International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Data Processing in Information Systems: Health Care Data Analysis

Srikanth Bethu[1*], B Sankara Babu[2], R. Aruna Flarence[3]

[1*, 2, 3] Deptartment of Computer Science & Engineering, GRIET College, JNTU, Hyderabad, India

[*] *Corresponding Author: srikanthbethu@gmail.com, Tel.: 9989707702*

Available online at: www.ijcseonline.org

Accepted: 24/Jul/2018, Published: 31/July/2018

***Abstract*— searching for medical information on the Web is popular and important. However, medical search has its own unique requirements that are poorly handled by existing medical Web search engines (WSE). The first online medical Web search engine that extensively uses medical knowledge and questionnaire to facilitate ordinary internet users to search for medical information. All existing medical WSEs assume that searchers can form appropriate queries by themselves. However, most Internet users do not have much medical knowledge. Frequently, a medical information searcher has only a vague idea about the problem that he is facing and does not know the proper way to clearly describe his situation in sufficient detail. As a result, appropriate guidance is highly necessary during the medical search process. This can be illustrated by an analogy to the medical diagnosis process.

In this paper we mainly focused on how health care data is analysed by a web user and how he is retrieving the information from the Data Processing Information Systems.

***Keywords*—Data Processing, Information Systems, Medical search Process, WSE**

I. INTRODUCTION

Online medical Web search engine that extensively uses medical knowledge and questionnaire to facilitate ordinary Internet users to search for medical information. It introduces and extends expert system technology into the search engine domain. It uses several key techniques to improve its usability and search result quality. First, since ordinary users often cannot clearly describe their situations due to lack of medical background, it uses a questionnaire-based query interface to guide searchers to provide the most important information about their situations. Second, it uses medical knowledge to automatically form multiple queries from a searcher' answers to the questions. Using these queries to perform search can significantly improve the quality of search results. Third, it structures all the search results into a multilevel hierarchy with explicitly marked medical meanings to facilitate searchers' viewing. Lastly, suggests diversified, related medical phrases at each level of the search result hierarchy.

Big Data is a normal data that are huge in size with lots of information in different format and lots of noise that cannot be mined using the traditional system. Sam Madden [1] stated that the data are too big, too fast, too hard and too complex to analyze with the existing system which is known as Big Data. The process of storing, analyzing, managing and visualizing the data is very difficult. According to Marko Grobelnik [1] Big Data is very similar to Small-data, Big Data requires a completely new tools and techniques to analyze and solve many real world problems in a better and an efficient way. The generation of huge volume of data (Big Data) leads to a development of an analytics called Big Data Analytics. Big Data Analytics is an sophisticated analytic technique used to analyze different types (structured, unstructured and semi structured) and size of data (terabytes to geophytes). The analytical process is used by many researchers, analysts and business people to make quick and accurate decisions. Big Data Analytics in healthcare contribute a major role in processing and analyzing the data in variety of forms to deliver suitable insights.

Social networking is one of the effective tool to make people aware of a particular product and easy reachable. The increased use of Social networking among public helps doctors to reach out to patients, guide them for treatments, provide counselling, creates close-knit support communities and faster recovery. Numerous blogs has been created and may users share large quantity of vision about many health care topics.

India with a population of billion plus people, one of the worlds's leading growing economies, 29.5% of population are underneath the estimated minimum level of income, and 46% of offspring are half-starved. Increasing population in India is a great threat to the health care structure of the country. The big data technologies are in early stage of its implementation and need further research and innovation to prove its reliability and effectiveness to stakeholders. This provides immense opportunities for researchers and industries to involve in early stage of its implementation. This collected data is in form of big data as the volume is large and in addition to that the data has high complexity, variety and also velocity as it is collected from different sources continuously. This leads to the need for a technology which can address above stated issues and provide effective solution, Hadoop technologies including Map Reduce, YARN comes in picture to achieve it. These technologies are in their early stages of their implementation; still it provides useful information to help the healthcare industry and reduces the cost.

