

Search Engines use for Effective Literature Search in Biomedical Research

SUBHASH B THAKRE, SAMIR H GOLAWAR, SUSHAMA S THAKR, AVINASH V GAWANDE

ABSTRACT

Introduction: Searching is hampered by the wide variations in terminology that may be common usage at any given time, which inhibit effective searching by both computers and people. Not enough knowledge and skills concerning literature search makes budding researcher incapacitate.

Objectives: This article discusses how to compare biomedical search engines and also covers the strengths and limitations of a particular search engine/database while searching for biomedical literature.

Methods: This research was a cross-sectional study by which five popular search engines in medical sciences were evaluated. PubMed, the basic search engine was used to search literature related to other search engines. Important and widely used web of search engines; were Google, Google scholar, PubMed and GoPubMed. This

article also focus on Quertle and KNALIJ, specifically developed to run easier and more effective searches within PubMed.

Results: Most widely practiced search engine is PubMed and Google scholar; however least was Knalij. Though the features and special features of search engines are very helpful in precise and conceptual search of biomedical literature. Every search engine has its own limitations. Moreover, the review provides a detailed summary for the recent advances in the field of biomedical literature search.

Conclusion: This review can directly serve both non-experts and expert users when they wish to find systems other than PubMed. Moreover, the review provides a detailed summary for the recent advances in the field of biomedical literature search. This is particularly useful for existing researcher for effective literature search.

Keywords: Google scholar, Knalij, Literature search, Pubmed, Quertle, Search engines

INTRODUCTION

Biologists currently waste a lot of time and effort in searching for all of the available information about each and small area of research. This is hampered further by the wide variations in terminology that may be common usage at any given time, which inhibit effective searching by both computers and people. Reviewing the literature requires the ability to juggle multiple tasks, from finding and evaluating relevant material to synthesising information from various sources, from critical thinking to paraphrasing, evaluating, and citation skills [1]. With the explosion of available health information, physicians increasingly rely on bibliographic databases for health information to guide the care of their patients. Unfortunately, physicians face challenges when trying to find the information they need [2-4].

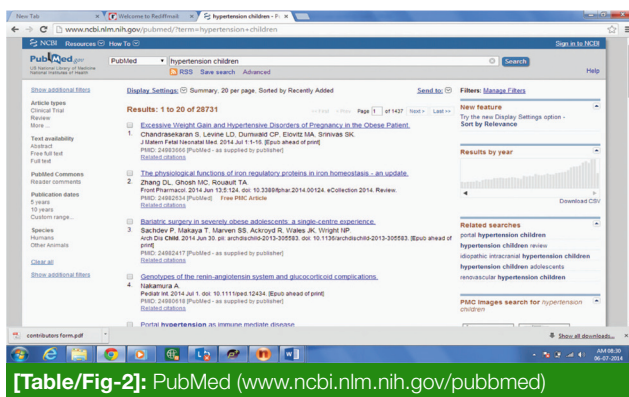
Yet, as many a doctors will point out, the bigger problem with medical knowledge today is not its paucity, but the difficulty of navigating what is there. Finding the right answer quickly for a patient is difficult, and perhaps nothing will replace a good medical librarian in finding that information [4-5]. Web search engine is a tool designed to search for information on the World Wide Web. The emergence of digitalisation and internet has increased possibility of making information available to

anyone, anywhere, any time and in any format [6-8] "What" information's (concepts), "Whom" are you looking for? (Person, institution, and place), "Where" (journals) and "When" (times) which are relevant to your query [9].

