ONTOLOGY-DRIVEN INFORMATION RETRIEVAL FOR HEALTHCARE INFORMATION SYSTEM : A CASE STUDY

Aidarus M. Ibrahim[1] and Hussein A. Hashi[2] Abdullalem A. Mohamed[2]

[1]Faculty of Computer Science and Information system, University Technology Malaysia (UTM), Malaysia
labiile@hotmail.com
[2]Faculty of Computer Science and Information system, University Technology Malaysia (UTM), Malaysia
oday_hashi@hotmail.com
[2]Faculty of Computer Science and Information system, University Technology Malaysia (UTM), Malaysia
aaldolah@yahoo.com

ABSTRACT

In health research, one of the major tasks is to retrieve, and analyze heterogeneous databases containing one single patient's information gathered from a large volume of data over a long period of time. The main objective of this paper is to represent our ontology-based information retrieval approach for clinical Information System. We have performed a Case Study in the real life hospital settings. The results obtained illustrate the feasibility of the proposed approach which significantly improved the information retrieval process on a large volume of data over a long period of time from August 2011 until January 2012

.Keywords

Ontology, Information Retrieval, UTM Clinic, Resource Description Framework(RDF)

1. INTRODUCTION

The field of Information Retrieval (IR) was born [1] to facilitate the fast and relevant retrieval. Fort the past forty years, the field has full-fledged significantly. Extensive users from different locations use the IR systems for their daily activity. However, finding relevant and useful information from large collections of data sources still poses some significant challenges. In this context, one of the substantial opportunities is to consider the semantics of the information using ontology. The use of ontology helps to overcome the limitations of keyword-based search [1] and puts forward as one of the motivations of the present work. While there have been contributions in this direction during the last few years, most achievements so far either make partial use of the full expressive power of an ontology-based knowledge, or are based on Boolean retrieval models, and therefore lack an appropriate ranking model [1] which is needed for scaling up to massive information sources.

2. RELATED WORK

In the recent past, several ontology-based approaches [1][2] have been proposed. The paper [1] illustrates reports on the methods, results and experience using a concept-based information retrieval approach.

DOI : 10.5121/ijnsa.2013.5105

Authors [2] evaluated the document adequacy with respect to a query using semantic proximities between ontology concepts and aggregating models. They described that the IR system depends on domain ontology to broaden the group of appropriate documents which are retrieved by employing graphical rending of query results to indulge user communications.
In paper [3], authors presented method for semantic query in out-dated relational database by creating ontological layer. A schema ontology mined from relational database schema and modified with global domain ontology is wrapped.

The paper [4] exposes an implementation of ontology and semantic approaches for accessing data from heterogeneous database. The main idea is to get the appropriate data and pass the relevant information to web users.

[5]This paper reports the experiment results of the UIUC-IBM team in participating in the medical case retrieval task of Image CLEF. It is experimented with multiple methods to leverage medical oncology and user (physician) feedback; both have worked very well, achieving the best retrieval performance among all the submissions.

In Article [6], described that the improvement of precision-recall of the search results is linked to semantic based information retrieval technique which needs to understand the meaning of the concepts in the user query. A huge amount of Relational dataset (RDB) that chains structuring data in a syntactic base is referred as traditional web. Unfortunately, to convert the available data stored in relational database into RDF format, which is mostly accepted standard proposed by the World Wide Web(W3C) is monotonous and error prone.[7,9,8].The authors build a wrapper that works as translator from semantic queries issued to the system into syntactic data available within these databases instead of immigrating available legacy data in relational database into RDF format based ontology. Moreover, another side effect, due to some limitations in RDF as purely text files [7, 8, and 9]; it can be benefited from the strong characteristics of relational model to RDF format. Although, the contemporary literature with regard to semantic search predominantly focuses on ontology query, while a mass of data is saved in heterogonous relational database. Hence, there is problematic issue which is how to use sematic query in database [12].

Perez and Conrad [11], showed “Relational OWL”, which routinely fetches the semantic of relational database and converts into RDF/OWL. This approach enables the usability of relational database in semantic applications based on the semantic of database structure on the database contents.

