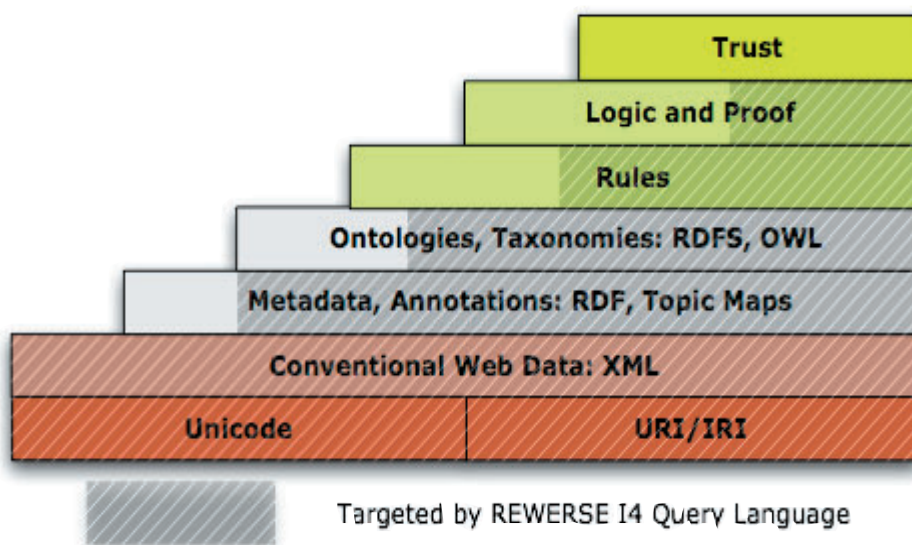


Reasoning-aware Querying

Abstract

Enabling more "meaningful" access to Web and Semantic Web data by integrating reasoning capabilities in a Web query and transformation language. Querying, i.e., the efficient and effective access to data, is one of the most essential enabling technologies for any information system. In the Web context, reasoning capabilities enhance traditional search and access technologies to be able to cope with heterogeneous, incomplete, and often even inconsistent information.



Mission

Effective and efficient access to information is recognized by industry leaders (e.g., Microsoft, Apple, Google) as a central topic for upcoming years. The query language developed by this working group aims at providing a unique solution to this problem in the Web context along three core principles:

1. integrated access to both standard Web data (in XML) and upcoming Semantic Web information (in RDF, Topic Maps, or OWL),

2. powerful but easy-to-use reasoning capabilities particularly for tackling the vast amount of heterogeneous and distributed data in the Web, e.g., by integrating formal taxonomies (named by Gartner among the top 10 technologies for the upcoming years) into the search,

3. easy and extensive extensibility of the query language to cope with the changing nature of the Web.

The objective of the working group "Reasoning-aware Querying" is to develop, implement, and test such a rule-based query language. A close integration with other REWERSE technologies is essential for realizing the improvements offered by such an approach to effective and efficient search.

Use Scenarios

Consider the organization of a press conference, technology awareness event, or similar event for a company offering a new technology in the area of "Web Services". Information about companies in a certain region that are interested in such a technology is not easily come by nowadays: Neither conventional search engines such as Google or AllTheWeb nor current yellow pages for companies (such as <http://yellowpages.superpages.com>) provide a sufficient precision in

answers even to finely tuned queries to automatically retrieve such companies: In the end, a human being has to investigate all the search results retrieved from Google or a yellow page service and manually collect the interesting companies from these results.

Our working group aims at developing a query language that can (at least partially) automate "filtering" from large selections of data (such as those returned by a search engi-

ne) and thus realizes a first step in the direction of the W3C's Semantic Web vision. The user formulates the search intent using either the textual query language, a visual rendering of the query language or a natural language interface. >

More information available at

<http://rewerse.net/i4>

The formalized search intent is used by the query processor to access the needed information resources: an appropriate formal taxonomy or ontology can be queried on how companies related to "Web Services" might be classified. This information in turn can be used to query a yellow pages service for all companies classified under any of the terms retrieved from the taxonomy.