We are the decision maker in regards, when to start and stop our search. A specific and well defined research query, appropriate search strategy, basic knowledge about type of literature, level of evidence and idea about various biomedical search engines is more than enough for quality literature search [8]. Search engine is a very powerful tool that will help you discover knowledge and understand huge data within a blink of an eye [10]. The goal of searching the literature is to find the right facts and the right references. Here are several different search engines for searching the biomedical literature. How can you find the one that is best for you? For Good quality search you need a search engine that covers the relevant content and enables you to quickly refine in on the papers you require. It is like finding a needle in the haystack; more hay just makes the task harder and longer. A researcher must keep in mind the strengths and limitations of a particular search engine/database while searching for a particular type of data. This article discusses how to compare biomedical search engines and also covers the strengths and limitations

Search query	No of studies retrieved	No of relevant studies
PubMed	17692	2261
Google search	2370	311
Google scholar	1268	153
GoPubMed	244	52
Quertle	7	3
Knalij	1	1

[Table/Fig-1]: Frequency of retrieval of relevant studies by use of PubMed



[Table/Fig-2]: PubMed (www.ncbi.nlm.nih.gov/pubmed)

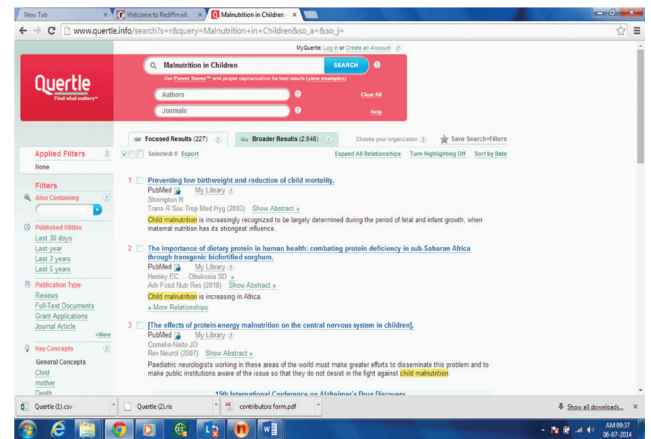


[Table/Fig-3]: Google scholar (http://scholar.google.co.in)

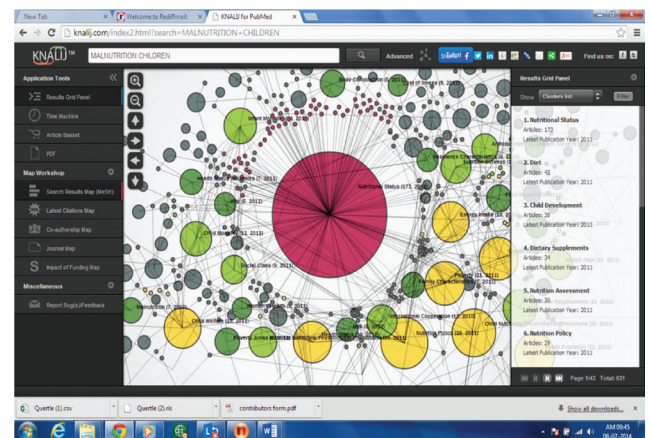
of a particular search engine/database while searching for biomedical literature.

METHODS

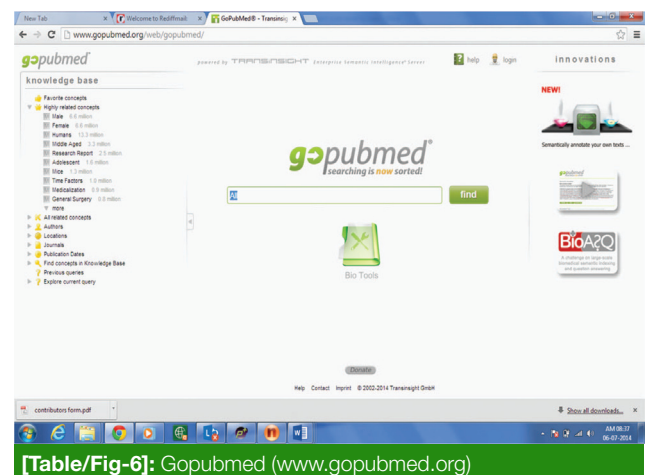
This research was a cross-sectional study by which five popular search engines in medical sciences were evaluated. PubMed, the basic search engine was used to search literature related to other search engines. To select keywords, medical subject heading (MeSH) was used. We entered given keywords in the search engines and after searching, all hits were evaluated and relevant entries pertaining to specific search query like PubMed, Google scholar, etc. were considered as related