II. Related Work

A. Online Medical Search Engine

Online medical search is a vertical WSE that crawls Web pages from a few selected, high-quality medical Web sites instead of the entire Web. In our experiments, we crawled 22GB of Web pages from WebMD, Health line, and Merck, three of the most popular medical Web sites[1] [2]. These Web sites cover the entire medical domain fairly comprehensively and include information on various topics such as symptoms, diseases, drugs, and treatments.

I Compared online medical search with two state-of-the-art medical WSEs: Google Health and Health line. We used both real medical case records from the Family Medicine Online Database (FMOD) and USMLE medical exam questions [3]. Correct diagnoses are available for both of them and serve as the ground truth for our evaluation. USMLE stands for the United States Medical Licensing Examination, whose exam question format is similar to the format of actual, well documented medical case records. Physicians have to pass this exam to obtain their licenses for practicing medicine. In our tests, each exam question is treated as a medical case. FMOD was developed by the College of Medicine of the Pennsylvania State University for educating medical students.

The FMOD records document patients' medical situations in great detail using mostly layman terms and can be easily understood by ordinary people.

I randomly selected 30 medical cases from the FMOD records and the USMLE exam questions. Since both USMLE and FMOD cover almost every aspect of medical practice [1][3], our random samples have a broad coverage of medical topics.

In my experiments, a user has up to 60 minutes to perform iterative search for each medical case. At the end of the search process, the user can list up to three diseases that he thinks best match the medical case's situation description. If any of these diseases is among the correct diagnoses accompanying the data set [4], the search is considered successful. I allow users to search for a relatively long time, because medical information searchers care about their health and often spend hours on searching. I allow users to list multiple diseases as their findings, because even doctors sometimes cannot make precise diagnosis without lab test results.

Ten colleagues served as assessors and users. None of them has formal medical training. For a medical case, each user randomly selected one of the three medical WSEs [5] (Online medical search, Google Health, or Health line) with equal probability to perform search. My experiments were performed on a computer with two 3GHz processors, 2GB memory, and one 111GB disk.

Similar to the TREC interactive track, I use two sets of measures as the performance metrics for medical WSEs: one set is objective while the other set is subjective. The objective performance measures include the success rate, the number of search iterations, the number of search result Web pages viewed, and the time spent on the search process. The subjective performance measures include the users' perceptions of ease of using the system, ease of understanding the system, usefulness of the search results[5][6], and overall satisfaction with the system. For online medical search, both the average usefulness of the overview Web pages for the top 10 diseases (for determining whether these diseases are related to the medical case's situation description) and the average usefulness of the top 10 suggested medical phrases at the first level of the search result hierarchy are also included. They were obtained from a brief questionnaire that users filled out after using the systems. For each objective or subjective performance measure, I average it over all the 30 medical cases and all the users, and report both its mean and its standard deviation when appropriate.

To give the reader a feeling of the contents returned by Online medical search, I present detailed results of the returned Web pages and the suggested medical phrases for a typical query scenario that corresponds to choosing "little or no sputum" and "no dispend" for the symptom cough. Table2.1 shows some relevant Web pages returned at the first level of the search result hierarchy. The suggested relevant medical phrases include silicosis (rank 1), smoking cessation (rank 2), pneumoconiosis (rank 3), oesophagi is (rank 4), respiratory system (rank 5), and bacterial infections. For a query scenario *Qs*, Online medical search generally can find several relevant web pages and medical phrases describing multiple topics related to *Qs*.

In this section, I present the overall experimental results. Online medical search is efficient at performing medical search. For all the 30 medical cases, the average time taken by online medical search to generate each part of the search result hierarchy is less than two seconds. Online medical search is much more effective than the other two medical WSEs in finding the correct diagnosis, where most of the user's time is spent on reading the search result Web pages. The objective performance measures in Table 2.1 shows that compared to the other two medical WSEs, Online medical search makes the user find results in fewer iterations, view fewer search result Web pages[7][8], spend less time on the search process, and achieve a higher success rate. All these differences are statistically significant.

