SWAD–Europe Deliverable 7.2: Databases, Query, API, Interfaces report on Query languages

Project name: Semantic Web Advanced Development for Europe (SWAD-Europe)
Project Number: IST-2001-34732
Workpackage name: Work package 7: Databases, Query, API, Interfaces
Deliverable title: Public report comparing existing RDF query language functionality, documenting different scenarios and users for RDF query languages
URI: http://www.w3.org/2001/sw/Europe/reports/rdf_ql_comparison_report/
Authors: Libby Miller
Abstract:

This report is part of SWAD-Europe Work package 7: Databases, Query, API, Interfaces. It is intended to compare existing RDF query language functionality, documenting different scenarios and users for RDF query languages (for example, scripters, programmers; data, schema). There are many existing documents covering this area. As part of the work on this deliverable, the project has begun to gather together people and materials for creating a repository of tests for RDF query. We are working with people who have their own query languages, creators of a repository of RDF query use cases, and with the creator of a comparison document for RDF query languages to hold a series of IRC and face to face meetings to discuss a common manifest format and results format for the tests. The aim is that the creators of similar RDF query languages can share testcases and thereby improve interoperability. More information is available on the RDF query FAQ.

Contents

- Introduction
- RDF query FAQ
- Recent Changes

Introduction

As part of the work on this deliverable, the project has begun to gather together people and materials for creating a repository of tests for RDF query. We are working with people who have their own query languages, creators of a repository of RDF query use cases, and with the creator of a comparison document for RDF query languages to hold a series of IRC and face to face meetings to discuss a common manifest format and results format for the tests. The aim is that the creators of similar RDF query languages can share testcases and thereby improve interoperability. More information is available on the RDF query FAQ.

RDF query Frequently Asked Questions (FAQs)

1. What is an RDF query?
2. What is the relationship between RDF query and rules?
3. Can someone explain if there is "an accepted" query language specification?
4. Why isn’t there a W3C working group or taskforce for RDF query?
5. How can I use XML Query for querying RDF?
6. How does RDF fit in with XQuery?
7. Could anyone please tell me where to find more information on RDFS Query Language?
8. How do query languages for RDF handle subclass and subproperty relations?
9. How do query languages manage RDF5?
10. Which RDF query languages can handle RDF schema-related queries?
11. Which RDF query languages can handle DAML+/OIL/OWL properties like inverse, unambiguousProperty?
12. How do I query DAML data?
13. What is the relationship between DAML+/OIL and RDF query?
14. How does RDF query handle datatypes?
15. Which query languages for RDF allow you to specify optional variables to bind?
16. Can you represent b-nodes in RDF queries?
17. Which RDF query languages can do substring matches? regex matching?
18. How do I query reified statements?
19. How can I query and return provenance information in RDF query?
20. Are there any recursive, template-matching, XSLT-like query languages for RDF?
21. Why not use an RDF graph with blanks for querying RDF?
22. How do I represent RDF graphs in a (RDBMS, SOL) database and query them?
23. Is there a programmatic query API?
24. How do I query generic RDBMS, SOL data using RDF query?
25. Are there any test suites or test cases for RDF query?
26. What other summaries of RDF query are available?
What an RDF Query? - There is no one official RDF query language - there has not been a W3C activity in this area, for example. However there are many RDF query languages in use. There is no explicit consensus about what an RDF query is, although there are many similarities in the available implementations. It is important to distinguish between the syntax of an RDF query and what it does. Many RDF query languages (but not all) have different syntaxes but do basically the same thing, that is, describe an RDF graph with parts missing, assign those parts variable names, and get a series of bindings to those variables. Examples of query languages that do this include RDFdb QL, Algae, Squish, RDQL, RDFql.

For example, a representation of a query which is a graph with parts missing might express the query:

"find me the name of the person whose email address is libby.miller@bristol.ac.uk, and also find me the title and identifier of anything that she has created"

This simple description is not sufficient to encompass all RDF query languages by any means. Some (in particular RQL) have a specific syntax for accessing RDF schema information (such as subclass, subproperty). Others have an XML Path-like structure that can recursively match subgraphs (Versa).

