INDEXING METHOD FOR SEMANTIC INFORMATION EXTRACTION FROM THE MEDICAL ONTOLOGY

Dr.Poonam Yadav*

Abstract

Semantic information extraction is kind of retrieval process to extract the user required information from the medical ontology. Once the domain ontology is constructed from the web documents, the ontology mining technique can be used to extract the information using the interesting measures. Accordingly, an indexing method is developed for semantic information extraction from the medical ontology. The medical ontology contains different attributes and relations in an organized way. The proposed approach of semantic information extraction contains three important steps such as, query parsing, indexing and matching. The input query obtained from the user is parsed and the indexing is done by matching to the predefined profiles. Then, direct matching of query with attributes of ontology is done using similarity measures. The performance of the proposed medical information retrieval is evaluated under different evaluation criteria. From the results, it ensures that the proposed indexing method provides improved effectiveness without compromising the computational time.

Keywords:- Information extraction, indexing, ontology, query, semantic data.

* D.A.V College of Engineering. & Technology, India

1. Introduction

In recent decades, the amount of Web information growth has further exacerbated user needs for efficient mechanisms like information and knowledge location, selection, and retrieval. How to gather useful information from the Web becomes a challenging issue for all the Web users. Capturing user's information through a given query is extremely difficult in most Web information gathering as users provide only short phrases in queries to express their information needs. Also, because of different personal perspectives, expertise, and vocabularies web users express their queries differently. These differences cause the difficulties in capturing user information needs. Thus, for capturing user information needs user's personal interests and preferences must be understood. For this, ontology can be created in personalized Web information gathering. Ontology represent the concept model consisting of relevant and non-relevant concepts which is obtained from their background knowledge. If concept models can be specified the user's interest, user information needs can be better captured, and therefore, meaningful, and personalized information can be gathered for Web users [6-10].

To simulate user concept models for gathering web information [2-5], ontologies are used. If ontologies can specify user background knowledge then more accurate user profiles can be acquired and thus the user information needs can be captured effectively [8]. Accordingly, an indexing method is developed for semantic information extraction from the medical ontology. Once the ontology is constructed, the relevant information can be retrieved based on the user query. Here, the retrieval algorithm is developed to obtain the most suitable information from the ontology. Here, two different types of information can be retrieved from the ontology, the disease information by putting the symptoms as query and drug information by putting the disease as query. The most supported information is extracted from the ontology and supplied to the user. The rest of the paper is organized as, the 2nd section contains existing works. The detailed proposed approach is written on the 3th section and results are produced in the 4th section. The 5th section consists of the conclusion part of the proposed approach.

2. Existing works

A handful of recent works are available in the literature for information extraction from the domain ontology. Accordingly, Chin-Ang Wu *et al.* [1] have proposed an active multidimensional association mining framework that incorporates with user preference ontology, which contains surrogate queries that represent frequently, used queries in the query history log. The representative power and the user preference of the surrogate queries are derived and expressed in fuzzy linguistic terms. The construction of the ontology is demonstrated also for mining search results. In this work, an indexing method is developed for medical information search using ontology. Here, two different types of information can be retrieved from the ontology, the disease information by putting the symptoms as query and drug information by putting the disease as query. For these two types of information retrieval, an algorithm is developed to obtain the information in an easy way. The proposed algorithm can be able to find the important information from the ontology, helping to the users to find some useful information about the disease and the drugs without much consulting with the doctors.

3. Indexing method for semantic information extraction from the medical ontology

This section presents the indexing method for semantic information extraction from the medical ontology. Initially, medical ontology is constructed from the document based on the user's requirement using attributes selection and relation finding. The medical ontology developed is then given to this algorithm for extracting information. The proposed method of extracting the information from the user is done with three important steps such as, i) Obtaining user query, ii) query indexing, iii) Query searching.

