International Journal of Engineering & Technology, 7 (2.33) (2018) 119-123

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



R2RS: schema-based relational databases mapping to linked datasets

Ju Ri Kim [1] *, Sung-Kook Han [2]

[1] College of Liberal Arts, WonKwang University, City Iksan, JeonBuk, 54538, Korea
[2] Department of Computer Engineering, WonKwang University, City Iksan, JeonBuk,54538, Korea
**Corresponding author E-mail: 1cyanic@wku.ac.kr, 2skhan@wku.ac.kr*

Abstract

Background/Objectives: The vast amounts of high-quality data stored in relational databases (RDB) is the primary resources for Linked Open Data (LOD) datasets. This paper proposes a schema-based mapping approach from RDB to RDF, which provides succinct and efficient mapping.
Methods/Statistical analysis: The various approaches, languages and tools for mapping RDB to LOD have been proposed in the recent years. This paper surveys and analyzes classic mapping approach and language such as Direct Mapping and R2RML. The mapping approaches can be categorized by means of their data modeling. After analyzing the conventional RDB-RDF mapping methods, this paper proposes a new mapping method and discusses its typical features and applications.
Findings: There are two types of mapping approaches for the translation of RDB to RDF: instance-based and schema-based mapping approaches. The instance-based mapping approaches generate large amounts of RDF graphs by means of mapping rules. These approaches causes data redundancy since the same data is stored in two ways of RDB and RDF. It is very easy to bring the data inconsistence problem when data update operations occur. The schema-based mapping approaches can effectively avoid data redundancy since the mapping can be accomplished in the conceptual schema level.
The architecture of SPARQL endpoint based on schema mapping approach consists of five phases:

- Generation of mapping description based on mapping rules.
- SPARQL query statements for RDF graph patterns.
- Translation of SPARQL query into SQL query.
- Execution of SQL query in RDB.
- Interpretation of SQL query result into JSON-LD format.

Experiments show the schema-based mapping approach is a straightforward, succinct and efficient mapping method for RDB2RDF.
Improvements/Applications: This paper proposes a schema-based mapping approach called R2RS, which shows better performance than the conventional mapping methods. In addition, R2RS also provides the efficient implementation of SPARQL endpoint in RDB.

Keywords: *RDF; Linked Data; Direct Mapping; RDB-to-RDF; SPARQL; SQL.*

1. Introduction

The Linked Data (LD) has emerged as a powerful enabler to extend the current Web of documents to a Web of interlinked data and, ultimately, into the Semantic Web. In the last decade the numerous best practices of Linked Data, including DBpedia, have been published by an increasing number of researchers, governments, public organizations and data providers, creating a global data space that interlinks billions of assertions: the Web of Linked Data 1. Therefore, the Linked Data has evolved from a practical research idea into a very promising technology that can realize the Web as a platform for an intelligent information system with semantic search and query capability. If the Linked Data paradigm in the Web is accelerated, the publication of machine-readable, open and linked datasets should take precedence [1-2].
Over years vast amounts of high-quality data are stored in relational databases (RDB), which are tremendous resources for populating Linked Data. Therefore, how to translate RDB into RDF (Resource Description Framework) that is standard model for Linked Data has attracted much attention. In addition, using RDF as basic format to represent relational data can achieve the data integration from heterogeneous DBMS, which has been a powerful and promising method for publication of Linked Data. The various RDB2RDF mapping methods play a key role for population Linked Data. In recent years, various mapping methods [3-6], mapping language [7-9] and tools [10] have been proposed. This paper proposes a schema-based mapping technique based on the analysis of current mapping techniques, which can realize the seamless mapping for RDB-to-RDF. This proposed schema-based mapping method from Relational Databases to Linked Datasets, called R2RS, has significant advantages contrast to other mapping methods, with the features of straightforward, flexible and efficient.
The remainder of this paper is structured as follows. Section 2 provides a survey and discussion of existing mapping methods and languages. Section 3 describes the R2RS mapping method in detail, including motivation, theoretical basis of mapping, mapping rules, specific mapping illustrations and architecture of

Copyright © 2018 Ju Ri Kim, Sung-Kook Han. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

SPARQL query endpoint. Section 4 summarizes the advantage of schema-based mapping and features of R2RS mapping.

