A Novel Search Engine to trace Medical Information Needs using Medical Domain Ontology

M.Revati, K.Nageswara Rao, M.N.V.Kiran Babu
Department of Computer Science and Engineering
Mother Theresa Institute of Technology and Sciences
Sathupally, India
revathi.cse.jrg@gmail.com, ksn_choudary@yahoo.com, kiranbabuonline@yahoo.co.in

Kolikipogu Ramakrishna
Department of Computer Science and Engineering
JNT University Hyderabad
Hyderabad, India
krkrishna.cse@gmail.com

Ch.Raja Jacob
Department of M.C.A
Nova College of Engineering and Technology
Jangareddygudem, India
rchidipi@gmail.com

***Abstract*— Information retrieval in medical domain is now sharing major part of the web search. Now a day's most of the people especially adults are browsing health care and medical information at their homes using internet. Medical Information Retrieval System (MIRS) through search engines providing positive information to the user based on the fixed questionnaires. In this paper we build a model for naïve users, who are having minimal knowledge to feedback the system by opting listed relevant questionnaire. Along with the framework, we also built an Intelligent Medical Search Engine (IMSE) for searching medical information on World Wide Web (WC3). The implementation setup of IMSE uses medical Ontology and questionnaire to facilitate naive internet users to search for medical information. IMSE introduces and extends expert system technology into the search engine domain. IMSE uses several key techniques to improve its usability and search result quality.**

Keywords-Medical Information Retrieval, Medical Ontology, Knowledge Network, Diagnosis Reports;

I. INTRODUCTION

Usage of the internet becomes so popular in almost every country. People use internet to share the information needs, means users may search for their information needs or they may share the information with trusted parties. Every day billions of people use internet for their personal information needs. As per the survey of H&HN DAILY magazine 80 percent of Internet users look online for health information, making it the third most popular online activity among those tracked by the study, trailing only e-mail and using search engines. Roughly 44 percent of Internet users look online for information about doctors and other health professionals; 36 percent look up information on hospitals and other medical facilities. Similarly the PWC survey found that only 14 percent of Americans currently access their medical records electronically, the upward trends in online health engagement suggest those numbers will climb dramatically in the next few years. By observing these facts we proposed to develop enhanced and user friendly (i.e. trusted) medical search engine. In order to capture major share of the users who periodically search their medical information needs, we build this novel search engine. As a road map we started looking at various Medical search engines so as to find the enhancements and requirements of the naïve users. Current medical search engines like OmniMedicaSearch medical Web[11.], HealthLine [8.], Medical World Search [10.], PogoFrog [20.] for physicians and healthcare professionals and MyHEalthCare [17.]

etc., search engines are unsupervised and untrained. These are meeting the needs for group of targeted people, but mostly lack in satisfying all categorical people who surf the net for disease diagnosis reports, root causes of diseases and other specific information needs. Since ordinary users often cannot clearly describe their symptoms due to lack of medical knowledge, we planned to use feedback query expansion approach by providing relative questionnaire options. This local query expansion is built to direct the user to provide known ill health feelings through an interface. Apart from this we use medical domain ontology knowledge structure to automatically form multiple queries from a searcher's answers to the questions. Using these queries to perform search can significantly improve the quality of search results. The search result is presented to the user in an organized way by ranking. It is also presented in a multilevel hierarchy with explicitly marked medical meanings to facilitate searchers' viewing. Our model finally suggests diversified, related medical phrases at each level of the search result hierarchy. Health diagnosis use medical knowledge, heavy dependence on medical knowledge leads to complex reasoning process. Basic key terms similarity matching techniques fails in handling most of the important issues in this reasoning process, including the symptom absence, some properties of specific symptom, age of the patient and quantitative test results. Good search results will be gained if it allowed knowledge structure support to search. (i.e. use of medical knowledge to transform these answers into appropriate keywords, and combine them with the other keywords that the searcher inputs into the text areas to form multiple queries). Instead of using single query, The Medical Search engines uses provided collection of queries to perform search. The system needs lot of exposure in all points of consideration like a doctor or professional diagnosis report. It means more explanatory information is needed to be supplied by the user to match all the combinations. In general naïve users are not good enough in vocabulary of the medical domain to provide detailed search query, hence the questionnaire supports the decision making process. Even though the user opts various questionnaires, the Medical Web search engine could able to provide unique diagnosis report as a result. It gives relevant and needful collection of results, in which user finally has to decide his relevance to the results. It is considered as successful search, if it gives relevant returns to the user. Hence, during the transformation from question answers into query keywords, we only need to maximize the probability that the resulting query keywords can facilitate searchers to find useful information. Searchers can refine their inputs multiple times after reading the search results This It reduces the burden on transformation process. Unlike general search engines, we wanted to give flexibility to the user to opt his choice and it is to be repeated till he or she satisfied with results. So as to reach this target, we proposed and built a novel Medical Search Engine by enhancing with Medical Domain Ontology. MedSearch automatically rewrites long queries into moderate-length queries by selectively dropping unimportant terms (i.e., words). Since unimportant terms not only appear in a large number of Web pages but also obscure the main theme of the query, dropping them can both greatly increase the query processing speed and improve the quality of search results[6.]. We organized this paper into five chapters where we briefed the Related Work in chapter 2, Chapter 3 with model and framework, Chapter 4 Implementation Setup, Chapter 5 Result analysis followed by conclusion and future Directions.