3.1 Obtaining user query

The major step as considered by the proposed approach is the extraction of the information from the constructed ontology. The information from the ontology is extracted based on a particular query assigned by the user. There are some important measures in accepting the queries. The proposed approach is using an indexing based approach for the retrieval of information from the ontology. The proposed approach gives the option to the user, the choice of entering either disease as the query or the symptoms. The indexing done based on the query given and is

accounted to the proposed method as a set, i.e. the proposed method accepts the input as a set of two values, which can be represented as,

$$query \rightarrow < d : I; S : I > \quad (1)$$

3.2 Query indexing

The query is accepted with their text given and the program automatically creates an index for the query with index 'I'. The index value is selected as 1 and 2, when diseases is entered as the query, then 1 is set as the index, while symptoms is given, the index I is set as 2. The query indexing is given to speed up processes of the ontology taxonomy traversal. If the query is given without the index value, the query word has to be compared with all the nodes in the ontology, which reduce the efficiency of the method because of the unwanted searches. So as to handle such condition the index values will be helpful. The index values act as identifier to the query and the nodes in the ontology and also help in matching the correct attribute to the relevant query. In the similar way, the ontology also indexed as per the indexing technique. We have used a forward indexing technique here.

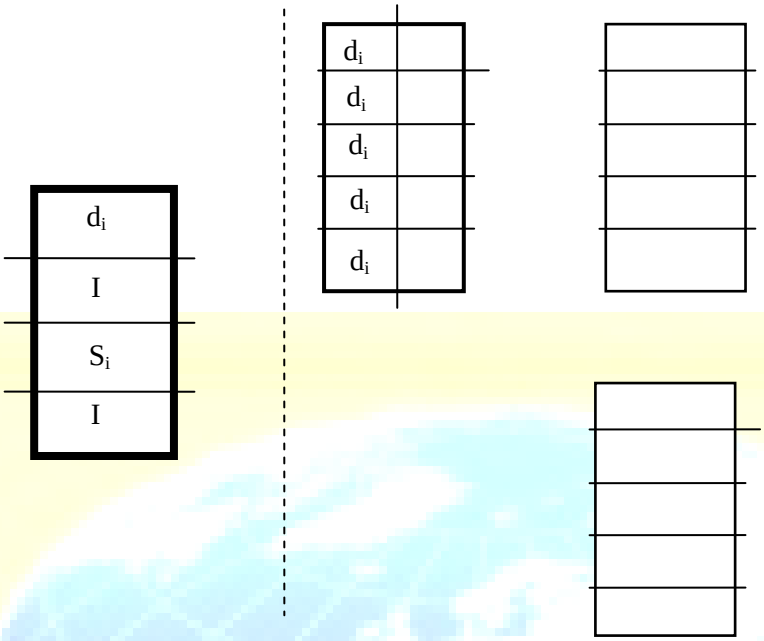

Figure 1. Indexing query and attributes

3.3 Query searching

Once the indexing is finished for the attributes, the information extraction has to be subjected based on the query. The query search for the relevant data regarding the index defined on the query. The search will continue upto the relevant information is extracted. The query initially identifies the relevant attribute, by comparing the index value of the top nodes of the disease set and the symptoms set. If the indexes are matched, then relevant attributes search for the associated attribute and that attribute will point towards the drug associated with disease. The process can be illustrated as follows.

The figure 1 explains the extraction information from the attributes based on the given user query. Initially the indexed query is selected and according to the index value, the relevant attribute is selected, either diseases or attributes. The selected attribute will search for the associated attribute. If disease is selected, then it will go for finding the associated symptoms and vice versa. Then, the proposed method triggers the search for the drug associated with attribute in selection. Once the drug is identified the relevant information regarding the query is retrieved as a tuple containing three values.

$$< d_i ; s_i ; r_i \ > \quad (2)$$

Where, the d_i represents the diseases, r_i represents the drug and the s_i represents the symptoms of the disease. The information is extracted based on the similarity score between each associated attributes. A set of information can be extracted from the attributes based on a query, but the information, which possess high similarity score is selected as the most relevant information regarding the input query. The query retrieval algorithm developed in this work is given in figure 2.