2. Related work

At present, RDB2RDF for the publication of RDF data on the Web and the integration of data from different RDBs has been a crucial research topic for Semantic Web and its enabler LOD. A myriad of approaches, techniques, mapping languages and corresponding tools for RDB2RDF have been proposed over the last decade [3-10]. The motivations, underlying principles, specifications, capabilities, and categorizations of RDB2RDF can be referred to these comprehensive surveys of proposed approaches [11-13].

Direct Mapping and R2RML (RDB to RDF Mapping Language) are typical mapping method and language, which are recommended by the W3C RDB2RDF Working Group. The core ideal of Direct Mapping is like that: subject forms from the concatenation of the base IRI/table name/primary key column name/primary key value; predicate forms from the concatenation of the base IRI/table name/column name; object forms from the column value. Direct mapping method is instance level, which means databases would generate large amount of RDF graphs. Each row of database table produces a group of triples with a common subject, and a simple table produces Row X (Column-1) RDF graphs. R2RML is companion of Direct Mapping method, which provides a language for describing customized mappings from RDB TO RDF. R2RML makes use of Logical Tables to realize the mapping description. The logical tables, being SQL query result, is to be mapped to RDF triples4.

The typical feature of Direct Mapping method is based on instances, and this method has obvious drawback which generates overfull RDF graphs for a Relational Database. All the data will be stored in two ways: one is relational database data; the other is RDF graph data, which is very easy to cause data inconsistency when data update operations are performed. In addition, R2RML, being mapping language, does not handle some of common problems that occur when translating RDB into RDF. For example, it does not use the existing common vocabularies. In addition, the R2RML language is complicated, but function is limited. Therefore, many improved mapping methods and languages have been proposed in the recent years. This paper proposes a new mapping method: R2RS, which is a schema-based mapping method and overcome the drawbacks of Direct Mapping method, and provide the data description format with JSON-LD.

3. R2RS: schema-based mapping of RDB to RDF

This section describes the concepts and techniques of schema-based mapping for RDB2RDF. We elaborate the details of the proposed schema-based mapping method called R2RS, and explain the motivation, theoretical basis, mapping rules, the architecture of SPARQL query. Some typical cases are used to show the mapping method and translation of SPARQL query

3.1. Motivation of R2RS

The goal of this paper is to propose an efficient mapping method that can translate RDB datasets into Linked Datasets, which can promote the population of Linked Data. As mentioned in the previous section, current many RDB2RDF mapping methods have significant drawbacks. This paper proposes R2RS mapping method that is expected to achieve the following objectives

- Separation of schema level and instance level to achieve the efficient mapping.
- Can accommodate mapping requirements from RDB to RDF.
- Can map 1:1, 1: M, M: N relationships and handle primary key and foreign key issues.
- Can provide SPARQL endpoint to query data in the relational database.

In order to achieve the above objectives, the R2RS mapping method need to provide a RDB2RDF transformation engine that can be able to process SPARQL queries and transform whole relational data into an RDF representation. The transformation engine can also validate and show the results of this approach.

3.2. Theoretical basis of R2RS

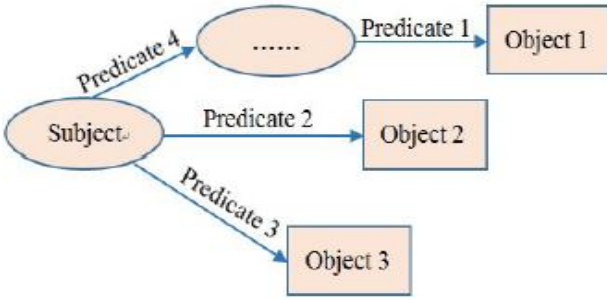


Fig. 1:. Entity Relation Model in RDB.

