ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 4, April 2013*

# A Variant Architecture Design for Intelligent Medical Search Engine (iMed)

Annu Anna Lal[1], Anna Alphy[2]

M.Tech Scholar, Dept of CSE, Rajagiri School of Engineering and Technology, Kochi, India[1]

Asst. Professor, Dept of CSE, Rajagiri School of Engineering and Technology, Kochi, India[2]

**Abstract**: Nowadays both ordinary internet users and doctors are increasingly using the web search engines for searching medical information. iMed is one of the intelligent medical search engines that use medical knowledge and questionnaire for searching. iMed introduces and extends expert system technology in to the search engine domain. This search engine uses many key techniques to improve its usability and search result quality. iMed provides a questionnaire based query interface to users. But it has some issues within its design. This paper is focused on certain issues raised during the detailed study on the search engine. In this an enhanced system design is also proposed. The recommended enhancements improve the working and efficiency of the medical search engine.

**Keywords**: iMed, content based image retrieval, Pattern Matching, medical knowledge

## I. Introduction

**S**earching medical information in the internet is relevant and common now. Medical search engines are developed for this purpose. Idea behind this is information retrieval which deals with the representation, organization, storage and access to information items. Organization and representation of the information items should provide the user with easy access to the information which he is interested. iMed [4] is one such medical search engine which is discussed in this paper.

iMed have a user interface and an answer interface that helps the user for better information retrieval. The other search engines follows the keyword matching technique and the user have to define the queries. The design of iMed considers the unique requirements of medical search. Instead of insisting searchers to form queries by themselves, iMed have a set of questionnaire. These features are attractive to ordinary users .so they greatly improves the user satisfaction by performing medical search efficiently and effectively. But the existing iMed have certain problems associated with it .The main issues are the generation of weak queries for searching purpose, lack of perfect user input and description and finally improper guidance due to lack of knowledge.

The rest of the paper is organized as follows. In sectionII of this paper a discussion about related works included. The other existing systems are discussed here. Then we move on to section III that gives an overview of the iMed medical search engine. This gives an idea about its working and design. Next comes section IV that defines the issues that are discovered during study of iMed and then section V includes an enhanced design for the existing search engine. The proposed system design reduces the user's difficulty and also provides better result. Then finally section VI concludes the paper.

## II. Related Works

In response to the huge needs in medical information many medical WSE's are developed since October 2005. All these search engines provides medical information for the users based on certain keywords. It mainly includes Healthline[10],Google Health [7], SearchMedica [11] and Medstory [6]. The health line includes symptoms search, body maps and also discusses top health topics. Google developed a web search engine named Google Health and it has certain sophisticated security techniques to keep the registered users data. It also provides certain forums and discussion pages. SearchMedica is a series of free medical search engines built by doctors for medical professionals, with localized versions for the United Kingdom, the United States, Spain and france.There are currently 4 localized versions for searchmedica. Microsoft developed the Medstory, a vertical search engine. Microsoft has bought Medstory because it's an intelligent and intuitive search

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 4, April 2013*

technology. In this the search software applies some artificial intelligence techniques to the medical and health information in government documents, medical journals and also on the Internet. It was released for beta testing on October 8, 2007 as a Live Search Health and is served as the front-end to Microsoft Health Vault Search. But all of this have poor user interface compared to iMed and user have to create their own query that plays a big role in the searching. In iMed they provide a questionnaire based interface that provides the users better choices and makes queries automatically according to this. This helps in efficient result generation.

### III. OVERVIEW OF IMED

This section gives an outline of iMed, its interfaces, working and the design details. They have a query interface, symptoms and Signe's page and the question pages.iMed hold its built-in medical knowledge in the form of diagnostic decision trees written by medical professionals (Collins 2002).

