

Health information retrieval through different approaches: A literature review

Mohammed AbduRabu Jafer¹, Mir Faeq Ali Quadri*²

^{1,2}Lecturer, Department of Dental Public Health, PO Box:- 114, Jazan University, Jazan, Kingdom of Saudi Arabia, 45142.
¹dr.mjafer@gmail.com, ²*faeq_ali@yahoo.com

Abstract

Background: The medical information in the internet especially the World Wide Web, is escalating at a tremendous rate. This has made a substantial amount of health information available to the public and healthcare professionals alike through internet. Thus Health information retrieval (HIR) has become a very important part of medical informatics. Due to the same reason there has been many approaches tried for health information retrieval. This paper analysis such approaches currently used.

Method: A literature review was performed in order to obtain research articles indicating the various health information retrieval methods. The databases that were searched included Medline, Meditext, Ebscohost, and Science direct. Keywords such as "health informatics, health information, health information retrieval" were used in combination with the truncations such as 'AND' & 'OR' wherever necessary.

Findings: With a thorough literature review it is seen that providing information indiscriminately to the clinicians is more likely to impair decision making rather than to enhance it. Quality of information acquired from patient-related literature is vital to the quality of care provided. The health care providers do have questions during their clinical care activities and seek internet for information but if the quality of the information retrieved is not good these questions go unanswered, even when the answers are available. These may have serious consequences on the outcome of the patients.

Improvements: The development of new HIR approaches and Unified Medical Language System are not only simplifying the process of HIR but also giving quality information in a short time. The future trend of HIR systems are towards a balance between their user friendliness and quality of information retrieved. Thus in conclusion it could be said that along with the development of HIR system it is very essential to build a good, strong, secure privacy and confidentiality policy to protect the vast data stored electronically from misuse.

Keywords: Health Informatics, Health information retrieval, Medical record system, Patient record system, Literature review.

1. Introduction

The Internet has taken the world by storm over past decade. It has had the major impact on every field of life including health and Health care [1]. The medical information in the internet, and especially the World Wide Web, is escalating at a tremendous rate. This has made a substantial amount of health information available to the public and healthcare professionals alike through internet [2]. In the past 10 years, powerful computers, software's, and technologies have been developed to enable health care organizations to automate the information available and decision making, so as to increase the time and quality of care of the patients [3]. The study of the collection, storage, retrieval, and analysis of data and information in health care to support such automation in clinical and administrative decision making is called Medical informatics [3].

Due to the vast amount of information and the rate at which they have become available, the collection and storage of them are not in order but scattered [4]. This has made difficult for locating the precise information by the user and most of the time the nature of information gathered is neither efficient nor effective [3]. Thus Health information retrieval (HIR) has become a very important part of medical informatics. Due to the same reason there has been many approaches tried for health information retrieval. This paper analysis such approaches currently used. The paper also given an account of information seeking behavior of patients and healthcare providers, importance of quality of health information for them, few new health information retrieval methods and privacy concerns.

2. Method

A literature search was performed in order to obtain research articles indicating the various health information retrieval methods. The databases that were searched included Medline, Meditext, Ebscohost, and Science direct. Keywords such as "health informatics, health information, health information retrieval" were used in combination with the truncations such as 'AND' & 'OR' wherever necessary. The research articles which did not fit the inclusion and the exclusion criteria were discarded. Only the research articles in English were searched. There was no time specification allotted for the search of the articles.

3. Findings

3.1. Information seeking behavior

3.1.1. Health care providers

A vast field of health care providers seeks information through internet. Public health professionals depend on medical information available online to make a quick and effective search for relevant reports and statistical information during public health emergency such as a disease outbreak [5]. The relative ease with which the information could be accessed and gathered online has made it popular among public health professionals, who are doing population research studies also [5].

Practicing physicians use the readily available and enormous amount of medical information to assist them in maintaining the level of their clinical skills and learn more about advances in medical diagnosis [6]. The neuroradiological and other medical images on the internet are used by the medical professional for both diagnostic as well as educational purposes to the young medical students [7].

The growing acceptance of evidence-based practice (EBP) in decision making by healthcare organizations has resulted in clinicians to be aware of recent published medical papers and reports. The medical information on internet has provided clinicians an easy way to keep up to date with their knowledge [8]. Clinicians also use internet to handle exceptional cases where in access to the full range of medical literature is required in a short time [9].

3.2. Patients and or Health Consumer

Over the past few years the general populations are more aware of the health and health problems. This has led to Health consumers of varying backgrounds to perform HIR for themselves as well as for friends and family to merely satisfy their curiosity [3]. Patients themselves also wish to be armed with knowledge about their medical problems and the readily available information online is very useful for them in this regards [10].