Table 2.1: Objective Performance Measures

Mean	Online medical search	Health line	Google health
Success rate	30%	23%	21%
Number of iteration	3.9	5.9	6.1
No .of search result	14	20	21
Time(min)	31	41	43

Table 2.2: Subjective Performance Measures

Mean	Online medical search	Health line	Google health
Ease of using	5.7	4.9	4.9
Ease of understanding	5.7	5.8	5.8
Usefulness	5.2	4.3	4.2
Satisfaction	5.0	4.2	4.2

Table2.2 shows the subjective performance measures. All the users are familiar with the traditional keyword query interface and the sequential order presentation of search results. It took these users a while to become accustomed to navigating the search result hierarchy in answer interface. As a result, users think that the traditional WSE user interface is slightly easier to understand than user interface, while the difference is not statistically significant. Nevertheless, once users understand Online medical search user interface [9], they can use it without difficulty. Online medical search answer interface has explicitly marked medical meanings and organizes together all the search results on the same topic or aspect so that users can find them easily. Users are also accustomed to using questionnaires in daily life. Consequently, users think that user interface is easier to use than the traditional WSE user interface. Overall, users think that online medical search produces more useful search results and is more satisfactory than the other two medical WSEs. These differences are statistically significant.

III. METHODOLOGY

Different types of modules are used to develop the online medical search engine. Those modules are,

3.1 Modules

- User Interface
- Search Techniques

3.1.1 User Interface

The user interface of online medical search engine contains two parts:

- The query interface
- The answer interface.

1. The Query Interface

In this section, it describes the query interface. In practice, I would expect most users of medical search to be ordinary Internet users without much medical knowledge, while medical professionals can also use this to help them accomplish their tasks. In designing of query interface, I adopt the following principles to provide the greatest convenience to medical information searchers:

- Principle 1: Minimize searchers' efforts.
- Principle 2: Be easily accessible to ordinary users without much medical knowledge.
- Principle 3: Be tolerant of imprecise user inputs.
- Principle 4: Allow incomplete inputs.

2. Query Interface Overview:

Fig. 3.1 shows the first screen of query interface. There are two possible cases:

- Case 1: If the medical information searcher knows the appropriate query keywords (e.g., the exact name of the disease, the medicine, the test, the procedure, or the treatment), he can use the traditional keyword search interface to find desirable search results. In this case, online medical search works in the same way as existing medical WSEs.
- Case 2: If the searcher does not know the appropriate query keywords, he can use the questionnaire-based interface that unique to online medical search to guide him through the search process. In this case, the techniques used in online medical search complement the techniques used in existing medical WSEs, online medical search uses medical knowledge to form keyword queries to perform search.

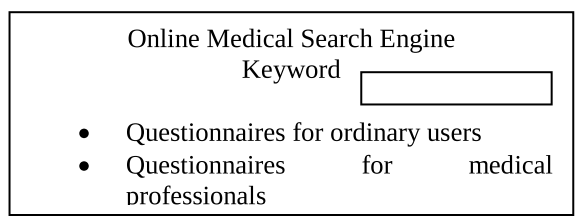

Fig 3.1 the First Screen of The Query Interface

It provides questionnaires for both ordinary users and medical professionals. Compared to the questionnaire for ordinary users, Questionnaires for Medical Professionals this questionnaire is more accurate for diagnosis purpose and can handle more difficult cases, while it often needs to ask more questions to significantly narrow down the list of possible diseases. In the rest of the paper, we focus on the questionnaire for ordinary users. The questionnaire for medical professionals is similar and omitted. In the questionnaire, the searcher first selects subjective symptoms (e.g., fatigue) and objective signs (e.g., hypertension), and then answers questions about their detailed descriptions. The searcher can also input other useful information that is not covered by the questions into text areas.

3. Symptoms and Signs:

In this questionnaire for ordinary users currently covers all the 267 symptoms and signs described in Collins. It would be overwhelming to display all these symptoms and signs to searchers on a single page. Instead, it organizes this questionnaire into two levels. As shown in Fig. 3.2, the first level of this questionnaire contains the 12 most frequently encountered symptoms and signs accounting for more than 80% of the chief complaints with which physicians are confronted, and an "others" option. All the other 255 symptoms and signs described in Collins are included in the "others" option as the second level of this questionnaire. To facilitate search, it classify those 255 symptoms and signs into multiple categories based on the affected body parts (e.g., general, head, neck, chest, abdomen, back, pelvic, extremities, skin). In most cases, the searcher can quickly find the appropriate symptoms and signs by checking only the first level of the questionnaire.