For more information, look at Enabling Inference, by R.V. Guha, Orly Lassila, Eric Miller, Dan Brickley, and for a list of available language implementations for RDF try Alberto Reggiori/Andy Seaborne - RDF Query and Rule languages Use Cases and Examples survey and Eric Prud'hommeaux's RDF query survey.

What is the relationship between RDF query and rules? - Numerous answers from a thread on www-rdf-interest@w3.org, September 2001.

Can someone explain if there is "an accepted" query language specification?

Why isn't there a W3C working group or taskforce for RDF query? - rdf/fig irc channel, 2003-02-27

libby: so, I'm working through my rdfqueryfaq, and I get to: Why isn't there a W3C working group or taskforce for RDF query?
libby: anyone know?
libby: not that I want one
danbri: Suggest W3C has not decided to Charter a WG yet. That's why there isn't one.
danbri: Re 'taskforce', this is a vague notion anyway. There is an IG-related mailing list, www-rdf-rules, which effectively services that purpose.

It seems possible that there will be one in the near future: see RDF interest group chatlogs for 2003-02-28, and work from the 2003 W3C All Groups Meeting and Technical Plenary - Semantic Web architecture annotated agenda and logs - Possible query work - Notes from RDF query IROPs (birds of a feather meetings) - rep report.

The W3C has information about the process of creating working groups and interest groups. There is a www-rdf-rules@w3.org list archive, which anyone can join - information about how to do that is at the bottom of that page.

How can I use XML Query for querying RDF?

How does RDF fit in with XQuery? - It is not possible to query generic RDF in XML using XQuery without preprocessing. This is primarily because the same piece of RDF can be expressed in many different syntactic formats in XML, making syntactic XML-based query formats like XSLT and XQuery difficult to use on arbitrary RDF. Attempts have been made to normalize RDF into a different XML syntax which can be queried with these tools, although none of these have been adopted officially by the RDF Core working group. Max Froumentin's XSLT RDF parser tackles similar issues, first normalizing and then parsing the RDF.

For more information, see the XML Europe 2001 paper by Robie at The Syntactic Web. Max Froumentin's XSLT RDF parser is documented here. There was also a thread about RDF and XQuery on www-ig@w3.org mailing list in June 2001.

Could anyone please tell me where to find more information on RDFS Query Language?

How do query languages for RDF handle subclass and subproperty relations?

How do query languages manage RDFs?

Which RDF query languages can handle RDF schema-related queries? - RQL can explicitly be used to retrieve RDFS (RDF Schema) information.

For simpler query languages, whether (for example) all the subclasses of a given class are also returned as well as the immediate subclass will depend on the underlying database. If the database implements RDF Schema (whether by explicitly adding all the subclasses and subproperties whenever a class or property is encountered or whether it uses rules to do this), then simple query languages will retrieve the same information. For simpler databases which do not compute the deductive closure of subclass and subproperty in this way, then the information will not be
Which RDF query languages can handle DAML+oil/OWL properties like inverse, unambiguousProperty, inverseFunctionalProperty?

How do I query DAML data?

What is the relationship between DAML+oil and RDF query?

- DAML+oil is a language defined in RDF where semantics are built on top of the RDF model. Therefore you can query DAML+oil data, just as you can query RDF Schema information using RDF query languages. However you will not get any DAML+oil-specific semantics from RDF query languages, so the result may be awkward to deal with.

There is also a DAML query language, DQL (DQL Semantics, DQL syntax), which is described using RDF and DAML+oil, but is essentially an RDF graph-matching language and does not express any DAML+oil specific query facilities.

Because Jena has a DAML API, the Jena team are often asked questions about the relationship between their RDF query language RDQL and DAML. Some answers are below.

"RDQL only returns Resource/Property/Literal objects - not the higher-level DAML objects that the DAML sub-system of Jena has that get via the DAML API. Also, it does not currently (to be changed) have access to the implications of the DAML inference even though these manifest themselves as virtual triples."