In paper [13] also presented a method to represent the schema and data stored in relational databases can be represented semantically. This showed how to figure out data items stored in that specific database. Although, with this approach, the schema and data items of each relational database are ready for semantic reasoning without the intervention of any expert. The method did not concentrate the ontological relations between tables and attributes.

The Article [14], The authors stored the RDF data in relational database to achieve the best possible query results where the actual query is performed. Later, they performed and adequate mapping of all RDF tipples to the relational data model. This mapping is specific to the needs of querying normal RDF triples, stored in a special way in relational database. The paper [15], writers introduced a working draft for an algebraic definition of RDF models. Unluckily, this work lacks the procedures required for enquiring the RDF graph.

In this paper, we proposed ontology based information retrieval approach resulted from a case study accomplished in University Technology Malaysia (UTM) Clinic. This paper is organized as follows: Section 2 presents the concepts and ontology based modeling for UTM Clinic. Section 3 is engaged to show the methodology to build the ontology based information retrieval. Section 4 demonstrates Query Execution and Information Retrieval Results. Finally, Section 5 concludes the paper.

3. Ontology based Modelling for UTM clinic

Ontology is the specification of concepts and relations. It plays a central role in semantic web applications by providing a shared knowledge about objects in real world, which promotes reusability among different modules. Therefore the semantic should be the first concern in any semantic applications.

For this Case Study, we modeled UTM Clinic ontology.Figure1 shows classes as identified during the case study. UTM Clinic and person are classes, while staff and patient are sub-class of person. Moreover, Doctor and Assistant are also sub-classes of Staff. The interesting point seen from the figure is patient is a sub-class of person. However, Person and Patient can be shared for some properties such as Fname, Iname, Height and Weight.

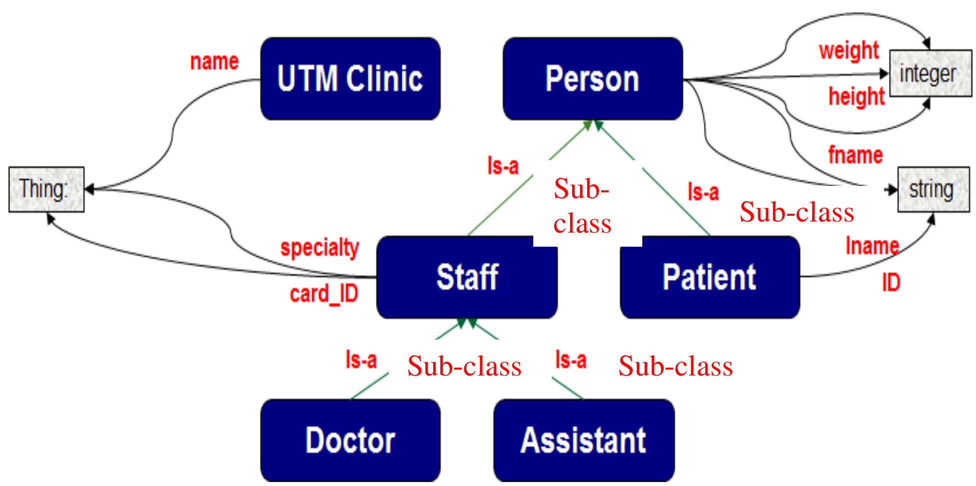


Figure 1: UTM Clinic Ontology-based Modeling

Figure 2 shows the whole idea for our Case-Study. The functionality and inter-dependency of classes and objects is figured out here. UTM Clinic has staff. Doctor and Assistant are sub-classes of staff. Due to the massive task of Doctor, cooperation is needed from the Assistant to take care of the patient. The four major duties performed by the doctor to see the patient can be either, medical checkup, consultation, diagnosis or suggestions. Patient can also be person while person can also a patient. For example, Doctor is a person; on the other hand patient can be a doctor.