Input : User query
Output : Extracted information
Procedure:

Start

Select the input query
Index the query

$$query \rightarrow < d : I ; S : I >$$

Compare query with attribute

$$q : I \Leftrightarrow < d_i : I ; s_i : I >$$

$$if\,(q : I \approx d_i : I)$$

$$then\,SELECT\,(d_i)$$

$$find\,associated\,s_i$$

Find the drug associated with d_i and s_i.

Extract the set of information regarding query

Select the extracted information with high mutual information score

End

Figure 2. Query retrieval algorithm

4. Experimental Results

The performance of the proposed medical information retrieval is evaluated under different evaluation criteria. The dataset includes the data including different diseases, symptoms and drugs. The data are extracted from the from the internet database, over 100 document are

extracted based on the medical data. The data are processed to extract the relevant details such as diseases, symptoms and drugs, in order to conduct research.

The performance evaluation of the proposed approach is conducted based on the time required for the extraction of information from the ontology as per the query given by the user. The query is based on index values and ontology also associated with index values. The time is the effective factor regarding the performance evaluation of the proposed approach. The time evaluation is taken by considering the time, which is required for the extracting information from the ontology based on the query. A number of queries are subjected to evaluate the performance of the proposed approach.

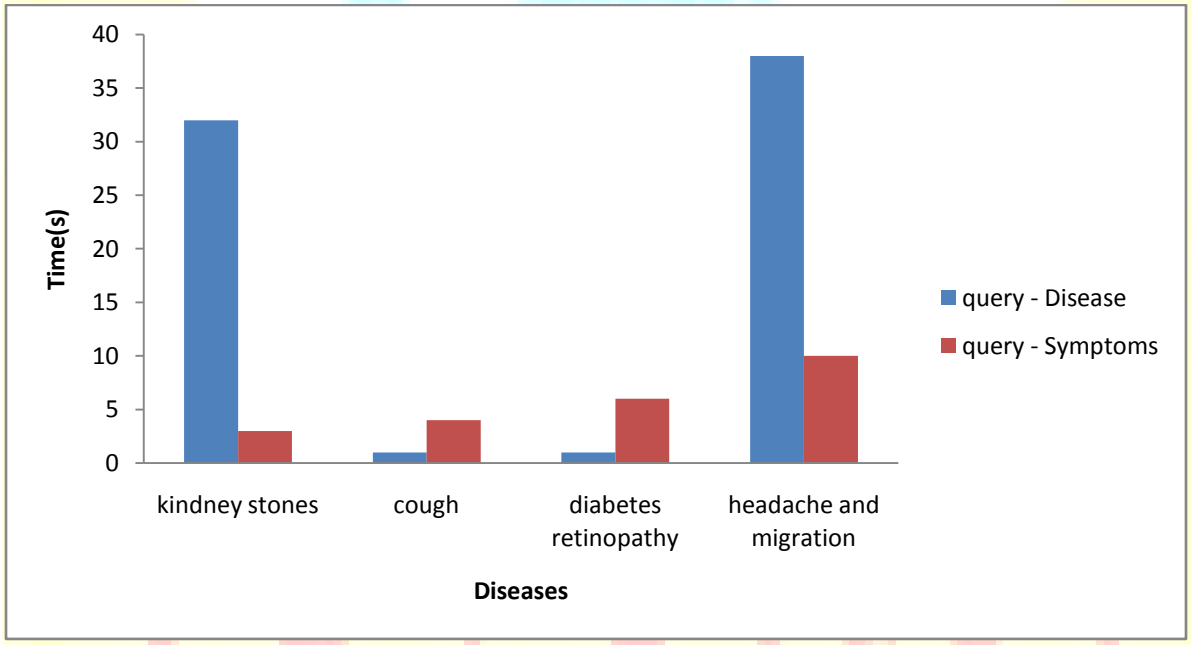

Figure 3. Performance evaluation

The figure 3 shows the performance evaluation of the proposed approach based on the time for information extraction. The process is done based on two different ways, by comparing the query with the disease and by comparing the query with the symptoms. We have considered four different diseases for the information extraction. When considering the responses marked in the graph, it can be identified that the query disease comparison shows both peak time and less time.

