Review Article

On the Performance of Medical Information Retrieval using MeSH Terms – A Survey

Swetha S*, Uma D, Suganya P, Nivedhitha V and Saravanakumar K

School of Information and Technology and Engineering, VIT University, Vellore, Tamilnadu, India

---

Abstract

Internet users have increased everywhere. Searching and retrieving documents is a common thing nowadays. Retrieving related documents from the search engines are difficult task. To retrieve correct documents, knowledge about the search topic is essential. Even though separate search engines are there to retrieve medical documents the users are not familiar with MeSH terms (Medical Subject Heading). So, both the search browser and the MeSH terms have to be integrated to make the search effective and efficient. To implement this integration, SimpleMed and MeSHMed were introduced. The MeSH terms have to be ranked to know how frequently it has been used and to know the importance of the MeSH terms. To rank it a semi – automated tool called MeSHy was developed. The terms were extracted, filtered, ranked and displayed to the user. Classifiers have to be constructed to label the documents as health and non – health. Three strategies were used to classify them. The errors that are commonly done by the users have to be found out. It was calculated based on the queries presented by the user to the search browser.

Keywords: Medical Information, MeSH, SimpleMed

---

1. Introduction

Information retrieval is the process of obtaining relevant information from collected information resources. The general task of information retrieval (IR) is searching for information in documents. Everyone has started to search information on-line which consumes less time and effort. Medical related information retrieval has been increasingly used. Medical information retrieval is the process of retrieving information based on the health issues queried by the user. According to the investigation done by the organization named Jupiter, 71% of people use the Internet to seek health information in 2007, and this percentage, which represents an estimated 160 million people in the United States, had increased by 37% since 2005[1]. In a survey conducted by Pew Internet Project in the year 2009, 83% of Internet users have searched for medical or health information. According to the center for studies on Information Technology and communication [2], it has been calculated that in the year 2009, around 33% of Internet users' activities in Brazil were associated to looking for medical information. The percentage of internet users rises to 55% in United States [3]. The search is mostly connected with disease, information about hospital, doctors and nourishment [4]. Queries are especially about particular diseases or medical problems to experimental treatments and medicine. An up to date survey reports says that people between the ages of 18-34 go online to find healthcare professionals while people above the age of 50 seek for

medical information online after consulting their general practitioner. Women usually searches for health related information for someone related to them while men search for medical information for friends. PubMed is a free database accessing primarily the MEDLINE database of references and abstracts on life sciences and medical science issues. NLM (National Library Medicine) at the National Institutes of Health maintains the database as part of the Entrez system of information retrieval. The users of PubMed are both medical and non-medical professionals. In case of medical users it is quite easy for them to perform search since they have some familiarity with the medical terms. The non- medical professionals are not aware of the medical terms and it is difficult to access and they are unaware of the accuracy of the result. To perform better information retrieval, innovative ideas have been proposed [5]. It is necessary to make the search easier and effective to both medical professionals and non- medical professionals. Medical thesauri such as Medical Subject Heading (MeSH) and related tools are there to help the consumers and they are integrated and synchronized [6]. To make the retrieval fast, accurate, reliable and user-friendly automated or semi-automated text-mining tools have been developed [7]. Interfaces were created to make the search effectively. MeSHMed was implemented based on search browser, MeSH tree browser and MeSH term browser. Text mining tool was developed [8]. Though many tools were developed to retrieve relevant information on the web, most of them are not efficient. Some on-line solutions are available to medical information retrieval such as MediAgent developed by NUS (National University of Singapore) [9]. MeSH term search and normal text word search were done in Medline to know the performance of two search strategies [10]. The two

---

• E-mail address: swethakeerthi123@gmail.com

search strategies complemented each other and should be used together for maximal retrieval. The users need more support in formulating the queries [11]. Studies were conducted to evaluate the effectiveness of a metadata based query suggestion interface for PubMed bibliographic search [12]. PubMed/MEDLINE offers an alternative to free-text searching on the internet, which allows the users to search using Medical Subject Headings [13].

