IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 40-45
www.iosrjournals.org

Retrieving the Relevant Information from the Merged Ontology File Created From Relational Database

Karishma Parab Gaonkar[1], Prof. Umesh A. Bapat[2]
[1]*(Computer Engineering Department, Goa College Of Engineering, India)*
[2]*(Computer Engineering Department, Goa College Of Engineering, India)*

Abstract: *As we know Database models, specially relational databases, enables information to be efficiently stored and queried, but it futile in the applications that require a more 'enriched' meaning i.e. Semantic information. So a successful new approach to represent semantic information has been defined in the last decade known as Ontologies. In the Semantic Web the information can be represented in the form of XML, RDF or OWL languages. Ontologies are the web documents generated to provide more accurate web content, thus by improving the performance of information retrieval. This paper has proposed an approach to automatically build an ontology file from a relational databases i.e. from mysql and from oracle. Then it performs merging on the files which are created from relational databases and then applying the SPARQL query on the merged ontology file we extract the relevant information which is needed for a user.*
Keywords: *Jena framework, OWL files, Relational Databases, Semantic Web, SPARQL Query.*

I. Introduction

World Wide Web is an interconnection of hypertext (HTML) documents shared through internet accessed by the web browser. As we know web browsers are used to extract the data in the form of text documents, images, videos and other multimedia formats. The Semantic Web is the addition of the current web (WWW). It means, "web with meaning". Ontologies are significant for defining the semantics (meanings) of web data. The Ontology languages (RDF,OWL etc) are used to represent the data in Semantic Web. The Ontologies are created from the existing databases such as mysql , oracle etc for Semantic Web applications. Some domains like banking, educational, medical and library management systems are implemented through the Ontology concepts.

Semantic Web provides knowledge to capture, share, and reuse structured and machine-readable domain specific knowledge and makes it accessible on web. Ontology is a depiction of domain-specific knowledge in order to share it with different applications. "Ontology is a defined specification of conceptual model. Machine and people also distribute information through ontology. Main components of ontology (web) are: classes, subclasses, properties (datatype property , object property), and individuals. Present ontology usages are digital library, Semantic Web, and information intelligent retrieval system, etc. Web Ontology Language (OWL) has been suggested by The World Wide Web Consortium (W3C) as a formal language for authoring ontology.

The OWL 2 Web Ontology Language, informally OWL 2, is an ontology language for the Semantic Web with properly defined meaning. OWL components such as classes, properties, individuals and data values are stored as Semantic Web documents. OWL ontologies can be used along with information written in Resource Description Framework (RDF). OWL ontologies themselves are mainly exchanged as RDF documents. OWL allows for greater machine interpretability of Web content than the content supported by XML, RDF and RDF Schema (RDF-S) by providing extra vocabulary along with a formal semantics.

The remainder of the paper is organized as follows. Section III represent constructing the ontology file from oracle and mysql database using transformation rules on this databases. Section IV represent the merging process, for the ontologies which are generated in section II are merged to single ontology file using merging techniques. Section V focuses on information retrieval process i.e. extracting the relevant information from merged ontology file, using SPARQL query.

II. Related Work

Carmen Martinez- Cruz, Ignacio J. Blanco, M. Amparo Vila proposed a paper "Ontologies versus relational databases: are they so different? A comparison" [2].The author proposed that ontologies provide a constraint-free framework to signify a machine readable reality, even in the Web. This framework assumes an open world in which information can be openly defined, shared, reused or distributed. Moreover, information can also be interchanged and used to create deductions or queries. Such representation is common and used by most of the community, which has converted some of the languages used into defacto standards (such as OWL or RDF). Semantic Web plays an important job in this respect because of the rising importance of Internet and the need to publish information therein. Obviously, the judgment to choose one or another technology depends on the final user's needs. If the information to be represented needs to be shared on Web then an ontology language should obviously provide a good solution. However, this choice probably would involve the use of both technologies, as a big amount of data needs to be stored and correctly managed. Ontologies provide an excellent way to represent actuality, but database is certainly the better method for storing such information when this is of significant size. Numerous mechanisms can be used for storing information represented in an ontology in databases: The two main trends involve the use of generic ontology environments which simply use databases as a repository and do not consider how the information is stored (OBDB) or signify a proper database schema and subsequently establish the equivalent mapping relation with an ontology, using any of the previously described approaches.

