Ontology and Semantic Web Approaches for Heterogeneous Database Access

Mohd Kamir Yusof, Mohd Nordin Abdul Rahman and Mat Atar Mat Amin
Fakulti Informatik, Universiti Sultan Zainal Abidin
Terengganu, Malaysia
mohdkamir@unisza.edu.my

Abstract

This paper presents about implementation of ontology and semantic approaches for accessing data from heterogeneous database. The purpose of this implementation is to retrieve relevant data and give relevant information to web users. The process of ontology and semantic are related. Two major processes in ontology are semantic mapping and extracting wrapper ontology. The main process for semantic web is to manipulate data into meaningful information. In this research, simple application was develop by using JAVA technology. JAVA technology was chosen because this technology have Jena library. This library is provide API and support SPARQL. After that, several experiments were done, and the results indicate implementation of ontology and semantic web approaches able to show relevant information to web users.

Keywords: Ontology, Semantic Web, Heterogeneous Database

1. Introduction

Web application provided lot of information to web users. The web users can search information from web based application. The information can help web users to make any decision. In organizations, decision maker will make a decision based on the information. Information is coming from extracting data in database. Data can be coming from single database or heterogeneous data. Most of web applications stored data into a single database. A single database stored lot of data. Increasing number of data will reduce the space size of storage database. In this case, performance of web query processing become slow to search and retrieve lot of data from a single database. This problem will effect to number of relevants information will display to web users. Most of web users aim to get relevants information from web application. In fact, very difficult to determine what the really users wants [7]. Based on the past researchers, several research effort has been attempt to solve this problem such as query refinement, etc. These effort is to help the user identify better terms for a query [1]. Nevertheless, there is no techniques that enables to identify when the results of the query are relevant [7]. In this research, ontology and semantic web approaches was applied by implementation of web based application. The purpose of this development is to evaluate capability of these approaches in delivering relevants information to web users.

2. Heterogeneous Database System

Heterogeneous database approach is differ compare to single database. The heterogeneous database is coming from more than one database. Based on past
researcher, HDBS was implemented in medical institutions. Hospital is one example using heterogeneous databases approach. The purpose of this implementation is to provide uniform access to electronic patient records in hospital computing environment that uses a MUMPS hierarchical database for storing patient demographic data and a Sybase relational database for storing patient laboratory results [14]. However, the core of the heterogeneous problem in hospital environment is that independently developed and maintained databases are heterogeneous with respect to their query models [14]. Heterogeneous databases approach also has been implemented in geospatial data. Implementation of heterogeneous databases in geospatial data involved three modules; a vector/vector integration module, a raster/vector integration module, and a databases module. Databases architecture was designed to preprocess inputs and to store and export results of the vector/vector and the raster/vector integration steps [8]. The unprecedented increase in the availability of information due to the success of the World Wide Web has generated an urgent need for new and robust methods that simplify the querying and integration of data [11]. A lot of research in the past focuses on developing methodologies for querying heterogeneous data sources. Integration data from existing databases in a distributed environment will give the impact of operations on the databases [9]. One of the approaches in database integration is unified global integration [11]. The purpose of this approach is to facilitate efficient global processing. However, this approach becomes hard to manage as the number and types of data sources increase. Another approach in database integration is system based on mediators and wrapper [11]. This approach is sophisticated that abstract the data sources from the users. The wrapper-mediator is remarkable scalable, and allows the integration of an increasing number of data sources. In this research heterogeneous database concept was implemented.

4. Motivation of Semantic Web

Semantic web is one approach can be implement in web based application. The purpose of this implementation is to produce more accurate information to web users. The main proses in the semantic web is to manipulate data from heterogeneous database into meaningful information. Figure 1 shows four major layers involved in the semantics web.

![Fig 1: Four Major Layers in Semantic Web](image)

XML is powerful technique for exchange and transporation data. XML as data structure language is allows users to add arbitrary structure to their documents. Programmer must know what the document creator means by each structure as XML alone does not capture information about the structures meaning [5]. RDF represent meaning, as users use metadata to describe Web resources and improve in RDF as encoded triples which states that particular
resources (subjects) have properties (predicates) with certain value [13]. The second top layers is ontology. Ontology provides a shared and common understanding of a domain that can be communicates between people and heterogeneous and distributed application systems.

