Chapter XIII

Adaptive Web-Based Database Communities

Athman Bouguettaya
Virginia Tech, USA

Boualem Benatallah
The University of New South Wales, Australia

Brahim Medjahed
Virginia Tech, USA

Mourad Ouzzani
Virginia Tech, USA

Lily Hendra*
Northrop Grumman Information Technology, USA

ABSTRACT

The evolution into the global information infrastructure and the concomitant increase in the available information on the Web, is offering a powerful distribution vehicle for organizations that need to coordinate the use of multiple information sources. However, the technology to organize, search, integrate, and evolve these sources has not kept pace with the rapid growth of the available information space. In this chapter,

*The work was prepared prior to being employed by Northrop Grumman Information Technology, and it does not reflect the views of the company.

This chapter appears in the book, Information Modeling for Internet Applications, edited by Patrick van Bommel. Copyright © 2003, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.
we present our work in the WebFINDIT project. WebFINDIT aims to achieve the scalable integration and efficient querying of Web-accessible databases through the incremental data-driven discovery and formation of interrelationships between information sources. WebFINDIT uses an ontological organization of the information space to filter interactions and accelerate service searches. More precisely, the information space is organized as domain-specific groups. Each group forms a database community to represent the domain of interest of the related databases. Additionally, WebFINDIT provides a monitoring mechanism to dynamically alter relationships between different database communities. This is achieved by using distributed agents that work as background processes. They continually gather and evaluate information about the intercommunity relationships to recommend changes. A prototype has been fully implemented in the context of a healthcare application.

INTRODUCTION

As a result of the rapidly growing number of organizations conducting business over the Web, a large number of heterogeneous information sources (e.g., home pages, tabular data, online digital libraries) is now readily available. The ability to efficiently and effectively share data on the Web is a critical step toward the development of the so-called information superhighway. Existing organizations would form online alliances to deliver integrated value-added information sources (e.g., e-catalogs, information portals).

The evolution into the global information infrastructure and the concomitant increase in the available information, is offering a powerful distribution vehicle for organizations that need to coordinate the use of multiple information sources. However, the technology to organize, search, integrate, and evolve these sources has not kept pace with the rapid growth of the available information space. The efficient sharing of Web data is especially challenging in environments where the information sources are largely autonomous and evolve dynamically. One of the key issues encountered frequently in large cooperative environments, such as data-intensive Web applications, is how users can efficiently query large, intricate, heterogeneous information sources.

Traditional techniques in multidatabases focused on data sharing among a small number of heterogeneous databases (Kim & Seo, 1991). Emerging techniques for querying data over the Web focused on information discovery and brokering in the context of unstructured or semistructured Web-resident data (Florescu, Levy, & Mendelzon, 1998). Our research aims at building a