A Framework To Convert Relational Database To Ontology For Knowledge Database In Semantic Web

Vishal Jain, Dr. Mayank Singh

ABSTRACT: Now days, a large amount of data is available on the internet in the form of structured, un-structured and semi-structured data. Basically, today lots of data is available offline and online in the form of data, collection of tables, rows and columns. But now in the era Web 2.0, we have one other mechanism to store the data in the form of ontology, which are very useful for a community as a way of structuring and defining the meaning of the metadata that are currently collected and standardized. In this paper, we are going to discuss various techniques for mapping between the database and ontology.

Keywords: Semantic Web, Ontology Mapping, OWL, RDF, Relational Database

————————————◆————————————

1. INTRODUCTION

Applications using ontologies become more intelligent since they can more deal with human background knowledge. More generally, ontologies are critical for applications which want to merge information from diverse sources [1]. A large amount of data is present on the web. It contains huge number of web pages and to find suitable information from them is very cumbersome task. There is need to organize data in formal manner so that user can easily access and use them. The advance of the Web has significantly and rapidly changed the way of information organization, sharing and distribution [15]. The majority of data underpinning the Web and in domains such as life sciences [NCBI resources] and spatial data management [Green et al., 2008] are stored in RDB with their proven track record of scalability, efficient storage, optimized query execution, and reliability [11].

2. INFORMATION RETRIEVAL

We view the future web as combination of text documents as well as Semantic markup. Semantic Web (SW) uses Semantic Web documents (SWD's) that must be combined with Web based Indexing. Current IR techniques are not so intelligent that they are able to produce Semantic relations between documents. Extracting information manually with the help of XML and string matching techniques like Rabin Karp matcher has not proven successful. To use these techniques normal user has to be aware of all these tools. So, keeping this in mind we have moved to concept of Ontology in Semantic Web. It represents various languages that are used for building SW and increase accuracy.

3. SEMANTIC WEB

The semantic approach comes to solve the polysemy problem which that the same word may have different meanings according to the context of sentences [19]. In spite of many efforts by researchers and developers, Semantic Web has remained a future concept or technology. It is not practiced presently. There are few reasons for this which is listed below:

    (a) Complete Semantic Web has not been developed yet and the parts that have been developed are so poor that they can't be used in real world.

    (b) No optimal software or hardware is provided.

"**SW is not technology, it is philosophy**" [5]. It is defined as collection of information linked in a way so that they can be easily processed by machines. From this statement, we conclude that SW is information in machine form. It is also called *Global Information Mesh* [6]. It is also known as framework for expressing information.

4. ONTOLOGY

Ontologies are nowadays ubiquitous. Their number, their complexity and the domains they model are increasing considerably [2]. Ontology is defined as a formal specification of a shared conceptualization of some domain knowledge [9]. It involves identifying and extracting relevant pages containing that specific information according to predefined guidelines. There are many IR techniques for extracting keywords like NLP based extraction techniques are used to search for simple keywords. Then we have AeroText system for text extraction of key phrases from text documents.

5. MAPPING BETWEEN DATABASE TO ONTOLOGY

The conversion of Relational Database to Ontology is one of the key areas of research. Many researchers have come up with numerous techniques and tools for converting relational databases to ontology. Some of the proposed approaches include RDBOnto, Data Semantic Preservation, DB2OWL, R2O, D2RQ, Semantic Bridge, DartGrid Semantic and Semantic interoperability among others. This paper presents a comparison of the already developed frameworks and tools. In addition, this paper identifies the

_____

- *Vishal Jain, Dr. Mayank Singh*
- *Research Scholar, Computer Science and Engineering Department, Lingaya's University, Faridabad*
- *Associate Professor, Computer Science and Engineering Department, Lingaya's University, Faridabad*
- *vishaljain83@ymail.com, mayanksingh2005@gmail.com*

problems and deficiencies of these tools. Finally, this paper proposes a framework that can be developed to overcome the deficiencies and problems posed by other tools and frameworks.

