Different RDB to RDF mapping languages

Priyanka Shukla
Department of Computer
Science and Engineering
RITM
Lucknow, India

Vaibhav Singh
Department of Computer
Science and Engineering
RITM
Lucknow, India

Akanksha Shukla
Department of Computer
Science and Engineering
RITM
Lucknow, India

Abstract: This document deals with the different techniques, mapping languages ,tools ,applications used for mapping Relational Databastes and Resource Description Framework.This document will serve as a guide for selecting a particular language for mapping .For the development of semantic web we need to map Relational Database to Resource Description Framework.Since most of the data on web is stored on RelationalDatabase and a conceptual gap is to be bridged between the Relational Database model and RDF to make this data available on web semantic.Many mapping languages and approaches have been found leading to the ongoing standardization of the World WideWeb Consortium(W3C) carried out in the RDB2RDF Working Group(WG).This paper would provide help and recommendations for selecting a mapping language.
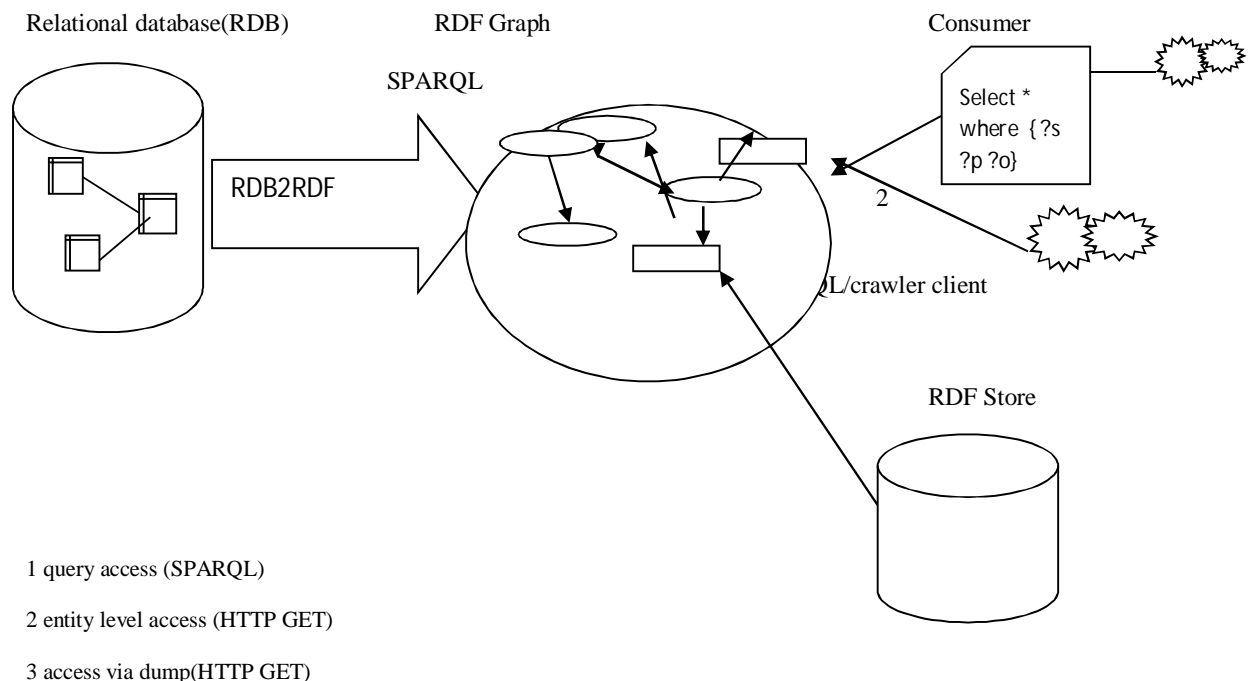
1.   INTRODUCTION

Mostly all the websites are backed by Relational databases.Most information is still modeled and are stored in Relational databases and hence out of reach for many web semantic applications. The success of semantic web is dependent on the translation of RDB to RDF and this is done throughdirect mapping.Direct mapping is a simple translation of RDB to RDF. As a consequence, such applications need to create a corresponding mapping between the relational and the semantic models for being able to access relational data. In this paper we study why we need these

mapping languages for making data available on web.Since we study different mapping languages we have to makesome comparative study for when and why to use a particular mapping language[9].This paper would describe the problems that occur with different mapping languages .So we must have certain classifications and categories which describes when to use which language.In this paper we also study a number of reusable mappings, which we define as RDB2RDF Mapping Patterns.