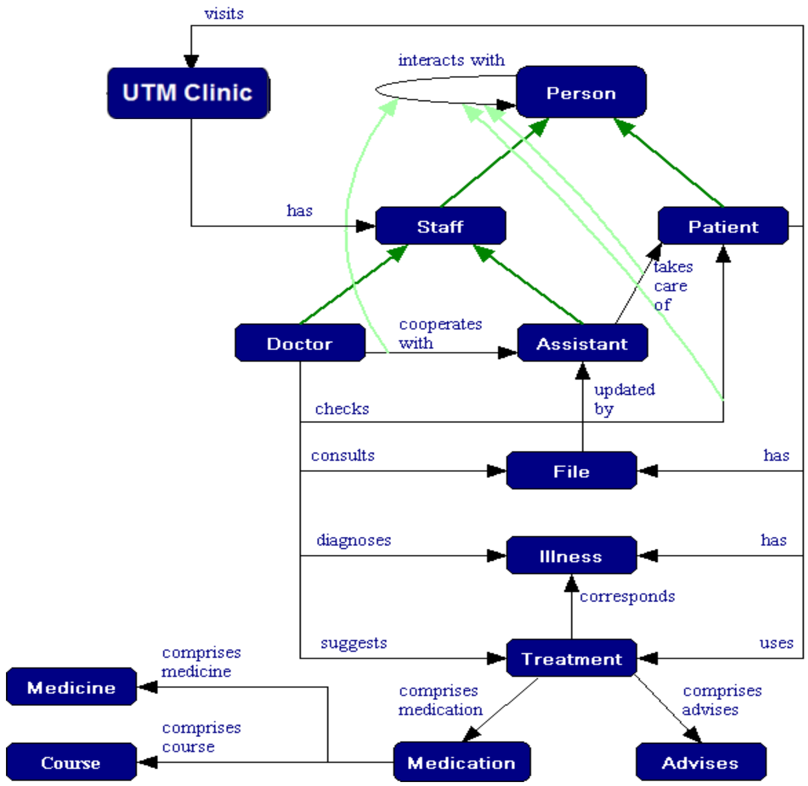


Figure 2: UTM Clinic Ontology Representation

4. Retrieval Mechanism

Information retrieval is used to satisfy user's information needs. In order, to achieve this goal, Information retrieval deals with representation, storage, organization of, and access to information items. Therefore, the information retrieval changes the users demand for information into machine processable query. Usually this query contains a set of terms that describes the Information needed. As information retrieval mainly deals with natural language, which might be semantically ambiguous, small errors concerning retrieved objects are inconsequential. The user may rather be interested in retrieving information about a subject.

In our system as shown in figure3, the user who wants to access the system by retrieving specific topic using the user interface through which the user can create query that brings the relevant Information. The user interface submits user query to the wrapper ontology to access the stored files in the database. The wrapper ontology wraps the relational database with schema ontology. The result generated by the query is translated using wrapper ontology to bring for the user. That is why users can issue queries on data semantics without having to use different ontology language.

