Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 4, April 2014, pg.1122 – 1129

RESEARCH ARTICLE

Efficient Semantic Information Retrieval System from Relational Database

Rajeshkumar T[1], Ramathilagam C[2], Dr. M.L.Valarmathi[3]

[1]Department of Computer Science and Engineering, Adithya Institute of Technology, Coimbatore, India
[2]Assistant Professor of CSE, Adithya Institute of Technology, Coimbatore, India
[3]Associate Professor of CSE, Government College of Technology, Coimbatore, India

Abstract— Now a days, the large amount of data available on the internet is in the form of HTML documents. Most of the web applications (about 75%) using a Relational database is available as a source of information. In the Semantic Web the data can be represented in the form of XML, RDF or OWL. The Ontology is a structured data on the Semantic Web to organize and share the knowledge about the specific domain. In this paper, a framework is developed for generating Ontology and extraction of knowledge from the generated information source. This framework has two major phases, first is to generate the Ontology from the existing Relational databases using set of construction rules, and second is to extract the knowledge from the generated Ontologies. The mapping process is between the Relational schema and the Ontology schema and the Ontology is represented in Web Ontology Language (OWL). The Relational database has set of finite relations, attributes and values for attributes. The Ontology has a finite set of classes, properties and instances and also has some semantic properties like symmetric property, inverse property, object property and data type property between the classes and instances. The information is extracted from Ontology using SPARQL query language. This framework provides the efficient results for information retrieval from the Ontology.

Keywords— Semantic Web; Ontology; Relational Database; Mapping Rules; OWL; SPARQL Query

I. INTRODUCTION

World Wide Web is an interconnection of hypertext (HTML) documents shared through internet accessed by the web browser. The web browsers are used to retrieve the data in the form of text documents, images, videos and other multimedia formats. The Semantic Web is the extension of the current web (WWW). It means, “web with meaning”. Ontologies are important for defining the semantics (meanings) of web data. The Ontology languages are used to represent the data in Semantic Web[1]. The Ontologies are created from the existing databases for Semantic Web applications. The relational databases are the mostly used information data resource in the world today. Several domains like, banking, educational, medical and library management systems are implemented through the Ontology concepts[6].

These Ontologies are generated by number of methods like mapping rules, extraction of metadata and so-called middle model for specific domain [2,5-10]. The major issue is to add semantics to database schemas. The relational databases are the important data resource widely used for web applications. These databases do not provide the semantic properties, such as symmetric property, inverse property and concept hierarchies. The metadata is extracted from these existing databases to construct the Ontologies [1].

The advantages of the above mentioned methods are generating the Ontology from existing databases for a specific domain. The rule based Ontology generation is one of the best method in Ontology generation from relational databases [2-7, 9, 11]. These methods are implemented and tested with Java, MySQL and JDBC.

The technology Java is used to easily create the interface using its components, Java Database Connectivity (JDBC) used for establish the connection between the data source and user interface. There are some frameworks to generate the Ontology, such as Jena, Sesame and tools available to view and manipulate the Ontology, such as Protégé, OWLGrEd and TopBraid composer [1, 15-17].

A simple vehicle Ontology is shown in Fig.1 below with the taxonomy.

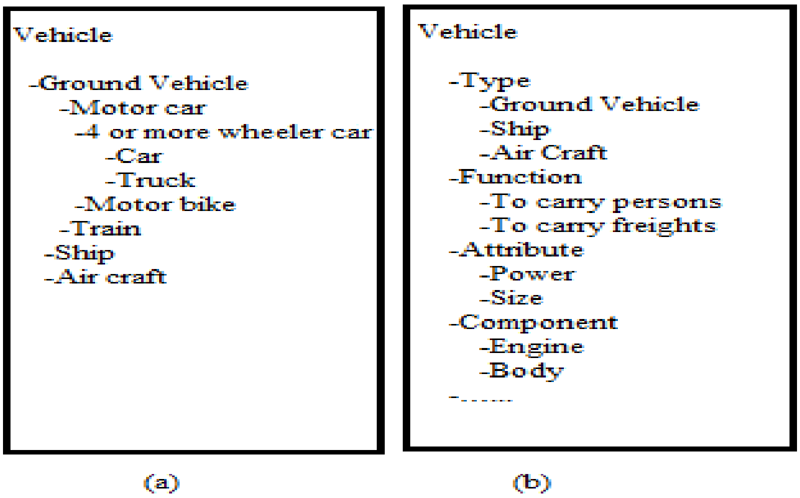

Fig 1: (a) A simple taxonomy of vehicle, (b) An Ontology of vehicle.