Mapping RDB to RDF is an active field of research . Many approaches were explored to make relational data available to Semantic Web-enabled applications.   These approaches introduced mapping languages that range from simple and pragmatic to highly specific or general-purpose.

Relational database(RDB)   RDF Graph   Consumer

1 query access (SPARQL)

2 entity level access (HTTP GET)

3 access via dump(HTTP GET)

2.   RELATED WORKS:

Satya S. Sahoo et al. has discussed different methods to generate mappings between RDB and RDF which are basically of two categories:Automatic Mapping Generation: This discusses a set of mappings between RDB and RDF namely:*A RDB record is a RDF node.The column name of a RDB table is predicate.*RDB cell is a value.An example of this approach is Virtuoso RDF View that uses the unique identifier of a record (primary key) as the RDF object, the column of a table as RDF predicate and the column value as the RDF subject. Other examples of similar tools are D2RQ and SquirrelRDF .[1] This approach also allows Semantic Web applications to query those RDB sources where the application semantics is defined in terms of the RDB schema. This approach is also called "Local ontology mapping". Domain Semantics-driven Mapping Generation: This approach incorporates domain semantics these are not captured in RDB schema .[10] Also, a mapping generated by using domain semantics also reduces the creation of triples for redundant or irrelevant knowledge. Mapping between RDB and RDF is represented by Xpath in XSLT stylesheet in a XML based declarative language.Two types of mapping implementations are–static andStatic ETL ,dynamic –query driven. ETL uses batch process to create RDF repository. Queries in systems mapping RDB to RDF may either be in SPARQL .SPARQL may be transferred into one or more sql query that are executed against RDB .

➢   Matthias Hert et al. has discussed a feature-based comparison of the state-of-the-art RDB-to- RDF mapping languages.This comparison framework is based on use cases and requirements for mapping RDBs to RDF.In this paper we apply this comparison framework and four main categories of mapping languages have been propsed .These are Direct mapping, Read-only general-purpose mapping, Read write general-purpose mapping, and Special-purpose mapping. In direct mapping, a direct approach for mapping RDBs to the Semantic Web is proposed[2]. It maps relational tables to classes in an RDF vocabulary and the attributes of the tables to properties in the vocabulary.The goal is to expose a RDB on the (Semantic) Web to make extra statements about it. The goal of R2RMLis to define a vendor-independent mapping language for read-only data access. R3M enables bidirectional RDF-based access to the RDB, *i.e.,* read and *write* access is supported. It employs a RDF-based syntax that contains the mappings of tables to classes and attributes to properties as well as information about integrity constraints.This paper provides guidelines for a RDB-to-RDF mapping language for a given applicationscenario and its requirements.

➢   Juan F. Sequeda et al. has discussed about the problems of directly mapping a Relational database to an RDF graph with OWL vocabulary .This paper shows that direct mapping is an automatic way of translating a relational database to RDF.This paper discusses that there are basically two fundamental properties of Direct Mapping :information preservation and query preservation. A direct mapping is information preserving if none of the   information is lost about the relational instance being translated, that is,there exists the ways through which original database instances may be recovered from the RDF graph resulting from the translation process[3]. A direct mapping is query preserving if every query over a relational database can be translated into an equivalent query over the RDF graph resulting from the mapping.It assures that every relational query can be evaluated using the mapped RDF data. To formally define

query preservation, we focus on relational queries that can be expressed in relational algebra and RDF queries that can be expressed in SPARQL .Additionally desirable properties are:monotonicity and semantics preservation.Monotonicity is desired to avoid recomputation of the entire mapping after updating databases .In general and practical scenario direct mapping is information preserving,monotone and query preserving only when relational databases contain null values.But unfortunately we found that no monotone direct mapping is semantic   preserving if foreign keys are considered.

