Semantic – Based Querying Using Ontology in Relational Database of Library Management System

Ayesha Banu[1] , Syeda Sameen Fatima[2] and Khaleel Ur Rahman Khan[3]

[1]Asso.Professor, Dept. of Informatics, Alluri Institute of Management Sciences, KU, Warangal.
ayeshaaims@gmail.com
[2]Professor, Dept. of C.S.E , Univ. College of Engineering ,OU , Hyderabad
sameenf@gmail.com
[3]Professor, Dept.of C.S.E, ACE Engineering College , JNTU, Hyderabad
khaleelrkhan@aceec.ac.in

ABSTRACT:

The traditional Web stores huge amount of data in the form of Relational Databases (RDB) as it is good at storing objects and relationships between them. Relational Databases are dynamic in nature which allows bringing tables together helping user to search for related material across multiple tables. RDB are scalable to expand as the data grows. The RDB uses a Structured Query Language called SQL to access the databases for several data retrieval purposes. As the world is moving today from the Syntactic form to Semantic form and the Web is also taking its new form of Semantic Web. The Structured Query of the RDB on web can be a Semantic Query on Semantic Web. The SPARQL is the Query Language recommended by W3C for the RDF(Resource Description Framework). RDF is a directed, labeled graph data format for representing information in the Web and is a very important layer of the Semantic Web Architecture. In this paper we consider the Library Management System (LMS) database, taking some tuples of the LMS Relational Schema. We discuss how the RDF code is scripted and validated using RDF Validator and how RDF Triples are generated. Later we give the graphical representation of the RDF triples and see the process of extracting ontology from the RDF Schema and application of the Semantic Query.

***KEYWORDS*:**

Semantic Web, Ontology, Semantic Query, Relational Database.

1. INTRODUCTION

Today's Web is a big pool of information stored in various forms. The World Wide Web has also changed the way people communicate with each other. Most of the web content today is suitable for human consumption with typical uses involving searching for the required information, reviewing the online stores and placing order for products and many more. The main tools people

DOI : 10.5121/ijwest.2011.2402

use to search today's web are the keyword based search engines like Google and Yahoo. Even if the search is successful the person has to still browse the retrieved documents further to extract the exact information he is looking for. The main obstacle to provide a better support to web user is that "the meaning of the web document is not machine-accessible".

There are tools which can retrieve texts , count number of words and check their spelling but when it comes to interpreting the sentences and extracting information useful to users the capability of current software's is limited. The solution for this is provided by the Semantic Web where traditional Web is focused on people, the Semantic Web is focused on machines.

The Web requires a human operator, using computer systems to perform the tasks required to find, search and aggregate its information. It's impossible for a computer to do these tasks without human guidance because Web pages are specifically designed for human readers. The Semantic Web aims to change it by presenting Web page data in such a way that it is understood by computers, enabling machines to do the searching, aggregating and combining of the Web's information — without a human operator. The organization of this paper is as follows: Section 2 presents the concept of Semantic Web and its major layers. Section 3 focuses on the Related Work for the process of building ontology based query for the relational database (Library Management System Database). Section 4 describes the Semantic Query in the RDB and Section 5 concludes the paper.

2.SEMANTIC WEB CONCEPTS AND BACKGROUND

In this section we describe the basic concepts of the Semantic Web, its layered architecture, different languages for the semantic web , ontology , RDF Triple RDS(Relational Database Scheme) and RDFS(RDF Schema).

2.1. The Semantic Web

The Semantic Web (SW) is an extension of the current web where the information is presented in a well-defined manner, better enabling computers and people to work in cooperation.[1] Berners-Lee suggested a layer structure for the Semantic Web as shown below:
[Figure Source:en.wikipedia.org/wiki/Semantic_Web_Stack]

