Introduction to the Databases Issue

by Suzie Allard

"How much information is there in the world?" According to Michael Lesk, who tackled this question in his 1997 paper of the same name, the huge amount of data would take several billion gigabytes or several thousand petabytes to store [1]. Modern database technology provides the means to store, manage and access such large amounts of information. However, databases are not a new phenomena. In fact, Herman Hollerith mechanized the storage of the 1890 Census in what some deem the first significant "computerized" database [2].

Today, the importance and impact of databases is unquestioned as government organizations, academic institutions, and business entities create and maintain extensive databases containing all kinds of information ranging from natural-language text documents, statistical tables, financial data, and multimedia objects to data of a scientific and technical nature. Many databases are composed of metadata, which means the records hold "data about data" such as information about the size and character of another database rather than primary source content such as a person's name and address. Database technologies, including architecture and access methods, are rapidly developing to keep pace with this demand for information management mechanisms.

Database designers and managers face many challenges that reflect the complexity of the burgeoning information environment. Database technologies must handle massive amounts of data, extract useful information from these repositories, and have the ability to reflect relationships between data maintained in different databases. In addition the architecture and system must provide integrity, recovery, concurrency, and security.

To answer these challenges, the three fundamental database models (hierarchical, network and relational) [3], have served as a foundation for developing more powerful and flexible data models, such as the extended-relational and object-relational models [4]. Well defined architecture and data schema assure efficient, logical data storage which increases database capacity and extends the capabilities of query languages and other access methods. In addition, data mining creates useful information by identifying related data within the vast stores [5]. Researchers now wrestle with the complexities of relational issues and interoperability [6]. For example, researchers are now able to use metadata more efficiently to improve data dissemination. Researchers are also able to use federated strategies for distributed databases [7].

Join us as Crossroads explores the fascinating world of databases!

References

1

Lesk, M. "How much information is there in the world?" http://www.lesk.com/mlesk/ksg97/ksg.html

2

Schneider, G.M. and Gersting, J.L. *An invitation to computer science*, 2nd ed. PWS Publishing , Pacific Grove, CA., 1998.

3

Encyclopaedia Brittanica. "database". http://www.britannica.com/bcom/eb/article/8/0,5716,29888+1+29424,00.html

4

O'Neill, P. and O'Neil E. *Databases: principles, programming and performance*. 2nd ed., Morgan Kaufman Publishers, San Francisco, CA., 2000.

5

Weiss, S. and Indurkhya, N. (contributor). *Predictive data mining: A practical guide.* Morgan Kaufman Publishers, San Francisco, CA, 1997.

6

Payette, S., Blanchi, C. and Lagoze, C., "Interoperability for digital objects and repositories". D-Lib Magazine 5(5): http://www.dlib.org/dlib/may99/payette/05payette.html

7

Mechanized Reasoning Group, Dipartimento di Informatica e Studi Aziendali Università degli Studi, "Federated Databases", http://mrg.cs.unitn.it/~mrg/distributed-intelligence/research-topics/fdb.html

Suzie Allard (slalla0@pop.uky.edu) is a doctoral student and Presidential Fellow at the University of Kentucky, where her research interests focus on digital libraries and knowledge management. She is currently chair-elect of the American Society of Information Science and Technology's Special Interest Group on Digital Libraries.