Contribution ID: **399**   Type: **Poster presentation**

Keyword Search over Data Service Integration for Accurate Results

Monday, 14 October 2013 15:00 (45 minutes)

Background: The goal of the virtual data service integration is to provide a coherent interface for querying a number of heterogenous data sources (e.g., web services, web forms, proprietary systems, etc.) in cases where accurate results are necessary. This work explores various aspects of its usability.

Problem: Querying is usually carried out through a structured query language, such as SQL, which forces the users to learn the language and to get acquainted with data organization (i.e. the schema) thus negatively impacting the system's usability. Limited access to data instances as well as users' concern with accurate results of arbitrary queries present additional challenges to traditional approaches (such as query forms, information retrieval, keyword search over relational databases) making them not applicable.

Solution: This paper presents a keyword search system which deals with the above discussed problem by operating on available information: the metadata, such as the constraints on allowed values, analysis of user queries, and certain portions of data. Given a keyword query, it proposes a ranked list of structured queries along with the explanations of their meanings. Unlike previous implementations, the system is freely available and makes no assumptions about the input query, while maintaining its ability to leverage the query's structural patterns - in case they exist. The system is discussed in the context of CMS data discovery service where the simplicity and capabilities of the search interface play a crucial role in the ability of its users to satisfy their information needs.

Summary

Primary authors: KUZNETSOV, Valentin Y (Cornell University (US));   ZEMLERIS, Vidmantas (Vilnius University (LT))

Co-author: KREUZER, Peter (Rheinisch-Westfaelische Tech. Hoch. (DE))

Presenter: ZEMLERIS, Vidmantas (Vilnius University (LT))

Session Classification: Poster presentations

Track Classification: Distributed Processing and Data Handling A: Infrastructure, Sites, and Virtualization