10. R2RML Mapping with D2RQ

The purpose of this lab is to let you learn how to publish RDF data from relational databases using R2RML with the D2RQ system.

R2RML and D2RQ

W3C R2RML[1] is a language for specifying customized mappings from relational databases to RDF datasets. Such mappings provide the ability to view existing relational data in the RDF data model, expressed in a structure and target vocabulary of the mapping author's choice.

D2RQ[2] is a system for accessing relational databases as RDF graphs using R2RML mappings. The access is provided by creating custom dumps of the database in RDF format for loading into an RDF store based on an R2RML mapping file.

R2RML Mapping over a Simple Football Database

Consider the Football database that consists of the following two tables.

id [PK] integer	name character varying	team_id integer
1	Lionel Messi	1
2	Cristiano Ronaldo	2

Figure 1: `player` Table

team_id [PK] integer	name character varying	city character varying
1	FC Barcelona	Barcelona
2	Real Madrid C.F.	Madrid

Figure 2: `team` Table

The `team_id` column of the `player` table is a foreign key pointing to the `team_id` column of the `team` table. An SQL file with commands to create that database is provided at: `http://bit.ly/13IgL8u`.

---

[1] `http://www.w3.org/TR/r2rml/`
[2] `http://d2rq.org/`

Task 1: Use R2RML mappings with D2RQ to transform the Football database into RDF.

The RDF output should look as follows: `http://bit.ly/1xf2w71`. The necessary steps to perform the mapping are as follows:

1. Download a version of D2RQ that supports R2RML mapping from `http://bit.ly/1AgYe0d`. There are two tools of D2RQ that you should use: `generate-mapping` and `dump-rdf`.

2. As in Lab 9, you have to set up a DBMS to be populated with databases. Load the Football database into your DBMS.

3. The `generate-mapping` tool allows you to auto-generate an R2RML mapping from a database. Use the tool to generate a skeleton R2RML mapping for the Football database. Execute the following command to run the tool:

   ```
   generate-mapping --r2rml -d $driver-class-name$ -u $db-user$
   -p $db-password$ $jdbc:url-to-database$ > $generated-mapping$
   ```

   The parameters are defined as follows:

   - `--r2rml`

     Generate a mapping file that is compatible with W3C R2RML.

   - `-d $driver-class-name$`

     The fully qualified Java class name of the database driver.

   - `-u $db-user$`

     The login name of the database user.

   - `-p $db-password$`

     The password of the database user.

   - `$jdbc:url-to-database$`

     JDBC connection URL for the database.

   - `$generated-mapping$`

     The mapping file that is generated.

4. Modify the generated R2RML mapping to output the required RDF data.

5. Dump the contents of the Football database into a single RDF file, using your modified mapping.

   The `dump-rdf` tool dumps the contents of the whole database into a single RDF file. This can be done with or without a mapping file. If a mapping file is specified, then the tool will use it to translate the database contents to RDF. Execute the following command to run the tool:

   ```
   dump-rdf -u $db-user$ -p $db-password$ -f $rdf-format$
   -j $jdbc:url-to-database$ $generated-mapping$ > $generated-rdf-dump$
   ```

   Some parameters are defined similarly to those of the `generate-mapping` tool, while some others are new:

- `-f $rdf-format$`

  The format of the RDF dump. Supported syntaxes are TURTLE, RDF/XML, RDF/XML-ABBREV, N3, and N-TRIPLE (the default).

- `$generated-rdf-dump$`

  The RDF dump file that is generated.

R2RML Mapping over a Books Database

At `http://bit.ly/1wzAT3H` you find a database about books that has the following schema.
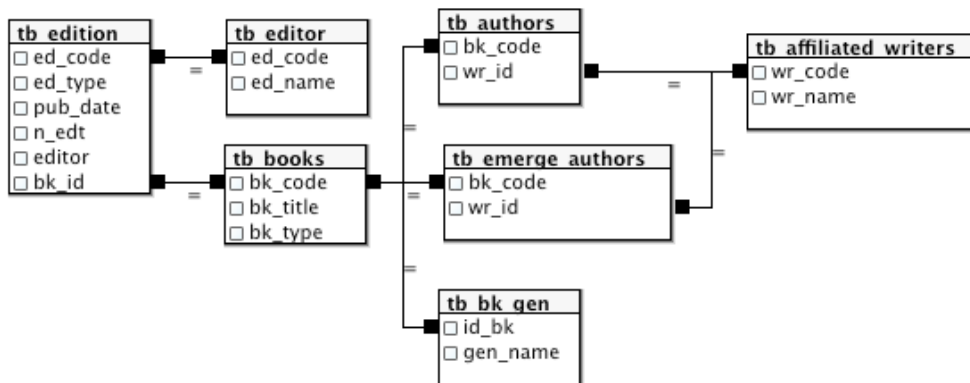

Figure 3: Books Schema

Task 2: Transform the Books database into RDF using similar steps as for Task 1. The generated RDF data should look like the one at `http://bit.ly/1BhkSnx`.

Task 3: Express and execute the following queries in SQL and SPARQL:[3]

1. How many writers are there?

2. Which books were authored by Kate Pearce?

3. What is the genre of the book titled Amberville?

4. Which books have the genre Horror?

5. Which books were published after 2009?

By way of comparison, discuss how difficult it is to formulate these queries

- in SQL

- in SPARQL over the data set generated with Direct Mapping in Lab 9

- in SPARQL over the data set generated with your R2RML mappings?

---

[3]Load your RDF data into Jena Fuseki first to make it queryable (as you learned in Lab 2).