



Munich Personal RePEc Archive

Testing Semantic-Web based Querying by Using Ontologies in Resource Description Format

necula, sabina-cristiana

Alexandru Ioan Cuza University of Iasi

April 2012

Online at <https://mpra.ub.uni-muenchen.de/51552/>
MPRA Paper No. 51552, posted 18 Nov 2013 21:06 UTC

Testing Semantic-Web based Querying by Using Ontologies in Resource Description Format

Sabina-Cristiana NECULA

Department of Research

Alexandru Ioan Cuza University of Iasi, Faculty of Economics and Business Administration

Iasi, Romania

sabina.mihalache@gmail.com

Abstract

This paper presents with examples some queries made on data sets using SPARQL. We treat the problem of available standards and tools. We show data tests resulted from querying different ontologies. Also the article treats the problem of describing data by using Resource Description Format.

Keywords: testing semantic-web based query, semantic-web, Enterprise search

Introduction

For a long period of time data has been considered unimportant in designing software. This fact has lead to massive data stored in software producers' formats, having the main effect that data is understand only by dedicated software. Collaborative work and mass information production changed the preoccupation from designing dedicated software to sharing, interoperability and semantic aspects.

Since Allen Newell (Newell, 1982) introduced the concept of knowledge level, many scientists and practitioners focused on the symbol level, considering that by optimizing the symbol level, the knowledge level is improved. Whereas the knowledge level is world oriented, the symbol level is system oriented.

The knowledge level consists of the information contained in data structures. The symbol level is represented by program's algorithms; data structures, and so on. We conclude from Newell findings that in order to improve semantic search, the semantics of data sources should be treated first.

There is a main problem and, in the same time, a controversy here. Data exists, data comes from different sources (internal and external to enterprise) differently described. There is an "open world" (the Web) and a "closed world" (the enterprise). If in the "open world" we discuss search engines and queries made by Internet users on existent data from www space, in the "closed world" we discuss databases and answers to queries that are priory represented. In the first case the answers to queries seems to be poorly represented in the aspect of semantics, in the second case the answers seems to be perfectly represented but not relevant. In this closed world of enterprise, usually decision-makers need answers to queries formulated by them and these queries are ad-hoc, very often cannot be anticipated and their answers need a mixture of information that comes from the both worlds. Therefore it is our goal to solve this problem by this article.

The scope of this paper is to present a method to improve search at the enterprise level. The main idea consists in using ontologies and semantic search technologies. The motivation is given by lack of interoperability and semantic consistency of different formats for the same content. In theory and practice there are many given solutions but as the field of research evolves any solution need time in order to be validated.

This paper has 4 Sections. Section 1 presents an introduction. Section 2 contains some aspects related to the current work on Semantic web. Section 3 presents with examples the uses of vocabularies, Resource Description Format (RDF), and SPARQL Protocol and RDF Query Language (SPARQL) for querying data sets. Section 4 treats the main conclusions.

Related Work

For the moment, we might say that there are two main approaches in the field of building semantic web applications: (1) using domain ontology; and (2) querying different ontologies. By making use of Semantic Web Stack, generally adopted in the field of semantic web research we present the two approaches using Figure 1. We present also, where it is situated our research from this article.