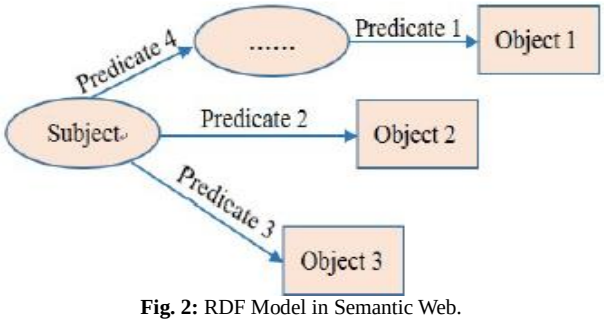


Fig. 2: RDF Model in Semantic Web.

The conceptual schema of relational is usually represented with entity-relationship diagram (ERD) as shown in Figure 1, which is almost identical to RDF data model as shown in Figure 2. Both of them are the models that describe the class, property and property value. Accordingly, it is reasonable that RDB2RDF mapping approach should be based on RDB schema of ERD. This provides a consistent way of mapping and makes it possible to preserve information and conceptual structures of RDB in the process of the mapping.

Since RDF data with triple structure (subject, predicate and object) are generated by means of the classes and properties of RDF Schema, RDB schema defined by ERD should be the starting point of RDB2RDF. This schema-based mapping provides the notable feature of separation between schema and instance. Focusing on conceptual schema, the mapping method can realize seamless translation of RDB to RDF at schema level. The mapping description also becomes more succinct, since it need not specify the mapping definition for each instance. Above all, schema-based mapping is expected to realize the requirements for RDB2RDF.

3.3. Mapping rules of R2RS

The schema-based mapping is applied to the primitive tables of RDB that are usually defined as an entity in ERD. The core mapping ideal of schema-based R2RS is straightforward as follows:

- Map table name into subject: the table name in RDB as a class corresponds to the subject of RDF model in semantic web.
- Map column name into predicate: the column name in RDB as an attribute corresponds to the predicate of RDF model in semantic web.
- Map cell value into object: the cell value in RDB corresponds to the object of RDF model.

In schema-based mapping, the primitive table schema is preserved and represented in RDF graph as shown in Figure 3. The mapped RDF graph shows the conceptual schema inherent in the table, not

RDF data graph of the instance triple. Note that the mapped object value is represented by the cell value term, 'table.column'. 'table.column' is used to represent the cell value. The instance data will be generated according to RDF graph when SPARQL query is executed against the RDB.

In schema-based mapping method, a RDB table as a resource is mapped into a subject with namespace. The table columns Ci are mapped into the predicates of RDF data that can be translated into the well-known ontology vocabularies, such as FOAF, DC, and metadata vocabularies in schema.org. The schema-based mapping can realize semantic interoperability of the mapped RDF data. The object value of the predicate is described with the cell value term, Table Name. Ci. The cell value term is used to translate SPARQL query into the equivalent SQL query. Noted that the mapped RDF graph is also schema-based, and it doen not generate triple instances and will provide RDF query pattern view for SPARQL query endpoint.

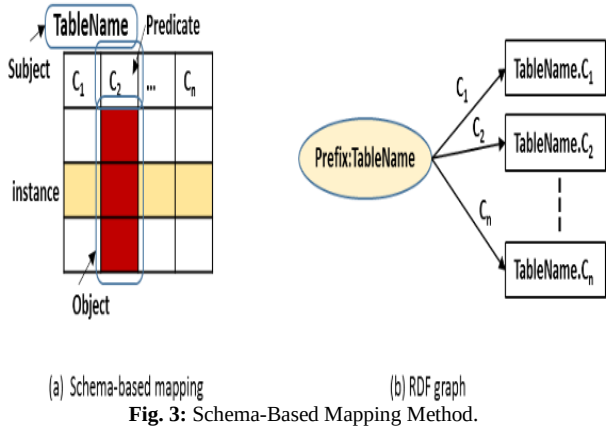


Fig. 3: Schema-Based Mapping Method.