Andy Seaborne, [jena-dev message]

"Once your data/ontology is DAML compliant, you will be able to do RDQL queries on it (actually, you can query without it being DAML compliant - it needs only to be legal RDF but then it might not be saying what you mean it to say :-)).

N.B. RDQL queries RDF as data - you would not get DAML inference as a result of an RDQL query."

Andy Seaborne, [jena-dev message]

"RDQL ("RDF Data Query Language") queries models at the RDF and so it is rather difficult, at the moment, to query at the DAML level with RDQL - you have to know how DAML uses RDF and there is no inference at all.

If you wish to access a DAML model in DAML terms you need to go through the DAML API. If you have a data-oriented requirement then RDQL, on an RDF model, can be used instead of a sequence of Jena core APIs calls."

Andy Seaborne, [jena-dev message]

How does RDF query handle datatypes?

- Datatypes are specified in the RDF Core concepts document and there is some informative explanation in the RDF Semantics document.

These documents are at [last call working draft stage]. "This document is in the Last Call review period, which ends on 21 February 2003. This document has been endorsed by the RDF Core Working Group. This document is being released for review by W3C Members and other interested parties to encourage feedback and comments, especially with regard to how the changes made affect existing implementations and content."

Datatyping has only been introduced comparatively recently, and I have not found any query languages that implement datatypes. This section will be updated when more information is available.

Which query languages for RDF allow you to specify optional variables to bind?

- Optional variables are very useful when the data to be queried is not completely consistent. For example if the data to be queried contains RDF about documents, and the RDF describes titles for some, descriptions for others, creators for some and contributors for others, then without optional variables, to access this information would require different queries for all these different variations. This is the case if all query variables must be matched in order for the query to complete successfully.

If optional variables are allowed, this information can be accessed in one query.

"There are two "forms" that XUL templates may be written in. The "simple" form, which is currently the most common form in the Mozilla codebase, and the "extended" form, which allows for sophisticated pattern matching against the RDF graph."

[XUL (in Mozilla)]

Here is an extended form primer, and the XUL reference. Algae by Eric Prud'hommeaux can also do optional variables. From a private communication about Algae's optional/required syntax:

optional terms:
{ask '}

(requiredP1 ?n1 ?n2)
(requiredP2 ?n2 ?n3)
(optionalP1 ?n3 ?n4)
(optionalP4 ?n4 ?n5)

Can you represent b-nodes [blank nodes] in RDF queries?

- "Blank nodes are treated as simply indicating the existence of a thing, without using, or saying anything about, the name of that thing. (This is not the same as assuming that the blank node indicates an 'unknown' uriref; for example, it does not assume that there is any uriref which refers to the thing. The discussion of skolemization in the proof appendix is relevant to this point.)"

[RDF Semantics, section 1.5]

For more information, see [primer introduction]. Because b-nodes do not have names, it is not usually possible to ask for them by name directly. In Jena for example it is possible to use b-nodes in RDQL queries, but only if you construct the query programmatically via the api, having first retrieved the b-node identifier using another method. See [brief chat with Andy Seaborne on IRC].
Which RDF query languages can do substring matches? regex matching? - partial answer

How do I query reified statements and application-specific schema?

select ?s, ?p, ?o from http://example.com/reification/example.rdf
using rdf for http://www.w3.org/1999/02-22-rdf-syntax-ns#

Jena seems to be a special case - see Jena-dev message from Dave Reynolds.

How can I query and return provenance information in RDF query? - Provenance information is just some information about where the RDF data came from. Many RDF implementations store this kind of information. However provenance information is external to the RDF model, meaning that there is no way of representing for each triple where it came from within RDF. Reification will not do it as reified triples are not asserted. There is a construct in N3 that allows grouping of RDF statements which may be used for provenance information, but this is not part of the RDF Core working group's RDF Model theory.

What this means is that RDF query languages which query only the RDF model (the triples) cannot directly access provenance information, even if the underlying store retains this information. However, if a query language is able to return objects (rather than strings) bound to the variables, then it may be possible to retrieve provenance information from this source.