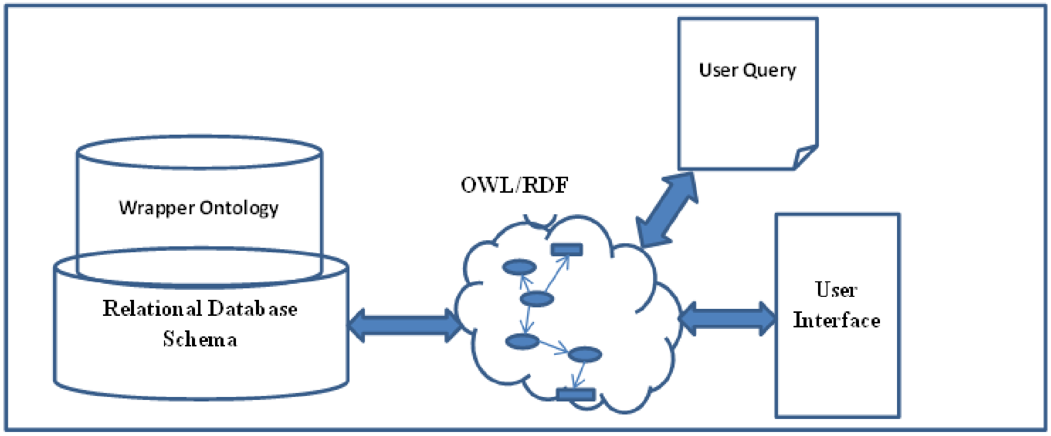


Figure 3: Retrieval Mechanism

The method implemented below exmins that any semantic retrieval algorithm must be tough enough to deal with the insuffeicency of KBs and Ontologies as they are not yet sufficiently developed to depend upon absolutely, and so they propose integrating a keyword based retrieval method in the overall system. Their semantic ranking method, brought here as an example for such methods is usually based on two steps:

a. Explaining documents with a weighted annotation, a graphic of the importance of the explained instance is add to the text. So, each instance is thus weighted,

$$d_i = \frac{freqd,i}{\max\{fr_{d'k}\}} x log \frac{N}{n_i}$$

Figure 4: Annotation Weighing

di - The weight of instance *i* in document *d*.

freqd, *i* - number of occurrences of *i* in *d*. *maxk{freqd,k}* - Maximum number of occurrences of any instance in *d*.

N –Total numbers of documents in the search space.

ni - number of documents annotated with *i*. The idea is that some annotations may bring us closer to the goal of evaluating the concepts in the document.

$$sim(d,q) =. \frac{d.q}{|d||q|}$$

Figure 5: document and Query relevance

d -The related document is weighted by the vector of annotation

q - A vector resembling the query. qi is the number of variables in the query which are related to instance i.

b. This approach of ranking is reworking of the classic Vector-space model used for correspondence evaluation. The final similarity result for the document is:

$$sim_T(d,q) = a.sim_o(d,q)+)(1-a).sim_k(d,q)$$

Where sim_o, is the ontology based on similarity index and sim_k is the keyword based similarity indices will be available to soften the error's implications.

4.1 Method of calculating the weight module

I. The more times the words appear in the document, more possibly it is the characteristic words;

II. . The measurement of the words will also mark the significance of words. Seemingly, one idea in the ontology is connected to other notion in that domain ontology. That also means that the association between two concepts can be dogged using the length of these two concept's connecting path in the concept lattice.

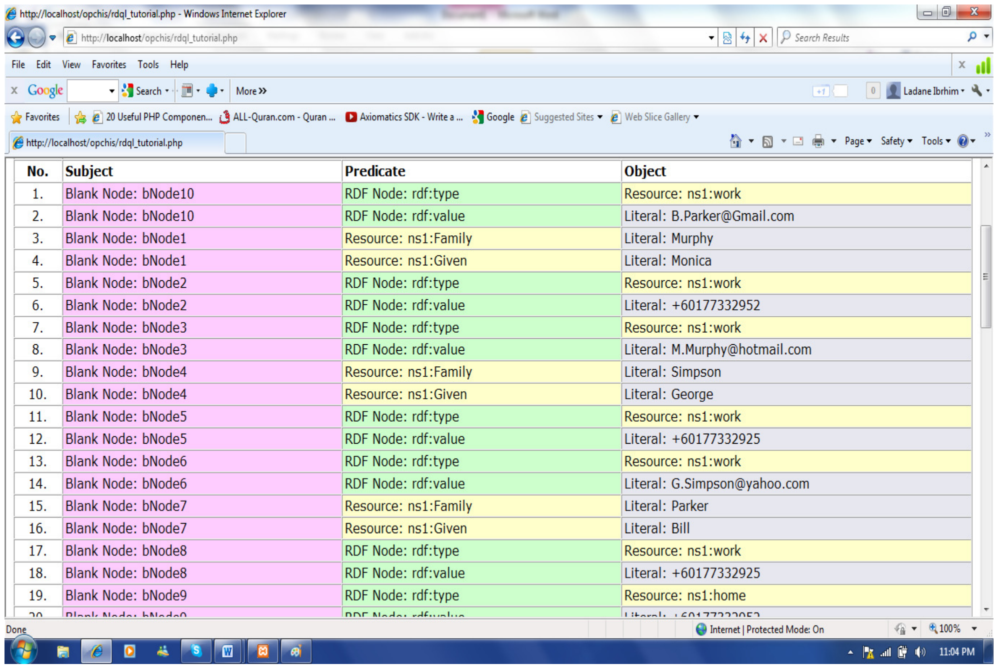
III. If the chances of one word are high, then the word will get additional weight. One word may be the characteristic word even if it doesn't appear in the document.