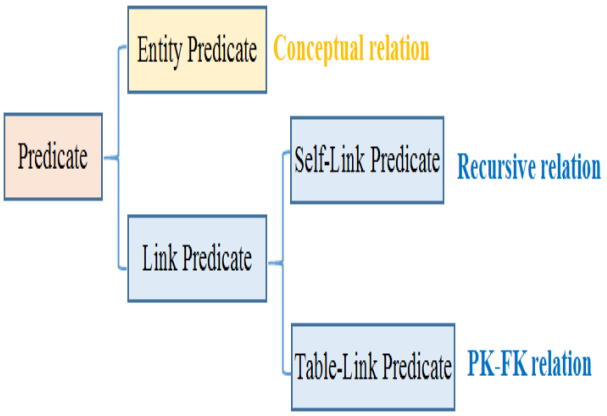
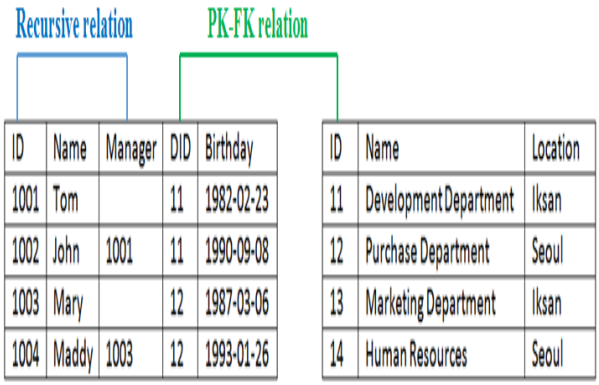


Fig. 4: Classification of Predicate in R2RS.

How to map 1:1, 1: M, M: N relationships is critical for RDB2RDF mapping methods. R2RS uses predicates to handle the relation mapping. In R2RS model, predicates consist of Entity Predicates and Link Predicates. The Entity Predicates correspond to the ordinary attribute columns, also known as conceptual relations. The Link Predicates correspond to the relations between tables in RDB. Furthermore, the Link Predicates include Self-Link Predicates and Table-Link Predicates. The Self-Link corresponds to the recursive relation and Table -Link corresponds to the PK-FK (Primary Key/Foreign Key) relation. The classification of predicates in R2RS is shown in Figure 4.



ID	Name	Manager	DID	Birthday
1001	Tom		11	1982-02-23
1002	John	1001	11	1990-09-08
1003	Mary		12	1987-03-06
1004	Maddy	1003	12	1993-01-26

ID	Name	Location
11	Development Department	Iksan
12	Purchase Department	Seoul
13	Marketing Department	Iksan
14	Human Resources	Seoul

Fig. 5: Example of Predicates.

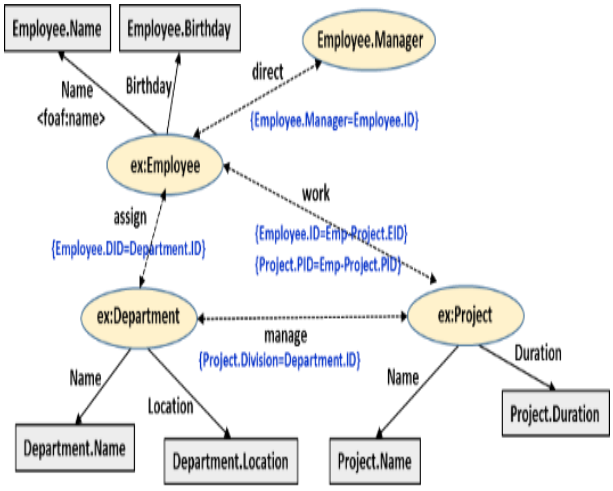


Fig. 6: The RDF Graph of Example in Figure 5.