5.1 COMPARISON OF ALREADY DEVELOPED TOOLS AND FRAMEWORKS

This section of the paper compares some of the well known frameworks used in the conversion of Relational databases to Ontology. The TDBToOnto is one of the renowned frameworks for converting relational databases to ontology. It is a vastly configurable technique that eases the process of designing and implementing ontology basing on a relational database [3]. TDBToOnto is a user-oriented device, which supports the access and input processes. The Asio Semantic Bridge is another renowned framework that uses the table and class approach [3]. The resulting ontology after the conversion consists of a class corresponding to each table in the database. The columns of tables appear as properties of the respective classes. The cardinality of a primary key is set to 1 while the cardinality of a nullable column is set to a maximum of 1. The developed ontology has rules that equate individuals depending on several primary key columns. The Semantic bridge approach helps to rewrite SPARQL queries to SQL and executes the SQL. Another renowned tool for converting relational database to ontology is the DataGrid Semantic Web kit. This tool has the capacity to map and query RDF generated from relational databases. The mapping process involves the conversion of a manual table into classes. This tool is one of the best tools because it has a visual tool that enables users to define mappings. The visual tool helps in the creation of SPARQL queries. The DataGrid Semantic tool then translates SPARQL queries into SQL queries depending on the defined mappings. Another powerful tool in the conversion of RDS to ontology is the DB2OWLtool. This tool is powerful because of its ability to automate the creation of new ontology from an existing relation database. DB2OWL creates ontology from a RDB by looking for particular cases of database tables. This helps to determine the component which will be created from a given database component. DB2OWL expresses the created ontology in the OWL-DL language, which uses Description Logics. The process of mapping begins by detecting particular cases for relations (tables) in the database schema. This tool converts each component (table, column and relation) of the database onto the relevant component (class, property and relation). The SOAM is a framework that represents tables as classes to predicate an approach for creating an ontology schema. This conversion process involves mapping the constraints of the relational model to those of the ontology schema. R2O is another well-known tool used in the conversion of relational databases to ontology. R2O is an XML-based language that helps to express mappings between the elements of relational databases and ontology. R2O mappings are useful in detecting any existing ambiguities and inconsistencies. Triplify is another useful tool in the conversion of relational databases to ontology. It helps to publish linked data and RDF from a relational database. Triplify accomplishes its purpose by mapping HTTP-URI request onto RDS queries expressed in a Structured Query Language. Triplify is a light-weight

application, which is easy to deploy and integrate in a wide range of web applications. Triplify does not support SPARQL, which other tools like Semantic Bridge and the DataGrid Semantic tool.

5.2 PROBLEMS AND DEFICIENCIES OF EXISTING TOOLS

From the above comparison, it is clear that there are numerous tools and frameworks for converting relational database to ontology. This indicates that new frameworks and tools are yet to be developed each new day. Each framework will come with its merits and demerits. In spite of the merits posed by each framework or tool, these frameworks and tools have some problems and deficiencies that need to be fixed. One of the key problems of these tools and framework is the difficulty in standardization. In spite of some few similarities, such as the use of structured query languages, these frameworks and, tools use different approaches, which are difficult to integrate and form a standard tool or framework. Most the tools and frameworks explicated in this paper use diverse representation formats, as well as tool-specific languages. In this case, only experts are able to re-use the crucial artifacts used in the mapping process. Some of the domain-specific problems are eminent in the BD2OWL framework that uses the OWL-DL language, which cannot be implemented in other domain-specific, concept frameworks, tools and applications. Therefore, existing frameworks lack the ability to help users re-use the artifacts. They do not use languages that the user can easily understand or learn. For instance, the use of SPARQL creates a hard task for users who have little knowledge of programming language semantics. This creates a need to integrate a language module that enables users to select and use the most convenient language for converting relational databases to ontology. There is a deficiency of a module that converts programming languages into a simple language for non-programmer to re-use in case the need to re-use code arises. Some relational database conversion tools such as R2O and D2R map use a manual mapping definition process, which is less preferable compared to an automatic process. Manual mapping is not reliable compared to automatic mapping. Therefore, these tools do not consider conversion time as a key factor in the conversion of RDB to ontology. It is extremely vital to address this deficiency by using a software module that automates the mapping process. This is an excellent mechanism of ensuring that the conversion process does not take a lot of time because of manual processes. It is a recommendable mechanism of ensuring that the user does not waste time, which can be allocated to other complicated processes such as querying and searching from a database. Another problem experienced in the current frameworks is the support for only few databases. For instance, BD2OWL supports Mysql and Oracle databases, leaving out other databases such as Microsoft Access. Therefore, there is a deficiency of a framework that can convert different databases from different database programs into ontology.