Symptoms and Signs

- ○ Abdominal Pain
  Chest Pain
- ○ Cough
  Vision Problem
- ○ Fever
  Weight Gain
- ○ Headache
  Weight Loss
- ○ Facial Pain
  Skin Problem
- ○ Pain in the foot
  Swelling of the legs

<<Previous Next>>

Fig 3.2 The First Level of The Questionnaires

The first level of the questionnaire, for each of the 267 symptoms and signs covered in the questionnaire, if its name is written in Collins in medical phrases unfamiliar to ordinary users, we use the consumer health vocabulary to annotate its name with layman terms.

For example, the symptom "haemoptysis'" is explained as "coughing up blood." As described in Zeng and Tse[10] , the consumer health vocabulary is constructed from medical WSE query logs. It provides a mapping between medical phrases and layman terms frequently used by medical information searchers. Ordinary users can easily understand all the symptoms and signs written in layman terms in the questionnaire. From all the 267 symptoms and signs in the questionnaire, the searcher can choose multiple of them reflecting his situation. Generally, when a doctor conducts medical diagnosis, he first identifies the chief complaints among all the patient's symptoms and signs (often there is only one chief complaint) and then performs analysis mainly based on these chief complaints. However, in medical search, ordinary users usually have no rigorous medical training and cannot correctly identify their chief complaints. To address this issue and to avoid missing important search results, it allows searchers to select multiple symptoms and signs without specifying their chief complaints.

4. Question Pages

For each of the 267 symptoms and signs covered in the questionnaire, Collins has a companion diagnostic decision tree *Td*. Each leaf node *N* of *Td* contains the disease names that are most relevant to the branching conditions (in the non-leaf, non-root nodes) leading to *N*. it uses these diagnostic decision trees to prepare questions for the symptoms and signs and also to transform question answers into query keywords. In this section, we show how questions are generated one by one using these trees. In Section, we show how to transform question answers into query keywords. After obtaining all the symptoms and signs chosen by the searcher, it will generate question pages to ask questions about their detailed descriptions. Each question page contains one or more questions. The questions in the next question page are selected according to the answers the searcher provides to the questions in the previous question pages, it traversing the corresponding diagnostic decision trees for these symptoms and signs. It can display all the used diagnostic decision trees on the answer interface.

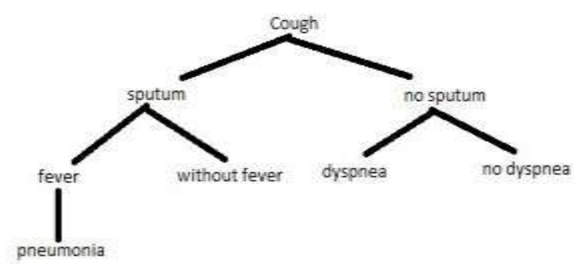

Fig 3.3 the Diagnostic Decision Tree For the Cough

Cough

Is there significant sputum production

- Yes
- No
- No An

Other inputs:

<< Prev next>> finish

Fig 3.4 the Question Page That is generated for the Cough

The diagnostic decision tree for the symptom cough that is described in Collins. If cough is the only symptom chosen by the searcher, the first question page generated will contain a single question "Is there significant sputum production?", as shown in Fig. 3.4. If the searcher answers "yes" to this question, next question will be "Is the sputum purulent?" Otherwise if the searcher answers "no" to this question, next question will be "Do you have difficulty breathing?" Fig. 3.3. The diagnostic decision tree for the symptom cough. Fig. 3.4. The first question page that is generated for the symptom cough.