II. RELATED WORK

Most of the Medical search engines were found limited in information Retrieval due to vocabulary mismatch. We made a survey on medical search engines and addressed various limitations. In this paper we proposed a novel model in-order to expand the user interaction so as to get improved search results. Some Medical Search Engines were found that they use the patient record for fetching the diagnosis reports, but it is clearly less effectively in search process. Thus the improvement offered by [19.] is attributable to its more nuanced weighing schemas as well as its use of the entire patient record as a query mechanism. Online search engines are notorious for overloading end users with irrelevant information. In healthcare settings, personalization of search to the individual patient can aid in filtering out unneeded results. Physicians, and in particular, physicians in training, need information that is clinically relevant to the patient under their care [6]. Personalization is done by re-ranking the results returned by the search engine using a more intensive natural language analysis of the documents [13.]. Portals and Directory Services could use GRID focused crawlers to automatically monitor and populate large fresh repositories on a variety of topics. More than one crawler that are actively exploring the same semantic topic on the Web and that therefore are bound to fetch an increasingly overlapping set of documents, wasting the parallelization effort [12.]. No existing Information System (IS) or Information Technology (IT) solution could effectively support the interpretation decision-making, esp. regarding the variant significance [21.]. Studies to what extent the content of different (medical) Web resources varies to provide the opportunity to create better health-specific search engines that can exploit this knowledge for ranking and sorting search results [15.] is to be specified. Health-related blogs are written by health care professionals or by patients. Patients share their experiences on the disease they are suffering from or on a particular treatment, or exchange other health-related information [10]. When a user wants to know about information on a very specific topic, they have troubles finding the correct search terms because the professional and lay medical vocabularies are not always well-matched. This problem can be solved by mapping user query to related medical concepts [1.]. HONSelect (Health on the Net Foundation) is a new search integrator that strictly works for health and medical queries; it integrates various collections like scientific articles, healthcare news, web sites and multimedia [9.]. HONselect is also having limitations in providing Information about the patient, available in the online patient record, can be used to provide a personalized response to a physician's search query. Re-ranking and summarization leverage a common representation of articles and patient record to make their tasks easier [14.]. Compared with a relevance-driven search engine, a well known for outperforming other search methods on MEDLINE search tasks [5.]. In this literature survey we analyzed most of the dimensions of the medical search and found they are not meeting the needs of information searchers. We built a novel framework to concise the retrieval results that are relevant to the end user.

III. MEDICAL SEARCH ENGINE FRAMEWORK

We model a framework to reduce the overhead in searching medical interested document with Medical ontology support. Figure 1 gives flow of steps to retrieve relevant documents from the web.