This paper mainly focuses on how to generate Ontology from existing database and information retrieval from the generated Ontology. The Ontology is generated from the existing database with the help of mapping or construction rules of Ontology. The relational databases are the important resource for web applications to share the information through the internet. The information retrieval is the process of getting the information related to the existing database. The advantages of this approach is generating the Ontology for any domain, size of Ontology document also reduced than the relational database size, retrieve explicit knowledge about the domain and graphical representation of the domain.

This paper is organized as follows: the section II explains related works of generating Ontology and information retrieval. Section III presents the mapping rules to generate the Ontologies from relational databases and SPARQL query language. Section IV presents the experimental setup and results of the proposed system. Section V presents the performance analysis of the proposed system. Section VI concludes this paper.

II. RELATED WORKS

An Munir, K., Odeh, M., and McClatchey, R. T. [4], presents the new approach to explicitly share the domain knowledge expressed in a specific domain Ontology, the end users can formulate the relational queries to retrieve the data from complex relational database. In this approach an Ontology-driven query formulation architectural framework, called OntoQF, there are two phases in this approach, one is pre-processing and the other one is translation phases. The pre-processing phase, an Ontology created from new database based on the transformation approach has been synthesized for specific domain Ontology and is populated and enriched with problem domain concepts and semantic relationship specified using OWL-DL. Once the Ontology generated for specific domain has been formulated, the end-users can write the queries translated, the process of translation phase converting the OWL language query to relational query statements. In this paper, the Ontology generated for the medical domain to check the correctness of the translation. The further research of this approach based on the OWL 2 syntax that includes data types, data ranges, cardinality restrictions and property chains, etc.

Trinh, Q., Barker, K., and Alhajj, R. [5], described a new framework for Semantic Web infrastructure to address the semantic interoperability between large scale environments of relational database systems and granularities in multiple levels. In a relational database systems, this framework describes a formal algorithm to use the relational database R`s metadata and structural constraints to construct specific domain OWL Ontology while preserving the constraints of the relational database systems. The generated Ontology is verified using a set of vocabularies defined in Web Ontology Language that describes the relational database systems on the semantic web. With the help of this framework the domain experts can easily generate the Ontology for a specific domain. The future research based on the merging Ontologies and sematic relationship between the multiple Ontologies.

Ouyang Dan-tong, CUI Xian-ji and YE Yu-xin [6], presents a set of axioms called Integrity Constraint(IC)-mapping axioms are stated. The integrity constraints are used to preserve the accuracy and consistency of the data in relational database systems. With the help of these axioms a special Ontology is generated with integrity constraints, which is used to adapt with domain knowledge to data in the relational database. The Ontology is generated through two phases; they are namely checking and modification of the data in the relational database. The checking phase process is to check the data is satisfies the integrity constraints and if not satisfied the modification phase changes the Unintegrity constraints to integrity constraints. The traditional mapping approaches are used map the relational database system to normal Ontology. In this papers the query performance is increased than the vertical and class decomposition approach.

Zdenka Telnarvoa [8] deals with mapping between relational data into Ontology, or filling Ontology with data from the relational database system. The issue in the mapping database schema into a common data schema expressed in the form of Ontology. There are number of methods available for acquiring Ontology from relational database systems, with the help of rules to generate the Ontology from relational database. In the mapping process the tables are mapped to classes or concepts and attributes are mapped to datatype properties and data are mapped to individuals in Web Ontology Language. The Ontology is the efficient way of data integration.