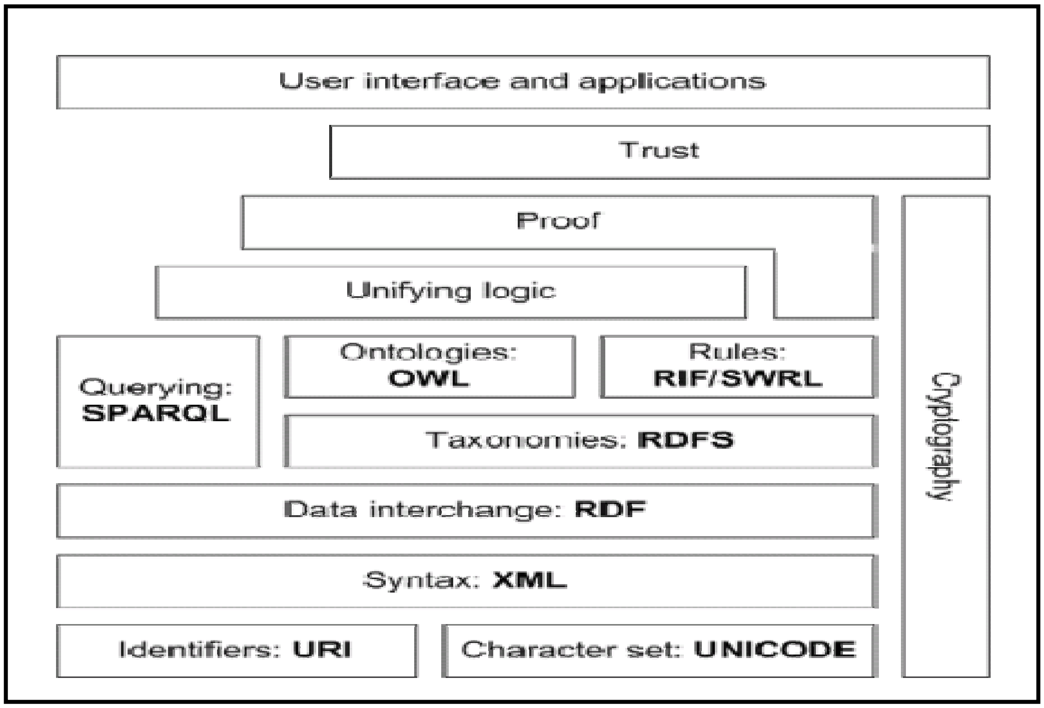

Figure 1: The Layered Architecture of Semantic Web

The architecture of the Semantic Web can be classified as four major parts

1. XML : The Representation Layer
2. RDF : The Knowledge Representation Layer
3. RDF Schema : The Ontology Layer
4. Agents : The Logic , Proof and Trust Layers

2.2. Languages for the Semantic Web

There are many languages proposed for semantic web where every language has its own utility and style of scripting and working. Some of the languages for Semantic Web can be [2]:

HTML (Hyper Text Markup Language): is the standard language in which Web pages are written. HTML was designed only for the purpose of displaying the data and it has no separation between the logical structure of the web page and the actual data of the page. This raised the difficulty in understanding the syntax and semantics.

XML (Extensible mark-up language): was designed as a language for mark-up or annotation of documents. XML separates the logical structure of the data from data itself. An XML object is a labeled tree and consists of objects with attributes and values. XML allows the definition of any kind of annotation, which opens the way to annotation with ontologies. XML Schema allows the definition of grammars for valid XML documents, but XML cannot recognize the semantics of information.

RDF (Resource Description Framework): To recognize the semantics of information RDF is proposed which represent information in a concept representation language. RDF is the widely accepted standard proposed by W3C Consortium for representing metadata. RDF documents consist of three types of entities: resources, properties, and statements. Resources may be Web pages or parts of Web pages. Properties are specific attributes, characteristics, or relations describing resources. Statements can the object–attribute–value triples.

RDF Schema (RDFS): defines a simple modeling language on top of RDF which includes classes, is-a relationship between classes and between properties, and domain/range restrictions for properties. RDF and RDF Schema are written in XML syntax.

OWL (Web Ontology Language): Like RDF and RDF Schema, OWL is a W3C recommendation, intended to support more elaborate semantics. OWL includes elements from description logics and provides many constructs for the specification of semantics, including conjunction and disjunction, existentially and universally quantified variables and property inversion.