5. Ontology Concepts in Web Query

A concept of the ontology may represent either semantic or linguistic information. The rationale is that ontology should represent the synonyms of each concept, when a concept is a compound name composed for two or more nouns. For instances, web users request about “WorkHard”. A semantic concept in ontology called WorkHard and the ontology should also contain linguistic information which indicate that its name a noun comprised of two words: Work and Hard, and is the denotation of the word Hard. The ontologies in web query processing can help user get more relevant information. Most of ontologies used to support web queries neither represent the instances of the conceptualization they are modeling, nor the linguistic information about their concepts. However, the problem about web users satisfaction still exist, because to improve correctly the user request, linguistic external repositories is needed to identify which concepts in the ontology are related with the terms provided by the user.

6. Implementation of Ontology and Semantic Web

In this section, ontology and semantic web approach will implement. These approaches are related in order to produce relevants information to web users.

6.1. Extracting SQL – Web User Query

In this component, four processes involved which are assign the initial query, exploit the initial query, assign a possible query and possible refinement query [15].

a) Initial Query

Query from web users through application or system is initial query. For example, web user sends a query about “Political Issues”. System automatically assign this query as a initial query.

A → IQ

Example 1

IQ = Political Issues
A = IQ (Political Issues)

b) Exploit Query

This process will exploit a query from web user. Once web user sends a query about “Political Issues in Malaysia”, this section will exploit into four words; Political, Issues, In, and Malaysia.

n = Number of words in A
M ∈ {n₁, n₂, n₃, ..., nᵢ}
M ∪ n
Example 1
A = Political Issues in Malaysia
A_1 = \{Political Issues\}
A_2 = \{in Malaysia\}
n = 4
N, number of words in A is 4, where data set in A exploit by space.

Example 2
A = Data Mining
A_1 = \{Data Mining\}
n = 2
N, number of words in A is 2, where data set in A exploit by space.

c) Assign a Possible Query

In this process, the query will refinement based exploitation query process.

Qn = Number of possible query
S \in \{Q_1, Q_2, Q_3, ..., Q_n\}

Example 1
A = Data Mining
A_1 = \{Data Mining\}
Assign any possible query;

\alpha_1 = \{Data Mining\}
\alpha_2 = \{Mining Data\}
Number of possible query is 2, where S \in \{Data Mining, Mining Data\}.

d) Refinement Query

Finally, this query will match keywords in data warehouse.

Start
Set initial variable is Q
Q \rightarrow \{Q_1, Q_2, ..., Q_n\}
Loop
\quad K = \{K_1, K_2, ..., K_n\}
\quad If (Q_i = = K_i)
\quad \quad \{Goto searching(Q_i,K_i);\}
While Q is null
K \in M
Loop
End

Fig 2: Matching Algorithm
Figure 2 shows the how the algorithm work with get the initial query until match to suitable keywords in data schema.

Example 1

\( \beta = \{K_1, K_2, K_3, ..., K_n\} \), where \( K_1, K_2, ..., K_n \) represents as a keywords. These keywords will store in temporary files at server side. Q

**Step 1**

Match \( Q_i \) with any keywords, \( K_i \) in temporary files.

**Step 2**

Hold data sources, \( DS_i \) location if found, otherwise keep new keywords, \( Q_i \) in temporary files.

**Step 3**

Repeat step 1 and step 2 until data set \( Q \) equal to null.

### 6.2. Ontology

In ontology, two processes involved which are semantic mapping and extracting wrapper ontology.

#### 6.2.1 Building Ontology-Based Query (Extracting Wrapper Ontology)

In building ontology-based query, several steps must be followed. First step in implementation of ontology is to create a model for query relational database on ontology. Two phases were considered in this model which are offline ontology extraction and online query issuing. In offline ontology extraction, system extracts the explicit classes and relations from the relational schema. In online query, web users can issue semantic query to the system. Figure 3 shows the query relational database on ontology model [10].