Juan F. Sequeda, Marcelo Arenas, Daniel P. Miranker[7], presents a solution, inspired by methods defined by the W3C where relational databases are directly mapped to RDF and OWL. Given a relational database schema and its integrity constraints, this direct mapping produces an OWL ontology, which, provides the basis for generating RDF instances. The semantics of this mapping is defined using OWL. Two fundamental properties are information preservation and query preservation. This paper proves that mapping satisfies both conditions, even for relational databases that contain null values and also consider two desirable properties: monotonicity and semantics preservation. The future work to extend direct mapping to rule based mapping on large relational databases. The mapping process repeated for each tuples in database. Each tuples considered as individuals and the column name of the each tuple considered as data type properties. The relationship between the tables is known as object properties in OWL. TABLE 1 shows the comparison of various related works that are mentioned above.

TABLE 1

COMPARISON OF RELATED WORKS

Title	**Method**	**Advantages**	**Limitations**	**Application Domains**
Ontology-driven relational query formulation using the semantic and assertional capabilities of OWL-DL	Preprocessing and Translation	Form relational query from the semantic data	Not efficient	Medical Domain
RDB2ONT: A Tool for Generating OWL Ontologies From Relational Database Systems	Metadata of the Database	Easy to implement using java	Merging Ontology	Education Domain
Mapping integrity constraint Ontology to Relational databases	Checking and Modification	Using integrity constraints is efficient	Changing Unintegrity to integrity	Telecommunicat ion domain
On Directly Mapping Relational Databases to RDF and OWL	Direct Mapping	Simple, information preservation, Query Preservation	Direct Mapping to consider datatypes of RDB under bag semantics	Aircraft domain
Relational Database as a source of Ontology creation	Mapping of filling ontology with data from RDB	Easy to create and update Ontologies	Needs correct mapping between RDB and OWL	University Domain

III. PROPOSED METHODOLOGY

There are several methods to generate the Ontology from relational database. One of the method is rule based Ontology generation. The drawbacks of the manual Ontology generation are mistakable work, time and cost is high. The automatic Ontology generation rectifies the drawbacks of manual Ontology creation.

TABLE 2

MAPPING BETWEEN RDB AND ONTOLOGY

Relational Database	**Ontology**
Table	Class
Column	Property
Data	Instances
Primary Key	Functional Property
Foreign Key	Object Property
Column Constraints	Property Restrictions

The generation process is done using mapping rules between RDB and Ontology as shown in Table 2. Each relational table is transformed into a class and each column is transformed into a property respectively. In addition, if the relational database table has foreign key references to other tables, these can be transformed to object property.

In the proposed system there are two major phases to reach the goal, Ontology generation from relational database and information retrieval from the Ontology. Figure 2 shows the architecture of Ontology generation and information system.

A. Ontology Generation Phase

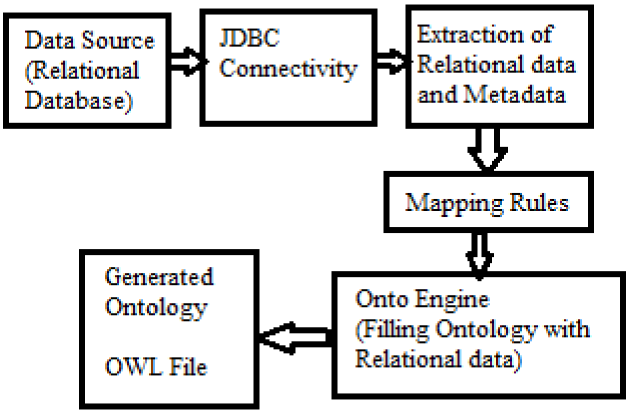

Fig 2: Ontology Generation System

The Fig 2. Shows that Ontology generated from the relational database. Through the JDBC connection can able to extract the relational data and metadata of the relational data. The mapping rules are framed to generate the OWL Ontology. The Onto Engine makes the OWL file with relational data. Finally the Ontology generated from RDB.

