Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ONTOLOGY MAPPING FOR INTEROPERABILITY REPRESENTATIONS IN HOTEL DOMAIN

[1]A. JUARNA, [2]R. REFIANTI, [3]A.B. MUTIARA, [4]D.I. PRATIWI

[1]Assoc.Prof., Faculty of Computer Science dan Information Technology, Gunadarma University, Indonesia

[2]Asst.Prof., Faculty of Computer Science dan Information Technology, Gunadarma University, Indonesia

[3]Prof., Faculty of Computer Science dan Information Technology, Gunadarma University, Indonesia

[4]Alumni, Faculty of Computer Science dan Information Technology, Gunadarma University, Indonesia

E-mail: [1,2,3]{ajuarna,rina,amutiara}@staff.gunadarma.ac.id, [4] dianirma.here @gmail.com

ABSTRACT

With the development of information technology today is the nature of today's information is also experiencing growth. These developments encourage a more dynamic source of information, autonomy, variety and large. Diversity occurs not only at the technical level but also at the level of representation of information. This difference can occur at syntactic, schematic and semantic level. This diversity also occurs in the hotel domains that do not have a standard that refers to the tourism domain. The existing traditional methods have not been able to overcome the problem of diversity to achieve interoperability in heterogeneous data sources. So with this method the obtained results are not necessarily relevant to the user wishes. One of the most viable approaches is the creation "Query Rewriting" based on ontology and mapping. As a mapping approach it is used GLAV (Global-Local-As-View) which is a combination of approaches GAV (Global-As-View) and LAV (Local-As-View). While the diversity of existing terminology in each source of data for specific domains basically state the same meaning as defined in the common ontology. In this paper it will be explained the illustrations of "Query Re-writing" and the steps to create mappings for the hotel domain using the Protege 3.4 as tool.

Keywords: *Interoperability, Mapping, Common Ontology, Semantic web, Query re-writing*

1. INTRODUCTION

Thousands of organizations and individuals are making their repositories available online. To exploit the full potential of these sources, modern information systems and Web applications must be able to retrieve, integrate, and exchange data. Unfortunately, the repositories and applications are developed by different people, at different times, with varying requirements in mind. Thus, the underlying data is inherently highly heterogeneous. To cope with the heterogeneity and achieve interoperability, a fundamental requirement is the ability to match and map data across different formats. These two tasks are found in the literature under the names *matching* and *mapping,* respectively. A match is associations between individual structures in different data sources. Matches are the required components for every mapping task. The mappings are the products of the latter. A mapping, in particular, is an expression that describes how the data of some specific format is related to data of another. The relationship forms the basis for translating the data in the first format into data in the second [1].

Mappings can be found in almost every aspect of data management. In information integration systems, mappings are used to specify the relationships between every local and the global schema. In schema integration, mappings specify how an integrated schema is constructed from the individual input schemas. In data exchange and P2P settings, mappings are used to describe how data in one source are to be translated into data conforming to the schema of another. A similar use is found in schema evolution where mappings describe the relationship between the old and new version of an evolved schema.

In this paper, we used the ontology based mapping in Tourism Information System. Tourism Information System utilized computer and information technology, commonly is known by the E-Tourism term. One of the characteristics of E-

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Tourism is a growing need for interoperability among the tourism organizations. The problem in interoperability is the diversity of information that makes the exchange of information becomes an obstacle that is difficult reconcile the absence of effective solutions.

As the sites of existing hotels, they do not have a standard that refers to the tourism domain. The source of the information provided has a different method of presentation even though the purpose of making the presentation of the site is the same. For example, if tourists want a holiday to Bali and want to stay at the Nikko Bali Hotel, then search for room availability of information through search engines by submitting a query "looking for available rooms for checkin checkout on 11 August 2011 on 15 August 2011 and room number 2", then the problem that arises is if the Nikko Bali Hotel use different terminology to query the user the desired results will not be obtained. Therefore, it is needed an ontology-based query rewriting and mapping by adapting concepts/terminology used in the data source is called the Query re-writing by using Semantic Web approach.

2. INTEROPERABILITY

Interoperability, as illustrated by the Institute of Electrical and Electronics Engineers (IEEE) [6], is "the ability of two or more systems or components to exchange information and use the information that has been exchanged".

