International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 4 365 - 369

Ontology Based E-Healthcare Information Retrieval System: A Semantic Approach

[1]Michael Raj T.F, [2]Priya Govindarajan, [3]Ravichandran. K.S, [4]Gayathri.M, [5]Uma.R
[1,2]Asst. Prof., Dept. Computer Science and Engineering, SRC, SASTRA University
[3]AD, ICT,SASTRA University
[4,5]PG Scholar, Dept. Computer Science and Engineering, SRC, SASTRA University
[1]*Email: tfmichaelraj23@yahoo.com*

Abstract- With the increase of data in the health care system provides a base for the development of an effective information retrieval system. The implementation of such information retrieval system integrates the heterogeneous information from the healthcare environment. Most of the existing information retrieval systems are syntactic based systems, which will provide inefficient results for the search queries. The objective of this approach is to design a semantic based E-Healthcare information retrieval system. The proposed approach uses an ontology to define the disease-treatment information and will be used for the effective information retrieval. The designated approach is evaluated with a web based tool and the results shows that there is an improvement in the approach.

Keywords:*Information retrieval, E-Health care system, Ontology system, WordNet*

*****

I. INTRODUCTION

The growth of information technology facilitates the medical field by providing the enormous data handling features and the information retrieval methodologies[24]. But still there exists some pitfalls in the data handling and in the information retrieval process [1]. Medical data are distributed across the web as heterogeneous contents.This heterogeneous medical data are to be integrated to provide valuable information to the society [27].

The aim of this article is to design and implement an efficient information retrieval system for the health care system. This takes significant challengesin identifying the data resources and methods for the effective information retrieval and integration [25]. The conventional approach is a keyword based approach has numerous limitations in information processing this can be rectified by the semantic approaches which uses the concept of ontology to describes the medical data [2].

Describing any information using ontology is a base for the development of semantic information retrieval system. Semantic based systems are using the concepts of RDF and other layers which are described in the following figure 1.

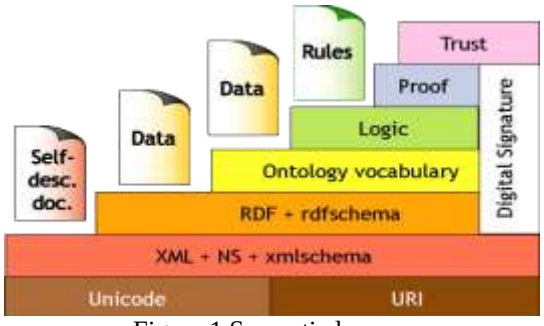

Figure 1 Semantic layers

II. RELATED WORK

Medical records are widely distributed as an autonomy of data. Sharing and accessing the medical data is a crucial process. It also increases the complexity of the information retrieval process when number of requestsare increased [3] [4]. These records are needed to be integrated by the ontology. This improves the health care system by taking them into new information retrieval process.A health care system faces the lot of challenges in its implementation process [25]. These challenges are listed below

- Centralized medical data repository
- Enabling privacy data security
- Universal ontology creation
- Disease-Treatment relationship
- Knowledge base
- Identifying effective information retrieval process
- Evaluation of the approach

A. Semantic web

Semantic web is the extension of the current web provides interlinked data. Conventional web will not resolve the anomalies in searches. The relation between the contents can be achieved by defining the semantic meaning among them. These data are machine processable and mergable to derive new knowledge base for the domain. The realization of the semantic web and its data is done through the concepts XML, RDF, OWL-S and URI [5].

B. Ontology

Semantic annotations are described by the ontologies, it conceptualize and gives explicit specification for the things in the semantic web. There are several definitions and meanings for the ontologies described in [6-8].

Ontologies are plays vital role in several fields which requires knowledge management, information processing and integration. With the help of Description Logic (DL) the ontology can be derived for the domain [9-11]. Figure 2 represents the sample ontology.