In generating questions, it uses the consumer health vocabulary to rewrite difficult medical phrases in diagnostic decision trees into layman terms. For example, "dyspnoea" in Fig. 3.3 is rewritten into "difficulty breathing." Also, online medical search may ask qualitative measures in the format of quantitative numbers and then convert these numbers into qualitative measures in order to traverse diagnostic decision trees. For instance, "Do you have fever?" can be asked as "What is your body temperature?" For each question asked by online medical search, the searcher can either answer it or provide no answer, it allows incomplete inputs. In the case that the searcher provides no answer to a question, it may use some "backup" question to replace as the diagnostic decision tree for a symptom or sign is generally not unique. For other useful information that is not covered by the questions, the searcher can input its keywords into the "other inputs" text area that appears on every question page. The searcher can stop answering questions and obtain search results at any time by clicking the "finish" button that appears on every question page. In general, the more questions a searcher answers, the more information has about his situation and the better the search results will be. A question page can contain more than one question in the following two cases. First, if the searcher chooses multiple symptoms and signs, it will ask questions about all of them. Second, some nodes in certain diagnostic decision trees have multiple descendant branches with non-conflicting conditions. When it reaches a node *N* in a tree, if the searcher either provides no answer to the corresponding question or selects multiple answers simultaneously, it cannot traverse along a single descendant branch of *N* and has to ask corresponding questions for all the (selected) descendant branches of *N*. When generating questions, it checks for redundancy to ensure that each same question is asked at most once. For example, at the first level of the questionnaire, if the searcher selects both symptoms cough and fever, fever will not be asked again when it generates questions for the symptom cough. Also, it only asks "consistent" questions. For instance, suppose the searcher selects a single symptom cough at the first level of the questionnaire. If he provides no answer to the question "Is there significant sputum production?", it will not ask questions about sputum properties, such as "Is the sputum purulent?" Instead, it treats all such questions as if the searcher provided no answer. All the redundancy and consistency checking in the question generation process is coded as rules. Most diagnostic decision trees written in Collins have depths smaller than five. Thus, it will usually stop asking questions and produce search results in fewer than five question pages. This fulfils Principle 1 of minimizing searchers' efforts.

3.1.2 Answer Interface

Fig.3.5. the hierarchical structure of the answer interfaces. For completeness, this section briefly summarizes answer interface, which is presented with more details in. In general, searchers prefer to simultaneously see various topics (e.g., disease names) that are potentially relevant to their medical situations.

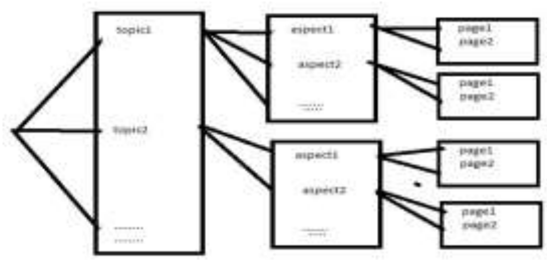

Fig 3.5 the Hierarchical Structure of the Answer Interface

For each such topic, searchers prefer to simultaneously see all kinds of aspects (e.g., symptom, diagnosis, and treatment) of it. As mentioned in Section 3.4, it uses diagnostic decision trees to find those topics that are potentially relevant to the searcher's medical situation. After obtaining the search results on those topics, it structures these search results into a three-level hierarchy that has explicitly marked medical meanings to fulfil the above requirements. This is shown in Fig. 3.5. At the first level of the hierarchy, all the search results are organized into multiple categories according to their topics (e.g., disease names). For each such topic *T*, an overview Web page *o PT* is provided to help the searcher determine whether this topic is related to his medical situation. At the second level, within each category, the corresponding search results are further divided into multiple sub-categories according to their aspects (e.g., symptom, diagnosis, and treatment).

3.1.3 Searching Steps

Online medical search is a vertical WSE that crawls Web page from a few selected, high-qualities medical Web sites rather than all the Web sites. Let *C* denote the collection of all the Web pages crawled by online medical search. As standard pre-processing steps in Web information retrieval, for the Web pages in *C*, (1)all the HTML comments, JavaScript code, tags, and non-alphabetic characters are removed, (2) stop words are dropped using the standard SMART stop word list, (3) noisy information is deleted using the frequent term sequence method described below, and (4) a forward index *If* and an inverted index *Ii* are built using the single-term vocabulary (i.e., the set of all the distinct words). In addition, another forward index *I ¢ f* that contains only medical phrases is built for the Web pages in *C*. it uses *I ¢ f* to suggest related medical phrases. In a given Web site, the useful information in the Web pages is often accompanied by a lot of noisy information, e.g., navigation panels, copyright notices, and advertisements. Removing this noisy information can greatly improve both the quality of search results and the search speed. We notice that a piece of noisy information usually appears in many Web pages and use the following frequent term sequence method to drop noisy information. A frequent term sequence is defined as a continuous sequence of terms that appears in many Web pages. For each Web page *P* in the Web site, we identify all the frequent term sequences and remove them from *P*. After obtaining the searcher's answers to the questions.

