry and in the research community. Given the wide range of applications for which XQuery is applicable, a wide spectrum of alternative techniques have been proposed for XQuery processing. Some of these techniques are only useful for certain applications, other techniques are general-purpose.

The goal of this seminar is to give an overview of the existing approaches to process XQuery expressions and to give details of the most important techniques. The presenters have experience from designing and building an industrial-strength XQuery engine [1]. The seminar will give details of that XQuery engine, but the seminar will also give extensive coverage of other XQuery engines and of the state of the art in the research community. The agenda for the seminar is as follows: (1) *Introduction to XQuery*: Motivation, XQuery data model, XQuery type system, Basic query language concepts; (2) *Internal Representation of XML Data*: DOM, SAX Events, TokenStream, Skeleton, Vertical Partitioning; (3) *XQuery Algebras*: XQuery Core vs. Relational Algebra, XQuery Algebras from Research Projects; (4) *XPath Query Processing:* Transducers, Automata, etc.; (5) *XQuery Optimization:* XML query equivalence, Rewrite Rules, Cost Models; (6) *XQuery Runtime Systems*: Iterator Models, Algorithms for XQuery Operators; (7) *XML Indexes*: Value and path indexes, others; (8) *XQuery Products and Prototypes*: XQRL/BEA, Galax, Saxon, etc. (as available); (9) *Advanced Query Processing Techniques, Related Topics*: Querying compressed XML data, Multi-Query Optimization, Publish/Subscribe and XML Information Filter, XML Data Integration, XML Updates, XML integrity constraints; (10) *Summary*.

Speakers

Daniela Florescu is a Senior Software Engineer in BEA Systems. She received her MS in Mathematics in 1990 from University of Bucharest, and her PhD in computer Science in 1996 from University of Paris VI. After being a researcher in ATT Research Center (New Jersey) for 2 years and another 2 years in INRIA, France, in 2000 she decided to try the industry path. She was a lead architect in Crossgain (Washington) and lead scientist in Propel (California), before starting her own company, XQRL, recently acquired by BEA Systems. Daniela has extensive experience in query languages and query processing. She is one of the editors of the standard XML query language, XQuery. Together with Donald Kossmann (also a founder of XQRL, Inc), she designed and implemented the streaming Xquery engine that is currently being shipped by BEA Systems as part of the Web Logic Integration 8.1 [1].

Donald Kossmann is a Full Professor for Computer Science at the University of Heidelberg in Germany. He received his MS in 1991 from the University of Karlsruhe and completed his PhD in 1995 at the Technical University of Aachen (RWTH). After that, he spent 18 months at the University of Maryland (College Park) and at the IBM Almaden Research Center, California, USA. From mid 1996 until 2000, he worked as a research associate at the University of Passau where he received his habilitation in 1999. From 2000 until 2003, he was an Associate Professor at the Technical University of Munich. He is a co-founder of two start-ups: i-TV-T AG and XQRL, Inc. His current research is focussed on the performance of database and information systems and on platforms for Web services.

References

[1] D. Florescu et al. The BEA/XQRL Streaming XQuery Processor. In *Proc. of VLDB Conf.*, September 2003, Berlin, Germany.

Proceedings of the 20th International Conference on Data Engineering (ICDE'04)
1063-6382/04 $ 20.00 © 2004 IEEE