Chapter 13
Topical and Spatio–Temporal Search over Distributed Online Databases

Nikos Zotos  
*Patras University, Greece*

Sofia Stamou  
*Patras University, Greece*

**ABSTRACT**

In this chapter, the authors propose a novel framework for the support of multi-faceted searches over distributed Web-accessible databases. Towards this goal, the authors introduce a method for analyzing and processing a sample of the database contents in order to deduce the topical, the geographic, and the temporal orientation of the entire database contents. To extract the database topics, the authors apply techniques leveraged from the NLP community. To identify the database geographic footprints, the authors first rely on geographic ontologies in order to extract toponyms from the database content samples and then employ geo-spatial similarity metrics to estimate the geographic coverage of the identified toponyms. Finally, to determine the time aspects associated with the database entities, the authors extract temporal expressions from the entities’ contextual elements and utilize a time ontology against which the temporal similarity between the identified entities is estimated.

DOI: 10.4018/978-1-61692-868-1.ch013
INTRODUCTION

A significant fraction of the hidden Web content pertains to data stored in online databases. Previous studies have shown that the size of the hidden Web is much larger than that of the surface Web and that its quality is relatively high (cf. The Deep Web). Although online databases offer large volumes of qualitative content, typical Web users are generally deprived of these data for three main reasons: first, because search engines, i.e. the predominant medium for accessing the Web data, do not index the contents of online databases (Xu, et al., 1998); second, because information seekers are not aware of all the available online databases that could potentially serve their queries, and third, because the majority of users are either reluctant or incompetent to specify different queries for a single information need so that these conform to the query syntax and/or language that different databases support.

To overcome the above difficulties, researchers have proposed several approaches, the majority of which aim at tackling the following issues: (1) how to enable search engines index the hidden Web content (Ntoulas, et al., 2005), (2) how to aggregate the database contents into content summaries so as to facilitate database selection for search queries (Ipeirotis & Gravano, 2002; Ipeirotis & Gravano, 2008), and (3) how to translate queries into appropriate formats for each database (Sugiura & Etzioni, 2000). Despite the success of the above approaches, there are still open issues with respect to searching for information in online structured data sources. One such issue concerns the multi-faceted representation of the database contents so as to allow users query different databases across multiple dimensions.

Currently, databases organize their contents thematically via the use of concept hierarchies. However, this kind of data organization enables searching the database contents in a single dimension, i.e. by topic. Unfortunately, searching by topic cannot accommodate all user needs and search behaviors, since a significant fraction of search queries aim at retrieving spatio-temporal data about a subject of interest. For example, consider a news archive (database). Some users might want to search the database contents by topic (e.g. economic crisis), others might want to search the database contents by location (e.g. job loses in France), yet others might want to search the database contents by time constraints (e.g. 2008 economic crisis). In this scenario, a monolithic topical organization of the database contents would hinder users from performing multi-dimensional searches, such as: impact of the 2008 economic crisis in France. Evidently, if we could organize the database contents across multiple dimensions, e.g. by topic [Economy → economic crisis], by location [Regional → Europe → France], and by time [21st century → 2008], we would not only enable multi-faceted online searches, but we would also save a lot of time from information seekers since we would help them locate the most relevant (topical, spatial and temporal) data sources faster. That is, in our example query the documents indexed under all matching facets, i.e. economic crisis, France and 2008, would be prioritized in the query results.

In addition to supporting multi-faceted searches over online databases, in many cases it is desirable that queries are simultaneously submitted to different databases that contain useful information. This is preferable in case the users are not aware of all the existing databases that could potentially serve their information needs, or in case the users do not want to bound their searches to a particular source of information. In this scenario, the search should be distributed over different databases and the relevant documents should be merged in a single ranked list of retrieved results. To support distributed searches over online databases, we can build metasearchers, which provide a uniform interface for querying multiple databases at once. A metasearcher performs three main tasks: upon issuing a query it selects the databases that contain relevant information, it translates the query in a
Related Content

Constraints in Authoring BIM Components for Optimal Data Reuse and Interoperability: Results of Some Initial Tests
[www.igi-global.com/article/constraints-authoring-bim-components-optimal/77815?camid=4v1a](www.igi-global.com/article/constraints-authoring-bim-components-optimal/77815?camid=4v1a)

Requirements for Model Server Enabled Collaborating on Building Information Models
[www.igi-global.com/article/requirements-model-server-enabled-collaborating/75132?camid=4v1a](www.igi-global.com/article/requirements-model-server-enabled-collaborating/75132?camid=4v1a)

3D Hilbert Space Filling Curves in 3D City Modeling for Faster Spatial Queries
[www.igi-global.com/article/3d-hilbert-space-filling-curves-in-3d-city-modeling-for-faster-spatial-queries/120062?camid=4v1a](www.igi-global.com/article/3d-hilbert-space-filling-curves-in-3d-city-modeling-for-faster-spatial-queries/120062?camid=4v1a)

A Multidimensional Model for Correct Aggregation of Geographic Measures
[www.igi-global.com/chapter/multidimensional-model-correct-aggregation-geographic/70451?camid=4v1a](www.igi-global.com/chapter/multidimensional-model-correct-aggregation-geographic/70451?camid=4v1a)