5.3 NEW PROPOSED FRAMEWORK WHICH CAN BE DEVELOPED

In order to overcome the problems of the existing frameworks and tools, it is critical to design an appropriate

framework. This paper proposes a framework similar to BD2OWL but has additional features that address the identified problems and deficiencies. This implies that the proposed framework can support many databases as possible and the most used programming languages. In addition, the new framework has the capacity to output information in different formats, which non-programmers can under re-use without the need for an expert. However, the proposed framework will not depend on particular table cases. It is a general framework that is applicable to all tables, whatever the case. The proposed mapping process involves converting tables into classes, which have several properties, as well as relationships. The conversion process will start when the user uses a well-designed user interface to sent queries to the database. The proposed visualization service must be able to present the required queries in a suitable manner. In order to consider the requirements if different users, including those who do not have programming skills, the visualization service should have an interface that has select option for users to key-in commands in a desired language. This should consider all the available programming languages as well as human language, which diverse users can understand. Therefore, it is extremely critical to include a module that translates the input text and instruction into different programming languages. In addition, the new framework ought to incorporate a module that enables users to export information into different formats apart from the default format. Users must be able to output information in the form of text files, tables and datasets among others. The table name (T_i) must match to the class name $(C_i.name)$ and each property (Y_i) must match with each column (L_i) of the table. In addition, The primary Key of each table $(P(T_i))$ must match to a class id $(C.id)$ in the generated ontology. The same applies for the foreign keys in order to ensure maximum referential integrity. The subscripted "i" refers to the number and, therefore, the subscript for each element depends on the total number of elements. For instance of there are five tables in the relational database, the each table (T) will have a subscript ranging from 0 to 4, and each table must have a unique subscript. Therefore, the number of classes created in the ontology will be the same as the number of tables in the relational database. In addition, the number of class properties will be the same as the number of columns in the database. The same applies for the primary keys and the foreign keys. In order to overcome the problems and deficiencies of existing, there will be a dynamic mapping mediator, which enables the user to convert data from multiple database sources and store it into easily readable text files. The new system ought to support heterogeneous data sources of data such as preformatted text files, MS Excel, MS Access, MySQL and Oracle among others. The new framework will have the capability to automate the translation of SPARQL queries using the mappings of a mediator class.

CONCLUSION

This paper emphasis on the concept of Ontology Mapping, discuss various approaches for converting relational database to ontology and vice-versa. It is evident the conversion of relational databases to ontology is a diverse process and the frameworks and tools used are diverse. These frameworks and tools have their merits and demerits. Data presentation and output formats and languages are crucial concerns. The proposed frameworks will ensure that there is maximum data integrity in after conversion. In addition, it offers users the ability to customize queries depending on their literacy level. Automation is also a critical part of the proposed framework.

ACKNOWLEDGEMENT

REFERENCES

[1].   http://conferences.telecom-bretagne.eu/data/odbis2008/cfp_ODBIS2008.pdf

[2].   Moussa Benaissa and Yahia Lebbah, "A Constraint Programming Based Approach to Detect Ontology Inconsistencies", The International Arab Journal of Information Technology(IAJIT), Vol. 8, No. 1, January 2011

[3].   Satya S. Sahoo, Wolfgang Halb, Sebastian Hellmann, Kingsley Idehen, Ted Thibodeau Jr, Sören Auer, Juan Sequeda, Ahmed Ezzat, " A Survey of Current Approaches for Mapping of Relational Databases to RDF", W3C RDB2RDF Incubator Group January 08 2009

[4].   Hadjila Fethallah and Chikh Mohammed Amine, "Automated Retrieval of Semantic Web Services: A Matching Based on Conceptual Indexation", The International Arab Journal of Information Technology (IAJIT), Vol. 10, No. 1, January 2013 61

[5].   Accessible from T.Berners Lee. "The Semantic Web". "Scientific American", May 2007.

[6].   Urvi Shah, James Mayfield," Information Retrieval on the Semantic Web", "ACM CIKM International Conference on Information Management", Nov 2002.

[7].   Michael Wick, Khashayar, Rohanimanesh, Andrew McCallum, AnHai Doan, "A Discriminative Approach to Ontology Mapping", VLDB '08, August 2430, 2008, Auckland, New Zealand

[8].   Zhuoming Xu, Shichao Zhang, and Yisheng Dong, Mapping between Relational Database Schema and Owl Ontology for Deep Annotation, WI'06: Proceedings of the 2006 IEEE/WIC/ACM InternationalConferenceon Web Intelligence, IEEE Computer Society, 2006, pp. 548-552.

