Comparing the Precision of Information Retrieval of MeSH-Controlled Vocabulary Search Method and a Visual Method in the Medline Medical Database

Nadjla Hariri[1], Somayyeh Nadi Ravandi[2]

[1.] Associate Professor, Department of Library and Information Science, Islamic Azad University, Tehran, Iran
[2.] Ph.D., Office Head of the supervision and Evaluation of Research Plans, Kashan University of Medical Sciences, Kashan, I.R. Iran

Corresponding Author:
Somayyeh Nadi Ravandi, PhD.,Office Head of the supervision and Evaluation of Research Plans, Kashan University of Medical Sciences, Kashan, Iran. Tel: +98.3155542999, Fax: +98.3155575057, Email: nadi_so@kaums.ac.ir

Abstract
Background: Medline is one of the most important databases in the biomedical field. One of the most important hosts for Medline is Elton B. Stephens CO. (EBSCO), which has presented different search methods that can be used based on the needs of the users. Visual search and MeSH-controlled search methods are among the most common methods. The goal of this research was to compare the precision of the retrieved sources in the EBSCO Medline base using MeSH-controlled and visual search methods.
Methods: This research was a semi-empirical study. By holding training workshops, 70 students of higher education in different educational departments of Kashan University of Medical Sciences were taught MeSH-Controlled and visual search methods in 2012. Then, the precision of 300 searches made by these students was calculated based on Best Precision, Useful Precision, and Objective Precision formulas and analyzed in SPSS software using the independent sample T Test, and three precisions obtained with the three precision formulas were studied for the two search methods.
Results: The mean precision of the visual method was greater than that of the MeSH-Controlled search for all three types of precision, i.e. Best Precision, Useful Precision, and Objective Precision, and their mean precisions were significantly different (P <0.001). Sixty-five percent of the researchers indicated that, although the visual method was better than the controlled method, the control of keywords in the controlled method resulted in finding more proper keywords for the searches. Fifty-three percent of the participants in the research also mentioned that the use of the combination of the two methods produced better results.
Conclusion: For users, it is more appropriate to use a natural, language-based method, such as the visual method, in the EBSCO Medline host than to use the controlled method, which requires users to use special keywords. The potential reason for their preference was that the visual method allowed them more freedom of action.

Keywords: Medline, controlled method, MeSH, visual method

1. Introduction

In the last 20 years, medical informatics has attracted increasing attention and its popularity and has grown considerably (1). This growth has facilitated biomedical research, which has the aim of developing knowledge and using it in the diagnosis, prevention, and treatment of diseases (2). The result of such growth is an unprecedented increase of information in the medical sciences and the creation of millions of databases for storing, organizing, managing, and presenting information to users. One of the most important databases is the Medline database which

is made available through different hosts. For example, more than 50 sites provided access to Medline in 2003, the most important of which were GratfullMed, Infotrieve, Dimdi, and BioMed Net (3). The number of such sites has increased in recent years. Some other hosts also provide Medline by subscription, and users who pay for the subscriptions receive more services than those who do not pay. Among these hosts are Ovid and Elton B. Stephens CO. (EBSCO) Host. At present, Medline, Biomed, and Pubmed, which are systems that provide access to Medline, have become a necessary part of researchers' efforts, and they are being used increasingly by physicians and patients as necessary tools for answering clinical questions (4). However, an important question concerns the extent to which a database, such as Medline, has fulfilled users' expectations in its ability to reply to clinical questions.

Some time ago, Medline, which is extraordinarily popular among researchers, included more than 19 million records from almost 5000 publications. However, in recent years, it has expanded such that it had 26 million records in January 2011 that include biomedical and health information starting in 1950 and going up to the present time, and it has been used for more than 2 million searches (4). It can be claimed that it has reliable and documented sources in almost all medical fields, but questions have arisen concerning 1) whether the existing search methods provided by different hosts of Medline really direct users to their desired sources and 2) whether the success rate and the rate of acceptance of the information provided are comparable among the different search methods. Looking at each one of these host sites and databases of Medline indicates that the different search methods use different search techniques and equipment. Thus, the quality of the searches, with their different strategies, specifications, and updating distance, are different for each site, and researchers must try to decide which methods can provide more accurate information in the shortest period of time (3).