2. Information retrieval VS Medical information retrieval.

Content skill levels of medical texts are essential for the comprehension of medical information. Users belong to many professions and so they need special skills to retrieve medical information [14]. Above all, the main difficulty is the formulation of query. Queries are formulated by the users in their natural languages. The approach used for query formulation is focused on the type of techniques used in the intelligence analyst process. Two types of query comparisons were done. One is Boolean query formulations against their "equivalent" VSM queries; and the other is Alternative Boolean query formulations against each other [15]. To achieve satisfaction of the users, a concept of a vector-model oriented IR-Tool based on the controlled vocabulary of the MeSH Thesaurus was used [16].

People have started to surf the information they want through the search engines. The main thing is that they have started to search for medical information also. It is important to take the search of medical information retrieval very seriously. If a user queries about his/her health condition in the normal search engine it may retrieve a set of documents. Very often the retrieved documents may be unrelated or irrelevant to the submitted queries. If the user is a medical professional he/she may know the medical terminologies and so their search would be effective when compared to others, but they too get struck with the retrieval. So, a special search engine called PubMed was introduced which is used only for retrieving medical information.

Each day, many people search for medical information on the net than visiting medical professionals. Medical professionals themselves are progressively using Web search engines to facilitate analysis because of the complexity in keeping up with the fast growth of medical knowledge. Many Internet users have searched for medical information on the Web. Most of these users thought they achieved useful information online, and were more ready to use Web search engines rather than going to a particular health-related Web site. Since medical information retrieval is very sensitive, it should be handled with care. The information retrieved should be very accurate as it has to be prescribed to the patients. The retrieval should also be fast, reliable and user friendly. It should try to give the related document in the first itself but it fails to give often. Without a proper plan medical information retrieval cannot be done efficiently.
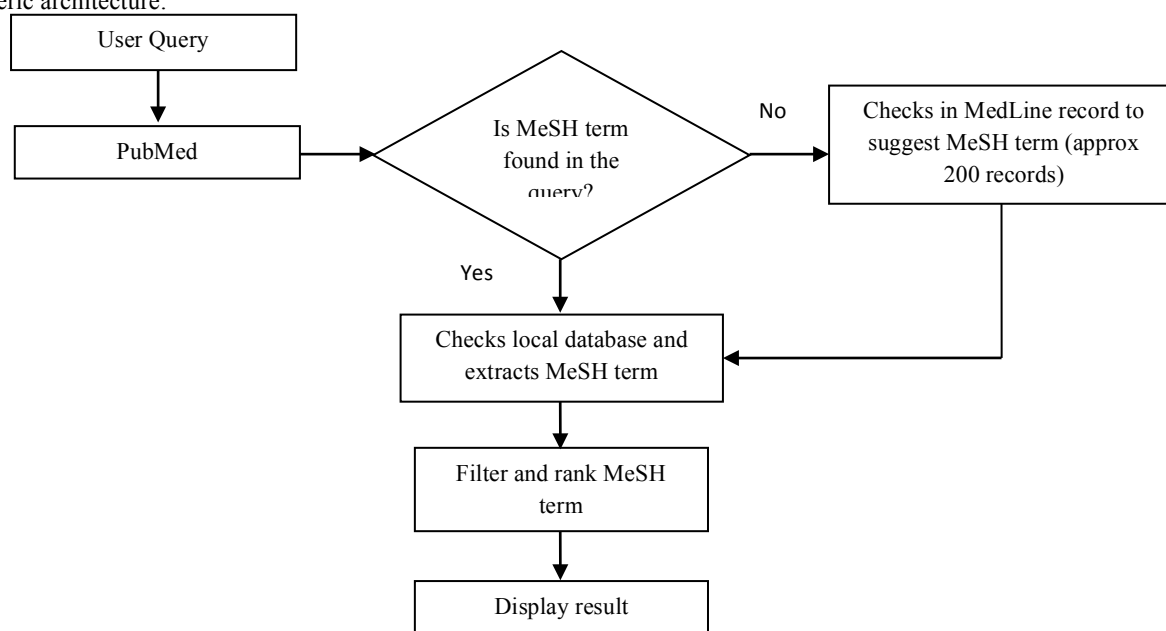
Generic architecture:

Fig. 1. Generic Architecture

An important and special term used for medical information retrieval is the MeSH term. MeSH means Medical Subject Heading. Search engines finds the documents only with the help of MeSH and the retrieved documents are related to the MeSH term only. If the appropriate term has been missed, then the related documents will not be retrieved even if it is available. The queries have to be formulated with more concern. Use of unrelated words may lead the search to become slower. If the retrieving person is not aware of the medical terms

he/she may think that the retrieved documents are correct. So, accuracy is the main thing in the retrieval of medical information.

3. MeSHMed

3.1 Objective.
When consumers search for health information the problem is that they are not familiar with medical terminology, medical thesauri such as medical subject heading. A tool

named "automatic term mapping" was created to overwhelm the unawareness of the medical term [17]. The drawback with the tool is that the users are not provided with the definitions of the terms. So tools are needed to integrate them.

3.2 Method.

Two interfaces were introduced namely SimpleMed and MeSHMed. SimpleMed was constructed with the help of a simple browser. MeSHMed was implemented using SimpleMed, MeSH tree browser and MeSH term browser. MeSH tree Browser is viewed in sequences of tree structure or hierarchical tree in terms of MeSH. After a query is given, related MeSH terms are displayed. If those MeSH terms are clicked it shows details and descriptions about the terms. When the query is entered in search browser, it will automatically inform the term browser to list correlated MeSH terms. The tree browser and the term browser are maintained with the MeSH expressions downloaded from National Library of Medicine (NLM).

3.3 Evaluation

45 participants were recruited to check the interfaces and they were undergraduate students and some graduated people. The data set used for it is OSHUMED [18]. OSHUMED is a medical oriented division that consists of 348,566 records having 270 medical journals. From the data set 106 topics were taken for the study. Out of 106 randomly six topics were selected. The selected search topics were based on the original OSHUMED queries. Users are prohibited to use the search system while answering the questions. Each answer was scored as wrong, partially correct, and correct. The scoring rules were developed by a board of four people: two usability study specialists and two domain specialists. The search was effective in MeSHMed when compared to SimpleMed.

3.4 Findings.

One of the main advantages of MeSHMed is the coordinated communication between the different components. After a query is presented in the Search Browser or the Term Browser, the Tree Browser will automatically zoom into related MeSH terms and tree structure is formed and the associated terms are visually displayed to the users. So it is more efficient from the view of users.

4 MeSHy

4.1 Objective

MeSH term are not presented readily to the users [19]. The extraction of relevant documents becomes a challenging one without the prior knowledge about MeSH terms. Many users are not satisfied with the documents retrieved and their queries go unanswered since the queries are not formulated properly [20]. So, it is an important thing to develop automated or semi-automated text-mining tools that are efficient and effective in extracting relevant document.

4.2 Method

A semi-automated text mining tool called MeSHy was developed. MeSHy is an algorithm that extracts MeSH terms from the documents and gives ranks to MeSH terms. MeSH term pairs are scored for every document. Using these scores it is easy to rank the retrieved documents. At last, the MeSH

term pairs are sorted based on the decreasing order of their total scores and offered to the user in an interactive table format. The relevant documents ID (PubMed document ID), year of publication and the type of the document are added to the result. Type includes case report, review etc.

4.3 Evaluation

The evaluation of MeSHy was performed by a set of skilled people. The set consists of one associate professor and three postgraduate students of Aristotle university of Thessaloniki. The result given by the evaluators was that MeSHy is proficient to process and present information from PubMed queries in an unexpected way. Customers' comment has also pointed out that MeSHy could make the writing of review papers easy, because it has the ability to point out popular and expected knowledge. Directly comparing MeSHy with other tools is difficult, as the internals of each method are different. The same queries were performed in FACTA [21] and Epiphanet [22] systems.