Nowadays, patients tend to be more informed before, during, and after consulting to their physicians. These patients also wish to take part in the decision making process for their health problems. For this purpose, patients are seeking information on internet as it is both easy and fast [10]. Patients also seek internet to verify what the physicians say, cross check the information and look for alternatives from various other sites [11].

3.3. Importance of quality of health information

3.3.1. Health care providers

Providing information indiscriminately to the clinicians is more likely to impair decision making rather than to enhance it. Quality of information acquired from patient-related literature is vital to the quality of care provided [6]. As stated above, health care providers do have questions during their clinical care activities and seek internet for information but if the quality of the information retrieved is not good these questions go unanswered, even when the answers are available. These may have serious consequences on the outcome of the patients [9].

For example, answers to questions about medications and its side effects in course of patient treatment could not only save clinicians from medical error but also save patient's life [12]. Quality information could also provide the physicians with the new diagnostic tests of which he could be completely unaware of. It could save lot of time and money in diagnosing patients and starting appropriate treatment to them [9].

Hence, the basic challenge of HIR is to provide the information, in a useful form when and where it's needed [6].

3.4. Patients and or Health Consumer

Patients and most citizens use general purpose search engines and web sites to seek the medical information which was most of the time found to be suboptimal [11]. Though web sites could satisfy patients curiosity the abundance of information and variable quality of its content may not give a clear response to patient's queries [13]. Also the credibility and authenticity of information in them is questionable and needs to be evaluated in detail, which most of the patients fail to do [14].

The potential harm from misleading and inaccurate health information could range from as simple as increased consultation time with physician to severe anxiety disorder [15]. According to Kawada (2003) inaccurate information may have bad effect on self – rated health of individual, which is found to be an independent predictor of control of ill health and survival [16]. Hence, the information received by the patients got to be of high quality.

3.5. Current health informational retrieval approaches

Information retrieval from the online sources is difficult both for health professionals and public because of distribution among various hosts, the speed of accumulation of new information, differences in organization and representation of data , unawareness of relevant sources, and the continuing change of existing data collections [17]. Hence, Health information retrieval (HIR) system has been designed to find relevant information among the numerous health information available online [4].

3.6. Current approaches for HIR

a) *Specialized guides*: This is the first step towards structuring the information available on line. These are developed and maintained by health care specialists and consists of structured information in a specific domain. Some examples of specialized guides in medicine are MedSurf (www.medsurf.com), CliniWeb(www.ohsu.edu/clinweb/), and Hardin Meta Directory (www.lib.uiowa.edu/hardin/md) [18].

Advantage: The main advantage of these guides is that the web pages referred to are often, accurate, reliable and relevant as they are maintained by medical experts [10].

Disadvantage: As they lack of automations, the evaluation of documents are done manual. The slow process of maintenance of these guides cannot cope with the growth rate of health information on the Internet [10]. Also, the new references are usually not systematically added. Hence, the approach is not satisfactory [18].

b) *Subject directory*: It is an extension to specialized guides where the information organized into a category tree by subjects. The users can add documents to the existing category or create a new one but to retrieve the information he has to go through the hierarchy [18]. Some of the examples are MedWeb (www.medweb.emory.edu/MedWeb) and Medical Matrix (www.medmatrix.org) [10].

Advantage: It is a good approach to add and maintain the relevant web page and the coverage is more extensive [10].

Disadvantage: indexes are maintained manually. Hence, to choose and structure categories becomes difficult as the information increases. Searches by different users for the same information often leads to different results [18].

c) *General or medical search engines*: Web pages are locally indexed by a software component called robot or a crawler. The so formed database is then explored to match the keywords by a process called search interface and relevant web pages are listed [10]. Examples of general search engines are: Google (www.google.com), AltaVista (www.altavista.com), yahoo (www.yahoo.com) and Hotbot (www.hotbot.com). These scan the full text of web pages and index as many pages as possible [18]. Examples of medical search engines are: MedHunt (www.hon.ch/Medhunt). This is used only to look for and indexing medical related web pages [18].

Advantage: it maximizes the number of documents returned for a given request and respond to their users quite fast [10].

Disadvantage: precision of the information gathered is low. This means that obtaining the domain specific information, such as the health domain, is difficult and will end up with lot of irrelevant information [18].

d) *Medical databases and search engines on these databases*: these are built for the medical professionals who are good in the terminology of medicine. For example: PubMed and MEDLINE database. PubMed is the official search engine on this database [18].