[Table/Fig-4]: Quertle (http://www.quertle.info)



[Table/Fig-5]: KNALIJ (http://knalij.com)



[Table/Fig-6]: Gopubmed (www.gopubmed.org)

articles. Citations related to literature search were considered as relevant citations [Table/Fig-1]. "PubMed" search engine was primarily used for retrieving published literature related to search queries "Google scholar", "Go PubMed", "Knalij", "Quertle" and "PubMed" itself. We attempted to include all

studies that have been published in PubMed indexed journals. [Table/Fig-1] reveals number of studies retrieved after specific search query. Further search was limited to Review, Core Clinical Journal Articles, Last Five Years, Full Text, and Human Species.

Literature resources: Search engines for internet-based search of medical literature include Google, Google scholar, Scirus, Yahoo search engine, etc., and databases include MEDLINE, PubMed, MEDLARS, etc. Several web-libraries (National library Medicine, Cochrane, Web of Science, Medical matrix, Emory libraries) have been developed as meta-sites, providing useful links to health resources globally [8]. A number of sources are now available in the biomedical field apart from PubMed, which provides free access to the largest biomedical resource available and is updated daily. Specifically, PubMed listed 6,615 journals (as on June 25, 2014) indexed from various countries [11]. Encouragingly, the national initiative by the Indian Council of Medical Research (ICMR) funded project, 'National database of Indian medical journals' by the National Informatics Centre Services Incorporated. (NICSI), New Delhi, provides among others, a bibliographic database of about 100 prominent peer-reviewed Indian biomedical journals indexed from 1985 onwards through indMED [12].

Web Sites: Web sites can be very good for finding quality information. However, you have to take the responsibility to rigorously evaluate each site for quality; anyone can post a Web page, regardless of their expertise or intentions.

Web search engines [13]: A web search engine is a software system that is designed to search for information on the World Wide Web. The search results are generally presented in a line of results often referred to as search engine results pages (SERPs).

Important and widely used web of search- engines; (Canadian medical association survey 2005) were Google, Google scholar, PubMed and Scirus [14]. Now, We shall focus on Quertle and KNALIJ, specifically developed to run easier and more effective searches within PubMed.

How to compare search engines? [15]

Quality, not Quantity - The goal of searching the literature is to find the right facts and the right references. In that endeavour, getting 100,000 hits is not better than retrieving 50,000 hits, when there were only 100 documents that were actually relevant. This means you need a search engine that covers the relevant content and enables you to quickly hone in on the papers you need. It is like finding a needle in the haystack; more hay just makes the task harder and longer.

Exploring the Results and Finding What You Need - When you do a literature search for the purpose of finding information, such as what genes are associated with a particular process, extracting the desired information can be another pain point. The easier the search makes that step, the better for you.

Ease of Understanding Why You Got the Results - Once you get a list of hits, the next thing most of us do is peruse the

results to see if the search indeed found relevant documents. You need a search engine that makes this easy since this can be a time-consuming task.

Precautionary Measures [16]: When you do a literature search to find an article to use as a reference, refining the results to get to the right reference can also be a hit-or-miss effort. Refining the search too much can exclude the key reference; not refining enough leaves you with a large haystack. Thus, the better search engine will support intuitive drill-down.

PubMed remains the most widely used resource for medical literature. [Table/Fig-1] PubMed offers number of special searching and limiting features [8,17] which help us in effective literature search.

PubMed [8, 15, 17] [Table/Fig-2]

PubMed is the US Government supported effort to provide a means to search the biomedical literature. Over the years, this amazing effort has been critical for essentially every research project and for most physicians.