It proceeds in the following steps:

Step 1: Find the potentially relevant topics.

Step 2: Construct the search result hierarchy.

Step 3: Suggest related medical phrases.

- **Finding Topics**

In the questionnaire-based query interface of online medical search, searchers do not input queries. Instead, forms queries automatically based on searchers' inputs. More specifically, it uses medical knowledge to transform the searcher's question answers into several potentially relevant topics (diseases). For each such topic, it forms multiple queries to construct the corresponding part of the search result hierarchy.

We first show how to find the potentially relevant topics.

The searcher chooses one or more symptoms and signs in the questionnaire, and it selects their diagnostic decision trees. For each such symptom or sign, it traverses to one or more branches in the corresponding diagnostic decision tree *Td* based on the searcher's answers to the questions. Each leaf node of *Td* contains several disease names. The disease names in the leaf nodes of all these branches form a first set *S1* of medical phrases, and the disease names in all the other leaf nodes of *Td* form a second set *S2* of medical phrases. A medical phrase *M* can appear in both *S1* and *S2* if the corresponding disease name appears in multiple leaf nodes of *Td*. In this case, *M* is dropped from *S2*. Consider a selected diagnostic decision tree *Td*. In general, all the branching conditions (e.g., symptoms, disease histories) in *Td* have false positives and false negatives in diagnosing diseases. Moreover, searchers without much medical background can answer questions incorrectly due to unawareness of the exact medical definitions of these branching conditions. According to the medical diagnosis principles the medical phrases in both sets *S1* and *S2* can be relevant to the searcher's situation. The medical phrases in *S1* are generally more relevant than the medical phrases in *S2*. Also, diseases not in *Td* are usually irrelevant to the searcher's situation. Now consider all the selected diagnostic decision trees. A patient can have multiple symptoms and signs concurrently due to the presence of one or more diseases. To avoid omitting possible diseases, it needs to consider the set *E* of medical phrases from all these trees. As a general differential diagnosis principle, a medical phrase is more relevant to the searcher's situation if it is related to multiple symptoms and signs chosen by the searcher and appears in their diagnostic decision trees simultaneously. To consider this factor, for each medical phrase *MÎE*, it computes *M*'s global weight as the sum of *M*'s local weights in all the selected trees. This global weight reflects *M*'s relevance to the searcher's situation. All the medical phrases in *E* are sorted in descending order of their global weights. In this way, the searcher's question answers are transformed into appropriately sorted medical phrases, and the searcher can find multiple relevant diseases (possibly for different symptoms and signs) simultaneously.

- **Constructing the Result Hierarchy**

In this section, we discuss how to construct the search result hierarchy of answer interface. One might consider using classification to do this. For example, all the Web pages retrieved for a topic can be classified according to their aspects. However, online classification of search results is time-consuming and generally unsuitable for an interactive medical WSE. Also, it is difficult to know how many Web pages need to be retrieved for a topic *T* in order to obtain a sufficient number of search result Web pages for each aspect of *T*. Actually, even if we use a query formed for *T* to retrieve a large number of Web pages, it is still possible that no Web page among them mentions certain aspects of *T*.