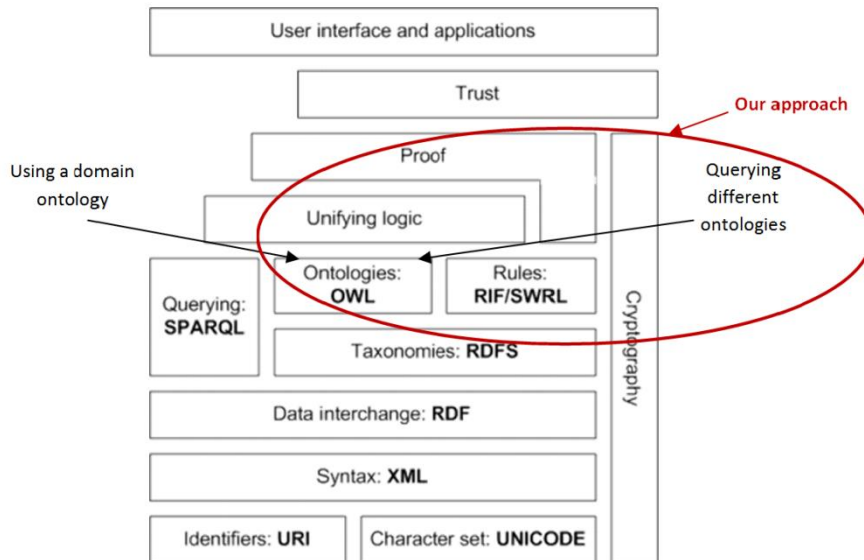


Figure 1. Semantic Web stack and the present research orientation in field of building semantic web applications

In order to discuss differences from our approach we present the related achieved work in the field of adopting semantic web standards and in the field of developing applications.

The Semantic Web is the extension of the World Wide Web that enables people to share content beyond the boundaries of applications and websites. It has been described in rather different ways: as a utopic vision, as a web of data, or merely as a natural paradigm shift in our daily use of the Web.

Enterprise semantic search related work

Looking at the business areas of early adopters registered in the W3C collection of Semantic Web Case Studies and Use Cases, we found that 36% are public institutions with an extension to 47%

for organizations that implemented e-government principles using semantic web technologies. Only 2% were finance applications. Semantic web technologies are by far most often used for data integration and for improving the search (V. Janev, S. Vraneš, 2010).

There are also many software applications in the commercial field but not sufficiently related to enterprise semantic search:

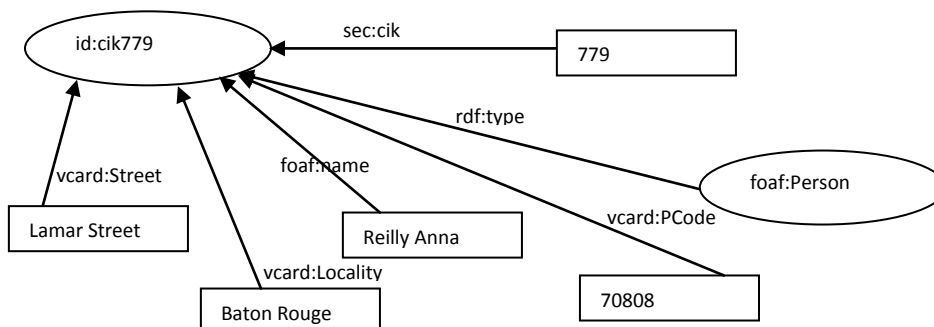
- World Wide Web Consortium (W3C) is the main contributory in the field of Semantic web research. On its page there are 44 *semantic web uses and case studies*. The main preoccupation remains for the moment developing tools and improving standard specifications in building ontologies.
- The commercial field bases primarily on Internet users and on links made by them. The research field bases primarily on contributed content shared by users realized by using dedicated semantic web open-source tools. It is an evolving field and its main contribution consists in building ontologies.
- From the commercial field, Google counts clicks made by internet users. Information is therefore organized depending on the interest manifested by users for some content. Yahoo! has ambitious results in the field of Semantic search. They realized a Search Monkey, where additional dynamic data from marked up fields can also be displayed on the search results page itself.

The role of RDF as an integration platform for data from multiple sources, primarily in form of relational databases (RDB), is one of the main motivations driving research efforts in mapping RDB to RDF. W3C RDB2RDF Incubator Group is concerned by standardization of RDB to RDF mappings. There are several studies or tools allowing mapping relational databases (RDBs) to RDF schema or OWL ontologies. Some of the most notable approaches of this kind are R2O (J. Barrasa, A. Gómez-Pérez, 2006), D2RQ, Virtuoso RDF Views (C. Blakeley, 2007) and DartGrid (W. Hu, Y. Qu, 2007).

A Practical Example in realizing SPARQL Queries

We start by describing using RDF graphs a datasets that we want to query.

We used Security Exchange Commission data sets available in n3 format at <http://www.rdfabout.com/demo/sec/>. Besides this data about companies we used some financial data available in Excel files at http://pages.stern.nyu.edu/~adamodar/New_Home_Page/data.html.