Mohammed Reda Chbihi Louhdi, Hicham Behja and Said Ouatik El Alaoui proposed a paper "Transformation Rules For Building Owl Ontologies From Relational Databases"[3].The author proposed that a method which consists in a set of transformation rules for building OWL ontologies from relational databases. The schema mapping uses the transformation rules to convert the components of the physical model into ontology's components. The data analysis is used to improve some disappeared aspects during mapping conceptual data model to the relational model (like disjointness and totalness in simple inheritance cases and the participating level of tables in n-ary relations).Relational Databases (RDB) are used as the backend database by most of information systems. RDB summarize conceptual model and metadata needed in the ontology construction. Schema mapping is a method that is used by all existing approaches for ontology construction from RDB. However, most of those methods use poor transformation rules that stops advanced database mining for structuring rich ontologies.

C.R. Rene Robin, G.V. Uma proposed a paper "A Novel Algorithm for Fully Automated Ontology Merging Using Hybrid Strategy"[7].The author's proposed algorithm brings together techniques in Lexical Matching, Semantic Similarity Matching , Similarity Check and Heuristics functions in order to present a fully-automatic merging framework for the purpose of getting better semantic interoperability in heterogeneous systems. Such ontology alignment means connecting entities of source ontology with those of target ontology based on different features of these ontologies and using different strategies. The proposed algorithm explores four dissimilar measures of similarity based on strings, linguistics, heuristics and structure, to support Merging. The merging task was performed manually for a long time, to simplify the process of merging semi automatic tools like Ontolingua, Chimaera were developed with lots of human involvement. The proposed algorithm using hybrid strategies improves the hope among the ontology engineers that it is possible to have a fully automatic ontology merging tool to merge ontologies irrespective of its size. The proposed algorithm uses varied strategies. Even if one of the strategies fails to admit the match the other strategies will. This makes the proposed system better. It is fully automatic. The user's only work is to give the owl files as input and a merged owl file will be produced as output.

A.M.Abirami ,P.Sheba Alice ,Dr. A.Askarunisa proposed a paper title "An Enhanced Method for the Efficient Information Retrieval from Resume Documents using SPARQL"[8]. K Dhanasekaran and Rajeswari Ramachandran proposed a way of classifying features related to plant domain and an information extraction has been used to obtain domain relevant features [9] aiming to predict various effects on plants. The proposed model has been implemented using Netbeans, Java Development Environment and the Protégé for defining the ontology. The user inputs are obtained and RDF file is created. The needs are entered on the admin side and the matched records are collected and displayed using SPARQL. JENA APIs are used to extract data from RDF.The proposed environment holds well for easy queries to more complex queries and not of much time variance is there for SPARQL to retrieve the matched records. The time taken by XPATH and SPARQL to retrieve information for various numbers of records from XML/RDF is found out and this obviously proves the efficiency of SPARQL in its retrieval process.

III. Ontology Generation Process

Generating the ontology file(.owl file) from oracle and mysql database by applying a transformation. Databases includes conceptual models and information resources that together can be taken as the conceptualisation storehouse of ontology. Based on analyses of the formal corresponding interaction between relational databases and OWL ontologies: a relational database contains several tables, a table contains several fields and records are the compilation of a field's value, where as an OWL ontology contains numerous classes, a class contains several properties and instances are the set of property values. The formal corresponding relationships between tables, fields and records in relational databases and classes, properties and instances in OWL ontologies make it possible to change one schema to another. The corresponding interaction between relational database components and ontology components are shown in Figure2.

The use of existing relational databases to produce ontology automatically is the main purpose of the proposed approach, in order to decrease the manual tedious work, save developing period and improve the efficiency of ontology. The building of a confined ontology design from a relational database is shown in Figure 1.In this tables are mapped to classes, table interactions are mapped to object property, columns are mapped to datatype property,and records are mapped as individuals in owl language.

