D2R Server: A Semantic Web Front-end to Existing Relational Databases

Richard Cyganiak and Christian Bizer
Freie Universität Berlin
richard@cyganiak.de, chris@bizer.de

RDF and the SPARQL query language and protocol provide a standardized way of exposing and linking data sources on the Web [Dod06]. Just as HTML front-ends to existing databases were a key factor in driving widespread adoption of the World Wide Web, RDF front-ends to these data sources will be important in bootstrapping the Semantic Web. We present D2R Server, a system that takes SPARQL queries from the Web and rewrites them to SQL queries via a D2RQ mapping. This on-the-fly translation allows Web access to the content of large databases without the need for replicating them into RDF.

1   Architecture

We use an evolution of the D2RQ mapping language presented in [BS04] to describe a declarative mapping from database tables and columns to RDF classes and properties. The language supports conditional mappings, mapping of multiple columns to the same property, the handling of highly normalized table structures where instance data is spread over multiple tables, and the usage of translation tables in the mapping process.

To support users in writing mappings, D2R Server includes a tool that automatically generates a default mapping by analyzing the database schema. The mapping can then be customized to use established RDF vocabularies and ontologies. A visual RDF browser supports this process. The browser itself is implemented as a SPARQL protocol client that communicates with the server through AJAX calls.

The following example maps a part of the Wordpress[1] database schema to RDF:

```
map:Posts a d2rq:ClassMap;
    d2rq:uriPattern "http://example.blog/post@@posts.ID@@";
    d2rq:class sioc:Post;
    d2rq:condition "posts.post_status='publish'";
    d2rq:dataStorage map:WordpressDatabase .
map:title a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:Posts;
    d2rq:property dc:title;
```

---

[1] http://wordpress.org/ (a weblog engine)

```
    d2rq:column "posts.post_title" .
map:author a d2rq:PropertyBridge;
    d2rq:belongsToClassMap map:Posts;
    d2rq:property dc:creator;
    d2rq:column "users.display_name";
    d2rq:join "posts.post_author=users.ID" .
```

A ClassMap is used to map each row of the `posts` table to one RDF resource identified by an URI like `http://example.blog/post42`, where 42 is the row's primary key value. Each of these resources is of `rdf:type sioc:Post`. A condition excludes unpublished posts. A PropertyBridge is used to attach a `dc:title` property to each resource, whose value is taken from the `posts.title` column. A second PropertyBridge attaches information retrieved from a different table via a join.

D2R Server is based on the following components: The Joseki server[2] implements the SPARQL [PS06] protocol. The ARQ query engine[3] parses SPARQL queries and breaks them down into simpler parts, which are then rewritten to SQL by the D2RQ engine.

2   Conclusions

D2R Server is a turn-key solution for RDFizing existing applications and for populating the Semantic Web with real-world data. It is available for download under the GPL license at `http://www.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/`.

We are currently creating mappings for a number of popular open-source web applications to explore how different systems can be mapped to the same RDF vocabulary. We are also extending the SPARQL-to-SQL rewriter to push a larger number of SPARQL constructs down into the database. Finally, we are working on incorporating results by other groups in the area of SPARQL query federation.

References

[BS04]   Christian Bizer and Andy Seaborne. D2RQ: Treating Non-RDF Databases as Virtual RDF Graphs. In *3rd International Semantic Web Conference (ISWC2004)*, Nov 2004. http://www.wiwiss.fu-berlin.de/suhl/bizer/pub/Bizer-D2RQ-ISWC2004.pdf.

[Dod06]   Leigh Dodds. SPARQLing Services. In *XTech 2006 Conference*, May 2006. http://xtech06.usefulinc.com/schedule/paper/61.

[PS06]   Eric Prud'hommeaux and Andy Seaborne. *SPARQL Query Language for RDF*, Apr 2006. http://www.w3.org/TR/rdf-sparql-query/.

---

[2]`http://www.joseki.org/`
[3]`http://jena.sourceforge.net/ARQ/`

3 Demo Description

Wordpress is a popular open-source weblogging engine backed by a database. In our demo, we will show how the metadata of a Wordpress blog – posts, comments, users, categories – can be made available as a SPARQL endpoint using established RDF vocabularies such as Dublin Core, FOAF and SKOS:

- A default mapping for the database is auto-generated,

- the RDF browser, shown in figure 1, is used to explore the database's contents,

- the auto-generated mapping is customized to remove unwanted parts of the data, use proper URIs, and use established vocabulary terms.

Having set up the endpoint, we can retrieve specific information from the database with complex SPARQL queries and show how these are rewritten to SQL.

We will also demonstrate a simple AJAX-based data integration "mashup" that lets users query multiple SPARQL data sources from a single interface.
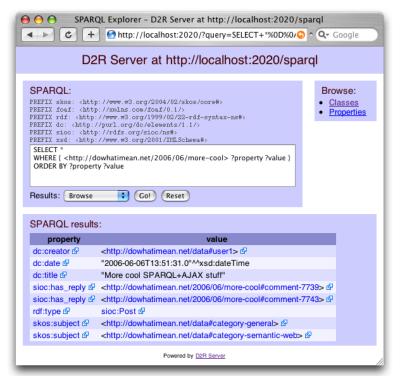

Figure 1: A blog entry, mapped from a Wordpress database, as seen in D2R's integrated RDF browser. The mapping has been customized to use popular RDF vocabularies like Dublin Core, FOAF and SKOS.