4.4 Findings

In MeSHy tool there is a chance of having MeSH term pairs without meaning since terms occurred in the document are only paired. MeSHy is highly dependent on queries. So if the query is poor, it may retrieve irrelevant documents and the performance of the tool may go down. The advantages of MeSHy are MeSHy uses pre-assigned MeSH terms and frequencies avoiding the usual process of stemming, tokenizing, etc. The user has to only input the query and the calculation done for frequency and all is not known to the user and so the users need not to be given training. There is some limitation also to use the algorithm. MeSH terms are assigned manually to documents by MedLine indexers. They may allocate unrelated terms or fail to spot related terms [23]. The use of MeSH terms could sometime direct to loss of information, because a complete document is represented as a restricted set of manually assigned terms.

5. InDECS

5.1 Objective

When users search for health related content in the internet the search tools available return web pages that are unrelated to it. The popular commercial search engine such as Google may also give correct and informative results [24, 25]. But it has some difficulties in retrieving relevant information [26]. Patients are gradually more accessing the internet before going to a health discussion [27]. So it is necessary to build up search engine that are medical specific which should be useful for non- medical public also [28]. To develop this kind of search engines it is essential to develop indexers for content. Variety of techniques can be used for it [29]. Investigation is done on classification technique [30].

5.2 Method

A classifier is constructed [31, 32] to classify whether the content is medical related or not. To investigate MeSH, three strategies were presented. Frequency of terms, occurrences of terms present in MeSH, InDECS were the three strategies. These three are used to generate vectors of characteristics. With the help of these vectors two databases were constructed namely training database and validation database. To construct training database, web pages from Portuguese version of the Merck manuals online medical library and web pages from the online version of Brazilian

newspapers were used. To construct validation database Brazilian web pages alone were gathered. For all the web pages the content was marked as health and non-health by four different evaluators who had training in health information. . A usual vector model for information retrieval is known as frequency of terms. In frequency of terms, the vector of characteristic was developed for each webpage in training and validation database. In Occurrence of terms analyses were made on each term on the webpage based on the term occurrence. InDECS was divided into two steps that are defined as constructing a table that tells the weight of MeSH term and with the help of the weight table a generation vector of characteristics is done for the page.

5.3 Evaluation
The vectors and their labels (health or non – health) were offered as illustration to a pattern classifier, with the plan of carrying out supervised training. A variety of pattern classifiers was tested for InDECS. The aim was not to examine the pattern classifier but to find which strategy is best and efficient for the approached problem. Two Naive Bayes models were used [33] the multivariate Bernoulli model and the multinomial model. Multivariate model was used to find whether the word occurred in the document or not. Multinomial model is used to count how frequently the word is used. In this pattern classifier the second model was used. After the training, a web content classifier that automatically finds whether a web page is medical related or not was obtained using vector of characteristics which is one of the strategy discussed. Dataset bootstrapping was created by taking a fixed number of health web pages and for each test randomly a non health web page was added. The web pages were presented to the validation database and sensitivity and specificity values were produced. Sensitivity is the ratio between number of pages labeled by the classifier as health and total number of validation database pages labeled as health. Specificity is the ratio of the same for non – health.

5.4 Findings
The sensitivity and specificity values calculated from validation database and the values are more for the InDECS strategy when compared to others. So, InDECS strategy is effective and efficient.

6. MeSH term error

6.1 Objective
It has become a challenging thing to retrieve medical information in an efficient way. PubMed offers an alternative way to search by using the keyword called MeSH. Though keyword based searching is done, it is difficult to retrieve information in a specified time. This study is based on the analysis of data that has been retrieved by universities of nursing department. The plan of the study was to illustrate an outline of the best, average and worst performers in the test.