Advantage: With good query words specific, authentic and scholarly information could be collected [18].

Disadvantage: As queries are usually detailed with medical terminology, it is hard for an ordinary person to understand. Therefore, medical databases are not suitable for an ordinary person to use [18].

e) *Meta-search engines*: In these, queries are automatically submitted to different search engines and the results so obtained are integrated to give extensive but specific information. Example of a general-purpose meta-search engine

is Dogpile (<http://www.dogpile.com/>), which searches many search engines, including Google, Yahoo!, MSN etc [19].

Advantages: Prevents tedious process of using queries in all the search engines [19].

3.7. Roles of Medical Subject Headings MeSH

MEDLINE is one of the world's largest and most indexed databases of medical information and it is maintained by the National Library of Medicine (NLM) [18]. The Medical Subject Headings (MeSH) was developed by the NLM with more than 100,000 concepts and has a hierarchy that goes 11 levels deep, to store information in a structured way [4]. In this, concepts are organized into hierarchies going from the most general on the top of the hierarchy to the most specific in the bottom of the hierarchy [4]. Though they provide additional descriptive flexibility to the organization of concepts but the polyhierarchies prevented one from a complete view concerning a specialty. This increased ambiguity in information retrieval and made it a bit difficult for consumers to use [20].

3.8. Roles of Unified Medical Language System (UMLS)

The project is focused on overcoming two important barriers to the development of information systems that can help health professionals make better decisions. These barriers are: 1) disparity in the terminologies used in different information sources and by different users, and 2) the sheer number and distribution of information sources that might be relevant to any user inquiry [21]. The UMLS supports the development of user-friendly systems that can effectively retrieve and integrate relevant information from variant sources. Essentially, it is the constructive approach with sophisticated knowledge representation in order to make an amenable an automation system [21].

3.9. New health information Retrieval approaches

To overcome the short comings of existing retrieval methods some of the following innovations are being tried.

a) World Reliable Advice for Patients and Individuals (WRAPIN): It is an advanced search engine with specialization in health domain online information. It is developed with a multidisciplinary collaboration of experts from various organizations [13]. The main purpose of WRAPIN is to help users assess the credibility of online medical information, using a reference base constituted exclusively of trustworthy documents [13].

b) MedicoPort is a search engine built to retrieve medical information from the internet. It is a general purpose engine built to help non-expert medical information seekers to acquire authentic health information from the web [10]. It is not for medical literature database search.

c) Health Information Query Assistant (HIQuA): Query formation is a major aspect of consumer HIR that is in need of improvement. All the current search engine depend on proper query to be selected. The limited medical vocabulary of the health consumers leads to simplistic queries and hence, less accurate information retrieved [3]. The Health Information Query Assistant (HIQuA) is developed to recommend alternative/additional query terms for the consumers. This system not only recommends medical concepts but also modifies an initial user query as building blocks to form more specific queries [3].

3.10. Privacy an issue

Privacy and confidentiality are emerging as one of the most critical issues of paramount importance in the era of networked information systems that store electronic patient records or institutional data. The use of electronically stored medical records, images, or body material are available to lot of unauthorized people and should be monitored very carefully [22]. Kagolovsky et al (1998) in fact question the permissibility of all or part of the data set to decision makers, even though they pursue a legitimated purpose [6]. The Mayo Clinic Foundation in United States of America has started to take a "broad informed consent" from all patients for future use of their data stored electronically. But it is interesting to see whether this sort of broad consent will fulfill requirements of specific informed consents required for most of the researches [22].

4. Conclusion

As the medical information available on the internet is increasing, the process of quality health information retrieval (HIR) from internet is becoming more and more difficult. Complicating HIR is both the varied needs of information and capabilities in using different search engines by practitioners, administrators, researchers and the general public from internet. The development of new HIR approaches and Unified Medical Language System are not only simplifying the process of HIR but also giving quality information in a short time. The future trend of HIR systems

are towards a balance between their user friendliness and quality of information retrieved. To conclude, along with the development of HIR system it is very essential to build a good, strong, secure privacy and confidentiality policy to protect the vast data stored electronically from misuse.