Finally, the results could be further improved by querying, e.g., the homepages of the resulting companies

a) to confirm that the company might be interested in the technology to be presented by looking at services and customers, b) to extract an executive summary of the company for further reviewing by a human being, and c) to find affiliates of the company potentially also interested in the technology therefore to be invited.

Due to the inherently heterogeneous and distributed nature of the Web, such a query will most likely still require human involvement, but by

presenting a more precise result set (and possibly additional information for judging the results) the involvement of the human is drastically reduced and simplified.

In this scenario, the crucial co-existence of standard Web and Semantic Web data is apparent: The standard Web data, e.g., the homepages of the companies or (to some extent) the yellow pages service, is enhanced by Semantic Web information.

Description of Research

As the goal of the "Reasoning-aware Querying" working group is the development of a query language that fulfills the objectives laid out above, our research focuses on

1. recognizing the requirements for such a language by an extensive survey of current proposals and interesting scenarios for using such a language. Here the cooperation with the application-oriented REVERSE working groups as well as with external experts on certain application areas is essential.
2. specifying a convenient syntax and formal semantics for the query language. This includes the investigation of formal properties of the language such as query and data complexity.
3. developing an efficient reference implementation for the query language.

Tools & Technologies

■ Xcerpt: Textual Query Language

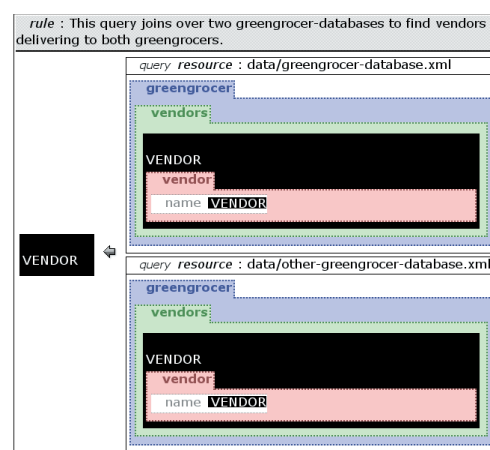
- <http://xcerpt.org/>
- prototypical reference implementation available
- carefully crafted syntax along the ideas of "query-by-example" where a query is written as a form to be filled with the results of the query

■ visXcerpt: Visual Query Language

- <http://visxcerpt.xcerpt.org/>
- visual rendering of Xcerpt using Web standards technology (XML, CSS, etc.)
- prototypical implementation of interactive features

■ VOXX: Verbalization of XML and Xcerpt

- controlled natural language subset (inspired by Attempto Controlled English, ACE)
- aims at expressing both data and queries in natural language
- intuitive interface
- well-suited for voice-only scenarios, e.g., automated call answering



Contact Person

Dr. François Bry, Professor
Ludwig-Maximilians-Universität
München
Institut für Informatik
Oettingenstr. 67
80538 München, DE

Phone: +49 89 2180 9310
Email: Francois.Bry@ifi.lmu.de

www.ifi.lmu.de/~bry/

Participant Representatives

Grigoris Antoniou (Heraklion);
Livia Badea (Bucharest);
François Bry (Munich);
Włodzimierz Drabent (Warsaw);
Thomas Eiter (Vienna);
Tim Geisler (webXcerpt);
Ian Horrocks (Manchester);
Claude Kirchner (Nancy);
Jan Matuszynski (Linköping);
Massimo Marchiori (Venice);
Wolfgang May (Göttingen);
Wolfgang Nejdl (Hannover);
Michael Schröder (Dresden);
Gerd Wagner (Eindhoven)

Complete member list at
<http://reverse.net/i4/people>

Impressum

webXcerpt Software GmbH
REVERSE Technology Transfer
Aurbacherstr. 2, D-81541 Munich
<http://reverse.net>

Contact: Andrea Kulas
ak@webxcerpt.com
Phone: +49 89 54 80 88 48

Responsible for the content:

Prof. Dr. Francois Bry
Ludwig-Maximilians-Universität München
Institut für Informatik
Oettingenstr. 67
D-80538 München
Francois.Bry@ifi.lmu.de
Phone: +49 89 2180 9310