SPARQL: SPARQL (pronounced as "sparkle") is an RDF query language, stands for **S**PARQL **P**rotocol **and R**DF **Q**uery **L**anguage. It was standardized by the RDF Data Access Working Group (DAWG) of the World Wide Web Consortium.

2.3. The Ontology

Ontology stands as the most important concept in Semantic Web, which can be defined as a collection of key concepts and their inter relationship providing an abstract view of an application domain. The Ontology enables both user and system to communicate with each other by the shared and common understanding of a domain [3]. For the web, ontology is about the exact description of web information or web resources and the relationship between them.

RDF (Resource Description Framework) can be used for describing the web resources and the relationship between them. The fundamental concepts of RDF are Resources, Properties and Statements.

Resources can be any object about which we would think. A resource may be an author or a book or a publisher and so on. Every resource has a Uniform Resource Locator (URL) or Web Address.

Properties are special kinds of resources which describe the relation between the resources. For example “title” (title of a book), “age” (age of any person) and so on. These properties in RDF are also identified by URLs.

Statements assert or declare the properties of resources. A statement in RDF is *Object-Property-Value* triple consisting of a resource (Object), property (the property of the resource) and value (is the property value of the resource) [4].

The graphical representation of a RDF triple can be shown as below:

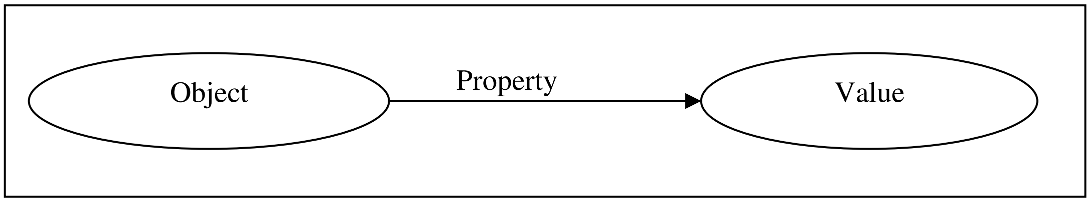

Figure 2: Graphical Representation of RDF Triple

3. RELATED WORK

To understand how the RDF triples are formed , how the RDF code is produced and validated and how the Graphical Representation is done, we consider the following example table of 2 books with its title , author , and price and published year given in the table.

Table 1: Book Details

Book Title	**Author**	**Price**	**PubYear**
Data Mining	Arun.K.Pujari	$56	1999
Semantic Web	Tim Berner's Lee	$65	2004

For the above data regarding books details the RDF code can be written as follows [8]:

```
<?xml version="1.0"?>

<rdf: RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:lib="http://www.libraryexample.cm/lib#">

<rdf:Description
rdf:about="http://www.libraryexample.cm/lib/DataMining">
  <lib:author>Arun.k.Pujari</lib:author>
  <lib:price>$56</lib:price>
  <lib:pubyear>1990</lib:pubyear>
</rdf:Description>

<rdf:Description
rdf:about="http://www.libraryexample.cm/lib/SemanticWeb">
  <lib:author>Tim Brener's Lee</lib:author>
  <lib:price>$65</lib:price>
  <lib:pubyear>2004</lib:pubyear>
</rdf:Description>

</rdf:RDF>
```

The above code can now be subjected against the RDF Validator and Converter to check for the syntax validation of the code and obtain the RDF Triples. Demo version of the RDF validator can be available at *http://www.rdfabout.com/demo/validator/* .The results of RDF Validator and Converter for the above code can be

RDF Validation Result

Syntax validated OK.

The document as Notation 3

@prefix lib: <http://www.libraryexample.cm/lib#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
<http://www.libraryexample.cm/lib/DataMining> lib:author "Arun.k.Pujari"; lib:price "$56";
lib:pubyear "1990".
<http://www.libraryexample.cm/lib/SemanticWeb> lib:author "Tim Brener's Lee"; lib:price "$65";
lib:pubyear "2004".