6.2 Methods
A brief introduction about PubMed and about MeSH term was given to the participants. There were 100 participants with different linguistic and educational background. The query formulations were performed in several steps. Initially the have to find a MeSH term related to the search question. To find the MeSH term the participants searched it in

PubMed module by entering a term. PubMed suggested more MeSH term and the participants selected related MeSH terms and sent it to the search box where they need to perform the search. These steps were repeated till the participants formulate a satisfactory query. If the correct term is found, the query is submitted to the PubMed and documents are retrieved. From the retrieved documents they were asked to select the documents that are relevant to the query. They were given 15 minutes of time to complete it. After that English knowledge test were conducted to measure their English skills. Then the participants were asked to give their feedback about the search engine. All the queries were collected and the errors made in the query are identified. The document that has been retrieved but considered as irrelevant are also calculated. The performers were categorised as worst, average and good performers based on the documents that were considered as relevant by the participants.

6.3 Evaluation
Even though the participants come from different linguistic and learning environment, a Kruskal–Wallis test pointed that there are no important dissimilarity in recall between these groups, so we can securely join them and use another classification for the reason of best, average and worst performers. With consider to PubMed experience, our test group was mixed: 44% had a detailed introduction into the use of the search engine, whereas others had only had a brief introduction (46%) and some (10%) had no introduction into PubMed. 97% use a computer a number of times a week to daily, but only 18% check with PubMed with the same occurrence. 40% of participants infrequently or by no means use PubMed to search for medical information. There were eight errors in formulating the queries. Analyzes were made on eight different error types on search presentation, and found that three of those error categories had an important collision on real and possible recall: incorrect MeSH terms, under specification, and the incorrect use of Boolean operators.

6.4 Findings
Incorrect MeSH terms error was made in about 1 out of 3 queries (29%). A total of 73% of the queries contains an incorrect MeSH, either because of empty result sets (33%), or since the results were unrelated to the search question (40%). The left over 27%, the search did surrender some applicable results, in spite of the use of a MeSH term that was not totally related for the hunt. Queries that contain a wrong MeSH term yielded less than one (0.78) relevant illustration on average. The error of under specification, is when queries contains only one or two terms and are therefore too wide, was made in 125 queries (40%). Underspecified queries yielded 0.41 relevant records on average. In incorrect use of Boolean operators 27 queries (8%), one or more Boolean operators were used wrongly. This error is mainly in the excessive use of AND (67%) and OR (33%). This error led to zero potential recall in 81%, yielding result sets in 37% of the cases, and yielding only records irrelevant to the search question in 44%. Totally 60 queries (19%) were devised correctly, with an average possible recall of above 4 records. It means that the participants who presented these queries could have selected an average of four related records, whereas they selected less than two.

7. Overall Observation

The non medical professionals are unaware of the medical terms that are used by the medical professionals. So two interfaces was created known as SimpleMed and MeSHMed. SimpleMed consists of search engine. MeSHMed consist of search engine, MeSH tree browser and term browser. Comparison of SimpleMed and MeSHMed was done and we found out that MeSHMed was effective when compared to SimpleMed. SimpleMed is too simple when compared with MeSHMed. The main advantage of MeSHMed is the addition of tree browser and term browser. MeSHMed supports natural language queries and also support users to view for medical terms and definitions and produce a MeSH tree. The remarks given by the users also confirm the effectiveness of MeSHMed. The synchronization of the components (tree browser and term browser) is the main advantage of the interface. When a query is given in the search browser the MeSH terms related to the query is displayed. When the MeSH term is clicked the details about the term is shown. We came to a conclusion that MeSHMed is the best since it is an interactive method that is effective for the users when compared to SimpleMed.

Extraction of related documents is not an easy job for non – medical professionals. So a semi – automated tool was developed called MeSHy to extract documents effectively and efficiently. MeSHy extracts MeSH terms and rank them. The main thing we found is that ranks for the MeSH terms are manually assigned. Since they are manually assigned users cannot predict whether the given rank is correct or not and decision about the paper cannot be predicted. Many MeSH term pairs are not meaningful. The tool is based on the queries that are given by the user. So if the query is wrong, the MeSH term pairs go meaningless and the ranking part is useless for the queries.