For example, there are two tables: Employee and Department. In Employee table, "ID" is primary key and "Manager" is the foreign key that refers to the primary key "ID". "ID" and "Manager" have recursive relation, which is mapped into Self-Link Predicate in RDF graph. In Department table, "ID" is primary key and "DID" in Employee table is foreign key that refers to this primary key "ID". Therefore, "DID" in Employee and "ID" in Department have PK-FK relation, which is mapped into Table-Link Predicate in RDF graph. The remaining column in the two tables such as "Name" in the two tables, "Birthday" and "Location", belong to conceptual relation that is mapped into Entity Predicate. Figure 5 shows the detailed classification of Link Predicate and Figure 6 shows the corresponding RDF graph.

In the RDF graph as shown in Figure 6, there are three subjects, which correspond to three table names in RDB. They are "ex:Employee", "ex.Department" and "ex.Project" respectively. In fact, "Employee.Manager" is also a subject, but it corresponds to the same table with "ex:Employee". The remaining nodes in the RDF graph are objects that correspond to the cell values in RDB. The predicates that are represented with solid lines are Entity Predicates, also known as conceptual relation. The predicates that are represented with dotted lines are Link Predicates. For example, the predicate "direct" is Self-Link Predicate. The predicates "assign", "work" and "manage" are Table-Link Predicates. Each Link Predicate has an object description of database join operation. For example, the connection attribute of predicate "assign" is "Employee.DID = Department.ID", which will be applied to the SPARQL-SQL engine.

In Figure 6, the comprehensive view of the whole conceptual mapping at the schema level is clearly represented without introducing any complex structures. The link predicates with the object expression (dotted line) resolve many complicated mapping issues efficiently, such as the foreign key relationships shown in the Employee-Project tables and Employee-Department tables. The common ontology vocabulary can be freely adopted as shown in

<foaf:name>. Therefore, R2RS is a straightforward and efficient mapping method for RDB2RDF.

3.4. SPARQL query of R2RS

Being schema-based mapping method, R2RS doesn't store data in RDF way, which means that all data are only stored in RDB. This method efficiently avoids data redundancy and inconsistency issues. R2RS only generates RDF instances when SPARQL query is executed against the RDB. R2RS provide the SPARQL query endpoint, which query against RDB by SPARQL-SQL query transformation engine. The architecture of SPARQL query is shown in Figure 7.

The preliminary work of SPARQL query is to realize RDB2RDF mapping by R2RS, which generates a schema-based RDF view. For the RDF view, the end-users execute SPARQL query and obtain the finally the query results by SPARQL-SQL query transformation engine.

The specific query procedure is shown as follows:

- R2RS maps RDB into RDF graph.
- End-users execute the SPARQL query which conforms to SPARQL grammar standard.
- SPARQL-SQL query transformation engine translates SPARQL query into SQL query.
- Execute SQL query against relational database and return the query result.
- SPARQL-SQL query transformation engine translates query result into JSON-LD format.

