Chapter LXVIII
A Web Query System for Heterogeneous Government Data

Nancy Wiegand
University of Wisconsin–Madison, USA

Isabel F. Cruz
University of Illinois at Chicago, USA

Naijun Zhou
University of Wisconsin–Madison, USA

William Sunna
University of Illinois at Chicago, USA

ABSTRACT

This paper describes a Web-based query system for semantically heterogeneous government-produced data. Geospatial Web-based information systems and portals currently are being developed by various levels of government along with the GIS community. Typically, these sites provide data discovery and download capabilities but do not include the ability to pose DBMS-type queries. One of the main problems in querying distributed government data sources is the difference in semantics used by various jurisdictions. We extend work in schema integration by focusing on resolving semantics at the value level in addition to the schema or attribute level. We illustrate our method using land use data, but the method can be used to query across other heterogeneous sets of values. Our work starts from an XML Web-based DBMS and adds functionality to accommodate heterogeneous data between jurisdictions. Our ontology and query rewrite systems use mappings to enable querying across distributed heterogeneous data.

INTRODUCTION

Governmental data are increasingly are being produced and distributed over the Web. Integrating and querying these data will greatly help improve governmental decision making. However, most current Web-based practices rely on keyword search engines and do not consider the heterogeneous problem in data sources. This paper presents a study of deploying powerful Database Management System (DBMS) technology to improve capabilities for data search and query and to use ontology approaches for data semantics. We give an overview of our system and address the problem of mediating between highly heterogeneous information developed independently by many different government units. We focus on geospatial data in general and on diverse land use coding systems in particular. To be able to map between heterogeneous values found in land use attributes, we extend work on semantic integration by resolving semantics at the *value* level of individual attributes in addition to the schema level.

Web sites are currently being developed by many levels of government to serve geospatial data. A recent example from the federal government is Geospatial One-Stop (2003), which was initiated to have all geospatial data and information from federal agencies along with many state, local, tribal, and private agencies accessible from one portal. Other examples of geospatial clearinghouses and portals at the national level include the Alexandria Digital Library and the Federal Geographic Data Committee (FGDC) Clearinghouse. In addition, there are many geospatial Web sites being produced at the state, county, and local levels. For example, state level examples for Wisconsin include the Wisconsin Land Information Clearinghouse (WiscLinc) and the prototype Wisconsin Land Information System (WLIS).

WLIS, for example, will allow central access to distributed data sets that remain under local government control and reside on local and county servers across the state (WLIS Project Team,

2000). The problem is that without state-mandated standards, data sets produced by local and county governments are highly heterogeneous. As one moves across jurisdictional boundaries, not only do database schemas vary, but also the definitions and acceptable values of attributes change significantly. In fact, most of the effort required to integrate diverse geospatial data lies with the nonspatial attributes. That is, integration problems caused by the use of different coordinate systems, for example, are more easily solved (WLIS Project Team, 2000).

We began our work in the context of WLIS but are extending it to a larger scale such as needed for Geospatial One-Stop. The typical purpose of WLIS and other geospatial Web sites is to enable search and download capabilities for geospatial data and services. However, such sites are limited or lacking in their ability to allow full DBMS-type querying over the content of distributed government data sources. Instead, querying in these sites is restricted to selecting a few metadata fields for the purpose of locating data sources. Sites that do allow DBMS querying on source data content provide an explicit connection over the Web to a backend DBMS containing the data of a single data source. Otherwise, to query data source content, a user has to download the data into a local Geographic Information System (GIS) or DBMS. To enhance these government Web sites, we are working to support full-fledged DBMS-type querying over distributed data sources.

This paper discusses our approach as embodied by a prototype Web-based query system in which we focus on resolving semantic heterogeneities between the attributes of related data sets. To design our system architecture, we first consider characteristics of government-produced geospatial data. Next, we discuss the semantic heterogeneity problem in general and then present our specific application problem concerning land use codes. We then discuss ontology integration methods including our semi-automatic solution. Our existing working system using land use data examples is described in the last section.

13 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/web-query-system-heterogeneous-government/21295

Related Content

E-Halli: An Opportunity to Become Rural Entrepreneurs
Swamy Tribhuvananda H.V.and Gopakumar K. (2009). *E-Government Development and Diffusion: Inhibitors and Facilitators of Digital Democracy (pp. 264-271).*
www.irma-international.org/chapter/halli-opportunity-become-rural-entrepreneurs/8986

Data Mining and Homeland Security
J. W. Seifert (2007). *Encyclopedia of Digital Government (pp. 277-282).*
www.irma-international.org/chapter/data-mining-homeland-security/11516

Level-Based Development of E-Government Services
Penelope Markellou, Angeliki Panayiotakiand Athanasios Tsakalidis (2008). *Electronic Government: Concepts, Methodologies, Tools, and Applications (pp. 1742-1752).*
www.irma-international.org/chapter/level-based-development-government-services/9820

Factors of Innovation Management Transformation in Digital Innovation Ecosystems of Russian Companies
Mikhail Khachaturyanand Evgeniia Klicheva (2022). *International Journal of Electronic Government Research (pp. 1-18).*
www.irma-international.org/article/factors-of-innovation-management-transformation-in-digital-innovation-ecosystems-of-russian-companies/315603

Big Data Analytics Capability and Governmental Performance: An Empirical Examination
Guido Ongenaand Arjen Davids (2023). *International Journal of Electronic Government Research (pp. 1-18).*
www.irma-international.org/article/big-data-analytics-capability-and-governmental-performance/321638