

The XOO7 Benchmark

Stéphane Bressan¹, Mong Li Lee¹, Ying Guang Li¹, Zoé Lacroix², Ullas Nambiar²

¹National University of Singapore
{steph, leeml, liyg}@comp.nus.edu.sg
²Arizona State University
{zoe.lacroix, mallu}@asu.edu

As XML becomes the standard for electronic data interchange, benchmarks are needed to provide the comparative performance analysis of XML management systems (XMLMS). Typically a benchmark should adhere to four criteria: relevance, portability, scalability and simplicity [1]. The data structure of a benchmark for XML must be complex enough to capture the characteristics of XML data representation. Data sets should be in various sizes. Benchmark queries should only be defined with the primitives of the language.

XML models present similarities with object-oriented data models. While XML is able to handle semi-structured data, it supports most of the features of complex object models. Classes, methods and inheritance are not defined in XML but classes can be expressed through element types and attributes. Thus in developing a benchmark for XML, we decided to use the well-established OO7 benchmark [3] designed for object-oriented database management system as a starting point. The XOO7 benchmark, an XML version of the OO7 benchmark, is a single-user based benchmark for XMLMS that focuses on the query processing aspect of XML.

The DTD and data set of XOO7 are directly obtained by mapping the OO7 schema and data set to XML [4]. OO7 does not model any specific application, but it intends to capture the characteristics of an object-oriented database. Additionally, in order to cater for the document centric view of XML, we extended the document object of OO7 to contain sub-elements mixed with text data. Thus, the *Document* element provides for a liberal use of free-form text that is "marked up" with elements. Therefore the XOO7 data set can capture all the characteristics of typical XML database applications. We provide a parameterized program to generate XML databases of various sizes and characteristics.

The XOO7 benchmark also extends and modifies the eight OO7 queries with a set of twenty-three queries. XOO7 provides relational, document and navigational queries that are specific and critical for XML database

applications. The queries test the primitive features and each query covers only a few features. XOO7 queries are defined to express the requirements published by the W3C XML Query Language working group. XOO7 queries are therefore supported by most XMLMS, which makes them very portable. Users can choose a subset of queries: data centric, document or navigational, to test on the features required in their applications [2, 5].

XOO7 has been implemented and used to evaluate two XML-enabled management systems: LORE, XENA and two native-XML management systems: Kweelt, DOM-XPath¹. The experiment results and the analysis can be found in [5].

The up-to-date information about the XOO7 benchmark can be found at the web site: <http://www.comp.nus.edu.sg/~ebh/XOO7.html>.

Reference

- [1] J. Gray. The Benchmark Handbook: For Database and Transaction Processing Systems, 2nd Edition, Morgan Kaufmann Publishers, Inc., 1993.
- [2] S. Bressan, M. L. Lee, Y. G. Li, Z. Lacroix, U. Nambiar. The XOO7 XML Management System Benchmark. NUS CS Dept Technical Report TR21/00, November 2001.
- [3] M. J. Carey, D. J. DeWitt, and J. F. Naughton. The OO7 benchmark. *ACM SIGMOD Int. Conf. On Management of Data*, pp. 12-21, Washington, 1993.
- [4] U. Nambiar, Z. Lacroix, S. Bressan, M. L. Lee, Y. G. Li. XML Benchmarks Put to the Test. Proceedings of the 3rd International Conference on Information Integration and Web-based Applications & Services, Linz, Austria, September 2001.
- [5] U. Nambiar, Z. Lacroix, S. Bressan, M. L. Lee, Y. G. Li. Efficient XML Data Management: An Analysis. Proceedings of the 3rd International Conference on Electronic Commerce and Web Technologies, Aix en Provence, France, September 2002.

¹ The name of the product is not published at the request of the developers