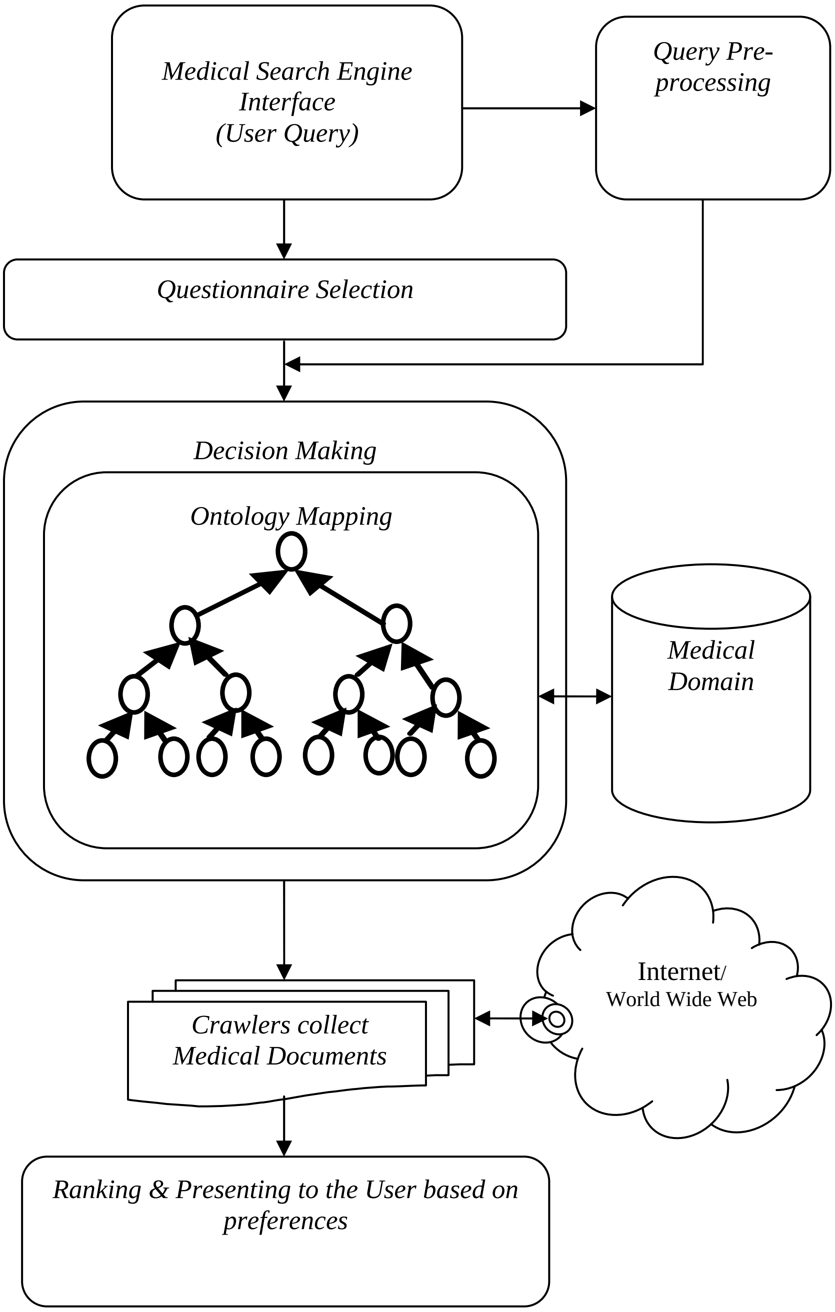

Fig.1 Medical Search Engine Framework

IV. MEDICAL DOMAIN ONTOLOGY

Naïve user sometimes fails in building good query to search due to lack of domain knowledge. To address this curtail of naïve user we proposed to use Medical domain ontology. Ontology is a co-related set of concepts with in a domain. This represents knowledge structure that explores the concept of a user query. Domain Ontology is a collection of terms with proper relationships of particular area, where Medical domain is collection of medical related terms with built-in relationships. To support the user for building a precise query, we embedded Medical Ontology to expand the user query.