B. Information Retrieval Phase

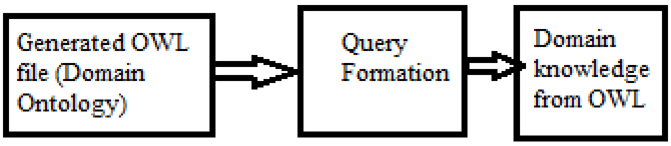

Fig 3: Information Retrieval

The Fig 3. Shows that information retrieval from the generated OWL file. The SPARQL query formed from the user interface as per the user need about the specific domain. The extracted knowledge can be stored in text file of image file. The protégé tool is widely accepted tool to manipulate and create the Ontologies.

C. Mapping Rules

Ontology as a “formal, explicit specification of a shared conceptualization”. The Ontology generated from the relational database with set of rules.

Rule 1: Mapping Tables

Creating the class for each table in the relational database. The name of the class is same as the name of the table in relational database.

$$Rel(r) \rightarrow O\ (Class\ (m)) \ldots\ldots\ldots (1)$$

Where Rel is a relational database and r is r1, r2,….r n relations in Rel, m is m1,m2…. mn class in Ontology O. The name of the r and m is same.

Rule 2: Mapping Columns

The columns are mapped into DatatypeProperty in Ontology. Each column belongs to a table in relational database and it has a type for data like Integer, String, Date, Float, and Time. The OWL uses datatypes from XML Schema.

$$Attr(x,r) \rightarrow DtP(p, dom, ran) \ldots\ldots\ldots (2)$$

Where Attr is Attribute, x is an attribute in relation r, DtP is DatatypeProperty, p is a DatatypeProperty with dom(domain) and ran(range).

Rule 3: Mapping Rows

The data values are mapped into individuals in Ontology. The data value specified with the type of data.

$$Val(v,r) \rightarrow Ind(i,m) \ldots\ldots\ldots (3)$$

Where Val is value, v is a data value in relation r, i is an Individual in class m,

Rule 4: Mapping Primary Key

If any column consist of Primary Key is mapped into InverserFunctionalProperty. The InverseFunctionalProperty describes two entities does not share the common value.

$$Pk(f,r) \rightarrow InvFP(p, dom, ran) \ldots\ldots\ldots (4)$$

Where Pk referes the Primary key, f is a primary key of relation r, p is InverseFunctionalProperty with specific domain and range.

Rule 5: Mapping Foreign Key

If any column consists of Foreign Key is mapped into ObjectProperty. The ObjectProperty links individuals to individuals and relationship between the classes in Ontology.

$$Fk(f,r) \rightarrow OP(p, dom, ran) \ldots\ldots\ldots (5)$$

Where Fk refers the Foreign key, f is a Foreign key of relation r, p is ObjectProperty with specific domain and range.

Rule 6: Class and Subclass

The subclass is created when any column in table is depend on the column in the other table.

$$Subclass(x, y) \ldots\ldots\ldots (6)$$

Where class x is subclass of class y.

D. SPARQL Query Language for Semantic Web

SPARQL (SPARQL Protocol and RDF Query Language) is a query language for Semantic Web data. It is similar to the SQL query for relational databases like MySQL, Oracle. The data in the relational database as stored in the form of subject, predicate and object. The RDF data also consists of RDF triples(subject, predicate and object).

Simple SQL Query for relational database,

SQL>Show tables;

TABLE 3

RESULT OF SQL QUERY

Tables_in_rkbank
account
branch

customer
depositor
loan

Simple SPARQL Query for Ontology,

SELECT ?subject ?object

WHERE { ?subject rdfs:subClassOf ?object }

TABLE 4

RESULT OF SPARQL QUERY

subject	**object**
Depositor	rkbank
Customer	rkbank
Loan	rkbank
Account	rkbank
Branch	rkbank

IV. EXPERIMENTAL SETUP AND RESULTS

The Ontology generation and Information retrieval system implemented in the platform of both Windows and Linux operating system with the specification of Intel core 2 duo 2.93GHz processor, 2GB memory and 256MB RAM. This system developed with the technology Java 1.7 and MySQL 5.5. The dataset taken from the following domains like banking, education, library management system and medical domain.

Table 5 analyses the feasibility of this developed approach tested with set of query statements from different datasets. The query statements are,

*Q*1: Get the no.of tables in the bank database.

*Q*2: Get the all records from the bank database, who having the account in the branch id=123.

