

Multilingual DGRC AskCal: Querying Energy Time Series in English, Spanish, and Mandarin Chinese

Andrew Philpot, Eduard Hovy, Lei Ding

DGRC
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
USA
{philpot,hovy,leiding}@isi.edu

Keywords; database query interface, NLP, database access

Large collections of heterogeneous data, such as government provided energy time series, can be difficult for users to navigate through in order to find the data of interest. User difficulties result from issues such as lack of familiarity with the terms of art in the domain, lack of expertise with the query mechanisms specific to particular end data sources, the need to combine or relate data from multiple locations in order to satisfy a single query, and the barriers posed when supplementary resources such as metadata, free-text documentation, and query construction tools are not accessible in a language the user understands well.

We demonstrate Multilingual AskCal, a system that leverages a general purpose ontological framework augmented with specific conceptualizations from the energy time series domain complete with terminological referents drawn from English, Spanish and Mandarin, as well as a query planner and various feedback modalities. AskCal is the natural language query interface front end to the Energy Data Collection (EDC) system, an NSF-funded project at the DGRC for the past 4 years. EDC contains over 50,000 tables of information about gasoline prices stored in a variety of formats, from live online websites to MS Access databases to tables on CDs.

In the demonstration, the user can explore the domain ontology, pose and refine energy data queries, execute the queries, and inspect the results using hierarchical, tabular, and graphical display tools. We show how a hybrid of natural language approaches, ATN parsing of free text queries, ad hoc modification of template queries, and feature-based menu selection relative to an as-parsed partial query, combine to provide the user with a high degree of query ease while supporting significant coverage of the important queries possible in the domain. We illustrate the utility of general and domain specific knowledge drawn from a multilingual ontology both for query planning and execution by the system and for effective navigation and control by the user.