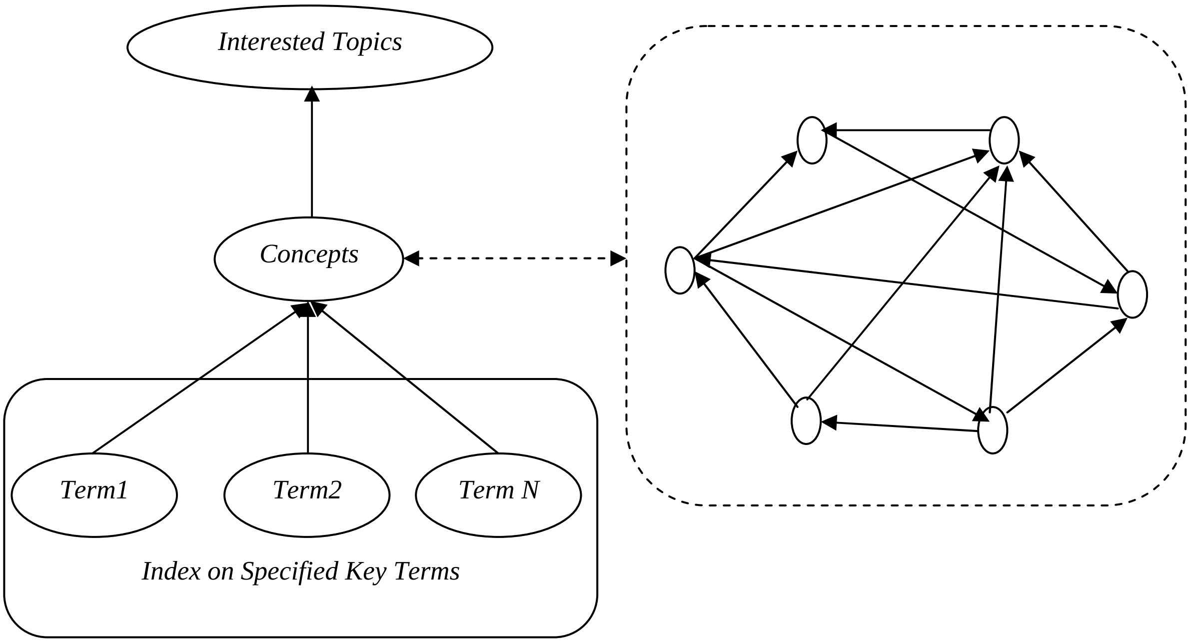

Fig.2. Concept Network

Concepts are derived by analyzing a huge number of web sites for particular domain by studying the initial key term's closed words. The user terms or concepts are preprocesses in order to select the most suitable terms by statistical similarity measures. The selected classes are finally incorporated to the ontology. The information regarding from where it was extracted, that is to be stored and processed recursively to build concept hierarchies and to find new related terms. The resulting collection of terms can be base for finding more complex ontological relations between concepts [3]. It is also used to guide a search for information or a classification process from a document corpus [7, 2, 4, 16]. The proposed ontology frame work is shown in Figure 1. Naïve user framework is constructed based on the knowledge of the domain. Knowledge structure of Medical domain has been considered here. The knowledge representation database, ontology, is organized into a three level hierarchical structure that is shown in Figure 2.

A. Topic Hierarchies

Root level topics share a parent child relationship, which provides generalization from a specific topic to a more general topic. These hierarchies of topics are stored in an n-ary tree structure with an exception that a node may have multiple parents as shown in Figure 3. In Medical domain the same topic may be represented as different topics. For example, in the domain of Medical, Agronomy and Medical represents same topics. Plant pathology is a subtopic of the topic Medical.

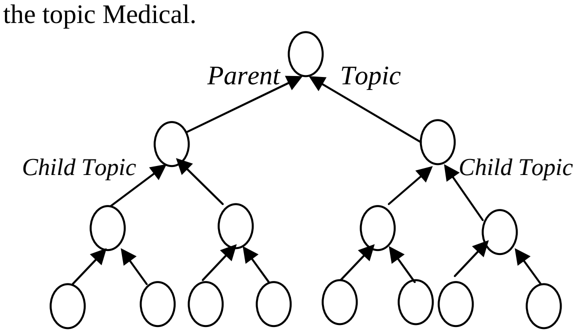

Fig 3. Topic Hierarchies

B. Concept Hierarchies

In general a topic may consist of many concepts, then single topic can be represented using multiple concepts which form the next level of the ontology and a concept may belong to one or more topics. A set of regularly used relations can be defined among the concepts. The relationships among such concepts includes *Is a Part of, Is a kind of, Contains, Causes, Acquired from etc.* The concepts in the domain are organized into a di-graph as shown in Figure 2. The existence of an edge between two concepts in the di-graph indicates that the concepts are related.

