D2R Server – Publishing Relational Databases on the Semantic Web

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

Abstract

D2R Server is a tool for publishing the content of relational databases on the Semantic Web. Database content is mapped to RDF by a declarative mapping which specifies how resources are identified and how property values are generated from database content. Based on this mapping, D2R Server allows Web agents to retrieve RDF and XHTML representations of resources and to query non-RDF databases using the SPARQL query language over the SPARQL protocol. The generated representations are richly interlinked on RDF and XHTML level in order to enable browsers and crawlers to navigate database content.

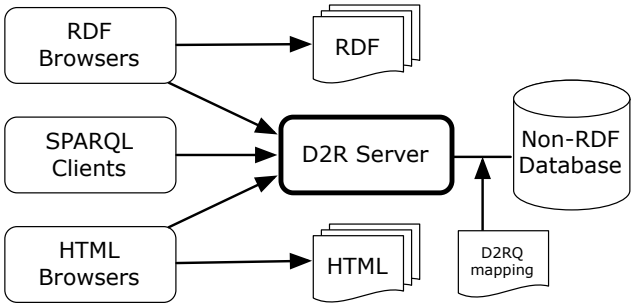
1 Introduction

The W3C recommendation *Architecture of the World Wide Web, Volume One* [Jacobs and Walsh, 2004] specifies the principles of the Web: Items of interest are called resources and are identified by URIs. Web agents may retrieve representations of resources by dereferencing URIs. The data format of a representation is determined by content negotiation relying on Internet media types. The main access paradigms to the Web are hyperlink navigation and search.

In this demonstration, we present an approach to publishing the content of relational databases on the Web which focuses on compliance with these principles. We introduce D2R Server, a system for publishing relational data on the Web. D2R Server enables RDF and HTML browsers to navigate the content of non-RDF databases, and allows applications to query a database using the SPARQL query language over the SPARQL protocol. The server takes requests from the Web and rewrites them to SQL queries. This on-the-fly translation allows the content of large databases to be accessed with acceptable response times. In the following, we describe how D2R Server handles the mapping from relational data to RDF, URI allocation, URI dereferencing, hyperlinking and search.

2 Mapping Relational Data to RDF

D2R Server uses the D2RQ mapping language [Bizer and Seaborne, 2004] to capture mappings between application-specific database schemas and RDFS schemas or OWL ontologies. A D2RQ mapping specifies how resources are identified and and how property values are generated from database content. The central object in D2RQ is the *ClassMap*. A ClassMap represents a mapping from a set of entities described within the database, to a class or a group of similar classes of resources. Each ClassMap has a set of *PropertyBridges*, which specify how resource descriptions are created. Property values can be created directly from database values or by employing patterns or translation tables. D2RQ supports conditional mappings on ClassMap and PropertyBridge level, the mapping of n:m relations, and the handling of highly normalized table structures where entity descriptions are spread over several tables.



D2R Server includes a tool that automatically generates a D2RQ mapping from the table structure of a database. The tool generates a new RDF vocabulary for each database, using table names as class names and column names as property names. The mapping can be customized afterwards by substituting auto-generated terms with terms from well-known RDF vocabularies.

3 URI Allocation

In ClassMaps, database entities are assigned URIs using URI patterns. For example, the pattern "`products/product@@Products.ID@@`" produces a relative URI like `products/product1134` by inserting a value from the `Products.ID` database column into the pattern.

D2R Server turns relative URIs into absolute URIs by expanding them with the server's base URI. This is the preferred

URI allocation mechanism, as it ensures that identifiers are within a URI space owned by the server operator. It also enables the server to answer HTTP requests about these URIs, making them dereferenceable.

If a database already contains URIs for identifying database content, for example in a table describing web documents, then these *external URIs* can be used instead of pattern-generated URIs.

4 Dereferencing URIs

D2R Server enables Web agents to retrieve RDF and XHTML representations of resources by dereferencing pattern-generated URIs. The data format to be sent is determined by content negotiation.