The underlying triples

<http://www.libraryexample.cm/lib/DataMining> <http://www.libraryexample.cm/lib#author> "Arun.k.Pujari".
<http://www.libraryexample.cm/lib/DataMining> <http://www.libraryexample.cm/lib#price> "$56".
<http://www.libraryexample.cm/lib/DataMining> <http://www.libraryexample.cm/lib#pubyear> "1990".
<http://www.libraryexample.cm/lib/SemanticWeb> <http://www.libraryexample.cm/lib#author> "Tim Brener's Lee".
<http://www.libraryexample.cm/lib/SemanticWeb> <http://www.libraryexample.cm/lib#price> "$65".
<http://www.libraryexample.cm/lib/SemanticWeb> <http://www.libraryexample.cm/lib#pubyear> "2004".

The Graphical representation of the RDF triples can result in a RDF Graph [9]. A part of RDF Graph for above example can be as shown

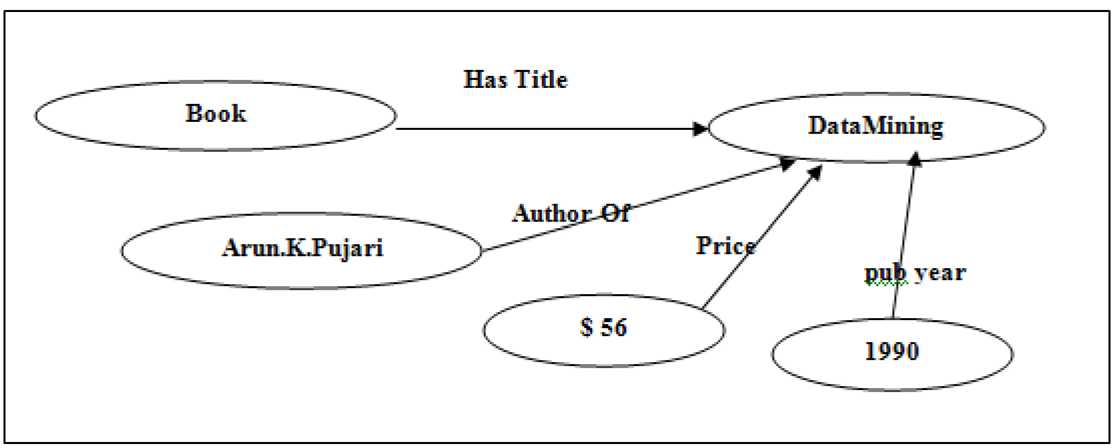

Figure 3: Graphical Representation of our Example RDF Triple

3.1. RDB and RDFS

Relational Databases are used more to store the content of the website. The integration of XML with relational database systems to enable the storage, retrieval, and update of data is of major importance [5]. But today many of the Web-based information systems do not aim at purely providing read-only access to their content which is simply represented in terms of web pages stored in the web server's directory but provide more access to their data. [6]. So Compared to XML and relational databases, which are structural-oriented, RDF takes into consideration a knowledge oriented approach that is designed specifically for the Web and that is extremely useful for the Semantic Web. One of the advantages of RDF over XML and relational model is that an RDF graph depicts in a unique form the information to be conveyed. Following table shows the relationship between the Entity Relationship Model (Structural model) and the ontology model (Semantic model) [3]

Table 2: The Relationship between ER-Model and Ontology Model

	Graphical Representation	**Implementation Schema**	**Instances**	**Modeling**
ER Model	ER-Model	Relational Schema	Relational Database	Data Modeling
Ontology Model	Ontology	RDF Schema	RDF	Knowledge Modeling

4. BUILDING ONTOLOGY BASED QUERY

To build the ontology based query for the Library Management System database we consider the framework proposed in[3] consisting of 2 phases, offline ontology extraction and online query issuing. In offline ontology extraction, the system extracts the explicit classes and relations from the relational schema. Then the domain expert will adapt the extracted ontology by adding the implicit relations to complete the ontology. In online query operation the user can issue a semantic query to the system, and the system maps that query into a related SQL query for the underlining relational database. The framework can be as shown:

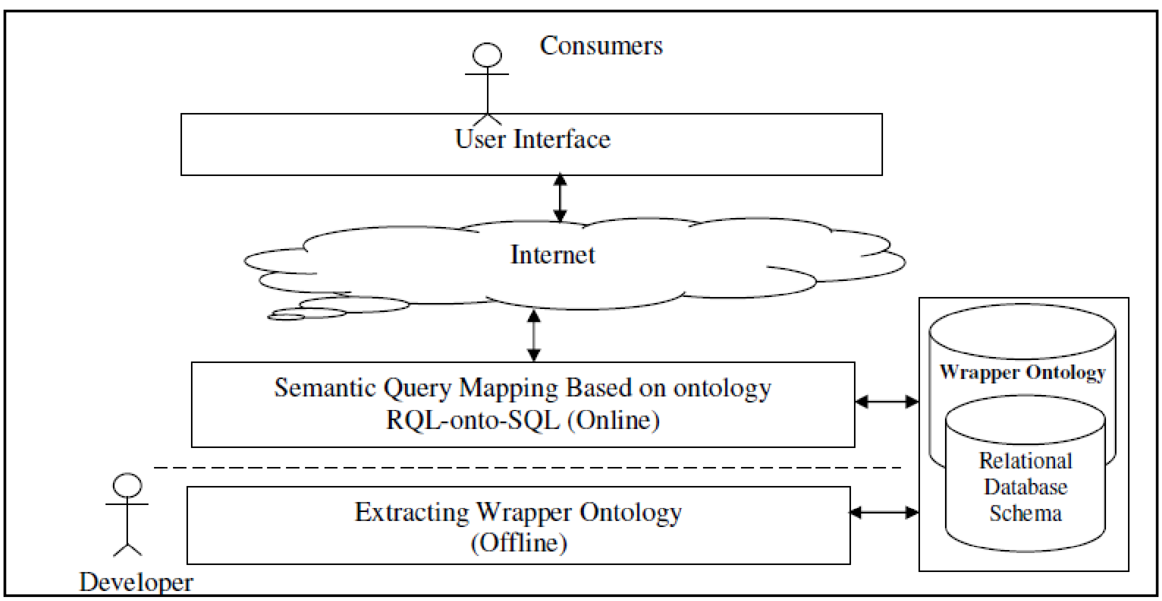

Figure 4: Framework for querying relational database based on ontology

4.1. Extracting Ontology from Relational Database

BOOK, MEMBER and PUBLISHER are the basic entities for the example database we use . Every Library has books with unique id, title, author, price and status (available or not). Members of the library borrow books and every member has member id, name, address and member type (student/professor/researcher). Library also maintains the details of publishers of the books including the publisher id, name and address. The database also maintains the complete book details and issue details. The complete Relational Database of the Library Management System considered can be shown in table below.

Table 3: A Relational Database for the Library Management System

Relation	Primary Key	Foreign Key
Book(Bid,Title,Author,Price,Status,Mid,Pid)	Bid	Mid ref. to Member Pid ref. to Publisher
Member(Mid,Mname,Addr,Mtype)	Mid	
Publisher(Pid,Pname,Addr)	Pid	
IssueDetails(Mid,Mname,Bid,Title,Doi,Dor)	Bid , Mid	Bid ref. to Book Mid ref. to Member
BooksDetails(Bid,Title,Price,Pid,Pname)	Bid,Pid	Bid ref. to Book Pid ref. to Publisher

In the semi-automatic process of extracting ontology from relational database we apply some rules like.

Rule1: If the primary key of any relation is unique and do not contain the primary key of any other relation then we consider such relation as on ontological class.

As per our Library example ***Member*** and ***Publisher*** are separate Ontological classes.

lib:Member (Mid,Mname,Addr,Mtype) **rdf:type** **rdfs:class.**

lib:Publisher (Pid,Pname,Addr) **rdf:type** **rdfs:class.**

Rule2: If the foreign key of any relation R1 is the Primary key of any other relation R2 then there exists an object property from R1 to R2 and the domain is R1 and range is R2.

As per our Library example members borrow books from library so Borrowerid from Borrow can be an object property with domain as Member and range as Book.

lib:Borrow rdf:type rdf:property.
lib:Borrow rdfs:domain rdf:Member.
lib:Borrow rdf:range rdf:Book.