*Q*3: Get all the records from bank database, who having more than 50000 loan amount.

*Q*4: Get the records from the bank database, who having balance less than minimum balance.

TABLE 5

EXECUTION TIME OF QUERY STATEMENTS

Query	SQL Query execution time in ms
Q1	0.9
Q2	1.0
Q3	1.0
Q4	1.8

Table 6 contains the size of the data in relational database and Ontology. The size of the database is differing in different technologies. The size is compared for different domains like bank, medical, education and University domains. From this table can determine the ontology is the best to store the data and infer the knowledge from the ontology.

TABLE 6

SIZE OF DATA IN RELATIONAL DATABASES AND OWL

Datasets	Relational database size in kb	OWL size in kb
Bank	224	12
Medical	256	13
Education	400	16
University	356	15

V. PERFORMANCE ANALYSIS

The performance analysis of different information retrieval algorithms query execution time is shown in Figure 4. The Fig 4 is the analysis chart for Table 5. The X-axis for different set of query statements and the Y-axis for execution time for each query statements. From this graph can determine the Semantic Web information retrieval approach has high efficiency and small scalability of query performance.

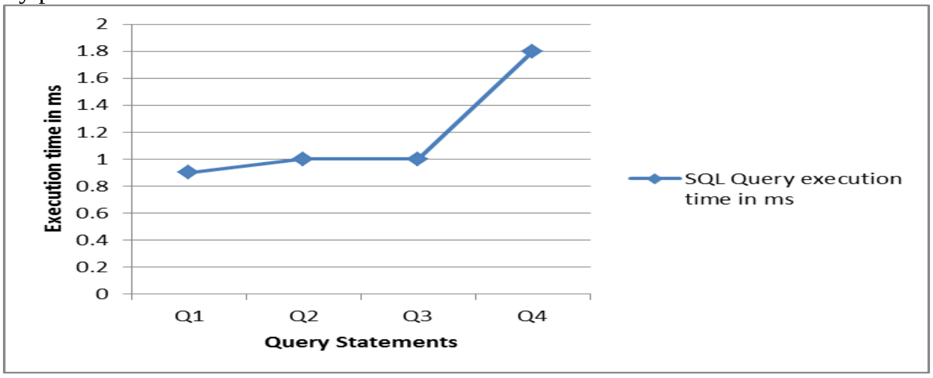

Fig 4: Execution time of Query Statements

The Fig 5 shows the best case to store the data in Ontology. This graph shows that best way to store the data in Ontology compared to relational database. The X-axis for the different domain datasets and Y-axis for size of the database in kilobytes (kb).

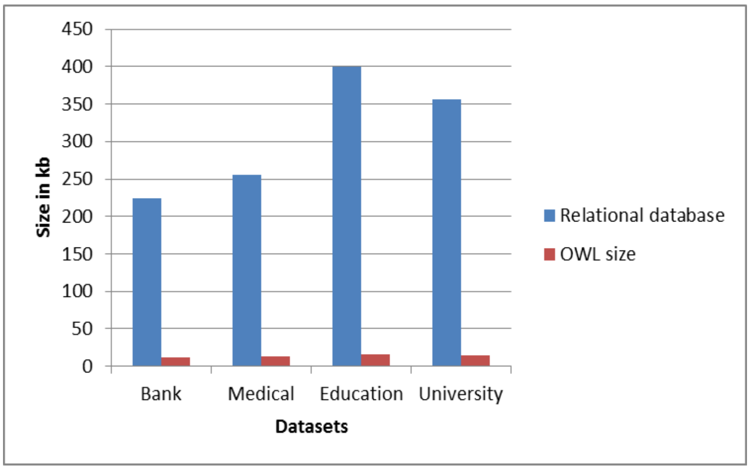

Fig 5: Dataset size in RDB and OWL

VI. CONCLUSION

This paper presents an approach for transforming the relational database into Ontology. The Ontology can be applied in several domains like banking domain, educational domain, university domain and medical domain. This transformation made by set of mapping rules for relational database to Ontology. The major advantage of this transformation is explicitly share the knowledge about the domain and size of the database also reduced than the relational database. The future work is based on merging two Ontologies for a specific domain.