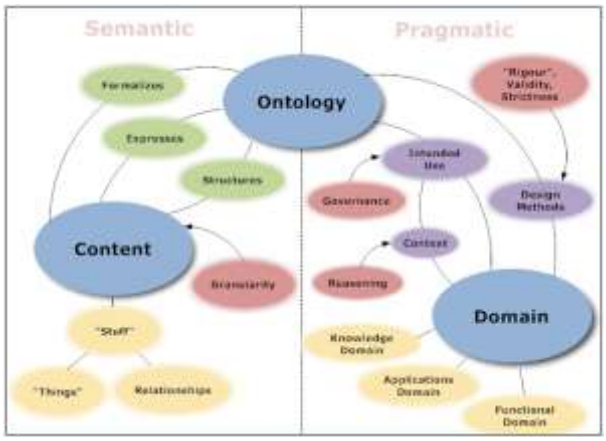

Figure 2 Sample ontology model

C. Information Retrieval approaches

Information retrieval (IR) is a process of retrieving the relevant information from the repositories which consists of documents of relevant and irrelevant domains. According to Lancaster the information retrieval is merely the retrieval of existing and non-existing contents of the documents about his or her request [12]. Based on the user requests the data are processed and converted into information then submitted to the corresponding users. The block diagram of information retrieval model is represented in the figure 3.

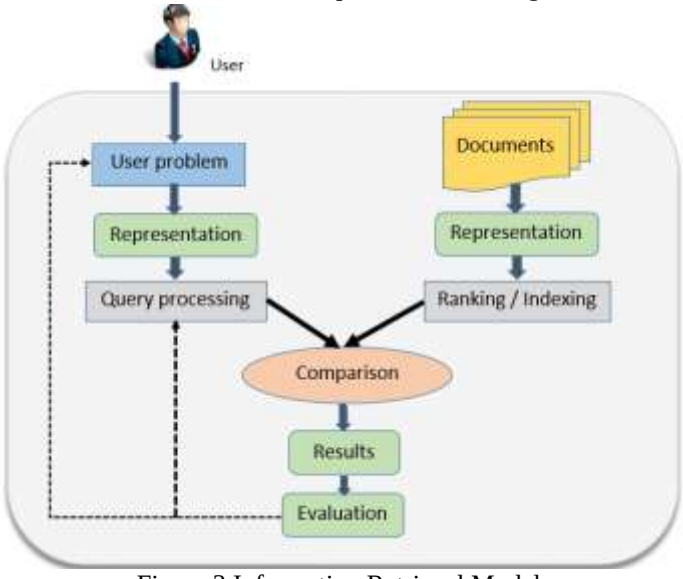

Figure 3 Information Retrieval Model

There are several information retrieval models are available. It can be classified three categories namely Boolean, statistical and linguistic and knowledge based approaches [24, 26]. There are more than one methods in each category depicted in figure 4. These models can be achieved in various situation to design and implement the designated IR systems [13].

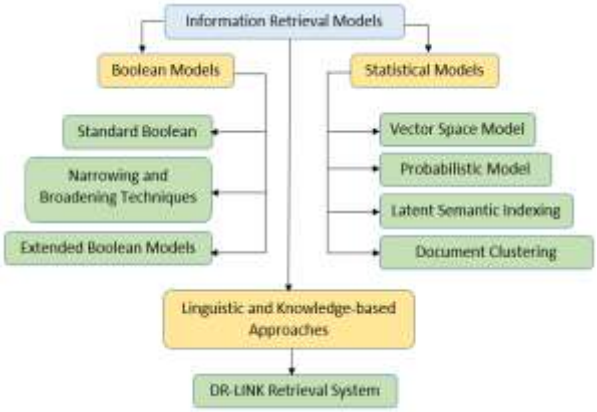

Figure 4 Information retrieval models

Semantic information retrieval becomes very essential part of any information processing engine. Ontologies are plays very vital role in describing the semantic information. Ontology removes the ambiguity in the specification of information. To build the ontology for the specific domain the Protégé tool is used [14].