C. Key Terms Relations

A concept is associated with several keywords or terms with specificity index [18.]. The specificity index stores the likelihood of the keyword representing a particular concept. These keywords are used to extract concepts from items and queries. The association of the key terms to the concepts has several advantages. Firstly, the different keywords having the same meaning are mapped to a common concept removing the synonymous ambiguity of keywords (Word Sense Disambiguation). For example “pain” keyword is disambiguated by adding “leg pain”, “nose pain” and “muscle pain”. Similarly in Medical field “pain” and “ache” are taken as single index terms and sometimes individual key term.

D. Item preprocessor

The Item Preprocessor takes input and pre-processes the natural language processing on source document set, which is divided into two steps, the term analysis and the content analysis.

E. Term Analysis

The term analysis uses a part-of-speech (POS) tagging system and a self-built term filter to collect a meaningful term set containing nouns. A measure of the relatedness between words can be a knowledge source for several decisions in natural language processing (NLP) applications. Our approach is to construct a linguistic context for each sense of the verb and noun, and to measure the number of nouns shared by the verb and the noun contexts. The purpose of content analysis is to find derivative words of each meaningful term. Therefore, according to all of the POS Tags, there are two kinds of sets, a document set D_i of RN collected Medical pages and a term set T_i of meaningful terms selected by term filter. A document set is expressed as $\{D_1, D_2 \ldots Dn\}$, where $n >= 2$ and a term set of meaningful terms is as $\{T_1, T_2 \ldots Tm\}$, where $m>=2$ ***then*** **the co-**occurrence of such terms is represented as

$$\text{Co-occurrence}(D_i, T_j) = \begin{cases} 0, \text{ when } Di \text{ co occurs with } Tj \\ 1, \text{ otherwise} \end{cases} \quad \text{------------ (1)}$$

With this formula (1) we use vector matrix to represent term relationships as M_{ij}:

$$M_{ij} = \begin{pmatrix} X_{11} & X_{12} & X_{13} \ldots X_{1m} \\ X_{21} & X_{22} & X_{23} \ldots X_{2m} \\ \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & X_{n3} \ldots X_{nm} \end{pmatrix} \quad \text{------------- (2)}$$

By the matrix in (2) for minimum global support used to find derivative words of each meaningful term and the result is stored in the repository.

F. Knowledge Management System

Knowledge Management Subsystem plays two important roles in this architecture. In first, is to get questions from users, to provide for users the answers to the questions they ask, and to provide tools for domain experts to validate the accuracy of the purpose built query-based partial Medical ontology. Users can ask questions in nature languages. The Second role is to find the answers to the questions.

V. EXPERIMENTAL SETUP

We developed a graphical user interface, associated through a browser interface. The GUI'S at the top level have been categorized into Administrative user interface and The operational user interface.

A. Administrative user interface

The administrative user interface concentrates on the consistent information that is practically, part of the organizational activities and which needs proper authentication for the data collection. The interfaces help the administrations with all the transactional states like Data insertion, Data deletion and Data updating along with the extensive data search capabilities.

B. The operational or generic user interface

The operational or generic user interface helps the users upon the system in transactions through the existing data and required services. The operational user interface also helps the ordinary users in managing their own information helps the ordinary users in managing their own information in a customized manner as per the assisted flexibilities. For implementation of the Framework, we divided task into modules.

a. Normal User

Any person who enters into our site will be a normal user. The user will have all the permissions to search based on two conditions i.e. Word Based Search and Questions and Answer.

b. Administrator

Administrator will be differentiated from other normal users by providing a username and password. Administrator will have the rights to enter or insert new search related data into the site. We built a system using Java, Servlets so as to simulate the nature of the Medical search engine.