CristianPérez de Laborda et al. in this paper it was discussed that main drawback of semantic web is the lack of semantically rich data,so an approach was presented to map legacy data stored in relational databases into the Semantic Web using virtually any modern RDF query language.It was suggested in this paper that web developer need not to learn and adopt a new mapping language, but he may perform the mapping task using his preferred RDF query language.In this paper a technique called Relational OWL was introduced that automatically transform relational data into representatable form. It converts the schema of a database automatically into an ontology and the data items as its instances, i.e. the data is described as it was in the database.It is a reasonable and acceptable technique because legacy data stored in relational database can be easily accessed by their built-in functionalities[4]. To perform such a mapping task, a Semantic Web developer does not need to learn and adopt a new mapping language, but he may perform the mapping task using his preferred RDF query language. For this purpose, data and schema components of the original relational database are first translated automatically into their Semantic Web representation based on Relational OWL. Then they may either be processed or mapped directly to a target ontology.Using virtual RDF query language results into RDF graphs as query results.

Juan F. Sequeda et al. in this paper has discussed that as we know for semantic web applications we need to map relational database to RDF .Since the W3C RDB2RDF presented two standards to map relational database to RDF .They are : Direct Mapping and R2RML mapping language. Direct Mapping is the default way of representing a relational database as RDF based on the structure of the database schema. R2RML is a language for expressing customized mappings from relational databases to RDF.Inthis particular paper different mappings have been compiled to present a non-exhaustive list of RDB2RDF   mapping patterns.These mappings were represented in R2RML[5] . We present four type of mapping patterns: Attribute Mapping Patterns, Table Mapping Patterns, Join Mapping Patterns and Value Translation Patterns[8]. Each pattern consists of a name, a question that defines the problem that is being addressed, description of the context, description of the solution in R2RML, an example R2RML mapping, a discussion and related patterns.In this paper fourteen mapping patterns have been presented .

Research paper	Technology	Language	Advantage	Disadvantage
SatyaS.Sahoo et al.	Automatic Mapping and Domain Semantics generation	SPARQL	Reduced tripples	-
Matthias Hert et al.		R2RML,R3M	Read and write access is supported	
Juan F. Sequeda et al	RDF graph with OWL vocabulary	RDF query,SPARQL	Semantic preservation and query preservation	Lacks monotone direct mapping if foreign keys considered
CristianP´erez de Laborda et al	Relational OWL	RDF query language	Legacy data stored in relational database can be easily accessed	
Juan F. Sequeda et al	Mapping Patterns	R2RML	-	Increased attributes results in increased query size

Table I: Comparison table for different approaches of RDB to RDF mapping language .
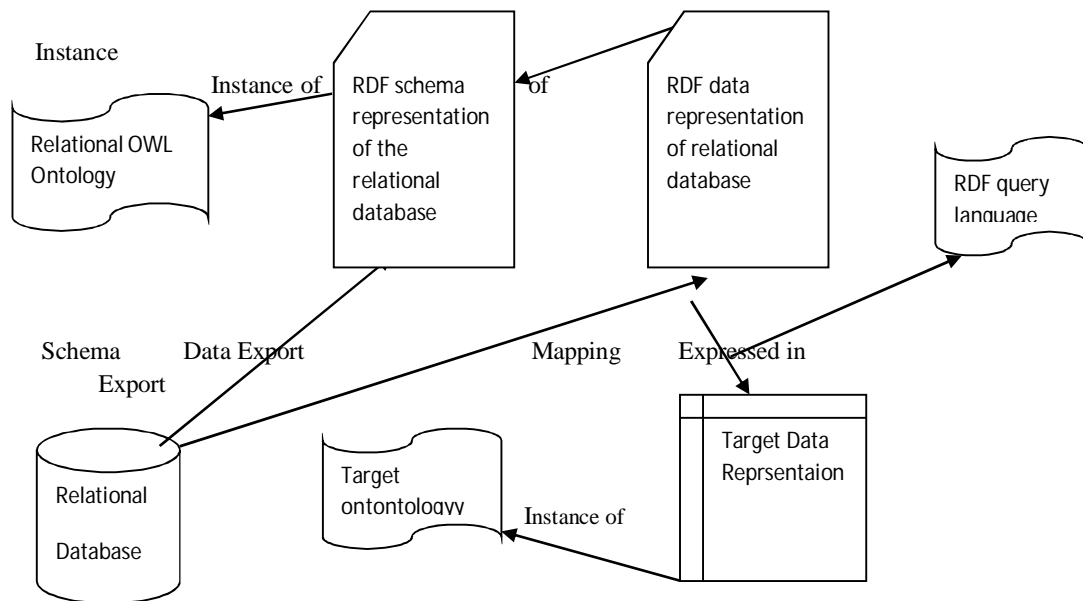

Figure 2.   Mapping Process

Figure .2 shows the complete relational database to RDF mapping process. Firstly the relational OWL representation of the data and data components of the original data source are generated. The schema representation is then converted into an instance of Relational OWL ontology.AS the relational OWL representation of the relational database is done the second step of actual mapping is performed. RDF query language are used to make queries for RDF model.