Search and ranking algorithms: PubMed uses the Entrez search engine for keyword matches of your search term against the text in the title, abstract and citation information of each document. There is no full-text document searching, although links to the full-text is provided on some of the separate document pages. The Entrez engine also matches your search terms against the MeSH headings to find documents. By default, results are sorted by date, and can be resorted by author, journal or title. No relevance sorting is available.

Content: PubMed covers the citation information for over 20 million journal articles.

PubMed Alternatives: Most of the alternative biomedical literature search sites are based on PubMed and use the NCBI Entrez search engine - even if the specific interface looks different. Hence, these PubMed "knock-offs" are not included in this comparative analysis.

Details about literature search techniques, search strategy, levels of evidence covered in earlier study [8] and special features in PubMed tutorial [17].

Google scholar [8,15, 17][Table/Fig-3]: Google scholar is a cross-disciplinary search engine for journal articles and other scholarly works. Google Scholar has been met with both enthusiasm and criticism since its introduction in 2004. [18]. Google Scholar does not offer the number and extent of special searching and limiting features available in PubMed. However, Google Scholar provides some advantages in that it is an easy place to begin a search to find an initial retrieval of possibly worthwhile articles. It also offers searchers the ability to find citations to older items that they would miss if they use only PubMed. Additionally, Google Scholar has the potential to provide access to the grey literature. This increased access to a part of the biomedical literature, which can be difficult to search, may have implications for the public health field [19].

Search and ranking algorithms: Google Scholar does keyword matching to find articles that contain all of your search terms. The results are presented in relevance order, where relevance combines the text of the article with citation, journal, and author information. The citation information appears to dominate, so that highly cited articles are listed first, not necessarily the articles that are relevant to your needs. Another option is search by date; latest publication appears first in summary. Google Scholar is collaborating with university libraries to develop a way to access full-text journals through institutional subscriptions, so that researchers and physicians affiliated with a university can go directly from a Scholar search to a full-text journal article if their university has a subscription to that journal. Also intriguing is the potential of future versions of Scholar to give free, efficient access to articles from commercial journals reproduced for open access on personal or institutional pages. When development is complete, Scholar may access the better quality sites now accessed by Google, supplemented by the electronic journal literature and additional reputable sources.

Content [6,15-16, 21]: Google Scholar's actual content is not well-defined, but essentially covers whatever Google can reach with its search engine indexing robots, limited to whatever Google decides is scholarly. As such, this search engine covers the most full-text content as most publishers allow Google to index their articles. This search engine also covers patents and legal opinions.

Limitations: While using Google Scholar, one has to keep in mind its limitations too. It does not index the majority of the scholarly materials indexed by commercial database vendors. Moreover, it does not provide a list of the sources it is indexing. It has never shared with the public their search algorithm which is why many times one may not find key pieces of literature in the database. Perhaps it could be used for citation verification rather than for true searching purposes. For quick clinical searches, Google Scholar returns twice as many relevant articles as PubMed and provides greater access to free full-text articles [21]. Although these two systems are difficult to compare, it is still important to explore the differences between them [22].

Quertle15: [Table/Fig-4] Like PubMed and Google Scholar, Quertle is freely accessible. The quertle work is simple: it finds relationships, not only keywords. So, all the terms in your query are found together in a meaningful way. It uses natural language processing to match your search against actual assertions made by the author, rather than simple keyword matching or proximity. The concept of relationship relies upon semantic triplet- "subject-verb-object" which represent a statement.