Many more rules of such type can be applied to extract the ontology from the relational database. Following figure shows the extracted ontology where all the class and the data type properties are defined but some object type properties are not yet defined.

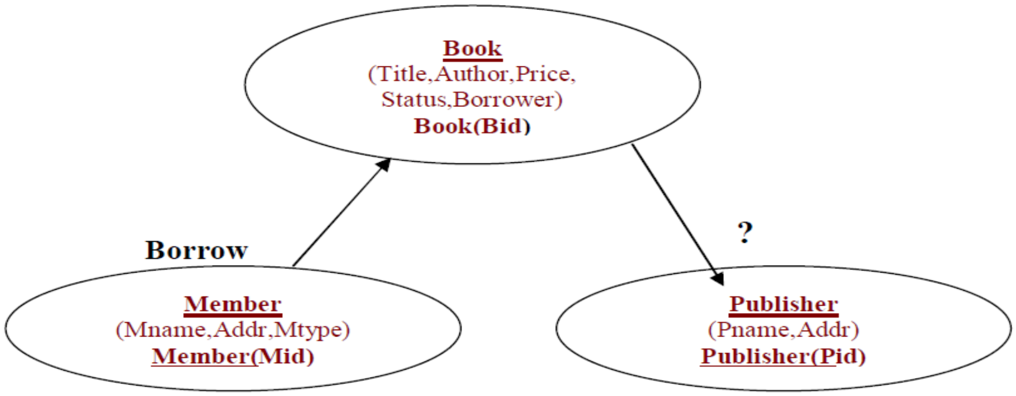

Figure 5: The Extracted Wrapper Ontology

The next step will be to adapt the extracted ontology to pre-defined domain ontology. This stage will add the explicit definition of the implicit relationships and adjust directions of the object properties between classes. For example the property (relation) between the Book and Publisher classes is not fully defined yet. Based on the domain ontology, the obtained full wrapper ontology is shown as:

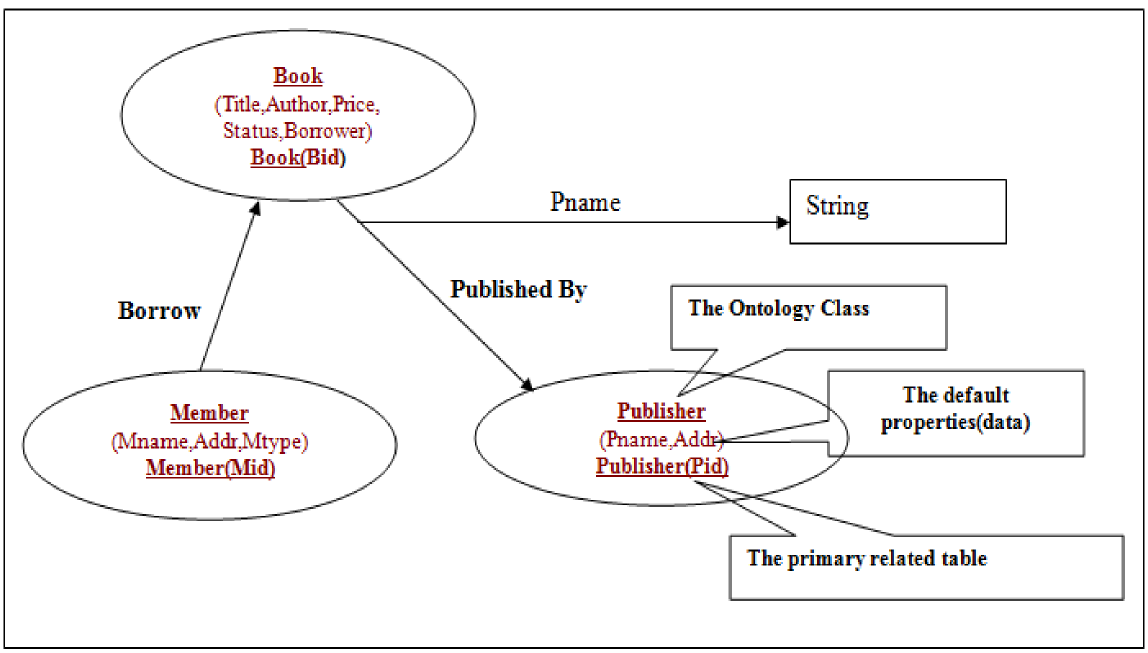

Figure 6: The Full Wrapper Ontology

The following RDF Schema Triples notation is the part of the whole schema.

