

HyperStorM - Administering Structured Documents Using Object-Oriented Database Technology

Klemens Böhm, and Karl Aberer

GMD - Integrated Publication and Information Systems Institute (IPSI)

Dolivostrasse 15, D-64293 Darmstadt, Germany; {kboehm, aberer}@darmstadt.gmd.de

The objective of the project HyperStorM ('Hypermedia Document Storage and Modeling') is to use object-oriented database technology to administer structured documents like SGML- and HyTime-documents. It is an asset of formats such as SGML to allow for the seamless integration of meta-information. HyTime provides a set of architectural forms, i.e., templates with a pre-defined semantics to be used in hypermedia documents for scheduling or hyperlinking, to give examples.

Requirements. We have identified the following requirements in the context of structured document storage: the database application has to administer documents conformant to arbitrary document-type definitions (DTDs). With regard to declarative access, it must be possible to formulate queries in a more precise way, e.g., by referring to documents' structure, in order to cope with the increase of documents in number and size. Updates on documents are advantageous as, first, it may be pointless to only insert entire documents when considering certain document types, such as encyclopaediae or codes of law. Second, formats such as SGML are in use to capture structured information going beyond the conventional notion of 'document'. - Document components' semantics should be available within the database application to ensure adequate performance, and to allow for querying based on such concepts. Finally, we think that, with regard to access via the WWW documents' conversion to HTML should take place at the server site: transformation of documents might not be straightforward, but may instead be driven by the database context. Then, the process is more efficient when carried out in the database.

System Overview. Due to the dynamic nature of the system, most structures comprising the document content have to be generated at runtime. As a first

step, the document-type definition is transformed to an SGML document on the syntactic level. By inserting this document into the database, generating those structures becomes a database-internal process. - A hybrid database-internal representation for documents has been chosen: while documents' structure within the database may be made explicit to a certain degree, the structure of the remaining components is obtained by interpreting "flat" strings. Before inserting a DTD into the database, it is enriched with knowledge on the semantics of documents' and document components' types. Thus, the database application is configurable.

SGML-/HyTime-specific indexing mechanisms and materialized views on documents are part of the database schema. Index structures have been constructed so that they do not have to be updated after small document modifications. Queries are optimized using application-specific semantics. - The framework has been implemented on top of the OODBMS VODAK [VOD95].

Demonstration Outlook. With the demonstration, we illustrate how the database configuration can be specified in a natural, SGML-compliant way, how documents of arbitrary type can be accessed both by navigation and declarative access (via the WWW), and how database technology can be exploited to ensure a higher quality of document presentation.

References

- [ABH94] K. Aberer, K. Böhm, C. Hüser: "The Prospects of Publishing Using Advanced Database Concepts", in *Proc. of Conf. on Electronic Publishing*: 469-480, Wiley Ltd., April 1994.
- [BöA94] K. Böhm, K. Aberer: "Storing HyTime Documents in an Object-Oriented Database", in *Proc. 3rd Int. Conf. on Information and Knowledge Management*: 26-33, ACM Press, Nov. 1994.
- [BAN95] K. Böhm, K. Aberer, E.J. Neuhold: "Administering Structured Documents in Digital Libraries", in *Digital Libraries*: 91-117, Springer LNCS, 1995.
- [VOD95] *VODAK V 4.0 User Manual*, GMD Technical Report No. 910, 1995.

Permission to make digital/hard copy of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copying is by permission of ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee.

SIGMOD '96 6/96 Montreal, Canada
© 1996 ACM 0-89791-794-4/96/0006...\$3.50