3. DISCUSSION

In this paper we studied different approaches for mapping RDB to RDF.Different approaches we studied have some advantages and disadvantages like using Direct Mapping could not be semantically preserved if the foreign keys are considered[8].On the other hand mapping using domain semantics reduces the creation of triples for irrelevant knowledge.Using the concept of Triplify results in a boost of semantic web applications.Triplify mappings are implemented as PHP scripts.There exist difficulties in querying RDF graphs by using   RDF query languages.All the queries must be expressed as if they are real XML documents not RDFgraphs Graphs[7].To overcome with such problems SQL syntax based languages were used in order to be easily understood and adoptable.But again such languages like RDQL have some drawbacks like the result of certain queries are not valid RDF triples.Thus to overcome with such difficulties we use different language SPARQL as representative of RDF query language .Use of different mapping patterns impacts upon the query performance.Sometimes if we increase the amount of attributes to be mapped, the size of SQL query would increase.

4. CONLCUSION

So,in this document we presented so many techniques or approaches for mapping RDB to RDF.These approaches are suggested and adopted to understand its mapping simplicity and implementing the language. These mapping languages are highly expressive.But this expressiveness sometimes results into increased complexity.So,different types of mapping languages are recommended according to the application scenarios.On the other side if we use RDF graph with OWL vocabulary for mapping RDB to RDF there exists certain problems.Combination of monotonicity with OWL vocabulary creates a problem in generating a semantically preserved direct mapping.This problem is solved by using non-monotone direct mapping.Most of the join and projection operations are not directly processed by RDQ query so they are passed to the underlying database with generated SQL query.To overcome the limitations of mapping patterns and solve performance issues we come up with new mapping patterns in areas such as Named Graphs, Blank Nodes for anonymous or sensitive data, Metadata, Languages, Data types[6].Finally we found that this is an area of research which needs to be focused and further researchers must be involved in working for the evolution of new mapping approaches to present web semantic.

5. REFERENCES

[1] BSatya S. Sahoo, Wolfgang Halb, Sebastian Hellmann, Kingsley Idehen, Ted ThibodeauJr, Sören Auer, Juan Sequeda, Ahmed Ezzat, Business Intelligence Software Division, HP"Survey of Current Approaches for Mapping of Relational Databases to RD". In W3C RDB2RDF Incubator Group January 08 2009.

[2] ChomsHert, Matthias; Reif, Gerald; Gall, Harald (2011).A Comparison of RDB-to-RDF Mapping Languages.In:Proceedings of the 7th International Conference on Semantic Systems (I-Semantics), Graz, Austria, September 2011.

[3] Juan F. Sequeda,MarceloArenas,Daniel P. Miranker "On Directly Mapping Relational Databases To RDF And OWL" in International World Wide Web Conference Committee (IW3C2) 2012.

[4] CristianP´erez de Laborda and Stefan Conrad "Database to Semantic Web Mapping using RDF Query Languages" in 25th International Conference on Conceptual Modeling, Tucson, Arizona, November 2006, Springer Verlag

[5] Juan Sequeda1, Freddy Priyatna, and Boris Villaz_on-Terrazas ," Relational Database to RDF Mapping Patterns" in W3C RDB2RDF Working Group(2012)

[6] RDB2RDF by "RDB2RDF Working Group" on 21-09-2012.

[7] Edgard Marx1,Percy Salas1,Karin Breitman1, José Viterbo2,andMarco Antonio Casanova1"RDB2RDF:A relational to RDF plug-in for Eclipse" in Wiley Online Library,3 july 2012.

[8] Nuno Lopes "An overview of RDB2RDF techniques and tools"in DERI conference,August 2009.

[9] Juan F. Sequeda,Daniel P. Miranker "SPARQL Execution as Fast as SQL Execution on Relational Data",University Of Texas,Austin

[10] Kate Byrne "Relational database to RDF Translation in the Clutural Heritage Domain"in School of Informatics,University of Edinburg.