The term interoperability includes some ideas / other ideas:

- Interoperability in engineering. Approval of the communications, transport, storage space and the necessary overview.
- Semantic Interoperability. Common ontology (Common Ontology) or thesauri are required to avoid the use of different terminology for the same meaning or similar terminology to their different meanings.
- Interoperability political. Resource sharing may involve a change in business processes of agencies / institutions.
- Interoperability between the groups. Partnerships and agreements between agencies /institutions are often required before information can be shared among them.
- Interoperability legally. Information sharing must follow legislation (e.g. resources where data protection laws apply).
- International Interoperability. Languages and cultural challenges may present new problems to solve.

3. SEMANTIC WEB

Semantic Web is a network that is able to understand not only the meaning of a word and concept, but also the logical relationships between them. Since most of the vocabulary knowledge is built upon two main pillars of semantic and mathematics, the Semantic Web has a great potential. On the Semantic Web, information will be tagged, so the computer can understand its meaning. And, a so-called intelligent software agent will understand the meaning and context of the word [2,6,8-9].

4. KOMPONEN-KOMPONEN SEMANTIC WEB

Developing the Semantic Web is made possible by the existence of a set of standards which are coordinated by the World Wide Web Consortium (W3C). The most important standards in building the Semantic Web is ontology, XML, XML Schema, RDF, OWL, and SPARQL. Here is a layer of the Semantic Web as recommended by the W3C (www.w3c.org)[8-9]:

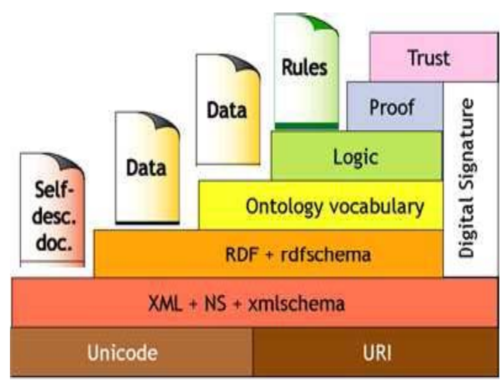

Figure 1. *Ontology Layer*

4.1. *Ontology Definition*

In [6], the definition of ontology is:

- One branch of metaphysics that focused on nature and the relationships between living things;
- Theories about the nature of living things.

Ontology is a theory about the meaning of an object, the properties of an object, and object relations that may occur in a domain of knowledge.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.2. Ontology Language

To be used, ontology must be expressed in real notation. An ontology language is a formal language of an ontology development. Some components of the structure of the ontology, among others:

- XML (Extensible Markup Language) Provide output syntax for structured documents, but have not been forced to XML documents using semantic constrains.
- XML Schema. Language for restrictions of XML documents structure.
- RDF (Resource Description Framework). Data model for objects ('resources') and its interrelation, providing a simple semantic for the data model, and data model can be expressed in XML syntax.
- RDF Schema is a vocabulary to describe properties and classes of RDF resources, with semantics for generalization hierarchies of properties and classes.
- OWL (Web Ontology Language). Adding some vocabulary to describe properties and classes, among others: the relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, various types of properties, characteristics of properties (e.g. symmetry), enumerated the classes.

4.3. Query Representation

Most of the data that exist in the world are found in the database, especially relational databases. Even larger amounts of data are in regular files, email archives, and other like kinds. Integration of all these data will provide tremendous benefits for organizations that have data.

RDF is a common language for the web that can represent uniform data from different sources. SPARQL, a query language for RDF, can combine data from different databases, as well as documents, or others who may claim to knowledge as a labeled directed graph.

SPARQL is a good language to unify the data in a relational database with other databases, as well as other data sources.

5. GLOBAL-AS-VIEW, LOCAL-AS-VIEW DAN GLOBAL-LOCAL-AS-VIEW

There are two important things in centralized data integration, i.e. system modeling and query processing. To model the relationship of data sources and common ontology, two basic approaches have been proposed [3,4]. The first approach, called Global-As-View (GAV), common ontology is expressed in terms of data sources. The second approach, called the Local-As-View (LAV), requires a common ontology defined independently of the source, and the relations between the common ontology and data sources are established by determining each data source is formed by defining every source as a view on the common ontology. In addition to two approaches above is the third approach, Global-Local-As-View (GLAV). GLAV approach is an approach that is a combination of LAV approach and GAV.

Comparison of the three approaches above can be shown in the table below.