Results

We compared our framework implementation results with existing Medical search Engines in terms of Relevancy of retrieved documents. This Medical search engine framework gave measurable improved results comparing with existing base Medical search engines. This is happened due to medical ontology support to the user to expand the root query.

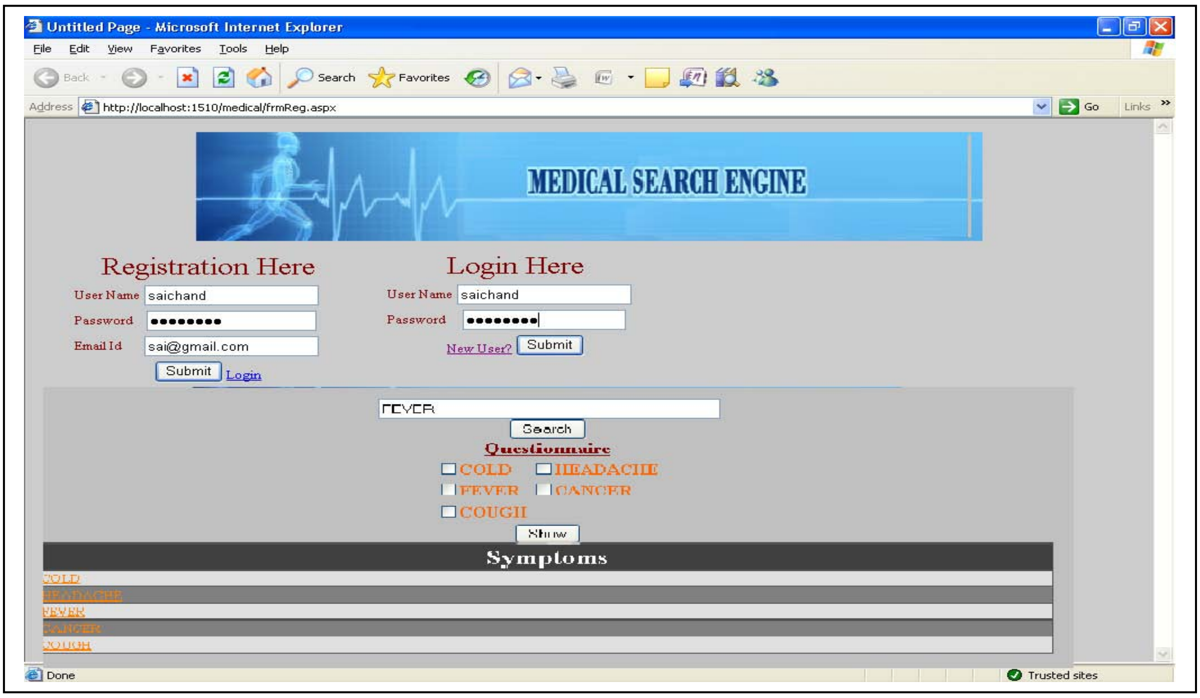

Fig.4.Registration and Login Interface

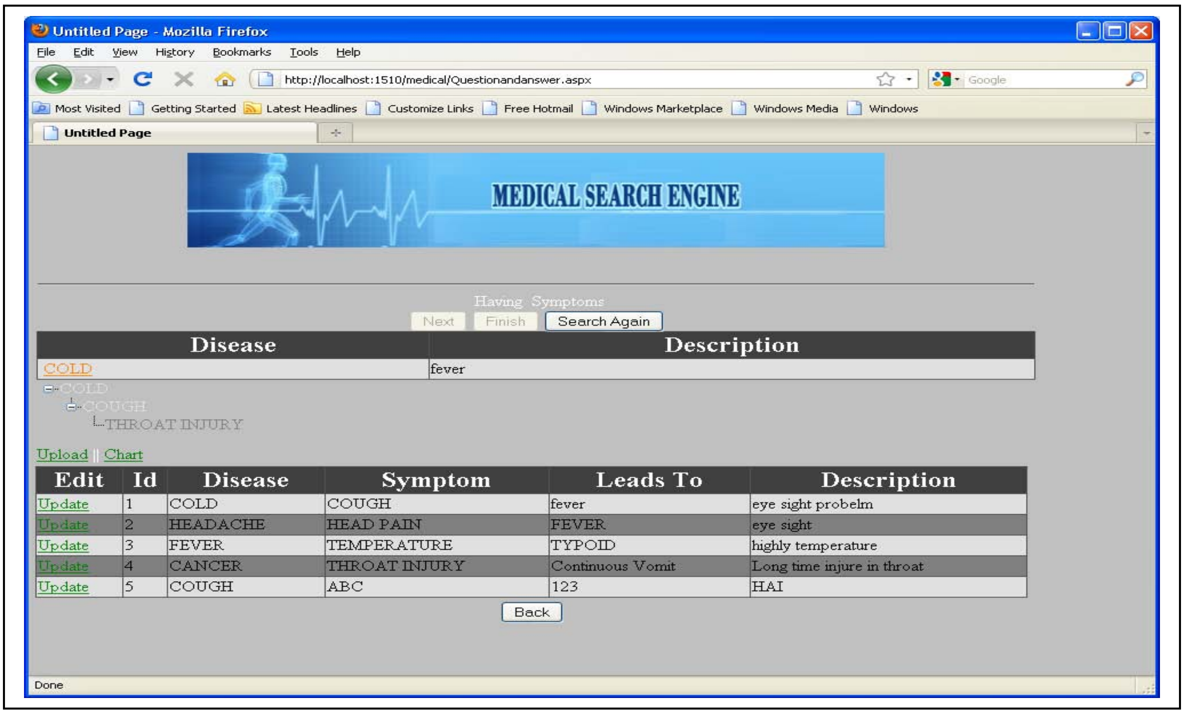

Fig.5.Customized User search