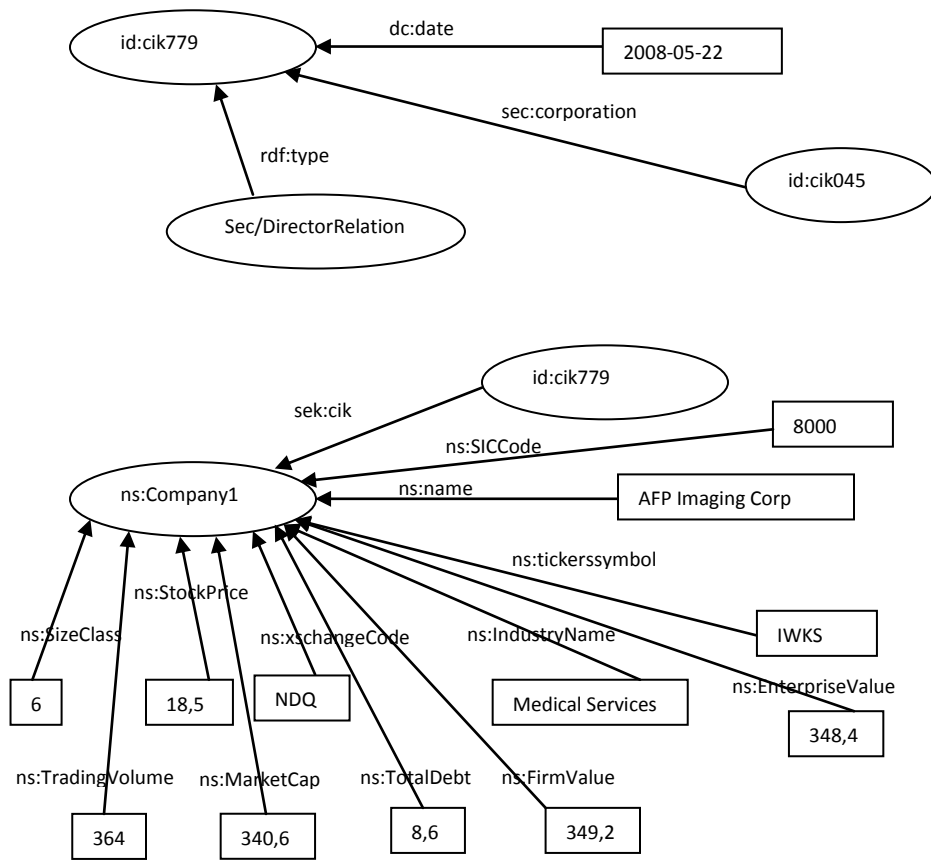


Figure 1. RDF classes, instances and properties

RDF triples can be described using turtle syntax. We present a simple example in Figure 2.

```
:company1 ns:name "China Infrastructure Invsmt"
```

Figure 2. A RDF triple in Turtle syntax

Writing queries by making use of SPARQL doesn't mean anything else than asking values for objects from subject-predicate—object triple. An example is given in Figure 3.

```
ns:company foaf:name ?name.
```

Figure 3. A SPARQL triple pattern, with a single variable

When writing queries all parts of a triple can be requested. An example is given in Figure 4.

```
?company foaf:name ?name.
```

Figure 4. A SPARQL triple pattern, with all parts as variables

On our datasets if we want to retrieve all variables from sec vocabulary that are of the Directorrelation kind of type we will write a query that looks like the one from Figure 5.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: http://www.w3.org/2000/01/rdf-schema#
```

```

prefix dc:      <http://purl.org/dc/elements/1.1/>
prefix vcard:  <http://www.w3.org/2001/vcard-rdf/3.0#>
prefix :       <http://example.org/company/>
prefix ns:     <http://sandbox.metadataregistry.org/uri/schema/fin>
SELECT *
WHERE {
    ?subject sec:cik ?cik;
           rdf:type sec:DirectorRelation.
}

```

Figure 5. A SPARQL query retrieving all variables

If we want to query what are the uri and the StockPrice of highest SockPriced companies we will write a query that look like query depicted in Figure 6.

