

Database Technology on the Web

atabases are as ubiquitous as the Internet. The Web's widespread growth has dramatically increased the number of data sources available (data warehouses, scientific data banks, digital libraries, and so on), and it has increased the opportunities to harness these data sources in unanticipated ways.

The Web was originally conceived for use by physicists, but high connectivity (spurred by the liberalization of Internet use in the early 1990s), ease of use, and inexpensive access have made it the de facto medium for publishing and disseminating information. The Web now encompasses all types of data — totally unstructured, semistructured, and highly structured — relating to all aspects of economic, social, and political life.¹

The information age revolution has highlighted the role of the database management system (DBMS) as a key enabling technology. DBMSs are currently the technology of choice for modeling, storing, managing, and querying large amounts of information.

A Resilient Technology

Database technology has always been

challenged and advanced by new uses and applications, and it continues to evolve along with requirements and hardware advances. From the early days of file systems and hierarchical and network databases to relational, object-oriented, and special-purpose databases, the technology has shown resilience in responding to the needs of changing computing environments (mainframe, client-server, desktop, and so on). The innovative relational technology, for instance, used set theory to provide more abstraction and a way to reason about data for optimization. Objectoriented technology answered the need for more powerful modeling and management techniques for advanced applications. Object-modeling capabilities now enable DBMSs to manage complex multimedia objects - modeling not only data but also functions, procedures, and methods.2

The early Web (the period from 1992 to 1996) provided users access to text-based pages through hypertext links. Nowadays, the Web provides access to a variety of data that can be multimedia-rich. Readily available information retrieval techniques such as inverted indices, which allow efficient keyword-based access to text, large-

Elisa Bertino
University of Milano

Athman Bouguettaya *Virginia Tech*

1089-7801/02/\$17.00°2002 IEEE http://computer.org/internet/ JULY • AUGUST 2002 31

ly enabled access to the exponentially growing Web. As pressure from users mounted to allow access to richer types of information and to provide services beyond simple keyword-based search, the database research community responded with a two-pronged solution. First, by using databases to model Web pages, information could be extracted to dynamically build a schema against which users could submit SQL-like queries. By adopting XML for data representation, the second proposed solution centered on adding database constructs to HTML to provide richer, queriable data types.

Database Challenges for the Web

Today's DBMS technology faces yet another challenge as researchers attempt to make sense of the immense amount of heterogeneous, fast-evolving data available on the Web. The large number of cooperating databases greatly complicates autonomy and heterogeneity issues and requires a careful scalable approach. We need better models and tools for describing data semantics and specifying metadata. Techniques for automatic data and metadata extraction and classification (ontologies, for example) are crucial for building tomorrow's Semantic Web.³ Query languages and query processing should also be extended to exploit semantic information.

Users also need adaptive systems to help them explore the Web and discover interesting data sources and interfaces that support different query and search paradigms. Data dissemination techniques and notification services must be developed to enable effective data delivery services. Web-centric applications such as e-commerce and digital government applications pose stringent organizational, security, and performance requirements that far exceed what is now possible with traditional database techniques. Recent XML-native or extended DBMSs still need to be fine-tuned and evaluated. Finally, we need new methodologies to support the design and development of data-intensive Web sites.

In this Issue

The large number of submissions to this special issue is testimony to the topic's importance and timeliness. The following articles cover a broad range of issues in using database technology on the Web. In "Managing Web-Based Data: Database Models and Transformations," Atzeni, Mecca, and Merialdo model each Web site as a database. Their technique extracts data from Web sites and stores it in databases that can be used to generate new sites.

In "A Generic Content-Management Tool for Web Databases," Kerer, Kirda, and Kurmanowytsch

use relational database techniques to generate Web-based update interfaces. They use extended entity relationship (EER) definitions to generate XML-based interfaces to update Web-based relational databases.

Nambiar et al. discuss the salient features of XML management systems in "Current Approaches to XML Management." They classify the systems using their X007 benchmark and present a comparative performance evaluation of native XML systems to XML databases.

In "Managing Scientific Metadata Using XML," Yang, Kafatos, and Wang present their experience storing and accessing earth science data. Their Distributed Metadata Server system allows researchers to define metadata drawn from heterogeneous sources.

In "The Debye Environment for Web Data Management," Laender et al. describe their Web site data management toolset. They use nested tables to represent Web data and allow users to query it from relational databases.

The selected set of papers provides an excellent summary of how database technologies are being leveraged for the Web. Both the theoretical and practical underpinnings are covered to illustrate how tomorrow's Web, with the help of database technologies, will look.

Reference

- A. Bouguettaya et al., "Managing Government Databases," Computer, vol. 34, no. 2, Feb. 2001, pp. 56-64.
- E. Bertino and G. Guerrini, "OODBMS History and Concepts," Succeeding with Object Databases, A.Chauduri and R.Zicari, eds., John Wiley & Sons, New York, 2000, pp.3-25.
- T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," Scientific Am., vol. 284, no. 5, May 2001, pp. 34–43.

Elisa Bertino is professor of database systems at the University of Milano, Italy, where she is currently the chair of the computer science department. She is a member of the *IEEE Internet Computing* editorial board and coeditor in chief of the *VLDB Journal*. Her main research interests are in database systems, security, and object-oriented technology.

Athman Bouguettaya is the program director in the computer science department at Virginia Tech. He is also director of the E-Commerce and E-Government Research Lab at Virginia Tech. He is on the editorial board of the *Distributed and Parallel Databases Journal*. His research interests are in Web databases, Web services, and workflows.

Readers can contact the authors at bertino@dsi.unimi.it or athman@cs.vt.edu.

32 JULY • AUGUST 2002 http://computer.org/internet/ IEEE INTERNET COMPUTING