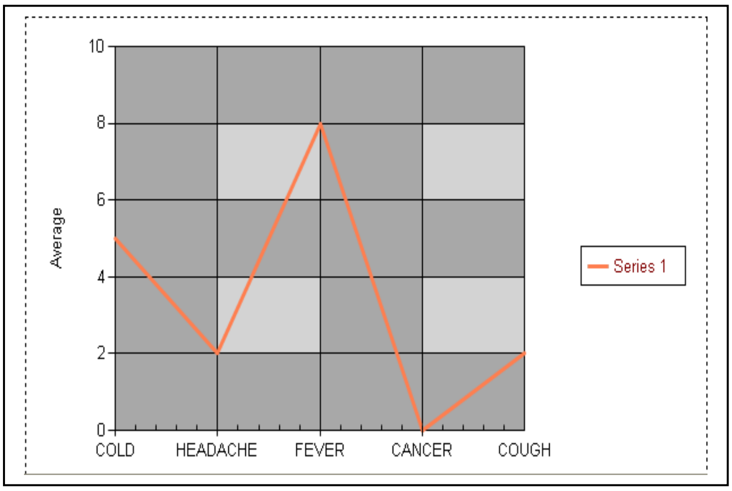

Fig.6.Diagnosis Report

This framework is tested by implementation. The tested model gave précised and exact user expected result as shown in fig.4,5 and fig.6.

ACKNOWLEDGMENT

First author expresses her sincere thanks to all the Staff members, Department of Computer Science and Engineering, Mother Theressa Institute of Technology and Sciences (MIST), Sathupally, A.P, India for their support and continuous encouragement to bring this work to the reality.

