Chapter VI

Keyword-Based Queries Over Web Databases

Altigran S. da Silva
Universidade Federal do Amazonas, Brazil

Pável Calado
Universidade Federal de Minas Gerais, Brazil

Rodrigo C. Vieira
Universidade Federal de Minas Gerais, Brazil

Alberto H.F. Laender
Universidade Federal de Minas Gerais, Brazil

Bertheir A. Ribeiro-Neto
Universidade Federal de Minas Gerais, Brazil

ABSTRACT

In this chapter, we propose an approach to using keywords (as in a Web search engine) for querying databases over the Web. The approach is based on a Bayesian network model and provides a suitable alternative to the use of interfaces based on multiple forms with several fields. Two major steps are involved when querying a Web database using this approach. First, structured (database-like) queries are derived from a query composed only of the keywords specified by the user. Next, the structured queries are submitted to a Web database, and the retrieved results are presented to the user as ranked answers. To demonstrate the feasibility of the approach, a simple prototype Web search system based on the approach is presented. Experimental results obtained with this system indicate that the approach allows for accurately structuring the user queries and retrieving appropriate answers with minimum intervention from the user.

INTRODUCTION

Online information services, such as online stores and digital libraries, have become widespread on the Web nowadays. Such services allow a great number of users to access a large volume of data stored in local databases, also called *Web databases*. Web users, however, are usually non-specialized and their interests vary greatly. Thus, two important problems are posed to designers of interfaces for Web databases: simplicity and uniformity. Interfaces for accessing Web databases are expected to be simple, since they are intended for laymen users. In addition, if an online service is to provide access to different types of information (i.e., many distinct databases), its interface should be as uniform as possible. Otherwise, users will be required to learn how to use a different interface for each distinct database.

The most common solution for implementing online services that access Web databases is the use of customized forms, navigation menus and similar browsing mechanisms. Although useful in some cases, this approach has some important shortcomings. Websites that provide access to multiple databases, such as Amazon.com (http://www.amazon.com), MySimon (http://www.mysimon.com) or Travelocity (http://www.travelocity.com), include dozens of different forms, one for each type of product, where each form might be composed of a large number of fields. From the point of view of a Web user, this type of interface might seem rather complex. From the point of view of a Web developer, it increases the development time and maintenance costs.

Another common inconvenience of query interfaces for Web databases is the fact that the answer set is frequently too large. In a traditional database system, appropriate tools and query languages are available to restrict the search results. In a Web search engine, document ranking (Baeza-Yates & Ribeiro-Neto, 1999) is used to deal with this problem. In Web database interfaces, however, such a method is usually not available.

In this chapter, we describe the use of keyword-based querying (as in a Web search engine) with Web databases and argue that this approach provides a suitable alternative to the use of interfaces based on multiple forms with several fields. Additionally, we show how to use a relevance criteria to rank a possibly large set of answers retrieved by a keyword-based query, as done in Web search engines.

Our approach uses a Bayesian network (Ribeiro-Neto & Muntz, 1996) to model and derive structured (database-like) queries from a query composed only of the keywords specified by the user. The structured queries are then submitted to a Web database and the retrieved results are presented to the user as ranked answers. This means that the user needs just to fill in a single search box to formulate a query. Our approach is thus able to provide online services with: (1) an interface that is simple and intuitive to Web users, and (2) the possibility of querying several heterogeneous databases using a single interface.

To demonstrate the feasibility of our approach, a simple prototype Web search system was implemented. Results obtained using this prototype on databases of three distinct domains (Calado, Silva, Vieira, Laender & Ribeiro-Neto, 2002) indicate that our approach allows accurately structuring the user queries and retrieving appropriate answers with minimum intervention from the user.

The remainder of this chapter is organized as follows. First, we briefly review the traditional paradigm employed for querying Web databases. We then discuss the use of

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/keyword-based-queries-over-web/9206

Related Content

INDUSTRY AND PRACTICE: The Dissemination of Good Practices in Database Development Work
Shirley Becker (1998). *Journal of Database Management (pp. 33-34).*
www.irma-international.org/article/industry-practice-dissemination-good-practices/51202

A Rhetorical Perspective on Localization and International Outsourcing
Kirk St. Amant (2005). *Encyclopedia of Database Technologies and Applications (pp. 570-574).*
www.irma-international.org/chapter/rhetorical-perspective-localization-international-outsourcing/11206

Semantic Attention Network for Image Captioning and Visual Question Answering Based on Image High-Level Semantic Attributes
Angelin Gladstonand Deeban Balaji (2022). *International Journal of Big Data Intelligence and Applications (pp. 1-18).*
www.irma-international.org/article/semantic-attention-network-for-image-captioning-and-visual-question-answering-based-on-image-high-level-semantic-attributes/313201

The Status Quo and Development Countermeasures of Venture Capital in the New Energy Economy Based on Big Data Analysis
Nan Feng, Yuguang Wang, Zhiguo Chenand Tingting Song (2023). *Journal of Database Management (pp. 1-23).*
www.irma-international.org/article/the-status-quo-and-development-countermeasures-of-venture-capital-in-the-new-energy-economy-based-on-big-data-analysis/322019

Theoretical Foundation
 (2019). *Information Systems Strategic Planning for Public Service Delivery in the Digital Era (pp. 1-19).*
www.irma-international.org/chapter/theoretical-foundation/233400