This problem is also accompanied by a more important concern, i.e., what methods and strategies provide better help and provide better and more accurate information as the volume of information continues to increase. Thus, users have the difficult decision of choosing among the large variety of services provided by these sites and hosts, especially when special hosts, such as Ovid and EBSCO, provide a lot of information quickly. For example, EBSCO provides four different search methods, i.e., Simple Search, Advanced Search and Medical Subject Heading (MeSH), and Smart Search in the Medline database and it has added the visual method to this database in recent years by adding clustering algorithms in information retrieval. The important question that researchers must consider is which of these methods and strategies is the best and the most accurate. Each of the methods has advantages and disadvantages and has special characteristics from the user's viewpoint. Perhaps, one cannot definitely determine whether one of these methods is preferred over the others and can fulfill the needs of users completely. However, some of the methods are more popular among users due to the types of algorithms that have been designed based on their own special needs and characteristics.

There are two classic criteria that are considered in such an evaluation, i.e., recall and precision. Recall means the ratio of the number of retrieved results relating to subject of search to documents relating to the subject in the entire database, and, generally, it is regarded as the success rate of information retrieval, which cannot be evaluated in the real world of databases due to impossibility of determining all documents relating to a given subject in a database. Precision also refers to the ratio of the number of the related search results to the total retrieved results, and it shows the acceptability of the evaluated documents and is measurable considering special conditions. The determination of precision can clarify the efficiency and effectiveness of these methods if it is determined by real users of information and their judgments at the time they were searching for real information they needed. The use of the MeSH-Controlled Vocabulary Search Method is available in the Medline database and in all of its hosts, and it has been used by researchers for years. The visual search method also is used by users in the EBSCO host. The researchers in this research decided to compare these two methods to determine which was more precise from the viewpoint of real users of Medline. One of the methods is based on the control of words and terms, and the other is based on the classification and clustering method in terms of precision classic criterion. Of course, the researchers made calculations to assess the two methods and also determined the three types of precision for each method based on the Best Precision, Useful Precision, and Objective Precision formulas, and the two methods were compared on the basis of these results.

The results of this research also clarified the comparative efficiencies of these methods for designers of different search methods in databases, and they also can help users choose the search methods that provide the best results for their specific needs. Based on their searches, the researchers concluded that the precision of the search methods in the medical databases had not been evaluated to date based on the three types of precision, i.e., Best Precision,

Useful Precision, and Objective Precision, which makes the results of this research more valuable. The following hypotheses were examined in this study:

  1- Best Precision of the retrieved sources in the Medline medical database (in the EBSCO host) in the visual search method is different from that of the MeSH-Controlled Method.

  2- Useful Precision of the retrieved sources in the Medline medical database (in the EBSCO host) in the visual search method is different from that of the MeSH-Controlled Method.

  3- Objective Precision of the retrieved sources in the Medline medical database (in the EBSCO host) in the visual search method is different from that of the MeSH-Controlled Method.

2. Materials and Methods
This research was conducted using a semi-empirical method, and two search methods were compared, i.e., the visual search method and the MeSH-Controlled Method using EBSCO as the host. Considering the precision that the researchers expected for any method and using the ratio comparison formula for both groups (5), a sample size of 287 was obtained for each method. To collect the information, first, 12 training workshops were held for students of higher education in three different majors, i.e., 1) nursing (intensive care nursing, geriatric nursing, and medical/surgical nursing), 2) health information technology, and 3) microbiology and parasitology. The participants were taught the two search methods of Medline.

To prevent bias towards the search results of one of the methods, the trainees changed the training order of the search methods in the different workshops and taught both methods in some cases. After training, the students searched using both methods in the desired fields and judged the title, abstract, and full text of the article (if available) in the first 20 findings. They even considered whether links were available in the articles to other related articles and scored them based on the relevance of the articles to the subject of interest, using the related checklists in Table 1. The users judged 300 searches for each method, and a total of 7101 titles of articles and abstracts and 3232 full texts of articles were studied and scored by the students. The reason for studying only the first 20 findings was that the users usually tended to review a small part of the findings of each search (between the first 10 to 20 findings), which were referred to in different research reports, such as those of Gwizdka and Chignell (6). At the end, researchers were asked to express their view about the methods. After collecting the students' checklists, scores were determined for each search, and then the precision of each search was determined using the Best Precision, Useful Precision, and Objective Precision formulas (Table 2). These formulas were taken from the work of Gwizdka and Chignell (6), which was based on the model proposed by Salampasis, Tait, and Bloor (7).

According to Gwizdka and Chignell' sprecision formulas, in the first formula (best precision), the numerator is the number of articles were awarded a score of 3 in every search (Table 1). In the second formula (useful precision), the numerator is the number of articles that were awarded a score of 2 in every search (Table 1). In the third formula (objective precision), the numerator is the number of articles that were awarded a score of 1 in every search. The denominator in all three formulas was 20because the first 20 searches were evaluated in every search. For the mean precision in all three formulas, the numerator is sum of the precisions obtained in the two methods, and the denominator is 300(the total number of searches in both methods). The mean average precision (MAP) for the 300 searches was calculated for both methods. To answer the research hypotheses, the results were entered into SPSS 11.5 software and analyzed using the independent Sample T-Test. The opinions of the people who conducted the searches about each method also were studied.