5. Query Execution and Information Retrieval Results

In this experiment, suppose that the user is looking to get single patient's information .So, the user builds the query and sends in to the system to get the exact information about the patient. As we can see from the following figure is the data generated by the system.

The first query process depends on the design of figure1 is extracting a SQL Query. To retrieve information a SQL statement must go through a process to extract the exact information about the patient needed by the user. In here, the execution of an RDQL query in RAP consists of four main steps. At first, the query string is parsed into a memory representation and subsequently passed to the corresponding engine. In the second step, the query engine searches for triples matching all patterns from the WHERE clause of the RDQL query and returns a set of variable bindings. Next, this result set is filtered by evaluating Boolean expressions specified in the AND clause. The last step is the processing of the final result.

In figure 6, illustrates the basic information related to the patient. We executed different queries based on out interest and the information that as user we want to retrieve.



No.	Subject	Predicate	Object
1.	Blank Node: bNode10	RDF Node: rdf:type	Resource: ns1:work
2.	Blank Node: bNode10	RDF Node: rdf:value	Literal: B.Parker@Gmail.com
3.	Blank Node: bNode1	Resource: ns1:Family	Literal: Murphy
4.	Blank Node: bNode1	Resource: ns1:Given	Literal: Monica
5.	Blank Node: bNode2	RDF Node: rdf:type	Resource: ns1:work
6.	Blank Node: bNode2	RDF Node: rdf:value	Literal: +60177332952
7.	Blank Node: bNode3	RDF Node: rdf:type	Resource: ns1:work
8.	Blank Node: bNode3	RDF Node: rdf:value	Literal: M.Murphy@hotmail.com
9.	Blank Node: bNode4	Resource: ns1:Family	Literal: Simpson
10.	Blank Node: bNode4	Resource: ns1:Given	Literal: George
11.	Blank Node: bNode5	RDF Node: rdf:type	Resource: ns1:work
12.	Blank Node: bNode5	RDF Node: rdf:value	Literal: +60177332925
13.	Blank Node: bNode6	RDF Node: rdf:type	Resource: ns1:work
14.	Blank Node: bNode6	RDF Node: rdf:value	Literal: G.Simpson@yahoo.com
15.	Blank Node: bNode7	Resource: ns1:Family	Literal: Parker
16.	Blank Node: bNode7	Resource: ns1:Given	Literal: Bill
17.	Blank Node: bNode8	RDF Node: rdf:type	Resource: ns1:work
18.	Blank Node: bNode8	RDF Node: rdf:value	Literal: +60177332925
19.	Blank Node: bNode9	RDF Node: rdf:type	Resource: ns1:home

Figure 6: Information of the patient

In figure 7, we sent a query related to the patient's name, although same names are already stored in the database, we can classify using their identity. As shown in the figure, all the names of the patient that are requested by the user are generated.

Query1: *'find the full name of all patients'*. **Output:**

No.	?fullName
1.	Literal: Bill Parker
2.	Literal: George Simpson
3.	Literal: Monica Murphy

Figure 7: Result of Query one

As shown in figure 8, as long as the user creates effective query the better result is generated. The query below shows, that the user has sophisticatedly based the query on criteria to obtain the relevant information.