Search and ranking algorithms: Take an example of a simple query based on information need: "What affects myocardial infarction"? In quertle you can put in the search box to run the query; the system automatically looks for all the synonyms like coronary heart disease, heart attack, ischemic heart disease,

etc. Results are shown by default via "relationship", with search term highlighted to provide immediate context. The label below search box gives you back more traditional search- "Keyword", with documents containing your terms anywhere also match with the MeSH terms. On the left side of the screen you find filter options going from traditional "published within" or "publication type" to innovate "key concepts". You can narrow your search by "action" implied by your search (in this case "causes") or by general concepts such as "acute, chronic, complication" and so on. In the cited example, focused results 4,911 and broader results are 8,204. By clicking on prevention focused results remained 342. They are well organized in order to help and foster navigation. The first key concepts directly stem from your query. There is a provision of refining your search by introducing a new key word in the upper left box "Also containing" filter. If you are looking for treatment initiation in myocardial infarction then you can use "Also containing" filter; search hit reduced to 22 focused results.

A dramatically innovative tip is the "Power term" feature, where the power term represents the entire classes of object. They are represented with a sign \$ before the term. You can use \$ disease, \$adverse effects, etc. Running a search with "Diabetes \$Drugs" means that system automatically explodes the query with types of diabetes and drugs known in the literature. Every time we did not have to write each drug. If you are looking for possible disease implication of hip replacement, typing "hip replacement \$ diseases" you will be shown results concerning osteoarthritis, pain, dislocation, necrosis without typing any one of them. By default, Quertle ranks the results based on relevance, where relevance is primarily determined by how many times the document supports the assertion. In addition to its advanced relationship-based results.

Content: Quertle's content covers all of PubMed plus a large number of full-text articles, the NIH RePORTER database of NIH grants, the TOXLINE database of toxic effects of chemicals, biomedical news reports, and whitepapers such as research reports from biomedical companies.

KNALIJ search engine [23-25] [Table/Fig-5]: "Knali" pronounced as "knowledge" is a visualization tool based on the assumption that the best pattern recognition system is persons own visual system. It is a very powerful tool that will help you discover knowledge and understand huge data within a blink of an eye. An exploration and visualization engine developed for the world's largest data sources in real time. It is the power of vision combined with the power of web. PubMed is the first world is really large repository that has been integrated with KNALIJ. Providing public access to millions of biomedical citations, online journals and online books. It supports all five browsers IE9+, Google chrome, Firefox, Opera, Safariplus and the iPad. It is a real-time infographic engine. KNALIJ PubMed draws upon more than 22 million abstracts as a unique knowledge discovery tool.

Search and ranking algorithms: Visualization engine provides search information in the form of cluster nodes and

article nodes (big discs) denote subject, author, journal or funding source (depending on the type of search). Clusters are groupings of related Articles. The article nodes (smaller discs) connect to clusters. These are Articles returned from searching PubMed. Searching with KNALIJ PubMed is exactly the same as with PubMed, having identical functionality and field tags. You can run a simple search within the search box or open the advance search and use PubMed search builder to combine different fields with Boolean operators. When you run a search the system first discovers the pertinent Article in PubMed, then maps complex relationship in real time and visualizes and identifies meaningful correlation and connections. The results are visually presented as interrelated colour knots. The cluster nodes (big discs) denote subject, author, journal or funding source (depending on the type of search). Clusters are groupings of related Articles. The article nodes (smaller discs) connect to clusters. These are the Articles returned from searching PubMed. Selecting any cluster from the 'Results Grid Panel' will focus and centre it on the displayed map. Click any cluster on the KNALIJ map to highlight it and its nearest neighbours in the foreground. All other clusters will be hidden. Articles for the visible clusters will be listed in the 'Results Grid Panel'. May features are similar to PubMed search like filters, time machine, advanced search, search by author, subject, and journal, etc. Search result map : Uses MeSH terms assigned by PubMed to cluster articles. Not included in it are articles without at least one assigned MeSH term. You can share your results via face book, twitter, etc.