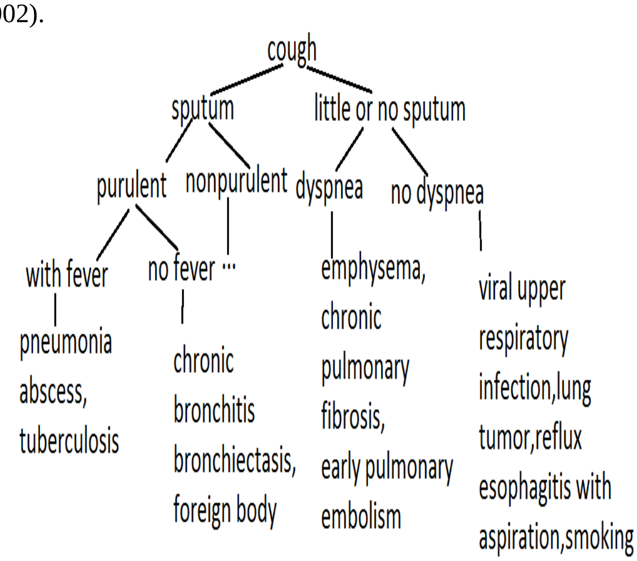


**Fig 1**: The diagnostic decision tree for the symptom "cough"

As shown in Figure 1, each diagnostic decision tree [4] corresponds to either a subjective symptom (e.g., fatigue) or an objective sign (e.g., hypertension). Each non-leaf, non-root node of a diagnostic decision tree corresponds to an answer to a question that iMed asks to the user. Each medical phrase in the leaf node of a diagnostic decision tree can become a query that iMed uses.

iMed uses diagnostic decision trees to help the searcher to form queries. At first the searcher selects one or more symptoms and signs from a list of known symptoms and signs. Then iMed asks questions related to these selected symptoms and signs. iMed navigates the corresponding diagnostic decision trees and automatically forms multiple queries with different weights based on the searcher's answers to the questions. All these formed queries are sorted in a descending order of their weights. Each query is used to retrieve some related Web pages. iMed combines the search results for all these formed queries together and returns them to the searcher in multiple passes. All those Web pages are sorted in the same order as their corresponding queries. In general, for a given symptom or sign, multiple diagnostic decision trees exist in different medical textbooks. Some of these decision trees have small depths whereas the others have large depths. Ideally these decision trees should be combined into a single one with a small depth. This combined diagnostic decision tree is stored in the knowledge base. In this way, a user only needs to answer a few questions, rather than many questions, before he can obtain search results. The backup options can be easily provided when a user does not know the answer to a question on the page. After obtaining the searcher's answers to the questions, iMed then proceeds in the following steps:

**Step 1**: Finds the potentially relevant topics.
**Step 2**: Construct the search result hierarchy.
**Step 3**: Suggest the -related medical phrases.

In the output interface [2] the entire search results are structured into a multi-level hierarchy that has explicitly marked medical meanings. More specifically, all the search results are organized into multiple categories according to their topics (e.g., disease names). Within each category, the corresponding search results are further divided into multiple sub-categories according to their aspects (e.g., symptom, diagnosis, treatment). This hierarchical search result output interface is intelligent in the sense that it automatically offers searchers what they want instead of waiting until they ask explicitly. iMed is a vertical WSE that crawls Web pages from a few selected, high-quality medical Web sites rather than all the Web sites. Here the Fig. 2 shows the architecture of an iMed [4].It includes four components forming the entire system.

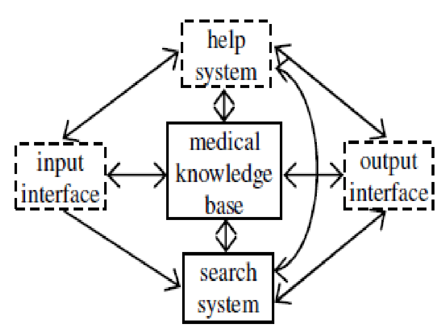


**Fig 2**. Architecture of an intelligent medical search engine

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 4, April 2013*

The components include the input interface, the help system, the medical knowledge base, the search system and the output interface. Interactions occur between almost each pair of components whereas most interactions are bidirectional. Users can see the input interface, the output interface, and the help system, whereas they cannot see the medical knowledge base and the search system.