Sometimes, the related documents are not retrieved properly. Classifiers have been constructed to separate the content as health and non – health. Three strategies were presented for it. They are frequency of terms, occurrences of terms present in the MeSH and InDECS. Vector of characteristics were developed with the help of these strategies. . Dataset bootstrapping technique was used to inspect the correctness of the classifiers when the number of information used for training database changes. The performance of these strategies was calculated in terms of sensitivity and specificity. Out of three strategies, InDECS was the best. It was proved by the specificity and sensitivity values calculated for the strategies. InDECS has the higher value (refer Table 1). These values were calculated based on the validation database and the training database.

Table 1. Sensitivity and specificity values of strategies

Strategies	Sensitivity	Specificity
Strategy 1	0.85	0.95
Strategy 2	0.89	0.93
Strategy 3	0.94	0.94

The performance of the queries was rated as best, average and worst based on the errors made by the user. The users were given a brief introduction about the PubMed and MedLine database. By the result produced by them it is clear that the common error made is the incorrect use of MeSH term and the Boolean operators. As the terms are incorrect the documents retrieved as not relevant to the search topic and the search is not successful.

8. Conclusion

The users have difficulty in retrieving relevant documents. Non – medical professionals are not aware of the medical terms used. If they are not familiar with the search topic they are unaware whether the retrieved documents are correct or not. To retrieve medical documents they have to know the MeSH terms to make their search effective and efficient. These terms are not given directly to the user while retrieving. So, to integrate the MeSH terms with the search engines interfaces have been introduced. The use of two interfaces SimpleMed and MeSHMed were seen. MeSHMed is the effective interface since it has two helpful components term browser and tree browser. With the help of these components the work can be done easily. The users' feedback too helped to come to a conclusion that MeSHMed is the effective interface. MeSHy tool was constructed to extract only the MeSH term from the document. The MeSH terms were extracted, ranked and presented to the user so that it is easy for the user to know the importance and rating of the documents. Classifiers were constructed to label the documents as medical related (health) and non – medical related (non – health) by using three strategies. Specificity and sensitivity values were calculated. InDECS strategy was the best among those since the result shown by the strategy was more when compared to others. The common mistakes done in the queries were found out. The errors done were the incorrect use of MeSH term and Boolean operators. When the user gives a query, and if they did not get relevant documents they keep on changing the query. To avoid it, MeSH terms related to the query are retrieved and the user can select the related term needed. With the help of the terms they can retrieve the documents. Further this study can be expanded by creating a common interface for medical and non – medical professionals. The interface should be an interactive interface that clears the doubt of the user. Before a query is submitted it has to give a brief introduction about the interface and the MeSH term and should give suggestion for writing effective queries.

References

[1] Mark S. Boguski, Online Health Information Retrieval by Consumers and the Challenge of Personal Genomics

[2] Brazilian Internet Steering committee, 2008, Survey on the use of information and communication technologies in Brazil.

[3] Pew Internet and American Life Project, 2000, The online health care revolution: how the web helps Americans take better care of themselves.

[4] Taha J, Sharit J, Czaja S, 2009, Use of and satisfaction with sources of health information among older internet users and nonusers.

[5] Anushia Inthiran, Saadat M. Alhashmi, Pervaiz K. Ahmed, 2012, Medical Information Retrieval Strategies: An Exploratory Study on the Information Retrieval Behaviors of Non- Medical Professionals. Vol 7, Issue 1.

[6] Xiangming Mua, Kun Lu , Hohyon Ryu, 2014 Explicitly integrating MeSH thesaurus help into health information retrieval systems: An empirical case study, Information processing and management pp.24-40.

[7] Krallinger M. Erhardt RA, Valencia A , 2005, Text mining approaches in molecular biology and biomedicine, Vol 10, pp. 439-445

[8] T. Theodosiou, I.S. Vizirianakis, L. Angelis, A. Tsaftaris, N. Darzentas, 2011, MeSHy: Mining unanticipated PubMed information using frequencies of occurrences and concurrences of MeSH terms. Journal of Biomedical Informatics pp.919–926

[9] Li Bin, K C Lun, 2001 The retrieval effectiveness of medical information on the web, Medical Informatics Programme, National University of Singapore, Clinical Research Centre, MD 11, 10 Medical Drive, Singapore 117597, Singapore.