Tabel 1. Comparison between GAV, LAV dan GLAV

	GAV	LAV	GLAV
Query Processing	More easy	More difficult	More easy
Modeling	Not able to model the situations of integrating where the source is lost in order to build a complete word view	Is designed to address something that is dynamics, in situation that source set is not perfect (adding, subtracting, or changing the content of data source)	Is easier to be expand-ed: Adding one new source doesn't need to change the mapping schema.
Work-example	TSIMMIS, Garlic, MOMIS, DIKE, Squirrel	Information Manifold, Info master	TIQS (Target based Integratio n Query System

6. QUERY RE-WRITING

Query can be generated by a combination of *class* or *property* of a given ontology. *Query re-writing* is a process to rewrite the original query into a new query by adjusting or terminology concepts used in each data source is incorporated in an integrated system. *Query re-writing* is done with respect to different concepts in the use of terminology to represent data from the users of each source [6].

6.1. Illustration of Query re-writing

Query re-writing is something important in the search data on a system with heterogeneous data sources. To facilitate understanding of the query re-

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

writing it can be used as an illustration as in Figure 2 which shows a UserView schema, mapping to a data source that is Nikko Bali Hotel. UserView use the terminology "**checkin**" and "**checkout**", while in the data source using the "**arrival**" and "**departure**".

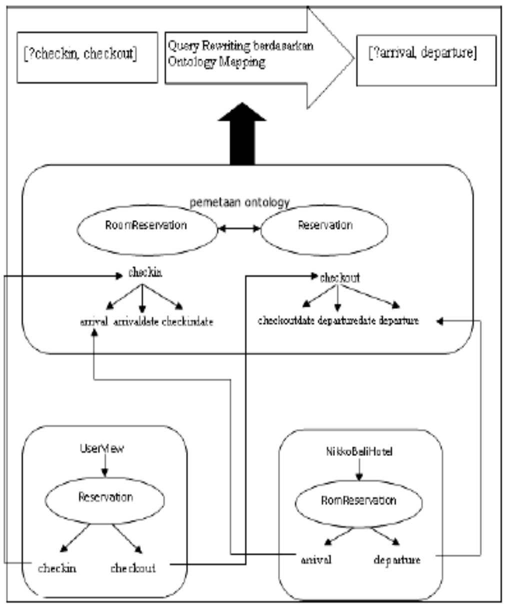

Figure 2. *Query re-writing* Processing using *Ontology Mapping*

In Fig.2 it is shown that the terminology used by UserView and data sources are different. Also in Fig.2 it is shown that "**checkin**" is mapped to "**arrival**" and "**checkout**" is mapped to the "**departure**", so the terminology is different but has the same meaning, and it can be bridged by a mapping.

In a data integration in which the source is described as UserView of common ontology, query processing is called view-based query processing which has two approaches, namely the view-based query rewriting and view-based-query-answering.

The approach used to perform mapping of the Global-Local-As-View (GLAV) is a combination of mapping approach GAV and LAV. Because there are two stages of mapping used. LAV is used for mapping of UserView to a common ontology, while GAV mapping approach used for mapping in the common ontology to the data source schema.

A common ontology used is a liaison between UserView to a source of data, for example if a user gives a query, the user's query will be sorted from the query structure which is given by the user into a UserView. After that it carried out a mapping from UserView into a common ontology; having obtained the appropriate terminology we have to do more of the common ontology mappings to data sources obtained in according to the terminology to obtain the desired results.

To handle the used queries we apply a query re-writing strategy. A query processing can occur in the two following directions:

- *UserView-to-Global Query rewriting*. Given a query q which is presented in a UserView. Extraction process is needed in order to find the terminology used by q. Extraction results are then mapped to the common ontology so we find the terminology that is known in the common ontology.
- *Global-to-local Query rewriting.* After the first mapping results in a terminology which is known in the common ontology, then the resulting terminology is used for the second mapping, i.e. the mapping of common ontology to the data source. The results of the second mapping will be used to reshape the query based on terminology that is known in the data source.

According to Wulandari [10] query rewriting process can be summarized with the following explanation: user presents a query q on a common ontology, the system rewrite q into a collection of sub-queries q, one for each data source. Then sub-queries are executed on the data sources to obtain the answers, which is then integrated to produce the answer of the query q. To clarify we can see the following example, suppose that a query from UserView containing "**looking for available rooms for arrival on 11 August 2011 departure on 15 August 2011 and room number 2**". If suppose that there are two sources of data, it will look like Fig. 3.