III. PROPOSED MODEL

The proposed model is to design and implement a framework for an E-Health care system. The system is ontology based and semantically retrieves the information from the repository. This framework integrates hospital, clinic, disease-treatment relations, doctors and the geographical locations of the hospital, travels and hotel information. The construction of the system is depends upon the following process

- Ontology creation for human disease-treatment
- Query pre-processing
- Retrieval of health care information
- Ranking the results
- Provide the results to users

A. Ontology creation

Ontology provides the knowledge base for the designated framework implementation. It relates different types of resources which are required to provide an integrated E-Health care system. Hierarchical information are defined in the ontology and provides cluster for the domain [14]. The created ontology links the resources to enable an integrated E-Healthcare systems.

B. Query pre-processing

Query pre-processing is very essential process in the IR. The query may contain relevant and irrelevant terms in the search context. Before retrieving the results for the query it should be processed to produce better response.Ruxendra explained the query processing approaches for heterogeneous environment [16].

Our domain specific IR system identifies the relevant terms for the target queries. These target terms are the input for the SPARQL queries and the general format of the SPARQL query is described in the figure 5. The SPARQL query consists of triple data <subject, predicate, object>[17]. It retrieves semantic data from the disease-treatment ontology.

Figure 5 General format of the SPARQL

C. Retrieval of health care information

IR for the Healthcare systems uses the ontologies to represent the symptoms, disease-treatment relationships [25], clinic information, doctors' information and also geographical information. The retrieval of information starts with user's query consists of either symptoms for a disease or disease information. These queries are parsed and the corresponding SPARQL triples are generated. Based on the query it gives different results described in the following figure 5. Similarities between the documents have been determined by the various methodologies namely ontology based, multi tree model and Bag of words concepts [18-20].

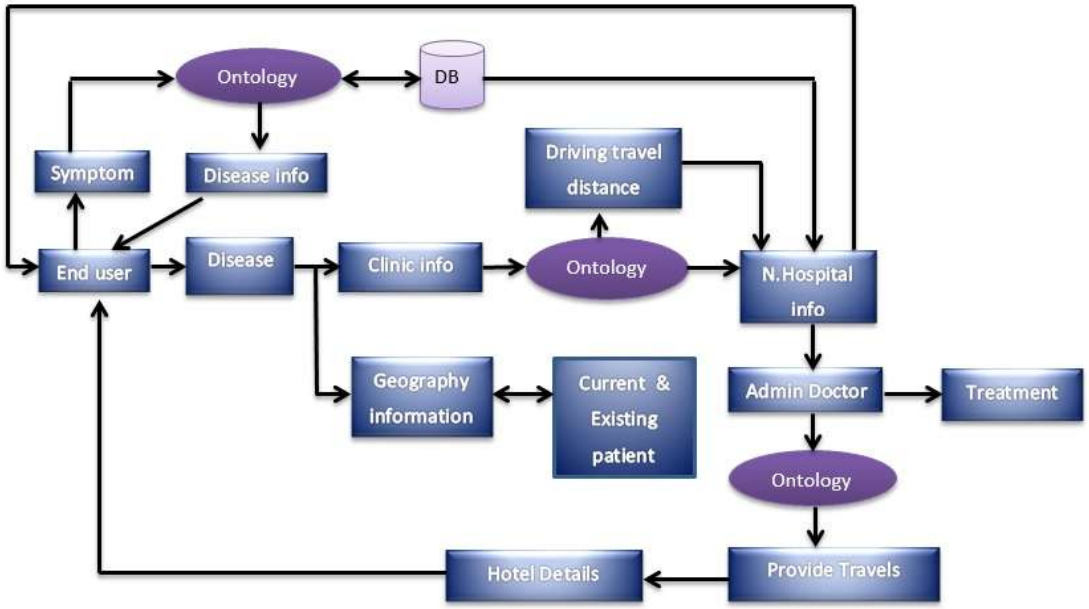

Figure 5 Integrated E-Healthcare system

This IR mechanism is used to facilitate the users to the disease treatment relationships based on the user's query which consists of symptoms, disease, clinic and doctors details. The objective of the system is achieved by the explicit specification of information about an E-Healthcare system which integrates disease-symptoms, clinic-doctors and transport-location. User request's contains the set of terms that specifies the information needed. Since the system is ontology based it can handle the terms which are semantically ambiguous.