Table 1. Description Subjective Relevance Scores by Gwizdka and Chignell (6)

Description	Relevance score
The most relevant	3
Partly relevant or contains a link to a page with score of 3	2
Somewhat relevant, for example, short mention of a topic within a larger page, technically correct (i.e., terms appear on a page – including metatags) or contains a link or links to a page ranked 2	1
No relevance, no query terms found (Metatags were examined as well) or a "bad" hit	0

Table 2. Formulas of Best, Useful, and Objective Precisions and Mean Precision

Formulas of mean average precision (MAP)	Pre Best (minFnHits) = $\dfrac{count\_of\ (score = 3)}{minFnHits}$
	Pre Use (minFnHits) = $\dfrac{count\_of\ (score \geq 2)}{minFnHits}$
	Pre Obj (minFnHits) = $\dfrac{count\_of\ (score \geq 1)}{minFnHits}$
Formulas of three types of precision provided by Gwizdka and Chignell (1999)	MAP of Best = $\dfrac{\sum preBest\ (minF20Hits)}{N}$
	MAP of Use = $\dfrac{\sum preUse\ (minF20Hits)}{N}$
	MAP of obj = $\dfrac{\sum preObj\ (minF20Hits)}{N}$

3. Results

The comparison of the averages of Best Precision in the visual search method and in the MeSH-controlled search method showed that best precision of visual search method was greater than that of the MeSH-controlled search method (0.1388 versus 0.1142). Statistical analysis of results using the independent sample T Testate the 99% confidence level showed that the mean precision of the first method was significantly different from that of the second method (P <0.001). The mean precision averages (MAPs) of both methods are shown in Table 3.

Our research findings indicated that the mean average of useful precision in the visual search method was equal to 0.3758, and, for the MeSH-controlled search method, the value was 0.3313. The independent sample T Test in the confidence level of 99% showed that there was a significant difference in useful precision between these two methods (P <0.001). Table 3 gives the mean average of useful precision and the mean precision percent for both methods. The comparison of the averages of objective precision of both search methods showed that the mean precision of the visual method (0.6040) was greater than that of the MeSH-controlled method (0.5380). The comparison of these two means indicated that there was a significant difference between the means of the precisions of the two methods (P <0.001)(Table 3).

Generally, the research findings indicated that60% of the researchers preferred the precision of the visual method of information retrieval for all three types of precision over the precision of the MeSH-controlled method. However, 65% of the researchers indicated that the control of keywords in the MeSH-Controlled Vocabulary Search Method gave them a clearer view about the keywords that should be used for conducting searches, so they were able to select more suitable keywords. Fifty-three percent of the researchers indicated that the use of a combination of the MeSH-Controlled Vocabulary Search Method and the visual method provided better results.

Table 3. Averages of precision of Visual search and MeSH-Controlled methods

Method	Mean of best precision	Mean of useful precision	Mean of objective precision
Visual	0.01388	0.3758	0.4060
MeSH-Controlled method	0.1142	0.3313	0.5380

4. Discussion

Studies conducted by the researchers didn't reach any conclusion for finding the studies that were conducted with the present research method in the medical databases and measured three types of precision in each search. It can be

said that the present research is different from other research efforts in this field from two perspectives: first, many studies have been conducted by specialists using fully predetermined subjects and questions to determine the precision of the search methods, such as the research efforts that were referenced later, but, in the present research, the participants were ordinary users of databases, i.e., students, and the subjects they searched were interesting to them or were based on their need for information. However, the traditional formula used to determine precision was used, but it has some problems. In the present research, the researchers used Best Precision, Useful Precision, and Objective Precision formulas that were presented by Gwizdka and Chignell (6) and applied by them to study the precision of two search engines, i.e., Hotbot, and Altavista. Ozel also designed a new kind of search method based on metadata, and he compared Best Precision, Useful Precision, and Objective Precision of these two methods (8).

