

The Michigan Benchmark: A Microbenchmark for XML Query Processing Systems*

Kanda Runapongsa

Jignesh M. Patel

H. V. Jagadish

Shurug Al-Khalifa

Department of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor, MI 48109, USA
{krunapon, jignesh, jag, shurug}@eecs.umich.edu

With the continuing increasing popularity of the eXtensible Markup Language (XML) as a representation format for a wide variety of data, and it is clear that large repositories of XML data sets will soon emerge. The effective management of XML in a database thus becomes a pressing issue. Several methods for managing XML databases have emerged, ranging from retrofitting commercial RDBMSs to building native XML database systems. There has naturally been an interest in benchmarking the performance of these systems, and a number of benchmarks have been proposed [?, ?, ?]. The focus of currently proposed benchmarks is to assess the performance of a given XML database in performing a variety of representative tasks. Such benchmarks are valuable to potential users of a database system in providing an indication of the performance that the user can expect on their specific application. The challenge is to devise benchmarks that are sufficiently representative of the requirements of “most” users. The TPC series of benchmarks accomplished this, with reasonable success, for relational database systems. However, no benchmark has been successful in the realm of ORDBMS and OODBMS which have extensibility and user defined functions that lead to great heterogeneity in the nature of their use. It is too soon to say whether any of the current XML benchmarks will be successful in this respect - we certainly hope that they will.

One aspect that current XML benchmarks do not focus on is the performance of the basic query evaluation operations such as selections, joins, and aggregations. A “microbenchmark” that highlights the performance of these basic operations can be very helpful to a database developer in understanding and evaluating alternatives for implementing these basic operations. A number of questions related to performance may need to be answered: What are the strengths and weaknesses of specific access methods? Which areas should the developer focus attention on? What is the basis to choose between two alternative implementations? Questions of this nature are central to well-engineered systems. Application-level benchmarks, by their nature, are unable to deal with these important issues in detail. For relational systems, the Wisconsin benchmark [?] provided the database community with an inval-

able engineering tool to assess the performance of individual operators and access methods. Inspired by the simplicity and the effectiveness of the Wisconsin benchmark for measuring and understanding the performance of relational DBMSs, we develop a comparable benchmarking tool for XML data management systems. The benchmark that we propose is called the Michigan benchmark.

In the Michigan benchmark, we primarily attempt to capture the rich variety of data structures and distribution possible in XML, and to isolate their effects, without imitating any particular application. The benchmark specifies a single data set against which carefully specified queries can be used to evaluate system performance for XML data with various characteristics. We have used the benchmark to test and analyze the performance of different XML database system implementations.

The most up-to-date information on the Michigan benchmark, and results using the benchmark, can be found at <http://www.eecs.umich.edu/db/mbench/>

References

- [1] T. Böhme and E. Rahm. XMach-1: A Benchmark for XML Data Management. In *Proceedings of German Database Conference BTW2001*, Oldenburg, Germany, March 2001.
- [2] D. J. DeWitt. The Wisconsin Benchmark: Past, Present, and Future. In J. Gray, editor, *The Benchmark Handbook for Database and Transaction Systems*. Morgan Kaufmann, second edition, 1993.
- [3] S. Bressan and G. Dobbie and Z. Lacroix and M. L. Lee and Y. G. Li and U. Nambiar and B. Wadhwa. XOO7: Applying OO7 Benchmark to XML Query Processing Tools. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, Atlanta, Georgia, November 2001.
- [4] A.R. Schmidt, F. Wass, M.L. Kersten, D. Florescu, I. Manolescu, M.J. Carey, and R. Busse. The XML Benchmark Project. Technical report, CWI, Amsterdam, The Netherlands, April 2001.

*This work was supported in part by the National Science Foundation under NSF grant IIS-0208852, by an IBM CAS Fellowship, and by a gift donation from IBM.