The output of the query processing module is given to the WordNet to identify thesynsetsin the query.WordNet is used to obtain the related terms which are semantically associated with E-Healthcare domain. This can be achieved by the WordNet API [21] [22]. The extraction of information from the ontology starts with built ontology. The ontology for the E-Healthcare system is designed using the open source tool Protégé [23]. The structure of the built ontology is given in the following figure 6.

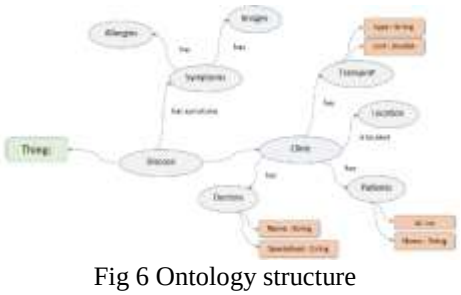

Fig 6 Ontology structure

IV. RESULTS AND ANALYSIS

The implementation of the system is done through a web application. The system receives an input from the users through the user interface design and the results are obtained based on the terms in the query. The performance of the system is determined by calculating the precision and recall values. The coverage of the system is determined by calculating the recall values. It is defined as the ratio between the relevant information retrieved by the system and the total amount of relevant information in the system.Precision is used to measure the accuracy of the system. It can be calculated by the ratio between the correct information retrieved and the total amount information in the system.

$$recall = \frac{Number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ relevant\ documents}$$

$$precision = \frac{Number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ documents\ retrieved}$$

The following table 1 shows the sample queries given by the users to the E-Healthcare system and the precision recall values are calculated. The graphical representation of the values are depicted in the figure 7.

TABLE 1 SAMPLE QUERIES AND PRECISION-RECALL VALUES

S. No	Sample Query	E-Healthcare IR System	
		Precision	Recall
1.	I had cough	0.68	0.44
2.	I have symptom of cough	0.62	0.41
3.	What can cause dry and persistent cough?	0.75	0.46
4.	What can I do about my chronic coughing?	0.6	0.55
5.	Have occasional cough and congestion feeling	0.68	0.46
6.	I am 12 weeks pregnant. Is it safe to take antibiotics for cough?	0.53	0.31
7.	What are the home remedies for cough?	0.68	0.5
8.	Why do we get morecolds and coughs in winter?	0.7	0.45
9.	How do you treat cough?	0.76	0.58

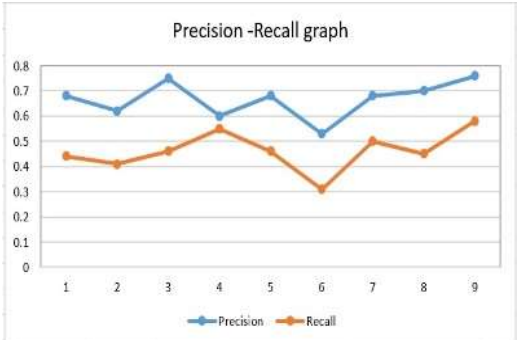

Fig 7 Precision-Recall values for E-Healthcare system

V. CONCLUSION

The objective of the system is to present an ontology-based information retrieval system for the healthcare domain. It calculates semantic similarity measures over the contents of the medical domain to determine the degree of likeness between two classes. Several efforts are reported on the use of ontologies in medical field. The main aspect of proposed approach is the use of ontology obtain the extract disease information with symptom based similarity. The proposed semantic search performed well in-terms of results than existing FCA (Financial Conduct Authority) couple search.