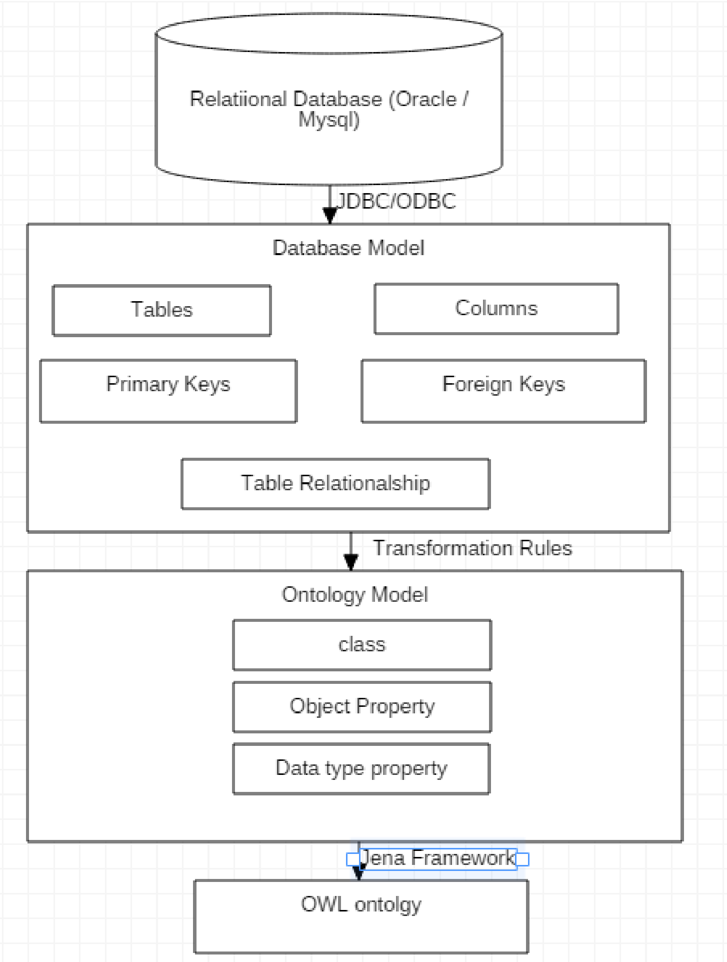

Fig 1: Construction of ontology from relational databases.

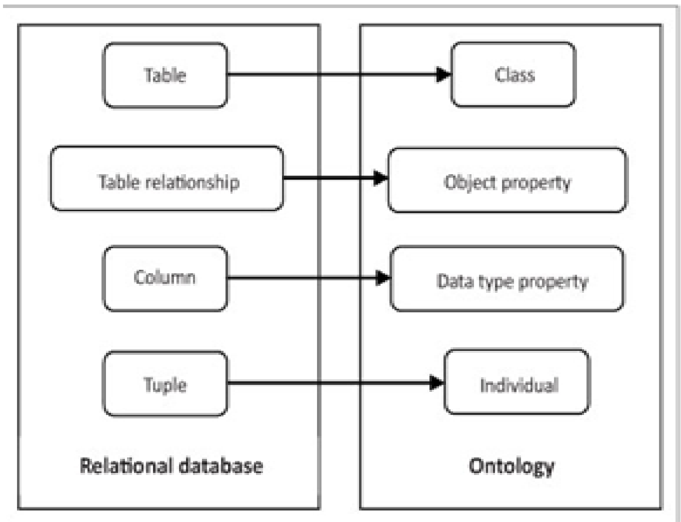

Fig 2: Relational Database component and their corresponding ontology component.