On the other hand the query – symptom shows constant time deviations. Thus, the information can be evaluated quickly by giving the query symptoms comparison strategy.

5. Conclusion

In this work, an indexing method is presented for semantic information extraction from the medical ontology. The medical ontology contains different attributes and relations in an organized way. The extraction of information from the medical ontology is done using three important steps such as, obtaining query, indexing and query matching. The input query obtained from the user is parsed and the indexing is done by matching to the predefined profiles. Then, direct matching of query with attributes of ontology is done using similarity measures. The experimentation is carried out on database extracted from the internet and the performance evaluation is conducted. The performance evaluation showed that the proposed approach is a time effective one. The maximum time required for query – diseases approach is high as compared with the query – symptoms approach.

References

[1] Chin-Ang Wu, Wen-Yang Lin, and Chuan-Chun Wu, 2010, An Active Multidimensional Association Mining Framework with User Preference Ontology, International Journal of Fuzzy Systems, 12(2), pp. 125-135.

[2] Bollegala D., Matsuo Y., and Ishizuka M., 2013, Minimally Supervised Novel Relation Extraction Using a Latent Relational Mapping, IEEE Transactions on Knowledge and Data Engineering, 25(2), pp. 419-432.

[3] Tak-Lam Wong, Wai Lam, 2010, Learning to Adapt Web Information Extraction Knowledge and Discovering New Attributes via a Bayesian Approach, , IEEE Transactions on Knowledge and Data Engineering, 22(4), pp. 523-536.

[4] Hatala M., Gasevic D., Siadaty M.. ; Jovanovic, J., Torniai, C., 2012, Ontology Extraction Tools: An Empirical Study with Educators, IEEE Transactions on Learning Technologies, 5(3), pp. 275-289.

[5] Wei Wang, Barnaghi, P., Bargiela, Andrzej, 2010, Probabilistic Topic Models for Learning Terminological Ontologies, IEEE Transactions on Knowledge and Data Engineering, 22(7), pp. 1028-1040.

[6] Hongbo Sun, Wenhui Fan, Weiming Shen, and Tianyuan Xiao, 2013, Ontology Fusion in High-Level-Architecture-Based Collaborative Engineering Environments, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(1), pp. 2-13.

[7] Mascardi V., Locoro A., Rosso P., 2010, Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation, IEEE Transactions on Knowledge and Data Engineering, 22(5), pp. 609-623.

[8] Chang-Shing Lee., Mei-Hui Wang., Hagras H., 2010, A Type-2 Fuzzy Ontology and Its Application to Personal Diabetic-Diet Recommendation, IEEE Transactions on Fuzzy Systems, 18(2), pp. 374-395.

[9] Yijia Zhang, Hongfei Lin, Zhihao Yang, Jian Wang, Yanpeng Li, and Bo Xu, 2013, Protein Complex Prediction in Large Ontology Attributed Protein-Protein Interaction Networks, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 10(3), pp. 729-741.

[10] Wang Peng., Xu, Baowen., Zhou, Yuming, 2010, Extracting semantic subgraphs to capture the real meanings of ontology elements, Tsinghua Science and Technology, 15(6), pp. 724-733.

 Dr. PoonamYadav obtained B.Tech in Computer Science &Engg. fromKurukshetra University Kurukshetra and M.Tech in Information Technology from Guru Govind Singh Indraprastha University in 2002 and 2007 respectively. She had Awarded Ph.D inComputer Science& Engg. fromNIMS University, Jaipur. She is currently working as Principal in D.A.V College of Engg. & Technology, Kanina (Mohindergarh). Her research interests include Information Retrieval, Web based retrieval and Semantic Web etc. Dr. PoonamYadav is a life time member of Indian Society for Technical Education