Chapter IX

Ubiquitous Access to Web Databases

Athman Bouguettaya, Brahim Medjahed, Mourad Ouzzani and Yao Meng
Virginia Tech, USA

ABSTRACT
With the emergence of the Web, there is a need to provide across-the-board transparency for accessing and manipulating data irrespective of platforms, locations, and systems. The challenge is to build an infrastructure to support flexible tools for information space organization, communication facilities, information discovery, content description, and assembly of data from heterogeneous sources. In this chapter, we describe the WebFINDIT system. WebFINDIT builds a scalable and uniform infrastructure for locating and accessing heterogeneous and autonomous databases in large and dynamic environments. One key feature of WebFINDIT is the clustering of Web databases into distributed ontologies. The main advantage of this ontological organization is filtering interactions and reducing the overhead of locating information. Another important feature is the large spectrum of heterogeneity being supported. Heterogeneity appears at different levels, including hardware (Sun and NT), operating system (UNIX and NT), database (Oracle, Informix, DB2, ObjectStore), and communication middleware (CORBA, DCOM, EJB, and RMI).

INTRODUCTION
The Web is emerging as the de facto medium for conducting all types of economic, political, and social activities. Nowadays, organizations publish data in various Web information sources (e.g., databases, HTML and XML repositories) without referring to any central control. Publishing information on the Web is no
longer a major issue. However, accessing Web databases in a uniform and efficient way is still a research goal. In general, Web users are novice. They are not expected to have experience dealing with databases or knowledgeable about all available databases. In this respect, requiring users to keep track of information such as locations, formats (or structures), content, and query languages of the growing number of Web databases is unreasonable. There is a need to provide across-the-board transparency to allow users to access and manipulate data irrespective of platforms, locations, systems, etc. Users must be provided with tools for the effective and efficient exploration of such systems. The challenge is to build an infrastructure that can support flexible tools for information space organization, communication facilities, information discovery, content description, and assembly of data from heterogeneous sources. Old techniques for manipulating these sources are not appropriate and efficient for Internet/Web-based databases. In this context, we distinguish the following key issues when dealing with Web databases:

- **Discovering** relevant information sources in a very large and dynamic information space.
- **Understanding** the meaning, content, and patterns of use of the available information sources. Users have a need to be incrementally educated about the information space.
- **Accessing** located sources and querying them for relevant information items.

To address the aforementioned issues, we developed the WebFINDIT system. WebFINDIT builds a scalable and uniform infrastructure for locating and accessing heterogeneous and autonomous databases in large and dynamic environments. One key feature of our proposed approach is the clustering of Web databases into distributed ontologies. An ontology provides an abstraction of a specific domain of information interest. For example, databases related to medical research would belong to the ontology called Research. The same database may belong to as many ontologies as it desires based on its own information interests. To deal with the dynamic nature of the Web information space, databases may join and leave ontologies at their own discretion. The main advantage of the proposed ontological organization is filtering interactions and reducing the overhead of locating information.

Moreover, ontologies and databases are linked through inter-ontology relationships. These links are dynamically formed based on users’ interests. They allow query “migration” when a query cannot be resolved locally. Sustaining the distributed ontologies requires a complex metadata support. For that purpose, we have associated with each database, a metadata repository called co-database. A co-database contains information about the underlying databases (e.g., DBMS and query language), their ontologies, and inter-ontology relationships. Most interactions with WebFINDIT require access to one or more co-databases.

A prototype of the WebFINDIT system has been fully implemented showcasing a healthcare application. An important feature of the system is the large spectrum of heterogeneity being supported. This heterogeneity appears at different levels including hardware, operating system, database, and communication middleware. A
Related Content

A Novel Approach to Managing the Dynamic Nature of Semantic Relatedness
Youngseok Choi, Jungsuk Oh and Jinsoo Park (2016). Journal of Database Management (pp. 1-26).
www.igi-global.com/article/a-novel-approach-to-managing-the-dynamic-nature-of-semantic-relatedness/165160?camid=4v1a

Map-Side Join Processing of SPARQL Queries Based on Abstract RDF Data Filtering
www.igi-global.com/article/map-side-join-processing-of-sparql-queries-based-on-abstract-rdf-data-filtering/230293?camid=4v1a
Integration of Data Semantics in Heterogeneous Database Federations
www.igi-global.com/chapter/integration-data-semantics-heterogeneous-database/11161?camid=4v1a

Using Ontology and Rule-Based Reasoning for Conceptual Data Models
Synonyms Detection: A Case Study
www.igi-global.com/article/using-ontology-and-rule-based-reasoning-for-conceptual-data-models-synonyms-detection/230292?camid=4v1a