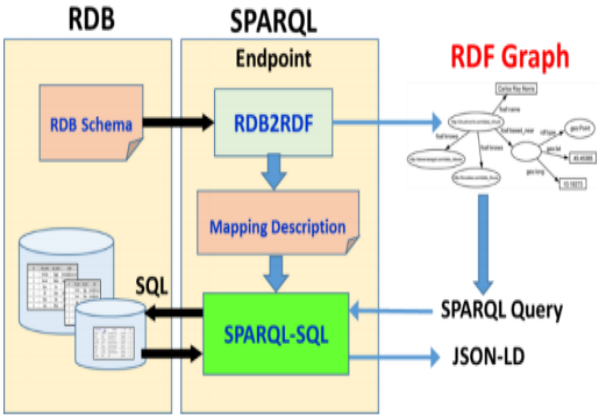


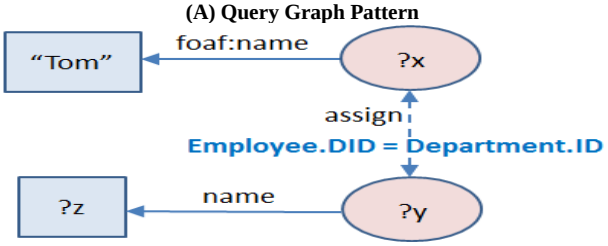
Fig. 7: Architecture of SPARQL Query Endpoint.

```
<!--@context-->
{
   "@id" : "http://example.org/Predicate#ex1001",
   "rdf-predicate" : "foaf:name",
   "db-column" : "employee.name",
   "@type" : "rdf:literal"
   "@language" : "en"
},
{
   "@id":"http://example.org/Predicate#ex1002",
   "rdf-predicate" : "employee.direct",
   "db-column" : "employee.manager",
   "object" : "employee.manager->employee.id"
   "@type" : "s-link"
},
```

Fig. 8: The Example of R2RS Mapping Language.

The SPARQL-SQL query transformation engine takes advantage of RDB2RDF mapping description to realize the transformations of query statements and query results. The R2RS mapping languge is JSON format as shown in Figure 8. The mapping description depicts the id, rdf-type, predicate-type, language, corresponding column name and join relation in RDB of nodes and edges in RDF graph. In addition, the mapping description make full use of the common ontology vocabularies such as "foaf:name"

(A) Query Graph Pattern



(B) SPARQL Query

```
SELECT ?z
WHERE
{
    ?x foaf:name "Tom".
    ?x ex:assign ?y.
    ?y ex:name ?z.
}
```

(C) SQL Query

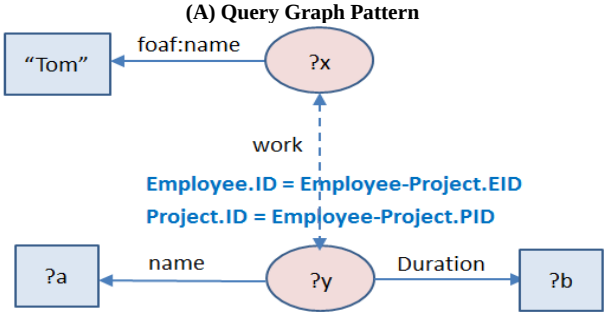
```
SELECT   d.Name as Department_Name
FROM     Employee e INNER JOIN Department d
         ON e.DID = d.ID
WHERE    e.Name = "Tom"
```

Fig. 9: SPARQL Query Example with 1: M Table-Link.

For all the RDB2RDF mapping method, how to mapping the relations of PK-FK is an obstacle that must be resolved. In this paper, the proposed R2RS mapping method takes advantage of table-link predicate to handle the relation mapping as shown in Figure 9.

In Figure 9, the query question is "what is the name of department where Tom is assigned to?" Figure (a) shows the query graph pattern which describes the relation like "Employee.DID = Department.ID". The transformation engine translate the SPARQL query into the SQL query according to the query graph pattern, just like figure (b) and (c).

(A) Query Graph Pattern



(B) SPARQL Query.

```
SELECT ?a ?b
WHERE
{
    ?x foaf:name "Tom".
    ?x ex:work ?y.
    ?y name ?a.
    ?y duration ?b.
}
```

(C) SQL Query

```
SELECT p.Name as Project_Name,
       p.Duration as Project_Duration
FROM employee e INNER JOIN employee_project ep
                INNER JOIN project p
     ON e.ID = ep.EID
        AND ep.PID = p.ID
WHERE e.Name = "Tom"
```

Fig. 10: SPARQL Query Example with M: N Table-Link.