[10] Elizabeth S. Jenuwine, PhD, MLS, Judith A. Floyd, PhD, RN, 2004, Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals.

[11] N.J. Belkin, 1980, Anomalous states of knowledge as a basis for information retrieval, Can. J. Inf. Sci. 5 pp.133–143.

[12] Muh-Chyun Tanga, Ying-Hsang Liub, Wan-Ching Wu, 2013, A study of influence of task familiarity on user behaviors and performance with a MeSH term suggestion interface for PubMed bibliographic search. International journal of medical informatics pp.832–843

[13] Klaar Vanopstal, Joost Buysschaert, Godelieve Laureys, Robert Vander Stichele, 2013, Lost in PubMed. Factors influencing the success of medical information retrieval. Expert Systems with Applications pp.4106–4114

[14] Rey-Long Liu, Yun-Ling Lu, 2009, Online assessment of content skill levels for medical texts. Expert Systems with Applications pp.12272–12280

[15] Xiangyu Jin, James French, Jonathan Michel, University of Virginia, Science Applications International Corporation (SAIC) Charlottesville, Virginia- Query Formulation for Information Retrieval by Intelligence Analysts.

[16] Georg Go-bel, Stefan Andreatta, Joachim Masser, Karl Peter Pfeiffer, 2001, A MeSH based intelligent search intermediary for Consumer Health Information Systems. International Journal of Medical Informatics pp.241–251.

[17] Yeganova, L., Comeau, D. C., Kim, W., & Wilbur, W. J. 2009, How to interpret PubMed queries and why it matters. Journal of the American Society for Information Science and Technology.

[18] Hersh, W. R., Hickam, D. H., Haynes, R. B., & McKibbon, K. A.,1994, A performance and failure analysis of SAPHIRE with a MEDLINE test collection. Journal of the American Medical Informatics Association

[19] Nelson SJ, Powell T, Humphreys BL, 2002, The Unified Medical Language System (UMLS) project. In: Kent Allen, Hall Carolyn M, editors. Encyclopedia of library and information science. New York: Marcel Dekker.

[20] Hoogendam A, Stalenhoef FH, de Vries Robbe PF, Overbeke A John, 2008, Analysis of queries sent to PubMed at the point of care: observation of search behaviour in a medical teaching hospital. BMC Med Inform Decis Making.

[21] Tsuruoka Y, Tsujii J, Ananiadou S. FACTA, 2008, a text search engine for finding associated biomedical concepts.

[22] Cohen T, Whitfield GK, Schvaneveldt RW, Mukund K, Rindflesch T. Epiphanet, 2010, an interactive tool to support biomedical discoveries. J Biomed Discov Collab.

[23] Trieschnigg D, Pezik P, Lee V, de Jong Franciska, Kraaij W et al, 2009, MeSH up: effective MeSH term classification for improved document retrieval.

[24] Bachrach CA, Charen T, 1978, Selection of MEDLINE contents, the development of its thesaurus and the indexing process.

[25] Aronson AR, Lang F, 2010, An overview of Metamap: historical perspective and recent advances.

[26] Chang P, Hou IC, Hsu CL, Lai HF, 2006, Are Google or yahoo a good portal for getting quality healthcare web information?

[27] Falagas ME, Karveli EA, Tritsaroli VI, 2008, The risk of using the Internet as reference resource: a comparative study.

[28] Mancini F, Falcao AE, Hummel AD, Costa T, Silva F, Teixeira F, et al, 2009, Brazilian health – related content web search portal: presentation on a method for its development and preliminary results.

[29] Qi X, Davison BD, 2009, Web page classification: features and algorithms.

[30] Dumais S, Chen H. Hierchical classification of web content, 2000, In: ACM SIGIR conference on Research and development in information retrieval.

[31] McLellan F, 1998, Like hunger, like thirst: patients. Journals and the internet. Lancet (British edition).

[32] Fava GA, Guidi J, 2007, Information overload, the patient and the clinician. Psychother Psychosom.

[33] McCallum A, Nigam K, 1998, A comparison of event models for Naive Bayes text classification.