REFERENCES

[1] A.A.Kogilavani1, B. Dr.P.Balasubramanie2, Design and Evaluation of the iMed Intelligent Medical Search Engine,International Journal of Recent Trends in Engineering, Vol. 1, No. 1, May 2009,546-549.
[2] A.Aldea, R.Bafiares-Alcantera,J.Bocio An Ontology bases knowledge management platform, IIWEB'03,IJCAI'03, 177-182,2003.
[3] A.Maedche, R.Volz,J.U.Kietz, A method for Semi-automatic Ontology Acquisition from a Corporate Intranet,EKAW'00 Workshop on Ontologieas and Texts,2000.
[4] A.Sheth, Ontology – driven information search, integration and analysis, Net object Days and MATES.2003.
[5] Aronson, A.R., Demner-Fushman, D., Humphrey, S.M., Lin, J., Liu, H., Ruch, P., Ruiz, M.E., Smith,L.H., Tanabe, L.K., Wilbur, J. (2005) Fusion of knowledge-intensive and statistical approaches for retrieving and annotating textual genomics documents. TREC, National Institute of Standards and Technology,TREC14,2005.
[6] Gang Luo1 Chunqiang Tang1 Hao Yang1 Xing Wei2, MedSearch: A Specialized Search Engine for Medical Information,WWW 2007, May 8–12,2007, Banff, Alberta, Canada.ACM 978-1-59593-654-7/07/0005,Pp:1175-1176.
[7] H.Alani, S.Kim, D.Millard, Automatic Ontology based Knowledge Extraction from Web Documents. 14-21, IEEE Intelligent Systems, IEEE CS, 2003.
[8] http://www.healthline.com.
[9] http://www.hon.ch/MesH.
[10] http://www.mwsearch.com.
[11] http://www.omnimedicalsearch.com.
[12] K.cerbioni,E.Palanca A GRID FOCUSED COMMUNITY CRAWLING ARCHITECTURE FOR MEDICAL INFORMATION RETRIEVAL SERVICES,2009.
[13] Kanter.A; F. Naeymi-Rad; I. E. Buchan. Right information, right patient, right time: Intelligent content searching supporting point-of-care In Proc. Annual Symp. AMIA, pages 403–applications.407, 2000.
[14] Kathleen R. McKeown Noemie Elhadad Vasileios Hatzivassiloglou, Leveraging a Common Representation for Personalized Search and Summarization in a Medical Digital Library,2000
[15] Kerstin Denecke, STUDYING CONTENT DIFFERENCES OF MEDICAL ONLINE WEB RESOURCES
[16] L.Magnin, H. Snoussi, J. Nie, Toward an Ontology-based Web Extration, 5th Canadian Conference on Artificial Intelligence, 2002.
[17] myhealthcare.com.
[18] P.L.Aban Kumar,Bhowmick, Ontology Based User Modelling For Personalized Information Access ,IJCSA, vol 7, No 1, 2010.
[19] Simone T, Personalizing Retrieval of Journal Articles for Patient Care,AMIA 2001,696-700.
[20] www.pogofrog.com.
[21] Yulong Gu1, James Warren2, Jan Stanek1, Graeme Suthers3, A System Architecture Design for Knowledge Management (KM) in Medical Genetic Testing (MGT) Laboratories,ICCSCWD - proceedings, 2006.

AUTHORS PROFILE

Author 1 : M.Revati, is a M.Tech (CSE) II year Scholar, Department of Computer Science and Engineering, Mother Theresa Institute of Technology and Sciences(MIST), Sathupally,India.

Author 2 : Mr.K.Nageswarao,M.Tech,(Ph.D) is Professor and HOD of the CSE Department, MIST,Sathupally. To his credit he has couple of publications both natinal and international conferences / journals. His area of interest includes Advances in Computer Applications.

Author 3 : Mr. M.N.V.Kiran Babu,M.Tech is working as Associate Professor in Department of CSE,MIST. His are of research includes Data Mining.

Author 4: Mr.Kolikipogu Ramakrishna received B.Tech (CSIT) and M.Tech (CSE) from JNT University. Currently he is pursuing Ph.D from JNT University, Hyderabad in the Faculty of Computer Science and Engineering. He is an active member of ISTE, CSI, IACSIT, IAENG. He has 8 years of teching experience in various engineering colleges. To his credit he has published 8-international journals and 4 – international conferences papers. He is an active member of Infosys campus Connect and IBM Grat Mind Challenges Programs. His area of Research Interest includes Informatin Retrieval systems, Artificial Intellegence, Data Mining, Natural Languages Processing, Matchine Learning.

Author 5 : Mr.Ch.Raja Jacob, M.C.A is working as Associate Professor and HOD, Department of MCA, Nova college of Engineering and Technology, His area of Interest includes Data Warehouse and Data Mining, Operating systems and other computer Applications.