……….
……….

lib: Book	rdf:type	rdfs:class
lib: Member	rdf:type	rdfs:class
lib: Publisher	rdf:type	rdfs:class
lib: PublishedBy	rdf:type	rdf:property
lib: PublishedBy	rdfs:domain	rdf:Book
lib: PublishedBy	rdf:range	rdf:Publisher

……….
(Here **lib** is the namespace for the extracted ontology)

4.2. SEMANTIC QUERY IN RDB (Relational Data Base)

After extracting and refining the wrapper ontology, the user can issue semantic queries based on extracted ontology concepts (Keywords) and these queries will be mapped onto plain syntactic SQL queries. The semantic queries will be based on SPARQL(pronounced as "sparkle") is an RDF query language, stands for **S**PARQL **P**rotocol **A**nd **R**DF **Q**uery **L**anguage. SPARQL can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware[7]. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports extensible value testing and constraining queries by source RDF graph. The results of SPARQL queries can be result sets or RDF graphs.

Following is the SPARQL syntax query to find the member details (Mname, Mtype, and Addr) who have borrowed the book titled "Data Mining" from the Library.

Prefix lib: http://www.libraryexample.cm
Select ?Mname ?Mtype ?Addr ?Title
Where {
?s lib:Mid ?p.
?p lib:Mname ?Mname.
?p lib:Mtype ?Mtype.
?p lib:Addr ?Addr.
?p lib:Borrow ?q.
?q lib:Bid ?r.
?r lib:Title?Title.
} ?Title="Data Mining"

The above SPARQL query is translated to SQL query by following procedure [3].

- The triples that share the same subject are grouped as they represent the same table information. So, each group represents some information about one concept in mediated ontology. For example, ?p subject group represent the Member information.
- Based on the mapping information, the translation algorithm replaces all predicates in the triples with corresponding columns name in relational databases tables. For example Mname, Mtype, and Addr are there in the Member table.

- If the predicate is not in the columns name, then it will be in object property names which are related to the linking tables. For example the Borrow property is not their in any table column. By inspection, it is in the mediated ontology linking the Member concept to the Book concept.
- For each separate group, a subquery clause is created, which consists of three parts: Select clause, From clause and Where clause. The Select clause is created according variable that occurs both in triple and in SPARQL select clause. The From clause is created according the column name in the triples. And the Where clause is created according the columns and mapping information. After all clauses are created, they are combined to construct the complete a query clause.

For the SPARQL query given above ,the SQL query translated is below.

Select Member.Mname , Member.Mtype , Member.Addr , Book.Title
From Member,Book
Where Member.Mid=Book.Mid and Book.Title=”Data Mining”;.

5. CONCLUSION

As we know that the traditional web includes a huge amount of Relational Databases (RDB) that support structuring data on a syntactic base. Converting available data stored in relational database into RDF format is tedious and error prone. So, instead of migrating available legacy data in relational database into RDF format based on ontology, we can build a wrapper that act as a translator from the semantic queries issued to the system into the syntactic data available within these databases. So this approach of extracting ontology from the RDB and allowing user to issue a semantic query and translating it into syntactic sql query and then transforming the results set obtained from relational databases into RDF triples using namespaces is more beneficial. There are many more concepts involved in mapping the semantic query using schema ontology into sql statements which can also be considered for implementation.