[9].   Zakaria Elberrichi, Abdelattif Rahmoun, and Mohamed Amine Bentaalah, "Using WordNet for Text Categorization", The International Arab Journal of Information Technology (IAJIT), Vol. 5, No. 1, January 2008

[10]. R. Ghawi and N.Cullot, "Database-to-Ontology Mapping Generation for Semantic Interoperability", 2007

[11]. Abdelkader Dekdouk, "Ontology-Based Intelligent Mobile Search Oriented to Global e-Commerce", The International Arab Journal of Information Technology, Vol. 7, No. 1, January 2010

[12]. Nadine Cullot, Raji Ghawi, and Kokou Ytongnon, Db2owl: A Tool for Automatic Database-to-Ontology Mapping, SEBD, 2007, pp.491-494.

[13]. Green, J., Dolbear, C., Hart, G., Engelbrecht, P., Goodwin, J., "Creating a semantic integration system using spatial data", in International Semantic Web Conference 2008 Karlsruhe, Germany.

[14]. Hu, W., Qu, Y., "Discovering Simple Mappings Between Relational Database Schemas and Ontologies", In Proc. of 6th International Semantic Web Conference (ISWC 2007), 2nd Asian Semantic Web Conference (ASWC 2007), LNCS 4825, pages 225- 238, Busan, Korea, 11- 15 November 2007.

[15]. Eduard Dragut Ramon Lawrence, Composing Mappings Between Schemas Using a Reference Ontology, International Conference on Ontologies, Databases and Application of Semantics (ODBASE), Springer, 2004, pp.783-800.

[16]. Yassaman Zand Moghaddam and Joe D. Horton, "RELATIONAL DATABASE SCHEMA TO ONTOLOGY MAPPING APPROACHES", 10th November 2010

[17]. Sidi Benslimane, Mimoun Malki, Mustapha Rahmouni, and Adellatif Rahmoun, "Towards Ontology Extraction from Data-Intensive Web Sites: An HTML Forms-Based Reverse Engineering Approach", The International Arab Journal of Information Technology (IAJIT), Vol. 5, No. 1, January 2008

[18]. Petros Papapanagiotou, Polyxeni Katsiouli, Vassileios Tsetsos, Christos Anagnostopoulos and Stathes Hadjiefthymiades, "RONTO: RELATIONAL TO ONTOLOGY SCHEMA MATCHING", AIS SIGSEMIS, 2005

[19]. Guntars Bumans, "Mapping between Relational Databases and OWL Ontologies: an Example", Scientific Papers, University of Latvia , 2010. Vol. 756 Computer Science and Information Technologies

[20]. C. Kavitha, G. Sudha Sadasivam, Sangeetha N. Shenoy, "Ontology Based Semantic Integration of Heterogeneous Databases", European Journal of Scientific Research, ISSN 1450-216X Vol.64 No.1 (2011), pp. 115-122

[21]. Fuad Mire Hassan, Imran Ghani, Muhammad Faheem, Abdirahman Ali Hajji, " Ontology Matching Approaches for eRecruitment", International Journal of Computer Applications (0975 – 8887) Volume 51– No.2, August 2012

[22]. Mohd Amin Mohd Yunus, Roziati Zainuddin, and Noorhidawati Abdullah, "Semantic Method for Query Translation", The International Arab Journal of Information Technology (IAJIT), Accepted in May 24, 2011

[23]. Mahamat Hassan and Azween Abdullah, "A New Grid Resource Discovery Framework", The International Arab Journal of Information Technology (IAJIT), Vol. 8, No. 1, January 2011

[24]. Bouchiha Djelloul, Malki Mimoun, and Mostefai Abd El Kader, "Towards Reengineering Web Applications to Web Services", The International Arab Journal of Information Technology (IAJIT), Vol. 6, No. 4, October 2009

[25]. Sidi Benslimane, Mimoun Malki, and Djelloul Bouchiha, "Deriving Conceptual Schema from Domain Ontology: A Web Application Reverse Engineering Approach", The International Arab Journal of Information Technology (IAJIT), Vol. 7, No. 2, April 2010

About the Authors

Vishal Jain has completed his M.Tech (CSE) from USIT, Guru Gobind Singh Indraprastha University, Delhi and doing PhD from Computer Science and Engineering Department, Lingaya's University, Faridabad. Presently he is working as Assistant Professor in Bharati Vidyapeeth's Institute of Computer Applications and Management, (BVICAM), New Delhi. His research area includes Web Technology, Semantic Web and Information Retrieval. He is also associated with CSI, ISTE.

Dr. Mayank Singh has completed his M. E in software engineering from Thapar University and PhD from Uttarakhand Technical University. His Research area includes Software Engineering, Software Testing, Wireless Sensor Networks and Data Mining. Presently he is working as Associate Professor in Krishna Engineering College, Ghaziabad. He is associated with CSI, IE (I), IEEE Computer Society India and ACM.