REFERENCES

[1] Grigoris Antoniou and Frank van Harmelen,, "A semantic Web Primer". MIT Press, Cambridge, England, 2003.

[2] Dejing Dou, Paea LePendu, Shiwoong Kim and Peishen Qi, “Integrating Databases into the Semantic Web through an Ontology-based Framework”, IEEE Computer Society, pp. 54-63, 2006.

[3] Munir, K., Odeh, M. and McClatchey, R., "Ontology-driven relational query formulation using the semantic and assertional capabilities of OWL-DL", Science Direct, Knowledge Based Systems, Vol. 35., pp.144-159., 2012.

[4] Quang Trinh., Barker, K.. and Alhajj, R., "RDB2ONT: A Tool for Generating OWL Ontologies From Relational Database Systems," Telecommunications, AICT-ICIW '06. International Conference on Internet and Web Applications and Services/Advanced International Conference on”,IEEE Computer Society, pp.170. 2006.

[5] Ouyang Dan-tong, CUI Xian-ji and YE Yu-xin, "Mapping integrity constraint Ontology to Relational databases", ScienceDirect, The Journal of China Universities of Posts and Telecommunications, vol. 17., Issue. 6., 2010.

[6] Juan F. Sequeda, Marcelo Arenas and Daniel P. Miranker, "On Directly Mapping Relational Databases to RDF and OWL", WWW '12 Proceedings of the 21st international conference on World Wide Web, ACM Digital Library, pp. 649-658., 2012.

[7] Zdenka Telnarvoa, "Relational Database as a source of Ontology creation", Proceedings of the 2010 International Multiconference on IEEE Computer Society, pp.135,139, 2010.

[8] Mukhopadhyay, D.; Banik, A. and Mukherjee, S., "A Technique for Automatic Construction of Ontology from Existing Database to Facilitate Semantic Web", In International Conference on Information Technology, IEEE Computer Society, pp. 246-251, 2007.

[9] Yutao Ren, Lihong Jiang , Fenglin Bu and Hongming Cai, "Rules and Implementation for Generating Ontology from Relational Database", In Second International Conference on Cloud and Green Computing, Vol 7, No 12, pp. 237-244, 2012.

[10] Boris Motik, Ian Horrocks and Ulrike Sattler, "Bridging the gap between OWL and relational databases", Science Direct, Web Semantics: Science, Services and Agents on the World Wide Web, Vol.7, Issue 2, pp. 74–89, 2009.

[11] Fuqi Song, Gregory Zacharewicz and David Chen, "An ontology-driven framework towards building enterprise semantic information layer", Science Direct, Advanced Engineering Informatics, Vol. 27, Issue 1, pp. 38-50, 2013.

[12] Seungseok Kang, Junho Shim and Sang-goo Lee, "Tridex: A lightweight triple index for relational database-based Semantic Web data management", Science Direct, Expert Systems with Applications, Vol. 40, Issue 9, pp. 3421-3431, 2013.

[13] Wei Hu and Yuzhong Qu, "Discovering Simple Mappings Between Relational Database Schemas and Ontologies", The Semantic Web Lecture Notes in Computer Science , Springer, Volume 4825, pp 225-238, 2007.

[14] Heru Agus Santoso, Su-Cheng Haw and Ziyad.T. Abdul-Mehdi, "Ontology extraction from relational database: Concept hierarchy as background knowledge", Knowledge Based Systems, ScienceDirect, Vol. 24, Issue 3, pp. 457-464, 2011.

[15] Dejing Dou, Paea LePendu, and Shiwoong Kim, Peishen Qi, "Integrating Databases into the SemanticWeb through an Ontology-based Framework", IEEE Computer Society, pp. 54-63, 2006.

[16] MAN LI, XIAO-YONG DU and SHAN WANG, "Learning Ontologies from Relational Databases", IEEE Computer Society, Vol. 6, 3410 – 3415, 2005.

[17] Zhou, S.; Ling, H.; Han, M. and Zhang, H., "Ontology Generator from Relational Database Based on Jena", Computer and Information Science, 3, pp.263-267, 2010.