The input interface provides users the symptoms and signs page and also allows users to input long queries, describing their symptoms and situations in detail in plain English.It also provides a guided interface to help the user input important information about his medical situation. Then iMed asks questions related to these selected symptoms and signs. In the case that all the important information about the user's medical situation is covered by the questionnaire, the user provides information by selecting symptoms, signs, and question answers without performing manual typing. The help system of the IMed provides different kinds of suggestions to facilitate the search process such as (1) suggesting diversified and related medical phrases that helps the user to quickly digest search results and refine his inputs, (2) suggesting the symptoms and signs that are related to the user's medical situation, and (3) suggesting the alternative answers to the questions asked by the WSE in case that the user answers questions incorrectly. The help system also provides explanations of symptoms, signs, asked questions, and suggested medical phrases in layman terms. An important feature of iMed is that its search system can automatically form queries. These queries are obtained using both the information the user provides in the questionnaire-based input interface and the medical knowledge stored in the knowledge base. The extensive use of medical knowledge is the key for making a medical WSE intelligent and also to quickly obtain high quality medical search results. An IMed stores in its medical knowledge base all its medical knowledge, which is used to support its input interface, output interface, help system, and search system. The search results are displayed in the output interface of system.In order to satisfy the various needs of different users simultaneously, the output interface needs to show the search results of all the potentially significant topics and their most common aspects. All the search results are structured into a multilevel hierarchy that has explicitly marked medical meanings.

## IV. Issues of Existing System

During the study of iMed we discovered certain limitations and drawbacks. iMed is a medical search engine whose efficiency is based on the input query that is developed automatically. The strength of query depends on the user's answers to the questions which in turn produce the desired results. The main issues are related with symptoms and signs selection. The other issue is the improper answers to the questionnaire. This leads to the creation of weak queries. Lack of proper medical knowledge is also a reason for this issue.

An ordinary user may not have good medical knowledge. So during the selection of symptoms and for the questionnaires they may sometimes give wrong answers. iMed develops the query based on this inputs and the above scenario generates weak query which in turn produces irrelevant output. This is one of the serious issues that degrade the efficiency of the search engine. The weak queries lead to the wrong information diagnosis in the decision tree and the user may get a wrong output. This generally reduces the overall performance and reliability of the system. Lots of garbage information's are produced in every turn.

## V. Proposed Solution

The above mentioned issues of iMed can be reduced by applying certain improvements on the existing system architecture. We note that addressing the challenging issues in intelligent medical search requires interdisciplinary knowledge of pattern recognition, expert system, and Web search. The main idea is to increase strength of the query developed by the search engine that in turn improves the result of the system. The strength of query can be improved by giving strong inputs to the system. The inputs strength can be increased by adding more solid information's or data to the system regarding the search. The following are the recommended improvements

(i) Introducing a provision for inputting medical images and test results.
(ii) Using better string pattern matching algorithm
(iii) Image retrieval technique.

Most of the issues are raised due to the lack of good input which is the reduced strength of user input to select the correct symbols and to describe their situation. Poor knowledge and description lead to wrong selection thus results in the wrong conclusion. So for an ordinary user the probability of getting correct results reduces. The above suggestions improves the strength of user inputs thus its helps WSA to guide the user in an efficient manner to reach the most desirable solution. It also suggests the symptoms and answers to questions during the searching process. User inputs are structured using multilevel hierarchy. They forms query automatically.

### A. *Proposed System Architecture*

The Figure 3 given below shows enhanced system

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol. 2, Issue 4, April 2013*

architecture for iMed. It consists of an Imaging system and Helping system in additional to the existing architecture. All these are connected to the output interface and to the input interface. Each of them briefly described in the following sections.

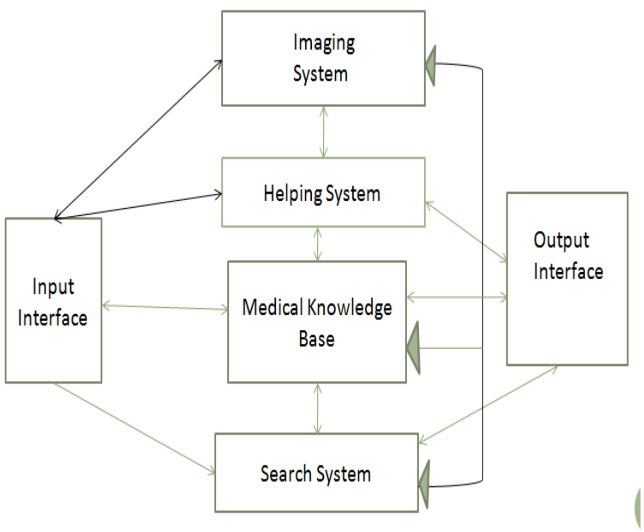


**Fig 3**. Enhanced Architecture of an intelligent medical search engine.

The imaging system uses the content based image retrieval system [5]. This system is used for searching images and to get the corresponding details based on match. The users have the option to give an image as input. Then searching of the corresponding details based on the given input image is done by the iMed. After that the particular image is matched against the one stored in database of search engine. When a match is found, the corresponding information stored with that image is passed to the user interface. Thus helps the user to get more details about the term.