```

PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX rdfs: http://www.w3.org/2000/01/rdf-schema#
prefix dc:   <http://purl.org/dc/elements/1.1/>
prefix vcard: <http://www.w3.org/2001/vcard-rdf/3.0#>
prefix :     <http://example.org/company/>
prefix ns:   <http://sandbox.metadataregistry.org/uri/schema/fin>

SELECT ?company ?StockPrice
WHERE {
    ?company ns:StockPrice ?StockPrice.
}
ORDER BY DESC(?StockPrice)
LIMIT 10

```

Figure 6. A SPARQL query that returns uri and StockPrice of the ten companies that have the highest Stock Price

In the next example we will query what are the companies that have created their uri between May 1st 2008 and December 12th 2011.

```

PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX xsd:  http://www.w3.org/2001/XMLSchema#
prefix dc:   <http://purl.org/dc/elements/1.1/>
prefix vcard: <http://www.w3.org/2001/vcard-rdf/3.0#>
prefix :     <http://example.org/company/>
prefix ns:   <http://sandbox.metadataregistry.org/uri/schema/fin>

SELECT ?name
WHERE {
    ?id dc:date ?date;
        sec:corporation ?corporation.
    ?company ns:name ?name.
    FILTER (?date > "2008-05-01"^^xsd:date &&
           ?date < "2011-12-12"^^xsd:date)
}

```

Figure 7. A SPARQL query that returns the name of the companies that created their uri between May 1st 2008 and December 12th 2011.

CONCLUSIONS

This paper presents with examples querying data sets by using Semantic Web technologies. Although not sufficiently treated by semantic web developers there are a lot of opportunities for those who intend to query public data sets like governments or companies or financial regulatory bodies.

We present in this paper the necessary SPARQL examples queries in order to observe what the potential for semantic search is. Our future work will refer to scalability in order that web

applications may work on big data sets. We mention that we used for our examples 89523 triples stored in our triple store and that working with bigger data sets represents a problem in accessing data for Joseki server.

Acknowledgments

This work was supported by CNCSIS-UEFISCSU, project number PN II-RU code 188/2010.

References

1. A. Newell, The Knowledge Level, *Artificial Intelligence*, 18(1):87-127, 1982.
2. J. Barrasa, A. Gómez-Pérez, Upgrading relational legacy data to the semantic web, In: *Proceedings of the 15th International World Wide Web Conference (WWW 2006)*, Edinburgh, United Kingdom, ACM 2006, pp. 1069–1070.
3. B. Blakeley, RDF Views of SQL Data (Declarative SQL Schema to RDF Mapping), OpenLink Software, 2007, <http://www.slideshare.net/rumito/rdf-views-of-sql-data-power-point-presentation-1-173180>
4. W. Hu, Y. Qu, Discovering Simple Mappings between Relational Database Schemas and Ontologies, In: *Proceedings of the 6th International Semantic Web Conference (ISWC 2007)*, 2nd Asian Semantic Web Conference (ASWC 2007), Busan, Korea, 11–15 November 2007, LNCS, 4825, pp. 225–238, <http://www.springerlink.com/content/e012x10284r17045/>
5. V. Janev, and S. Vraneš, Applicability assessment of Semantic Web technologies, *Information Processing and Management*, Elsevier, 2010, doi:10.1016/j.ipm.2010.11.002
6. T. Berners Lee, <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>
7. <http://www.w3.org/2001/sw/sweo/public/UseCases/>
8. <http://www.w3.org/2005/Incubator/rdb2rdf/>
9. D2RQ Platform, <http://www4.wiwiss.fu-berlin.de/bizer/D2RQ/spec/>
10. OpenLink Virtuoso Platform, Automated Generation of RDF Views over Relational Data Sources, <http://docs.openlinksw.com/virtuoso/rdfviewgnr.html>
11. Securities Exchange Commission RDF Data, <http://www.rdfabout.com/demo/sec/>
12. ^ "W3C Semantic Web Frequently Asked Questions". W3C. <http://www.w3.org/2001/sw/SW-FAQ>. Retrieved March 13, 2008.
13. ^ I. Herman, "W3C Semantic Web Activity", W3C. <http://www.w3.org/2001/sw/>, Retrieved March 13, 2008.