Content: Quertle and KNALIJ are two innovative tools created to search PubMed in an easier and more effective way. They dramatically reduce time to discover meaningful results. Quertle allows a semantic search in multiple biomedical databases (PubMed included) and runs a query via relationships between concepts, so that you retrieve at ease more pertinent results and can navigate them by "key concepts". KNALIJ is a visualization tool which searches PubMed and presents the results in the form of visual, interactive maps you can zoom, scale, and explore according to new paths. Knalij provides latest citations till 2012.

Go PubMed [25] [Table/Fig-6]: www.gopubmed.org; the semantic search engine for the life sciences, has been recognized with the 2009 red dot: best of the best award in the category communication design – graphical user interfaces and interactive tool. Trans insight has been recognized with the German Innovation Prize IT for its outstanding developments in Enterprise Semantic Intelligence at CeBIT 2011.

GoPubMed allows significantly faster finding information needed through the use of biomedical background knowledge. It doesn't rank, the user does! GoPubMed retrieves PubMed abstracts for your search query and sorts relevant information to the 4 top level categories.

Search and ranking algorithms: Searching by concepts, authors, journals, etc. is done using the same search strategies

as in PubMed. "What", "Who", "Where" and "When" show the concepts, persons, institutions, places, journals and times which are relevant to your query search. Further you can refine features shown in the tree to the left of the term. Simply click on it! "Require" and "Exclude" are filters to refine the search. They select all documents mentioning or excluding one concept and its children of a category.

CONCLUSION

By our three selection standards, a total of five Web systems were included in this review. They are comparable to PubMed given that they are designed for the same purpose and make use of full or partial PubMed data. We first provided a general description of PubMed including its brief description of search engine, search ranking algorithm, content, unique characteristics and limitations if applicable. Next, according to their different features, we further described each of them in greater detail and showed their differences. Finally we reviewed the five systems as a whole and discussed their innovative aspects with respect to searching, result analysis and enrichment, and user interface/usability. This review can directly serve both non-experts and expert users when they wish to find systems other than PubMed. Moreover, the review provides a detailed summary for the recent advances in the field of biomedical literature search. This is particularly useful for existing researcher for effective literature search.

REFERENCES

- [1] Budgen D, Brereton P (2006) Performing systematic literature reviews in software engineering. Proc 28th IntConf Software Engineering, ACM New York, NY, USA, pp. 1051–1052. doi:10.1145/1134285.1134500.
- [2] Ely JW, Osheroff JA, Ebell MH, Chambliss ML, Vinson DC, Stevermer JJ, Pifer EA. Obstacles to answering doctors' questions about patient care with evidence: qualitative study. *BMJ*. 2002; 324(7339):710.
- [3] Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. *Health Info Libr J*. 2007; 24(2):78–94.
- [4] Shariff HZ, Bejaimal Shayna AD, Sontrop JM, Lansavicuchus AV, RHaynes RB, Mathew AW, Gerg AX. Retrieving Clinical Evidence: A Comparison of PubMed and Google Scholar for Quick Clinical Searches. *J Med Internet Res*. 2013; 15(8): e164.
- [5] Al-Ubaydli M. Using search engines to find online medical information. *PLoS Med*. 2005; 2:e228.
- [6] Jain NC. Open access ensures effective information retrieval of medical literature in e-databases. *Indian journal of Community Medicine*. 2013; 38(1):1–3.
- [7] Casebeer L, Engler S, Bennett N, Irvine M, Sulkes D, DesLauriers M, et al. A controlled trial of the effectiveness of internet continuing medical education. *BMC Med*. 2008; 6:37.
- [8] Thakre S B, Thakre S S, Thakre A D. Biomedical literature search for budding researcher. *J ClinDiagn Res*. 2013; 7(9):2033-37.
- [9] Searching is now sorted. [Web document]. Available on: <http://www.gopubmed.com/web/gopubmed/www/GoPubMed/Search/index.html#HowToUseGoPubMed>.
- [10] www.knalij.com. [Web document]. Available at: <http://knalij.com/index2.html?search=SEARCH#>; Last accessed on 21-07-2014.
- [11] Currently Indexed Journals in PubMed. [Web document]. Available from: <http://www.ncbi.nlm.nih.gov/nlmcatalog?term=currentindex%5Ball%5D>. [Last accessed on 2014 June 25].
- [12] indMED/ medIND. Available from: <http://indmed.nic.in/>. [Last accessed on April 25, 2014].
- [13] Web Search engine- Wikipedia. [Web document]. Available at : http://en.wikipedia.org/wiki/Web_search_engine. [Last accessed on 2014 May 20].