Figure 10 shows a complex M: N relation mapping. The query question is "What are the name and duration of the project that Tom involved in?" In relational database, it is a join operation for three tables. However, in the RDF query graph pattern, the Table-Link Predicate "work" accompany with a object expression "Employee.ID = Employee-Project.EID Project.ID = Employee-Project.PID", the transformation engine translate the link into join operation by the description of the object expression.

Many complex relation-mapping issues in RDB can be translated into a simple RDF query graph pattern in the SPARQL query Endpoint of R2RS. R2RS provides a clean and efficient way to realize query translation against RDB because the almost identical conceptual schema is used in both SPARQL and SQL.

4. Conclusion

Since large amount of quality data are stored in diverse relational databases over years. RDB2RDF mapping methods have been an efficient to populate Linked data. RDB2RDF mapping methods not only integrate data from heterogeneous DBMS, but also provide the unified RDF framework to access the data. In the recent years, RDB2RDF mapping have attracted much attention and become a research hotspot.

This paper proposes a new and instinctive RDB2RDF mapping method at the conceptual schema level. Since the conceptual schema of RDB is similar to ontological domain modeling of RDF, the proposed schema-based mapping R2RS can achieve more coherent mapping than the conventional direct mapping approaches by dissolving the operational differences, such as graph pattern matching and JOIN operations. The mapping description is straightforward on account of the compatible conceptual structures and can accommodate the complex relationships in an effective manner. So, the schema-based mapping R2RS provides an efficient way to implement SPARQL endpoint into RDB, which is vital to disseminate Linked Data.

5. Acknowledgment

- This paper was supported by Won Kwang University in 2017

References

[1] Bizer, C.: The emerging web of linked data. IEEE Intelligent Systems, 2009, 24 (5), pp. 87–92.

[2] Smith Yoshimura, K.: Analysis of International Linked Data Survey for Implementers. D- Lib Magazine, 2016, 22, pp. 7-8.

[3] Būmans G, Čerāns K. RDB2OWL: a practical approach for transforming RDB data into RDF/OWL[C]//Proceedings of the 6th International Conference on Semantic Systems. ACM, 2010, pp. 25.

[4] Sitharamulu V, Babu B R. A Novel Proposal for Bridging Gap between RDB-RDF Semantic Web using Bidirectional Approach [J]. International Journal of Applied Engineering Research, 2016, 11(6), pp. 4456-4460.

[5] Malik K R, Ahmad T. Technique for Transformation of Data from RDB to XML Then to RDF [M]//Web Semantics for Textual and Visual Information Retrieval. IGI Global, 2017, pp. 70-91.

[6] Sequeda, J., et al.: Relational Database to RDF Mapping Patterns. Proceedings of the third International Conference on Ontology Patterns (WOP'12), 2012, 929(3), pp. 97-108.

[7] Hert M, Reif G, Gall H C. A comparison of RDB-to-RDF mapping languages[C]//Proceedings of the 7th International Conference on Semantic Systems. ACM, 2011, pp. 25-32.

[8] Das S, Sundara S, Cyganiak R. R2RML: RDB to RDF mapping language [J]. 2012.

[9] Čerāns K, Būmans G. RDB2OWL: a RDB-to-RDF/OWL mapping specification language [J]. Information Systems, 2011, pp. 139-152.

[10] Neto L E T, Vidal V M P, Casanova M A, et al. R2RML by assertion: A semi-automatic tool for generating customised R2RML mappings[C]//Extended Semantic Web Conference. Springer, Berlin, Heidelberg, 2013, pp. 248-252.

[11] Hert, M., Reif, G., Gall, H.C.: A Comparison of RDB-to-RDF Mapping Languages. In Proceedings of the seventh International Conference on Semantic Systems, ACM, New York, 2011, pp. 25-32.

[12] A Survey of Current Approaches for Mapping of Relational Databases to RDF, W3C RDB2RDF Incubator Group January 08 2009. Technical report, 2009.

[13] Michel F, Montagnat J, Faron-Zucker C. A survey of RDB to RDF translation approaches and tools [D]. I3S, 2014.