6. REFERENCES

[1] Gerd Stumme, Andreas Hotho, Bettina Berendt , “Semantic Web Mining State of the Art and Future Directions: A Survey” Web Semantics: Science, Services and Agents on the World Wide Web Volume 4, Issue 2, June 2006, Pages 124-14

[2] Bettina Berendt, Andreas Hotho, Dunja Mladenic,Maarten van Someren, Myra Spiliopoulou, Gerd Stumme , ” A Roadmap for Web Mining: From Web to Semantic Web”

[3] Mostafa E. Saleh , “Semantic-Based Query in Relational Database Using Ontology” , Canadian Journal on Data, Information and Knowledge Engineering Vol. 2, No. 1, January 2011.

[4] Grigoris Antoniou, Frank Van Harmelen , “A Semantic Web Premier” , Second Edition, MIT Press.

[5] Gerti Kappel, Elisabeth Kapsammer and Werner Retschitzegger, " Integrating XML and Relational Database Systems", World Wide Web: Internet and Web Information Systems, 7, 343–384, 2004, Kluwer Academic Publishers.

[6] D. Florescu, A. Levy, and A. Mendelzon, “Database techniques for the world wide web: a survey,” ACM SIGMOD Record 27(3), 1998.

[7] Chunying Zhou,, Yaowen Zheng, “Query Rewriting from SPARQL to SQL for Relational Database Integration” , IEIT Journal of Adaptive & Dynamic Computing, Vol. 1, No. 1, 1-8

[8] Ora Lassila, and Ralph R. Swick, “Resource Description Framework (RDF) Model and Syntax Specification” W3C Recommendation, 1999

[9] Dan Brickley, and R.V. Guha, “Resource Description Framework (RDF) Schema Specification”, W3C Recommendation, 1999.

[10] Peng Liu, Xiaoying Wang, Aihua Bao and Xiaoxuan Wang , “Ontology Automatic Constructing Based on Relational Database” Grid and Cooperative Computing (GCC), IEEE 9th International 2010.

[11] Xu Zhou, Guoji Xu andLei Liu , “An Approach for Ontology Construction Based on Relational Database” , International Journal of Research and Reviews in Artificial Intelligence Vol. 1, No. 1, March 2011.

[12] Olivier Corby, Leila Kefi-Khelif, Hacène Cherfi, Fabien Gandon and Khaled Khelif , “Querying the Semantic Web of Data using SPARQL, RDF and XML”

Authors

Ayesha Banu is a Post Graduate in M.Sc (CS) from Kakatiya University in 2002 and M.Tech (CSE) from JNTUH in 2009 and she is pursuing her PhD from JNTUH in the Area of “Semantic Web Mining”. She worked as a Lecturer in Masterji PG college from 2003 to 2006 and as Asst. Professor in Alluri Institute of Management Sciences from 2007 to 2009 and presently working as Associate Professor in Department of Informatics in same college since May 2010. She is a Life Member of ISTE and APSMS. She has delivered Guest Lectures in ACE Engg. Colege ,MGIT , Deccan college of Engg and GNITS. On Research issues in data mining , clustering techniques.

Dr. Syeda Sameen Fatima obtained B.Tech., Electronics and Communication Engineering from JNTU in 1982 , M.Phil., Computer Methods, University of Hyderabad, India in 1983 , M.S., Computer Science, University of Massachusetts, Amherst, USA in 1993 and Ph.D. in Computer Science and Engineering, from Osmania University, India in 2004. Her Research interests Include Information Retrieval Systems, Data Mining ,Artificial Intelligence, Machine Learning. She ha 25 years of teaching experience and is presently working as Professor and Head , Dept of CSE , Univ. College of Engineering , OU , Hyderabad. She has published papers in various National and International Journals and Conferences.

Khaleel Ur Rahman Khan obtained B.E. (CSE) from Osmania University in 1993 and M.Tech (CS) from JNTU in 1998. PhD in Computer Science from Osmania University in the area of Wireless Mobile Ad Hoc Networks. He is presently working as Professor and Dean at ACE Engineering. His research interests include Heterogeneous Networks, Opportunistic Networks, Transaction Management in Ad Hoc and Sensor Networks, Data and Web Mining. He has published over 30 papers in various Peer Reviewed International Journals and Conferences. He has teaching experience of over 18 years.