To address the above problem, we use a novel automatic query formation method to construct the search result hierarchy. Our main observation is that the medical domain is a closed one. In the desired search result hierarchy, we can know the keywords for all the topics and their corresponding aspects. As a result, for each part of the search result hierarchy, we can use a different, specifically formed query to obtain the corresponding search result Web pages. More specifically, for each topic *T*, the overview Web page is retrieved using a query specifically formed for *T*. Also, for each aspect *A* of *T*, the corresponding search result Web pages are obtained using a query specifically formed for *A* of *T*. When forming these queries automatically, we use medical knowledge and consider the different roles that various levels play in the search result hierarchy. This can expedite the speed that searchers find their desired information. The resulting search result hierarchy fulfils all the requirements mentioned. In our query formation method, we could form the complete set of queries for all found topics and all their aspects, use these queries to retrieve all the search results, and construct the entire search result hierarchy in a single batch. Nevertheless, this approach puts unnecessary burden on online medical search and is undesirable. Searchers often skip completely many topics and aspects that they think are irrelevant to their medical situations at their first glance. Hence, there is no need to generate the search results for those topics and aspects. Moreover, searchers prefer to see outputs as soon as possible instead of waiting until the entire search result hierarchy has been constructed. To reduce the load on online medical search and to maximize the speed that searchers can see outputs, it constructs the search result hierarchy one part at a time. Each part of the hierarchy is generated only at the time that it is needed. If a part is never needed, it is never generated. More specifically, at the beginning to constructs only the first level of the search result hierarchy. If the searcher clicks a button and asks for more information about topic *T*, then it constructs for *T* the corresponding part of the second level of the search result hierarchy. Similarly, if the searcher clicks a button and asks for more information about aspect *A* of *T* at the second level, then it constructs for *A* of *T* the corresponding part of the third level of the search result hierarchy. When constructing the search result hierarchy, we frequently encounter the case that multiple formed queries share a few common terms. In this case, we share the inverted list union computation task that is common to processing these queries. Consequently, processing these queries together is much faster than processing these queries separately.

- **Suggesting Medical Phrases**

In this section, we describe how to suggest related medical phrases. We focus on the first level of the answer interface. The other levels can be handled in a similar way. In general, good medical WSEs should automatically suggest diversified, related medical phrases to help searchers quickly digest search results and refine their inputs. These suggested medical phrases should be ordered by their relevance to the searcher's inputs. It extracts and ranks medical phrases based on multiple sources: the Mesh ontology, the collection *C* of crawled Web pages, and the formed queries.

IV. RESULTS AND DISCUSSION

In this chapter we discussed the designing and implementation of intelligent medical search engine. The results were discussed according to their function.

Fig 4.1 Registration Screen

In this Fig 4.1 shows Registration Page, When the User enter into our website, user can start the registration process then only user can enter into our website. After successful completion of registration process, user must save their username and password for future use. By using those username and password only he can login into our website.

Fig 4.2 Login screen

In this Fig 4.2 shows login Page, here the user will enter user name and password, user name and password will be checked. After the validation if user name and password is entered correctly user will be allowed to gain access to system, if user will not enter the correct user name and password then user will not be allowed to gain access to system, an error message will appear, suggesting that to enter correct username and password.

Fig 4.3 Key Word and Questionnaire Interface Page

In this Fig4.3 shows When the user enter into this page, user have two types of options to search the medical information's first one is keyword interface option to search medical information and second one is questionnaire based interface. User can select any one of the interface. In this page user select the keyword interface to search medical information. In this interface user literally known about their situation and their problem.

Fig 4.4 Questionnaire Interface Page

In this Fig 4.4 shows user select the Questionnaire based query interface options. When the user select that option, it will guide the searchers(user) to provide the most important information about their situations according to the selected option and next question will be generated in an appropriate manner.

In this below Fig 4.5 shows questionnaire based query interface a question page will be open. User should answer that question; according to the answer next question will be generated. At last user find appropriate answer to that question.

Fig 4.5 Questionnaire Based Query Interface Page

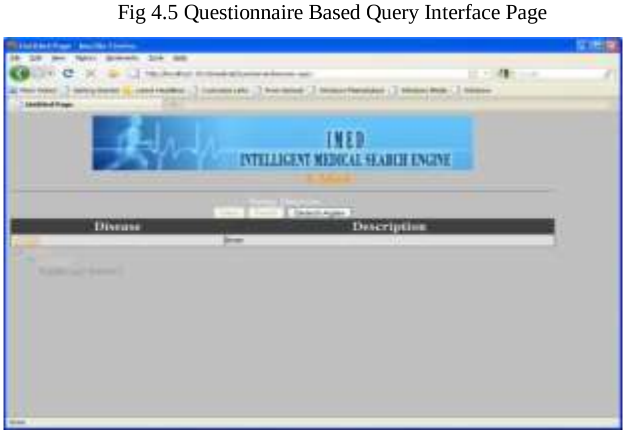

Fig 4.6 Answer Interface Page

Fig 4.6 shows that online medical search engine uses medical knowledge to automatically form multiple queries from a searcher' answers to the questions. By using these queries to perform search can significantly improve the quality of search results.