- [14] Giustini D, Basky E. A look ag Google Scholar, PubMed, and Scirus comparison and recommendations. *JCHL/JABSC*. 2005;26:85–89. Available on: <http://pubs.chla-absc.ca/doi/pdf/10.5596/c05-030>. (Last access on July 20, 2014).
- [15] Which Biomedical Search Engine is best for you?[Web document]. : <http://www.biomedicalsearchengines.com/> [last accessed on May 25, 2014].
- [16] National Library of Medicine (US) Key MEDLINE Indicators. [Web document]. Available from: [webcitehttp://www.nlm.nih.gov/bsd/bsd_key.html](http://www.nlm.nih.gov/bsd/bsd_key.html). [Last accessed on May 25, 2014].
- [17] PubMed Home page. [Web document]. Available on: <http://www.ncbi.nlm.nih.gov/pubmed>. [Last accessed on April 15, 2014].
- [18] Google Scholar help: about Google Scholar. [Web document]. Mountain View, CA: *Google*, 2005. Available on: <http://scholar.google.com/intl/en/scholar/about.html>.
- [19] Turner AM, Liddy ED, Bradley J, and Wheatley JA. Modeling public health interventions for improved access to the gray literature. *J Med Libr Assoc*. 2005. Oct; 93(4):487–94.
- [20] Jim Henderson. Google Scholar: A source for clinicians? *CMAJ*. Jun 7, 2005; 172(12): 1549–50.
- [21] Sharif SZ, Benejaimal SA, Sontrop JM, Lansavichus AV , Haynes RB, Weir MA, Gargt AX. a comparison of PubMed and Google Scholar for quick clinical searches. *J Med Internet Res*. 2013 Aug 15; 15(8):e164. doi: 10.2196/jmir.2624.
- [22] Gardner S, Eng S. Gaga over Google? Scholar in the social sciences. *Library Hi Tech News* 2005; 8:42–45.
- [23] www.knalij.com. [Web document]. available on: <http://knalij.com/index2.html?search=SEARCH#>; last accessed on 21-07-2014
- [24] Giglia V. Quertle and KNALIJ: searching PubMed has never been so easy and effective. *Eur J PhysRehabil Med*. 2011 Dec;47(4):687-90.
- [25] How to Use Go PubMed. Searching is now sorted. [Web document]. Available on <http://www.gopubmed.com/web/gopubmed/www/GoPubMed/Search/index.html#>.

AUTHOR(S):

1. Dr. Subhash B Thakre,
2. Dr. Samir H Golawar,
3. Dr. Sushama S Thakr,
4. Dr. Avinash V Gawande

PARTICULARS OF CONTRIBUTORS:

1. Associate Professor, Community Medicine, Government Medical College, Nagpur, India.
2. Assistant Professor, Community Medicine, Government Medical College, Nagpur, India.
3. Associate Professor, Community Medicine, Indira Gandhi Government Medical College, Nagpur, India.
4. Assistant Professor, Community Medicine, Government Medical College, Nagpur, India.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:

Dr. Subhash B Thakre,
Associate Professor, Community Medicine,
Government Medical College, Nagpur, India.
Ph : 9822366532
E-mail : drsubhasht@rediffmail.com

FINANCIAL OR OTHER COMPETING INTERESTS:

None.

Date of Publishing: Dec 01, 2014