![Figure 3: Query Relational Database on Ontology Model](image-url)
Table 1 below shows the primary key, foreign key and their relationship. The purpose of this table is to show relation among tables.

**Table 1: Relational Database for Book Store**

<table>
<thead>
<tr>
<th>Relation</th>
<th>Primary Key</th>
<th>Foreign Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book(BookID, Name, Price, ISSN, SupplierID, CategoryID)</td>
<td>BookID</td>
<td>SupplierID, CategoryID</td>
</tr>
<tr>
<td>Supplier(SupplierID, BookID, SupplierName, Address, Tel, Email, CSupplierID)</td>
<td>SupplierID</td>
<td></td>
</tr>
<tr>
<td>BookType(BookTypeID, Name)</td>
<td>BookTypeID</td>
<td></td>
</tr>
<tr>
<td>CategorySupplier(CSupplierID, Name)</td>
<td>CSupplier</td>
<td></td>
</tr>
</tbody>
</table>

**Rule 1:** If primary key of more than one relation is the same, merged in one ontological class and their attribute should be merged.

Ex: Book rdf:type rdfs:class
Book(BookID, Name, Price, ISSN, SupplierID, CategoryID)

**Rule 2:** If the primary of one relation is unique for that relation, and not contain the primary key in another relation, then that relation will be considered as one ontological class.

Ex: BookType rdf:type rdfs:class
BookType(BookTypeID, Name)
Ex: CategorySupplier rdf:type rdfs:class
BookType(CSupplierID, Name)

**Figure 4: The Full Wrapper Ontology**

Table 2 below shows the detailed about ontology.
6.2.2 Semantic Query in Relational Database

Semantic query will implement after defining wrapper ontology. The purpose of implementation semantic query is to help user issue semantic query based on extracted ontology concept (based on keywords), and these queries will map onto plain syntactic SQL queries. SPSQL will be used to issue either schema query or data query. Schema query focuses on querying RDF schemas. Data query is related will filter instances. SPARQL-syntax query below shows how to translate into SQL.

**SPRSQL-Syntax**

```sparql
Select ? name ? price ? ISSN ? Supplier Name ? WHERE {
?s ex: BookID ? x.
?x ex: name ? bname.
?x ex: ISSN ? bISSN.
?x ex: Supplier Name ? Sp.
) ? name="Data Mining"
```

This query is to find the book information (name, price, ISSN) where book name =”Data Mining”.

**SQL Syntax**

```sql
SELECT Book.name, Book.price, Book.ISSN, Supplier Name FROM Book, Supplier WHERE Book.SupplierID = Supplier.SupplierID and Book.Name="Data Mining";
```

After executing this SQL statement, system will retrieve all related information about data mining from databases and display to web users.

7. Experimental Results

In this section described about implementation of ontology and semantic web approaches, sample queries and analysis the performance of ontology and semantic web approaches.
7.1. Implementation

In this implementation, web based application was developed using JAVA, HTML and four different Database Management System (DBMS). JAVA is programming language to develop a web application system based on architecture in figure 5. JAVA programming was choose because it is a powerful language and make a web application is portable and easily to accessable through World Wide Web. In JAVA, many RDF libraries are provided. The most complete library is Jena. Jena was develop to provide API that was designed specifically for the JAVA programming especially for web application development. Jena also provides ontology API and rules engine for basic inference RDF schemas. It is also support SPARQL by calling ARQ module [10].

7.2. Sample Query

Figure 5 show the web user interface. The web user just entered the keyword and press “submit” button for query process. In this experiments, suppose the web user is looking to buy books. Assume that a web user is looking to buy “data mining” book.

