

XML Databases: Principles and Usage

Jaroslav Pokorný

Department of Software Engineering
Charles University, Faculty of Mathematics and Physics,
Malostranské nám. 25,
118 00 Praha,
pokorny@aksi.mff.cuni.cz

Abstract

Originally XML was used as a standard protocol for data exchange in computing. The evolution of information technology has opened up new situations in which XML can be used to author, maintain, and deliver content and consequently, new applications of XML occurred. XML serves as a data model and background for databases of XML documents as well as for applications beyond today's data models (hierarchical structures, recursive structures, regular expressions). XML also plays a significant role as a technological platform for Semantic web.

A motivational power for XML databases has roots in application demands like

- processing external data (Web pages, other text databases, structured data),
- E-commerce: lists of products, personalized views of these lists orders, invoices in e-commerce, e-brokering,
- integration of heterogeneous information sources (e.g. integrated processing data from Web pages and from tables of a relational database).

To store XML data in a database means to manage large numbers of XML documents in more effective way. Although this idea looks attractively there is also skepticism from the side of XML DB developers. For example, M. Kay (Software AG) says: I generally argue that XML is designed primarily for information interchange, and that the requirement for storage is secondary.

In the world of databases we can distinguish entities like a database model, database schema and query languages. During development XML technology, which includes also XML databases, many database-like approaches have been worked out. Some of them are standardized, e.g. remind standards: XPath 1.0 (W3C Recommendation, 1999), XPath 2.0 (W3C Recommendation, 2007), XQuery 1.0 (W3C Recommendation, 2007),

XSLT 1.0 (W3C Recommendation, 1999), XSLT 2.0 (W3C Recommendation., 2007), XML Schema 1.1 (W3C, 2007).

In practice, a use of these languages is generally too complicated. For example, real schemas in XML Schema only rarely use advanced constructs of the language. Most of them are structurally equivalent to a DTD specification.

Interesting questions appear with languages XQuery and XSLT. Although completely different, they have the same computational power. Choosing when use each one is not always easy. Integration of relational and XML data resulted in development of SQL/XML language (ISO/IEC 9075-14:2003). It allows relational data to be published in an XML form that can then be queried using XQuery.

Considering development of XML databases we can recognize two main directions:

- to map the documents into data structures of the existing DBs (XML-enabled DB),
- to develop a DBMS with a native XML storage (native XML DB, shortly NXD).

An implementation of native NXD is undoubtedly a new challenge both for developers and researches of database systems. Bourret registers 43 products in July 2007 [Bo07]. In database architectures, NXDs provide a nice example when a DBMS needs a separate engine [Po07]. A great debate concerns question when to use these approaches and even what is the purpose of XML native databases [Bo07b]. Concerning enterprises, content management provides a good example. According to a recent study by ZapThink, producers of content spend over 60% of their time locating, formatting, and structuring content and just 40% for creating the content. As XML separates formatting data from XML content, a new trend is to build content management system on the top of NXD.

The goal of the paper is to highlight both alternatives in XML databases development and focus on a discussion of practice of XML databases, mainly in enterprises.

References

- [Bo07a] Bourret, R.: XML Database Products, 2007.
<http://www.rpbouret.com/xml/XMLDatabaseProds.htm>.
- [Bo07b] Bourret, R.: Going native: Use cases for native XML databases, 2007
<http://www.rpbouret.com/xml/UseCases.htm>.
- [Po07] Pokorný, J.: Database Architectures: Current Trends and Their Relationships to Requirements of Practice. In: INFORMATION SYSTEMS DEVELOPMENT Series, Advances in Information Systems Development: New Methods and Practice for the Networked Society, Proceedings Information Systems Development Springer Verlag, 2007, pp. 269-279.