A content based image retrieval system [5] works more efficiently than the other techniques. In this the search will analyse actual content of the image rather than the meta data with the image. The content of image refers to the colour, shape, textures or any other information derived from image. Here image distance measure is used to compare similarity between two images. It is actually the measure of similarity between images. Then the search results are stored based on the distance to the queried image. If the distance is zero it indicates that the exact match of images. Querying by image regions is also possible that is an outline, shape or part of an image can be used for matching or searching. Query technique removes the difficulties that can arise when trying to describe images with words. Based on the above results the following are possible now.

(i)An easy way to suggest chief symptoms and signs to the user
(ii)Helps in answering the questions during the searching
(iii)To get better knowledge about the current situation

This can be done without much training .User just needs to give an input image to the imaging system. He can use this image to get guidance from system or to search similar images. Each image is stored in database along with its details of diseases and the corresponding webpage links. Then matching process is performed and the details along with links are provided to the user .This can be used as a reference for users during search and inputting.

### *B. Pattern Matching*

The time performance of an exact string pattern matching can be greatly improved by using an efficient algorithm. The String Pattern Matching algorithm should be based on the distinctive features of medical language. To choose the most appropriate algorithm for pattern matching, the distinctive features of the medical language must be taken into account. The characteristics of the medical language are emphasized in this manner, the best algorithm of those reviewed is proposed, and the detailed evaluations of time complexity for processing medical texts are provided.

The time performance of exact string pattern matching can be greatly improved if an efficient algorithm is used. By considering the growing amount of text handled in the electronic patient records, it's worth implementing this efficient algorithm. The **Boyer-Moore-Horspool algorithm** [5] achieves the best overall results when used with medical texts. This algorithm basically performs at least twice as fast as the other algorithms tested. So it is better the ordinary pattern matching algorithm used by the system.

## VI. CONCLUSION

Various approaches to improve the performance of IMed are discussed in this paper. One of the drawbacks observed is that, they lack perfect user input and the generation of weak queries. TheBoyer-Moore-Horspool algorithm is recommended to improve the pattern matching technique. These features are very attractive to majority of Internet users who have little medical background. The experiments with a wide range of medical scenarios demonstrate that iMed greatly improves user satisfaction by performing medical search effectively and efficiently.

## REFERENCES

[1] Gang Luo and Yu P.S (2010) "Lessons Learned from Building the iMed Intelligent Medical Search Engine" in IEEE Transactions.
[2] G. Luo (2008), "Intelligent output interface for intelligent medical search engine" in Proc. AAAI'08, 2008, pp. 1201-1206
[3] C. Tang and G. Luo (2008), 'On iterative intelligent medical search," in Proc. SIGIR'08, 2008, pp. 3-10
[4] Gang Luo M.S (2009), "Design and Evaluation of the iMed Intelligent Medical Search Engine", in Proc.ICDE'09,pp 1379-1390
[5] Wikipedia- pattern matching algorithms, image retrieval techniques , Boyer-Moore-Horspool algorithm
[6] Medstory Homepage http://www.medstory.com, 2007
[7] GoogleHealthhompage (2008) http://www.google.com/Top/Health
[8] R. A Baeza – Yates and B.A. Ribeiro-Neto Modern Information Retrieval, ACM 1999
[9] G. Luo and C. Tang, "Challenging issues in iterative intelligent medical search," in *Proc. ICPR'08*, 2008, pp. 1-4.
[10] Health line homepage http://www.healthline.com, 2007.
[11] Searchmedica http://www.searchmedica.com, Wikipedia

## BIOGRAPHY

**Annu Anna Lal** is currently pursuing her Master of Engineering in Computer Science and Engineering with Specialization in Information Systems from Rajagiri School of Engineering and Technology, Kochi (India) under M.G University.