A RDF representation of a resource is retrieved by dereferencing the resource URI with a HTTP request that asks for content type *application/rdf+xml*. A XHTML representation of the resource is retrieved by dereferencing the same URI with a HTTP request that asks for content type *text/html* or *application/xhtml+xml*.

XHTML representations are currently a fairly simple human-readable rendering of the RDF representations. They are rendered using Velocity templates in order to allow customization. Future version of D2R Server might employ Fresnel lenses to improve resource display.

According to [TAG, 2005], only *information resources* (i.e. documents) can have representations served on the Web over HTTP. When URIs that identify other kinds of resources, such as a person, are dereferenced, then the HTTP response must be a 303 redirect to a second URI. At that location, a document describing the real-world resource is served. D2R Server implements this behaviour.

5 Hyperlinking

The classic navigation paradigm on the Web is following hyperlinks. D2R server supports hyperlink navigation by providing links on RDF and XHTML level.

Any RDF triple whose object is a dereferenceable URI can be seen as a hyperlink [Berners-Lee, 2006]. This is how resources published by D2R Server are interlinked with other databases and external RDF documents.

To aid discovery of related resources, D2R Server includes an `rdfs:seeAlso` triple with every resource description that points to an RDF document containing links to other resources produced by the same ClassMap. If resources are identified with external URIs, then an additional `rdfs:seeAlso` link points to a local RDF/XML document that contains everything the database knows about the resource. By dereferencing the external URI and by following the `rdf:seeAlso` link, RDF browsers can retrieve both authoritative and non-authoritative information about the resource.

RDF-level hyperlinks serve as "breadcrumbs" for RDF crawlers and RDF browsers such as Tabulator [Tim Berners-Lee et al., 2006] which allows a user to interactively explore the Web of interlinked RDF documents.

All RDF-level hyperlinks are also available in XHTML representations. Additional XHTML hyperlinks lead to navigation pages containing lists of other resources produced by the same ClassMap, and to an overview page that lists all of these navigation pages. This overview page provides an entry point for crawlers of external Web search engines to index the content of the database.

6 Search

D2R Server allows applications to query non-RDF databases using the SPARQL query language over the SPARQL protocol. Queries are executed against a virtual RDF graph representing the complete database. Query results can be retrieved in the SPARQL Query Result XML Format and the SPARQL/JSON serialization.

7 Conclusions

Most structured data is stored in relational databases today and, in spite of progress in the area of RDF and XML storage, will keep on being maintained primarily in relational databases in the mid-future. Therefore, we believe that providing Web access to existing relational databases is crucial for populating the Semantic Web with relevant real-world data.

D2R Server is available under GNU GPL. More information about D2R Server is found on the D2R Server website `http://www.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/`.

8 Acknowledgments

This work is part of the Knowledge Nets project within the InterVal-Berlin Research Centre for the Internet Economy and is funded by the German Ministry of Research BMBF.

References

[Berners-Lee, 2006] Tim Berners-Lee. Linked data, 2006. `http://www.w3.org/DesignIssues/LinkedData.html`.

[Bizer and Seaborne, 2004] Christian Bizer and Andy Seaborne. D2rq: Treating non-rdf databases as virtual rdf graphs. In *3rd International Semantic Web Conference (ISWC2004)*, 2004. `http://www.wiwiss.fu-berlin.de/suhl/bizer/pub/Bizer-D2RQ-ISWC2004.pdf`.

[Jacobs and Walsh, 2004] Ian Jacobs and Norman Walsh. *Architecture of the World Wide Web, Volume One*, 2004. `http://www.w3.org/TR/webarch/`.

[TAG, 2005] W3C Technical Architecture Group TAG. *httpRange-14: What is the range of the HTTP dereference function?*, 2005. `http://www.w3.org/2001/tag/issues.html#httpRange-14`.

[Tim Berners-Lee et al., 2006] Tim Berners-Lee et al. Tabulator: Exploring and analyzing linked data on the semantic web. In *Proceedings of the 3rd International Semantic Web User Interaction Workshop*, 2006. `http://swui.semanticweb.org/swui06/papers/Berners-Lee/Berners-Lee.pdf`.