Alberto Reggiori's Perl implementation of RDQL in RDFStore has an optional fourth argument to each query triple. This fourth argument is the url from which the data to be retrieved originated. This is useful where the database is very large and the original source of the data to be searched for is known. more information about RDF Store and provenance.

Are there any recursive, template-matching, XSLT-like query languages for RDF - Versa, David Allsop's query language ('reachable'), RDF Objects (Alex Barnell).

Why not use an RDF graph with blanks for querying RDF? - It is an example of how you might do this, by Dan Brickley.

People have begun to discuss this as a possible method for describing an RDF query in a neutral way for query test cases. There are various problems, for example:

- although it seems reasonable to use RDF blank nodes to represent variables in a query, predicates cannot be blank in RDF, and so queries with predicate variables cannot be expressed
- it is not clear how to express the distinction between the variables to be bound and the variables which play a part in the query but whose bindings are not required to be returned. This is an important distinction which will govern the time taken for the query to complete and the number of results returned.

Here are various other answers from the www-rdf-rules list.

How do I represent RDF graphs in a relational (RDBMS, SQL) database and query them? - The SWAD-Europe report Mapping Semantic Web Data with RDBMSes describes several schemas used by different database systems for storing RDF in SQL databases. There is also a somewhat older survey by Sergey Melnik.

For RDF queries, an interesting approach is Matt Biddulph's query rewriting. This takes a query described in a simple query language for RDF and rewrites it into the SQL required for retrieval from a one-table store. Matt has a PHP version, and there is also a Java version for two-table SQL schema for RDF, based on Matt's work.

The advantage of this approach is that applications need only make one hit on the database, rather than one for each part of the query.

Alternatively it may be possible to query ordinary relational database tables as RDF.

Is there a programmatic query API? - There is no single RDF query language and therefore no one query API.

Jena does have a programmatic query API for its query language RDQL. See Jena javadoc.

How do I query generic relational (RDBMS, SQL) data using RDF query? - RDF Access to Relational Databases, Eric Prud'hommeaux's new tool for generating specific tables from a generic triple store

Eric Prud'hommeaux is a message about alone and even being able to query a relational database with an application-specific schema.

There are some related threads from www-rdf-rules@w3.org January 2003, www-rdf-interest March 2003, and www-rdf-interest May 2003.

Are there any test suites or test cases for RDF query? - Many implementations have their own test suites, however there is no cross-implementations set of tests available as yet.

There is now an RDF query testcases repository on the W3C site which contains proposals for RDF query manifest and result set formats, and details of ongoing IRC (internet Relay Chat) and face to face meetings. IRC meetings are currently being held regularly on this topic: everyone interested is welcome.

Preliminary discussions on this subject are held on the wwwrules mailing list and summarised on the SWAD-Europe wiki. An IRC meeting was held in February 2003 on the topic of deciding a manifest format for tests (summary of discussion).

What other summaries of RDF query are available? - W3C query languages workshop 1998

A workshop held in 1998 which produced a number of very influential papers on RDF query (in particular Enabling Inference, by R.V. Guha, Ora Lassila, Eric Miller, Dan Brickley).

Eric Prud'hommeaux's RDF query survey


A summary of RDF query language characteristics, and sample queries from various implementations.

"This document is intended to provide an understanding of the concepts and issues related to querying semantic web data. Further it provides a survey of implementations. Web service-related examples come from a strawman WSDL RDF model proposed in another document."

Alberto Reggiori/Andy Seaborne - RDF Query and Rule languages Use Cases and Examples survey
A database of use cases and sample queries that can be added to online, including a very useful use-case based query comparison document, and a schema for use cases.


A RDF Query Birds of a feather meeting (BOF), July 2001

Eep3 Alpha: CWM Clone and SW API - Sean Palmer

Eep3 is a general Semantic Web API written in Python, with various features:
- A basic RDF API including various stores, a query engine, and an NTriples/N3triples parser.
- A CWM clone capable of naming builtins (which can return lists!) in filler/apply/think modes.
- New distributions of n3s (the Notation3 preprocessor) and norm.py (an NTriples normalizer).
- Chumpette: an IRC bot capable of maintaining a weblog using RDF input from IRC.
- A Notation3 parser - this is the only "external" package, since I just modified DanC's Yapps grammar for N3 (there were quite a few bugs in the original, plus this new version is no longer dependent on CWM).
- An rdfdiff program (WIP).