3.1 Transformation Rules

Rule 1: The tables that have only simple columns (without foreign key constraint) are transformed into simple classes into the ontology.Example:

```
<owl:Class rdf:ID="PERSON"/>
```

Rule 2: Tables that have foreign key constraint are transformed into simple classes in the ontology. Each foreign key is mapped into two Object-Properties (mutually inverse). The first one has the class equivalent to current table as domain, and its range is the referenced table by the foreign key.The second one (inverse of the first Object-Property) is declared as inverse functional.Example:

```
<owl:Class rdf:ID="ACTIVITY"/>
<owl:ObjectProperty rdf:ID="activityHasProject">
	<rdfs:domain rdf:resource="#ACTIVITY" />
	<rdfs:range rdf:resource="#PROJECT" />
</owl:ObjectProperty>
<owl:InverseFunctionalProperty rdf:ID="project'sActivity">
	<rdfs:domain rdf:resource="#PROJECT" />
	<rdfs:range rdf:resource="#ACTIVITY"/>
	<owl:inverseOf rdf:resource="#activityHasProject" />
</owl:InverseFunctionalProperty>
```

Rule 3: Identify simple inheritance relationships from tables. All tables in this category are sub-tables in hierarchies. Each sub-table is changed into a class in the ontology and is declared as a subclass of the table referenced by the foreign key (which is also the primary key of each sub-table).Example:

```
<owl:Class rdf:ID="PERSON" />
<owl:Class rdf:ID="STUDENT">
	<rdfs:subClassOf rdf:resource="#PERSON" />
</owl:Class>
<owl:Class rdf:ID="TEACHER"/>
	<rdfs:subClassOf rdf:resource="#PERSON" />
</owl:Class>
```

Rule 4: The tables containing a composite primary key (two or more columns) which is also a foreign key whose fields are referencing just two tables, are mapped into two Object-Properties mutually inverse.
Example:

```
<owl:ObjectProperty rdf:ID="hasProject">
	<rdfs:domain rdf:resource="#ENGINEER" />
	<rdfs:range rdf:resource="#PROJECT" />
</owl:ObjectProperty>
```

Rule 5 :Each record in the database is changed to individuals in OWL language and the values are transformed to literals in case of OWL language.

IV. Merging Process

Merging is the process of building coherent ontology from two or more current dissimilar ontologies of identical domain. The merged ontology will be equal to the source ontology, which is unaffected. The process of merging is nothing but creating a new ontology from different ontologies of the same domain. Ontology creation starts with, relating the domain in a hierarchal way. There are diverse approaches to depict the domain in a hierarchal form like top-down and bottom-up. The ontology must be identical or overlapping domain, so it is not suitable for ontologies of dissimilar domain and also mapping process depends on the ontology merging application.

The system proposed uses varied strategies. Even if one of the strategies fails to acknowledge the match the other strategies will. This makes the proposed system improved. It is fully automatic. The user’s only work is to give the owl files as input and a merged owl file will be produced as output. The strategies considered are Lexical Matching, Semantic matching using Wordnet, Similarity Checking of properties.

4.1 Linguistic Comparison

The linguistic matcher finds the probable pairs of term from two ontologies. The similarity is computed based on the Jaro Winkler distance. Similarity score are assigned to every pair if it matches. If the similarity score is bigger than the threshold then the similarity of each pair is determined.

The Jaro distance d_j of two given string s_1 and s_2 is

$$d_j = \begin{cases} 0 & if\ m = 0 \\ \frac{1}{3}\left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}\right) & otherwise \end{cases} \quad (1)$$

Where:
$|s_i|$ is the length of the string s_i;
m is the number of matching letters;
t is half the number of transpositions.

The Jaro Winkler distance d_w is

$$d_w = d_j + (l_p(1 - d_j)) \quad (2)$$

Where:
d_j is the Jaro distance for strings s_1 and s_2;
l is the length of common prefix at the start of the string up to a maximum of 4 letters;
p is a constant, The standard value for this constant
in Winkler's work is p=0.1

4.2 Syntactic Similarity
Based on Levenshtein's edit distance technique we calculate the syntactic similarity measure for two strings. The Levenshtein distance among two strings a,b is given by $lev_{a,b}(|a|,|b|)$ where

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & if\ \min(i,j) = 0, \\ min\begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & otherwise \end{cases} \quad (3)$$

4.3 Semantic Similarity
There are many types of semantic similarity process, which are used to measure the similarity between two concepts. Semantic similarity is calculated using wordnet in java.