Fig 4.7 Administration Interface Page

Fig 4.7 shows Administrator Interface page, into the website by entering his username and password. Second level Administrator can upload the images and information regarding the problem. Regularly check the suggestion send by the user. Any extra information are required users then only he can add the information or provide similar information to the user.

V. Conclusion and Future Scope

Online medical search engine that extensively uses medical knowledge and questionnaire to facilitate ordinary Internet users to search for medical information. The design of online medical search takes into consideration the unique requirements of medical search. Instead of asking searchers to form queries themselves, it uses a questionnaire-based query interface to guide searchers to provide the most important information about their situations. It requires no special user training, forms queries automatically, structures all the search results into a multi-level hierarchy that has explicitly marked medical meanings, and suggests related medical phrases. These features are attractive to ordinary Internet users who have little medical background. Our experiments with a wide range of medical scenarios demonstrate that greatly improves user satisfaction by performing medical search effectively and efficiently. For the future work in the same way we are interested in using our techniques to build intelligent search engines for other domains.

The main conclusions of the study may be presented in a short Conclusion Section. In this section, the author(s) should also briefly discuss the limitations of the research and Future Scope for improvement.

REFERENCES

[1] A.L. Komaroff, Harvard Medical School Family Health Guide, Free Press, 2004.

[2] C. Sherman. (2005) Curing medical information disorder. [Online]. Available:http://searchenginewatch.com/showPage.html?page=3556491

[3] D.L. Kasper, E.Braunwald, and A. Faucietal, Harrison's Principles of Internal Medicine, 16th ed., McGraw-Hill Professional, 2004.

[4] John Wiley & Sons, American Medical Association Family Medical Guide, 4th ed., 2004.

[5] J.G. Carbon ell and J. Goldstein, "The use of MMR, diversity-based re-ranking for Reordering documents and producing summaries," in *Proc. SIGIR'98*, 1998, pp. 335-336.

[6] P.J. Bickel and K.A. Dorsum, Mathematical Statistics: Basic Ideas and Selected Topics, Vol. 1, Prentice Hall, 2001.

[7] P.M. Healey and E.J. Jacobson, *Common Medical Diagnoses: An Algorithmic Approach*, 2nd ed., W.B. Saunders, 1994.

[8] R.A.Baez a-Yates and B.A.Ribeiro-Neto.*Modern InformationRetrieval*,ACM Press/Addison-Wesley, 1999

[9] R.D. Collins, Algorithmic Diagnosis of Symptoms and Signs: Cost-Effective Approach Lippincott Williams & Wilkins, 2002.

[10] S.E. Robertson, S. Walker, and M. Hancock-Beaulieu, "Okapi at TREC-7: automatic ad hoc Filtering, VLC and interactive," in Proc.TREC'98, 1998, pp. 199-210.

[11] A survey on Internet of things for health care and medication management. IJSRCSE, Volume 2, Issue 1, January – February 2017, Pages 1 – 5.

Authors Profile

SRIKANTH BETHU has completed M.Tech in Computer Science and Engineering in 2011 at Osmania University, Hyderabad, India. He was currently working as Assistant Professor in Department of Computer Science and Engineering at Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad. He was very fond of Data Mining concepts. He has published several research papers in the areas of Data Mining, Data Engineering and Big Data. Attended and Presented several research papers in various National and International conferences.

Dr. B. Sankarababu was a doctorate from Aacharya Nagarjuna University, Guntur, Andhrapradesh, India, and completed PhD in 2016. Currently working as a Professor in Department of Computer Science and Engineering at Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad .He has done his research on Data Mining. He has published several research papers on Data mining concepts.

