

# XMach-1: A Multi-User Benchmark for XML Data Management

**Erhard Rahm, Timo Böhme**

University of Leipzig, Germany  
<http://dbs.uni-leipzig.de>

The specification of XMach-1 (XML Data *Management* benchmark, Version 1) was developed at the University of Leipzig in 2000 and published at the beginning of 2001 [BöRa01]. It was the first XML database benchmark. The benchmark defines a database of XML documents and a set of operations covering important characteristics of XML processing and querying. Key features of XMach-1 are scalability, multi-user simulation and the evaluation of the entire data management system. It has been successfully implemented for a variety of native XML database systems and XML-enabled relational and object-relational DBMS.

The benchmark is based on a web application in order to model a typical use case of a XML data management system. The system architecture consists of four parts: the XML database, application servers, loaders to populate the database and browser clients. The application servers run a web (HTTP) server and other middleware components to support processing of the XML documents and to interact with the backend database. The XML database contains both document-centric and data-centric XML documents. The largest part is document-centric consisting of semi-structured documents with larger text portions such as books or essays. These documents are synthetically produced by a parameterizable generator. In order to achieve close-to-reality results when storing and querying text contents, text is generated from the 10,000 most frequent English words, using a distribution corresponding to natural language text. The documents varies in size (2-100 kB) as well as in structure (flat and deep element hierarchy). The second part of the database is a data-centric directory containing the metadata of the other documents such as document URL, name, insert- and update time. All data in this document is stored in attributes (no mixed content) and the order of element siblings is free. Compared to structured data in relational databases it shows some semi-structured properties such as variable path length using recursive elements or optional attributes.

The database can be scaled by increasing the number of documents, e.g. from 1000 text-oriented documents to 10 million documents. The metadata document scales proportionally with the number of text documents. A distinctive feature of XMach-1 is that documents may be schema-less or schema-based and that we increase the number schemas with the number of documents. This allows us to test a database system's ability to cope with an increasing number of different element types and to test query execution across multiple schemas. Additionally the benchmark supports evaluating schema-less and schema-based document storage.

The XMach-1 workload mix consists of 8 query operations and 3 update operations. They cover a wide range of processing features such as complete reconstruction of complex documents, full text retrieval, navigational queries, queries using sorting and grouping operators etc. In addition to the textual specification there exists a XQuery formulation of the queries [XM02]. Update operations cover inserting and deleting of documents as well as changing attribute values. Having update operations defined is unique across the XML benchmarks. Despite the missing data manipulation language for update operations it is insightful to test this functionality.

Since XMach-1 is a multi-user benchmark the primary performance metric is throughput measured in Xqps (XML queries per second). This value is calculated from the workload mix which defines firm ratios for each operation. According to its simulated domain this mix emphasizes the retrieval of complete documents whereas update operations have only a minor share. Nevertheless the latter one can have a significant impact on the execution of the query operations since cache and transaction management have to provide current data. Of course, the benchmark can also be applied in single-user mode, e.g. to determine the response times of the various queries as a reference point.

## References

- BöRa01 Böhme, T.; Rahm, E.: XMach-1: A Benchmark for XML Data Management. In Proceedings of German database conference BTW2001, pp. 264-273, Springer, Berlin, March 2001. Available under [XM02]
- XM02 XMach-1 web page. <http://dbs.uni-leipzig.de/en/projekte/XML/XmlBenchmarking.html>, 2002