CWM

Cwm is a general-purpose data processor for the semantic web. It is a forward chaining reasoner which can be used for querying, checking, transforming and filtering information. Its core language is RDF, extended to include rules, and it uses RDF/XML or N3 serializations as required. Originally, from "Closed world machine" because it processed information in a limited space, cwm does not make any assumptions about a closed world. Think of it as defined area but with openings - like a valley. Cwm is written in python.

RDF for "Little Languages" Query, Transformation and Report Generation - Graham Klyne

Algae - Eric Prud'hommeaux

Algae is a constraint-based query interface based on algernon.

Versa - Uche Ogbuji
"I am one of the developers of Versa, a query language for RDF. There are many other query languages for RDF, and probably will be at least until the community agrees to standardize. The Versa developers tried most of these and found them unsuitable for various practical reasons. In particular, Versa is designed to be integrated into other programming languages and systems. It is inspired in many ways by XPath, the very successful query language (of sorts) for XML."

**RDQL for Jena - Andy Seaborne**


- **Announcement**
  
  "RDQL is an implementation of an SQL-like query language for RDF. It treats RDF as data and provides query with triple patterns and constraints over a single RDF model. The target usage is for scripting and for experimentation in information modelling languages."   

**SquishQL/Inkling**

- [http://swordfish.rdfweb.org/rdquery/](http://swordfish.rdfweb.org/rdquery/)

- **Announcement**
  
  "This is an RDF query engine, written in Java, which can take SQL-like query strings and which uses the JDBC API."   

**RDFStore - RDQL**

- [download], [demos], [further demos]

  "RDFStore implements the RDQL language to query RDF repositories directly from Perl. The toolkit consists of a Perl API, a streaming SiRPAC parser and a generic hashed data storage custom designed for the RDF model. The storage sub-system allows transparently storage and retrieval of RDF nodes, arcs and labels, either from an in-memory structure, from the local disk or from a very fast and scaleable remote storage. The latter is a fast networked TCP/IP based transactional storage library that uses multiple single key hash based BerkeleyDB files together with an optimized network routing daemon with a single thread/process per database. The data indexing model is general enough to retrieve RDF subgraphs and properties using free-text and statement-group sensible matching. Each literal value gets indexed in its full Unicode form and in-memory data structures or objects can also be serialised on disk. The API supports bNodes (blank Nodes or anonymous-resources) but the storage internally does treat them like any other resource. Being in Perl, an un-typed language, the toolkit at the moment does not treat typed literals in any special way; all query filtering operations on the values are processed using pure Perl regular expressions and eval constructs." ([Three Implementations of SquishQL, a Simple RDF Query Language, 2002](http://lists.w3.org/Archives/Public/www-rdf-interest/2002May/0063.html))

- **Announcement**

**Edutella query languages**


**RDF querying syntax inspired by XSLT**


**RDF db query language**


**RDF query in RDF - Patrick Stickler**

- [http://lists.w3.org/Archives/Public/www-rdf-interest/2002May/0063.html](http://lists.w3.org/Archives/Public/www-rdf-interest/2002May/0063.html)

**RDF QL - RDF Gateway - Geoff Chappell**


**TAP - Guha et al**

- [http://tap.stanford.edu/overview.html](http://tap.stanford.edu/overview.html)

**Joseki - Andy Seaborne**


**RQL and RDQL in Sesame**

- [http://sesame.administrator.nl/](http://sesame.administrator.nl/)

- **Announcement**

**DQL**

- [a proposed syntax], [model]

**SeRQL**

- **Announcement**

**Prolog and RDF rules**
A Prolog engine written in Java (refactored from XProlog) which is loaded with RDF rules. At present the knowledgebase can be entered as Prolog-format triple(\(X, Y, Z\)) or as N-Triples. It has a simple command-line style UI.

More information.