In Fig.3 it is shown that the first it would be done a segregation of the structure of a given user query. Then the terminology used will be taken where this terminology suggests the property is used to form a query from the user side. Then by applying the function changes in terminology (e.g. arrival ≈ check-in), it is obtained terminology in the common ontology. The results of the mapping between UserView with the common ontology are the terminology that will be mapped to each source. In Fig.3 it shown that arrival will be mapped to the arrival ≈ arrivaldate and checkindate through the arrival ≈ arrivaldate change function, where terminology checkindate is a element of source1 and arivaldate is an element of source2.

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

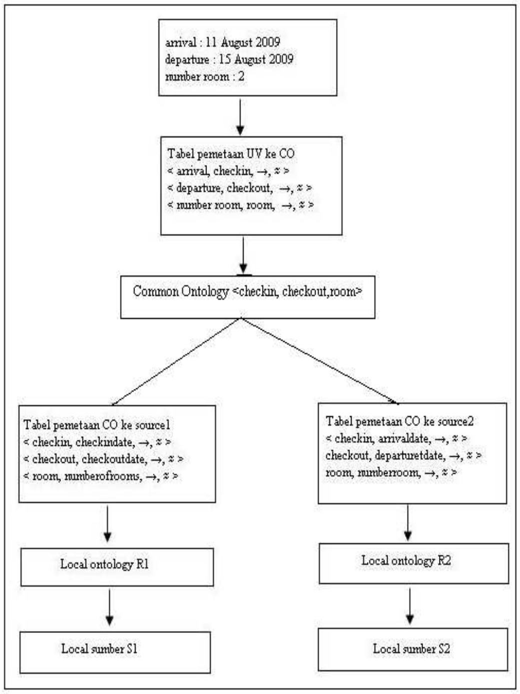

Figure 3. Example of flow of mapping process

In this paper, Common ontology mapping and data sources in RDF/OWL is expressed by the owl: equivalent*Property* which shows a word equivalent that has the same meaning. In the example of this mapping "checkin" is mapped to the terminology "checkindate", arrival and arrivaldate. Terminology checkout is mapped to the term "checkoutdate", departure and departuredate. While the terminology "numberrooms" is mapped to the terminology rooms, numberofrooms, numbersuites, totalroom and quantityofrooms. Examples of these mappings can be shown in Fig. 4.

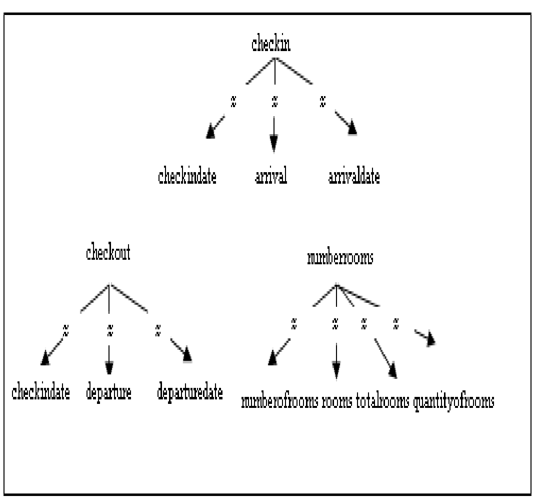

Figure 4. Example of mapping between *Common Ontology* and data source

7. STEPS TO BUILDING SIMULATION OF *QUERY RE-WRITING*

In this section it is explained the steps to build a simulation of query re-writing. The steps are as follows.

1. The first step is to see the sites from which data sources will be used to look at the terminology used in each hotel. This terminology will be used to build a common ontology. A common ontology terminology used must be complete so that the terminology in all data sources can be represented.
2. The second step is the formation of the terminology used on the user side (UserView), establishment of common ontology and rebuilding the data from the source data in the OWL representation using software tools Protege 3.4.
3. The third step is the stage of mapping. The mapping is to connect the user terminology and terminology in the data source. In this mapping, two phase mapping was carried, i.e. the mapping between the view of the user (UserView) with a common ontology. The second mapping stage is the stage of mapping from common ontology to data sources. The process of mapping done using a software tool Protege 3.4.
4. AWK programming to process query rewriting. AWK programming has an advantage in the manipulation of strings. Due to represent re-query the origin of the user, the query command should be modified in accordance with the usage in the data source [9]. Prior to obtaining a complete query, the command-forming elements must be made first by doing string manipulation of the previous command, and string manipulation of the mapping.
5. Representation of Query rewriting using SPARQL. SPARQL is a formal query language.