V. Information Retrieval Phase

Extracting the information from the merged ontology file by applying the SPARQL query on the merged file. SPARQL query has many advantages over the other query languages like it wants no database connectivity and every time the records may not be entered manually. Its only need is to state an RDF format and each time this RDF file may be restructured with user records.

SPARQL queries can be nested within JAVA languages and only duty is to import packages enhancing the SPARQL to run in java environment. This enables SPARQL to be run in NETBEANS / ECLIPPSE. So that no special software need to be installed.

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

VI. Conclusion

The Ontology can be useful in several domains like banking domain, educational domain, university domain and medical domain. This transformation completed by set of mapping rules for relational database to Ontology. The major benefit of this transformation is explicitly share the knowledge about the domain and size of the database also condensed than the relational database. In case of ontology merging, generate a common repository of knowledge base and to eliminate overlaps in existing ontologies, we go for Ontology Merging. Two or more Ontologies are merged to build greater ontology for complete knowledge. The user's only job is to give the owl files as input and a merged owl file will be produced as output. In case of Information retrieval, we extract the information from merged ontology file by applying the SPARQL query on the merged owl file. SPARQL query has many advantages over the other query languages like it needs no database connectivity and each time the records may not be entered manually.

References

[1]. Rajeshkumar T, Ramathilagam C, Dr. M.L.Valarmathi, Efficient Semantic Information Retrieval System from Relational Database , *International Journal of Computer Science and Mobile Computing, Vol.3 Issue.4, April- 2014, pg. 1122-1129* .

[2]. Carmen Martinez-Cruz, Ignacio J. Blanco and M. Amparo Vila, Ontologies versus relational databases: are they so different? A comparison, *Artif Intell Rev (2012)38:271–290 DOI 10.1007/s10462-011-9251-9 Published online: 15 June 2011 ©Springer Science+Business Media B.V.2011*.

[3]. Mohammed Reda Chbihi Louhdi, Hicham Behja and Said Ouatik El Alaoui, Transformation Rules For Building Owl Ontologies From Relational Databases, *Sundarapandian et al. (Eds) :ICAITA, SAI,SEAS, CDKP, CMCA-2013 pp. 271–283,2013. © CS & IT-CSCP 2013.*

[4]. 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.*

[5]. C.Ramathilagam and M.L.Valarmathi, A Framework for OWL DL based Ontology Construction from Relational Database using Mapping and Semantic Rules , *International Journal of Computer Applications (0975 – 8887) Volume 76– No.17, August 2013* .

[6]. Natalya Fridman Noy and Mark A. Musen," Algorithm and Tool for Automated Ontology Merging and Alignment".

[7]. C.R. Rene Robin and G.V. Uma,A Novel Algorithm for Fully Automated Ontology Merging Using Hybrid Strategy,*European Journal of Scientific Research ISSN 1450-216X Vol.47 No.1 (2010), pp.074-081 © EuroJournals Publishing, Inc. 2010* .

[8]. A.M.Abirami, P.ShebaAlice and Dr. A.Askarunisa proposed a paper title "An Enhanced Method for the Efficient Information Retrieval from Resume Documents using SPARQL".

[9]. K Dhanasekaran and Rajeswari Ramachandran, "Design and Analysis of Feature Classification Model using Information Extraction in Tomato Growing Environment", International Information Institute (Tokyo). Information, 17(8), 2014, pp.3947-3959.

[10]. W3C, the semantic web activity, Available at : http://www.w3.org/2001/sw/

[11]. Grigoris Antoniou and Frank van Harmelen, A Semantic Web Primer Second Edidtion(The MIT Press Cambridge, Massachusetts London, England).

[12]. [Tutorial] Jena.doc

[13]. Vishal Jain and Dr. S. V. A. V. Prasad,Mapping Between RDBMS And Ontology: A Review, *international journal of scientific & technology research volume 3, issue 11, november 2014.*