REFERENCES

[1] PrashishRajbhandari, Rishi Gosai, Rabi C Shah, Pramod KC, SEMANTIC WEB IN MEDICAL INFORMATION SYSTEMS, International Journal of Advances in Engineering & Technology, Nov. 2012.

[2] 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.

[3] Networking Health; Prescriptions for the Internet. Tech report TCD-CS-2001-01, Dept. of Computer Science, Trinity College Dublin; http://www.cs.tcd.ie/publications/tech-reports/.

[4] R.S. Dick et al, The Computer Based Patient Record: An Essential Technology for Healthcare,National Academy Press, Washington, D.C., 1997

[5] Berners LEE, Godel, Turing, "Thinking on the Web", Wiley, pp 39

[6] W. Borst. Construction of Engineering Ontologies. PhD thesis, Institute for Telematica and Information Technology, University of Twente, Enschede, The Netherlands, 1997.

[7] T. R. Gruber. A Translation Approach to Portable Ontologies. Knowledge Acquisition, 5(2):199–220, 1993.

[8] R. Studer, R. Benjamins, and D. Fensel. Knowledge engineering: Principles and methods. Data & Knowledge Engineering, 25(1–2):161–198, 199

[9] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, 2003.

[10] Franz Baader and Ulrike Sattler. An overview of tableau algorithms for description logics. StudiaLogica, 69(1):5–40, 2001.
[11] Giuseppe De Giacomo and Maurizio Lenzerini. Boosting the correspondence between description logics and propositional dynamic logics. In Proc. of the 12th Nat. Conf. on Artificial Intelligence (AAAI'94), pages 205–212, 1994.
[12] Lancaster, F., Elliker, C. and Colonell, T. H., 'Subject analysis', Annual review of information science and technology , 24, 34-84.
[13] Faloutsos, Christos, and Douglas W. Oard. "A survey of information retrieval and filtering methods." (1998).
[14] Shah, N.H., Cole, T. and Musen, M.A., 2012. Analyses using disease ontologies. PLoSComputBiol, 8(12), p.e1002827.
[15] Silberschatz, Abraham, Henry F. Korth, and ShashankSudarshan. Database system concepts. Vol. 4. New York: McGraw-Hill, 1997.
[16] Domenig, Ruxandra, and Klaus R. Dittrich. "Query Preprocessing for Intergrated Search in Heterogeneous Data Sources." In Datenbanksysteme in Büro, Technik und Wissenschaft, pp. 154-163. Springer Berlin Heidelberg, 2001.
[17] http://www.linkeddatatools.com/querying-semantic-data
[18] Oleshchuk V., and Asle P. (2003), Ontology Based Semantic Similarity Comparison of Documents, Proc. of IEEE 14th workshop on database and expert systems applications.
[19] Thiagarajan R., Manjunath G., and Markus S. (2008), Computing semantic similarity using ontologies ISWC 08, the International Semantic Web Conference (ISWC), Karlsruhe, Germany.
[20] Hajian B., and Tony W. (July 2011), A method of measuring semantic similarity using a multi-tree model proceedings IJCAI 2011 - 9th Workshop on intelligent techniques for web personalization & recommender systems (ITWP'11) Barcelona, Spain, 16.
[21] George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.
[22] https://wordnet.codeplex.com/
[23] http://protege.stanford.edu/
[24] PriyaGovindarajan, Ravichandran K.S., "Data mining- an evolutionary arena", Research Journal of Applied Sciences, Engineering and Technology, 7(22), pp. 4749-4753 (2014)
[25] PriyaGovindarajan, Ravichandran K.S., "A Framework For Identifying Disease-Treatment Relationship Using An Updated Database",International Journal of Applied Engineering Research,10(16), pp 36802-36805(2015)
[26] PriyaGovindarajan, Ravichandran K.S., "A Combined Schema for Surveillance", Research Journal of Information Technology, 6(4), pp.406-412 (2014)
[27] PriyaGovindarajan, Ravichandran K.S.,"Comparative study of soft-computing methodologies and its medical applications", Biomedical Research, 27 (2): 368-371 (2016).