Query2: '*find the given name and age of all patients over 30.*' **Output:**

No.	?givenName	?age
1.	Literal: Bill	Literal: 33 **(rdf:datatype="http://www.w3.org/TR/xmlschema-2/integer")**
2.	Literal: George	Literal: 41 **(rdf:datatype="http://www.w3.org/TR/xmlschema-2/integer")**

The same goes in figure 9; we generated a query related to patient's information. Handling the different queries for different purposes makes the system very powerful. All the queries as seen in the figures are compiled successfully and generated the relevant results associated to the patient's information requested by the user.

Figure 8: Result of Query 2

Query3: '*find the private telephone number of the person whose office number is '+60177332925' and return additionally his given name and age*'. **Output:**

No.	?givenName	?age	?telNumberHome
1.	Literal: Bill	Literal: 33 **(rdf:datatype="http://www.w3.org/TR/xmlschema-2/integer")**	Literal: +60177332952

Figure 9: Result of Query 3

5. In conclusion

In conclusion, the paper presents case study performed in UTM Clinic for ontology based information retrieval approach. Web based approach was developed using PHP, RDF API PHP (RAP). PHP is a programming language used to develop web based applications. It is powerful language which can make the web applications portable and easily accessible. RAP is a library for parsing, querying, manipulating, serializing and serving RDF models. It is developed to provide API for ontology and RDF schemas the case study focused on producing more relevant information on UTM clinic system. The main attainment of this case study was to generate a result from single patient's information useful for information retrieval.

REFERENCES

[1] Bevan Koopman, Peter Bruza, Laurianne Sitbon2, Michael Lawley"a concept-based information retrieval approach" Australian e-Health Research Centre, CSIRO" AEHRC & QUT at TREC 2011.

[2] Vishal Jain" Information Retrieval through Multi-Agent System with Data Mining in Cloud Computing" IJCTA | JAN-FEB 2012.

[3] Mohameth-François Sy, Sylvie Ranwez1Jacky Montmain1, Armelle Regnault, Michel Crampes1, Vincent Ranwez"User centered and ontology based information retrieval system for life sciences." Sy et al. BMC Bioinformatics 2012, 13(Suppl S4http://www.biomedcentral.com/1471-2105/13/S1/S4.

[4] Mohd Kamir Yusof, Mohd Nordin Abdul Rahman and Mat Atar Mat Amin"Ontology and Semantic Web Approaches for Heterogeneous Database Access" International Journal of Database Theory and Application Vol. 4, No. 4, December, 2011.

[5] 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

[6] Parikshit Sondhi1, Jimeng Sun, ChengXiang Zhai1, Robert Sorrentino,Martin S. Kohn2, Shahram Ebadollahi, Yanen Li1Medical Case-based Retrieval by Leveraging Medical Ontology and Physician Feedback: UIUC-IBM at ImageCLEF 2010.

[7] M. Barathi and S.Valli"Ontology Based Query Expansion Using Word Sense Disambiguation." (IJCSIS) International Journal of Computer Science and Information Security,Vol. 7, No. 2, February 2010.

[8] Atsutoshi Imai, and Shuichi Yukita, "RDF Model and Relational Metadata", 17th InternationalConference on Advanced Information Networking and Applications (AINA'03).

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

[10] Dan Brickley, and R.V. Guha, "Resource Description Framework (RDF) Schema Specification", W3CRecommendation, 1999.

[11] Donglin Chen, Liying Li, Rui Wang, An Ontology-based System for Semantic Query over Heterogeneous Databases, World Congress on Software Engineering, IEEE 09.

[12] Cristian Pérez de Laborda and Stefan Conrad, "Querying Relational Databases with RDQL" BerlinerXML Tage 2005.

[13] G. Karvounarakis, V. Christophides, D. Plexousakis, and S. Alexaki, "Querying RDF Descriptions for Community Web Portals", In BDA'2001, Agadir, Maroc, pages 133–144, 2001.

[14] Sergey Melnik. Algebraic Specification for RDF Models.http://wwwdiglib.stanford.edu/diglib/ginf/WD/rdfalg/rdfalg.pdf, 1999. Working Draft.