8. CONCLUSION AND FUTURE WORK

Ontology mapping is a mapping process of the terminology-terminology that is used so that the different terminology to a data source can be bridged with the mapping of UserView to a common ontology that is a combination of ontology-ontology that is used in the data source and mapping of common ontology to the data source in order to find the desired terminology in accordance with the user desires.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Ontology mapping is designed to address the problem of interoperability in the domain of HOTEL that is the diversity of terminology that is used at each hotel. Different terminology used each hotel are obtained after an average of 5 observations terminology that is checkindate-arrivaldate-arrival-checkin, checkoutdate-departure-departure-date-checkout, quantityofrooms-rooms-numbersuites, numberrooms, roomsname-room-type-suiterequired , and roomrates-suiterates-room rate-raterates. This ontology mapping will be used as the formation of a Query re-writing.

At this paper, we have limited our work to the mapping. For further development it may take a few steps away to form a query re-writing. So it is better if the mapping is continued until the stage of query-rewriting using AWK programming language or other language which has the advantage in string manipulation.

REFERENCES:

[1] Z. Bellahsene, A. Bonifati, F. Duchateau, and Y. Velegrakis, *On Evaluating Schema Matching and Mapping*. In: Schema Matching and Mapping, Z. Bellahsene et.al. (eds), Springer, Berlin-Heidelberg, 2011, pages 253-292.

[2] Anonim. ONTOLOGY: BAHASA DAN TOOLS PROTEGE. http://paperwgdbis.abmutiara.info/tutorial/Bahasa_tool_ontology.pdf, download July 2010.

[3] D. Briukhov, K. Leonid, and M. Dimitri. Source Registration and Query Rewriting Applying LAV/GLAV Techniques in a Typed Subject Mediator. In *Proceeding of the 9th Russian Conference on Digital Libraries RCDL '2007*. Pereslavl, Rusia.

[4] Cali, D. Calvanese, G. Giacomo, and M. Lenzerini. Models of Information Integration: Turning Local-As-View into Global-As-View. In *Foundation of Models for Information On Line proceedings*. http://www.fmldo.org// FMII-2001.

[5] D. Calvanese, G. Diacomo, M. Lenzerini, and M. Y. Vardi. Answering Regular Path Queries using Views. In *Proc. Of the 16th IEEE Int. Conf. on Data Engineering (ICDE 2000)*, pages 389_398.

[6] M. C. Daconta, L. J. Obrst, and K. T. Smith. *A Guide to the Future of XML, Web Services and Knowledge Management*. Wiley Publishing, Indianapolis, Indiana, 2003.

[7] IEEE. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York, NY, 1990.

[8] W3C. http://www.w3.org/tr/2004/rec-owl-features-20040210/, 2004.

[9] W3C. RDF and SPARQL: Using Semantic Web Technology to Integrate the World's Data. http://www.w3c.org// 2007/03/VLDB/, 2007.

[10] L. Wulandari. Query Rewriting Berbasis Ontology dan Mapping Untuk Mencapai Interoperabilitas Pada Sumber Data Heterogen Di Tingkat Semantic. Dissertasi, Universitas Gunadarma, 2009.

AUTHOR PROFILES:

Achmad Benny Mutiara was born in Jakarta, in 1967, and is a Professor of computer science at Gunadarma University. He received the B.S degree in Dept. of Physics from University of Indonesia and Dept. of Informatics Engineering from Gunadarma University, Indonesia, in 1991. He also received the M.S and PhD degrees in Computation from Universitaet Goettingen, Germany, in 1996 and 2000, respectively. He is Dean, Faculty of Computer Science and Information Technology at Gunadarma University. His current interests are Computer Modeling and Simulation, parallel computing and Computational Science.

Rina Refianti is with Faculty of Computer Science and Information Technology at Gunadarma University. She is a Assistant Professor, received the B.S. and M.S. degree in Dept of Information system from Gunadarma University, in 1991 and 2003, respectively.

A. Juarna is an Assoc.Prof and combinatorlist at Faculty of Computer Science and Information Technology, Gunadarma University, Indonesia. He got his Ph.D *dual degree* in Combinatorics from Universite de Bourgogne-France under supervising of Prof. Vincent Vajnovszki and from Gunadarma University under supervising of Prof. Belawati Widjaja. Some of his papers were presented in some conference such as Words-2005, CANT-2006, GASCom-2006, and some others are published in some journals or research reports such as CDMTCS-242 (2004), CDMTCS-276 (2006), The Computer Journal 60(5)-2007, Taru-DMSC 11(2)-2008.