The first process based on architecture in figure 1 is extracting a SQL statement. A SQL statement must through 4 sub processes in extracting SQL statement process. If the keyword is already exists in data schema, this information will submit to next process. The next process is ontology. Two sub processes in the ontology process are semantic mapping and extracting wrapper ontology. In these processes, searching and retrieving process from data sources occurred. The purpose of these processes is to search and retrieve only relevant information to the web users. In figure 6, all relevant information will display to web users based on a keyword enter by web users.
Table 3: Results and Comparison

<table>
<thead>
<tr>
<th>Sample Query</th>
<th>Number of Relevance Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Data without...</td>
<td></td>
</tr>
<tr>
<td>without Implementation of...</td>
<td></td>
</tr>
<tr>
<td>and Semantic Approach</td>
<td></td>
</tr>
<tr>
<td>Data Mining</td>
<td>20</td>
</tr>
<tr>
<td>Data Warehouse</td>
<td>45</td>
</tr>
<tr>
<td>Information Retrieval</td>
<td>12</td>
</tr>
<tr>
<td>Database Integration</td>
<td>7</td>
</tr>
<tr>
<td>Education</td>
<td>8</td>
</tr>
<tr>
<td>University</td>
<td>7</td>
</tr>
<tr>
<td>Access Data by Implementation</td>
<td></td>
</tr>
<tr>
<td>of... and Semantic Approach</td>
<td></td>
</tr>
<tr>
<td>Data Mining</td>
<td>20</td>
</tr>
<tr>
<td>Data Warehouse</td>
<td>34</td>
</tr>
<tr>
<td>Information Retrieval</td>
<td>10</td>
</tr>
<tr>
<td>Database Integration</td>
<td>7</td>
</tr>
<tr>
<td>Education</td>
<td>5</td>
</tr>
<tr>
<td>University</td>
<td>6</td>
</tr>
</tbody>
</table>

7.3. Analysis for Implementation of Ontology and Web Semantic Approach

Table 3 shows the comparison between web based application with implementation on ontology and semantic web approaches and without implementation of ontology and semantic web approaches. This comparison based on number of relevance data. Several sample queries have been tested in experiments. Based on table above, the result indicates number of relevance data for “Data Mining” is 20 in both of these architectures. In “Data Warehouse”, number of relevance data for proposed heterogeneous database architecture was deducted from 45 to 34 compared to heterogeneous database architecture. In others queries sample in
table 3, such as Information Retrieval, Education and University, the results indicates number of relevance data for proposed heterogeneous database architecture was decreased about 5% to 10% average compared to heterogeneous database architecture. However, the result indicates number of relevance data for “Database Integration” is same for both of these architectures. In conclusion, the result in table 3 indicates implementation of ontology and semantic web approaches was improved the web application performance in term of number of relevance data.

8. Conclusion and Future Work

In conclusion, this paper was presented implementation of ontology and web semantic approaches in accessing data from heterogeneous database. This research focuses on producing more relevant information to web users. Based on experiments and results above, these approaches are able to improve number of relevant information to web users. The main advantage in this research, the implementation of ontology and semantic web is very efficient and suitable for two numbers of words (such as Data Mining, Information Retrieval, etc). In future work, new architecture with implementation of ontology and semantic web for accessing data from heterogeneous database will propose to improve web query processing in term of response time.

Acknowledgement

Special thanks to Universiti Sultan Zainal Abidin (UniSZA) was provided funds to implement of this research and Che Mat Ismail for support and advice.

References


Authors

Mohd Kamir Yusof obtained her Master of Computer Science from Faculty of Computer Science and Information System, Universiti Teknologi Malaysia in 2008. Currently, he is a Lecturer at Department of Computer Science, Faculty of Informatics, Universiti Sultan Zainal Abidin (UniSZA), Kuala Terengganu, Terengganu, Malaysia. His main research areas include information retrieval, database integration and web semantics.

Mohd Nordin Abdul Rahman obtained her PhD in Computer Science from Universiti Malaysia Terengganu in 2008. Currently, he is a Lecturer at Department of Computer Science, Faculty of Informatics, Universiti Sultan Zainal Abidin, and Terengganu, Malaysia. His main research areas include web services, cloud computing and software engineering.

Mat Atar Mat Amin obtained his Master of Computer Science from Universiti Teknologi Malaysia in 2006. Currently, he is a Lecturer at Department of Multimedia, Faculty of Informatics, Universiti Sultan Zainal Abidin in. His main research areas animation and creative multimedia.