5. References

1. L. N. Pealer, S. M. Dorman. Evaluating Health Related Website. *Journal of School Health*. 1997; 67 (6), 232-235.
2. V.T. Burton, S.A. Chadwick. Investigating the practices of student researchers: Patterns of use and criteria for use of internet and library sources, *Computer Composition*, 2000; 17(3), 309–28.
3. Q.T. Zeng, J. Crowell, R.M. Plovnick, E. Kim, L. Ngo, E. Dibble. Assisting Consumer Health Information Retrieval with Query Recommendations, *Journal of the American Medical Informatics Association*, 2006; 13(1), 80-90.
4. L. F. Soualmia, S. J. Darmoni. Combining different standards and different approaches for health information retrieval in a quality-controlled gateway, *International Journal of Medical Informatics*, 2005; 74 (2-4), 141 -150.
5. D. Revere, A. M. Turner, A. Madhavan, N. Rambo, P. F. Bugni, A. M. Kimball, S. S. Fuller. Understanding the information needs of public health practitioners: A literature review to inform design of an interactive digital knowledge management system, *Journal of Biomedical Informatics*, 2007; 40 (4), 410-421.
6. Y. Kagolovsky, D. Freese, M. Miller, T. Walrod, J. Moehr. Towards improved information retrieval from medical sources, *International Journal of Medical Informatics*, 1998; 51 (2-3), 181-195.
7. S. Dominich, J. Goth, T. Kiezer. NeuRadIR: Web-based neuroradiological information retrieval system using three methods to satisfy different user aspects, *Computerized Medical Imaging and Graphics*, 2006; 30(4), 263-272.
8. D. P. Lawrence, A. Spink. Assessment of preferences for classification detail in medical information: is uniformity better? *Information Processing & Management*, 2003; 39(3), 465-477.
9. L. M.M. Braun, F. Wiesman, H. J. van den Herik, A. Hasman, E. Korsten. Towards patient-related information needs, *International Journal of Medical Informatics*, 2007; 76(2-3), 246-251.
10. A.B. Can, N. Baykal. MedicoPort: A medical search engine for all, *Computer Methods and Programs in Biomedicine*, 2007; 86(1), 73-86.
11. G. Eysenbach, C. Kohler. How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews, *British medical journal*, 2002; 324, 573-577.
12. S.M. Maviglia, C. S. Yoon, D. W. Bates, G. Kuperman. Knowledge Link: Impact of Context-Sensitive Information Retrieval on Clinicians' Information Needs, *Journal of the American Medical Informatics Association*, 2006; 13(1), 67-73.
13. A. Gaudinat, P. Ruch, M. Joubert, P. Uziel, A. Strauss, M. Thonnet, R. Baud, S. Spahni, P. Weber, J. Bonal, C. Boyer, M. Fieschi, A. Geissbuhler. Health search engine with e-document analysis for reliable search results, *International Journal of Medical Informatics*, 2006; 75(1), 73-85.
14. J. Brophy, D. Bawden. Is Google enough? Comparison of an internet search engine with academic resources, *Aslib Proceedings: New Information Perspectives*, 2005; 57(6), 498–512.
15. P. Kim, T. R. Eng, M. J. Deering, A. Maxfield. Published criteria for evaluating health related websites: review, *British Medical Journal*, 1999; 318, 647-649.
16. T. Kawada. Self-rated health and life prognosis, *Archives of Medical Research*, 2003; 34(4), 343-347.
17. J.R. Griffiths, P. Brophy. Student searching behavior and the Web: Use of academic resources and Google, *Library Trends*, 2005; 53(4), 539–554.
18. L. Bin, K. C. Lun. The retrieval effectiveness of medical information on the web, *International Journal of Medical Informatics*, 2001; 62(2-3), 155-163.

19. B. J. Jansen, A. Spink, S. Koshman. Web searcher interaction with the Dogpile.com metasearch engine, *Journal of the American Society for Information Science and Technology*, 2007; 58(5), 744-755.
20. S. Schulz, U. Hahn. Medical knowledge re-engineering – converting major portions of the UMLS into a terminological knowledge base, *International Journal of Medical Informatics*, 2001; 64(2–3), 207–221.
21. B. L. Humphreys, D. A. B. Lindberg, H. M. Schoolman, G. O. Barnett. The Unified Medical Language System An Informatics Research Collaboration, *Journal of American Medical Informatics Association*, 1998; 5(1), 1-11.
22. J.P. Vandenbroucke. Maintaining privacy and the health of the public should not be seen as in opposition, *British Medical Journal*, 1998; 316(7141), 1331-1332.

The Publication fee is defrayed by Indian Society for Education and Environment (www.iseedyar.org)

Cite this article as:

Mohammed AbduRabbu Jafer, Mir Faeq Ali Quadri. Health Information retrieval through different approaches: A literature review. *Indian Journal of Medicine and Healthcare*. 2015; 4(2), July.