The absence of similar studies made the discussion and comparison of results difficult. For this reason, researchers have relied on making comparisons of the results of their research with the results obtained by other researchers who measured the precision rate using the traditional technique. In this research, we studied the precision of two search methods from the users' perspective. The first search method used the Medical Subject Heading of MeSH combined with vocabulary, and the second method was the visual method, which is a cluster search method in which searching can be done by combining the keywords of users and the user interface-proposed clusters of EBSCO. Findings of the present research confirmed all of the three research hypotheses, and, as the researchers expected, the visual method in EBSCO Medline had significantly greater precision than the MeSH-controlled search method. The mean of the objective precision obtained in the MeSH-Controlled method was 53.8%, which was in good agreement with the results of previous researchers in this field, i.e., Poulter, Rubin, and Altman(9), who showed precision rates of 44% and 50% in the cut-off points of 10 and 5 times, while the results of research conducted by Saka et al. (10) showed that the precision of search in Pubmed, one of the free versions of Medline that uses the MeSH-Controlled search method, was 68%, while it was 70% in the ordinary (natural language) search method. The research results of Saka et al. (10) showed that there was no significant difference between the precision rates of the two methods they studied.

Research by Zohour, Asadi Garakani, and Sarabi (2003) showed that the use of the natural method was more precise than the MeSH-Controlled method. Research by Hersh & Hickam (11), which was conducted to study the precision and recall of the sources searched in Medline using three search methods, i.e., the MeSH-Controlled search method, textual indexing, and indexing based on words and events showed that textual method was more effective than the other two methods. The reason Kostoff et al. (12) obtained such research results for the MeSH-Controlled method was that a lot of the conceptual information presented in Abstracts was ignored by the indexers in the MeSH method; although a subject heading of the MeSh method may be perfect, it cannot be expected to present all of the concepts of interest to the potential users. Saka et al. (10) also referred to these problems and expressed the belief that authors should try to use suitable keywords for their articles, use standard terms, and prepare their Abstracts more precisely. It was concluded in all of these efforts that have been referenced that the MeSH indexing process should be improved so that it could perform better, because many keywords used in the MeSH-controlled method and for searching don't fulfill all of the users' needs. Of course, this does not mean that the use of the MeSH vocabulary is not suitable for searching or that searchers should not use it, because each search method has its own advantages. For example, in the present research, 65% of researchers believed that the use of MeSH helped them have a clearer understanding of the subject of the search so they were able to use more suitable keywords for their searches. Thus, it can be concluded that, in most cases, the use of combined methods, rather than a special search method, can give better results.

5. Conclusion
This research showed that visual method in EBSCO host had better performance than the MeSH-Controlled method. Then findings of this research are important because, unlike the traditional method, three types of precision were measured, i.e., Best Precision, Useful Precision, and Objective Precision, for the first time in a medical database from the perspective of real users. This new method can provide an improved approach for the evaluation of methods that are used to search databases. Future research should focus on comparing this type of precision with different searching methods available in Medline database or other medical databases. The results of such research can lead to the improvement of the evaluation of the methods used to search databases, and they can help designers produce better, more appropriate search algorithms.

Conflict of Interest:
There is no conflict of interest to be declared.

Authors' contributions:
Both authors contributed to this project and article equally. All authors read and approved the final manuscript.

References:
1. Chen H, Fuller SS, Friedman C, Hersh W. Medical Informatics: Knowledge Management, Data Mining, and Text Mining in Medical Informatics. New York: Springer, 2005
2. Cohen A. A Survey of Current Work in Biomedical Text Mining. Brief Bioinform. 2002 March 6(1):57-71. Doi: 10.1093/bib/6.1.57
3. Zohour A.R, Asadi Garakani F, Sarabi Ershad R. Evaluation of Five Free Medline Sites in Internet. JRMS. 2003 Sep [Cited 2013 Jun 14];8(3): [about 1 p]. Available from: http://journals.mui.ac.ir/jrms/article/view/2975 [Abstract]
4. Wang J, Cetindil I, Ji Shengyue, Li Ch., Xie X, Li G, et al. Interactive and fuzzy search: a dynamic way to explore MEDLINE. Bioinformatics. 2010 July 12; 26(18):2321-7 Doi:10.1093/bioinformatics/btq414
5. Mousavi Jarrahi, AR; Katoozian, H. Consulting Research Software. Supplied in Vice Chancellor of shahid Bheshti Medical University. 2001 Version1
6. Gwizdka J, Chignell M.Towards Information Retrieval Measures for Evaluation of Web Search Engines. Available: http://www.imedia.mie.utoronto.ca/files/webIR_eval1_99.pdf
7. Salampasis M, Tait J, Bloor C. Evaluation of information-seeking performance in hypermedia digital libraries. Interacting With Computers 1998 10: 269-84. Doi: 10.1016/S0953-5438(98)00010-1
8. Ozel, SA. Metadata-Based And personalized Web Querying. [Dissertation]. [Ankara]: Bilkent University; 2004.137p
9. Poulter GL, Rubin DL, Altman RB, Seoighe C. MScanner: a classifier for retrieving Medline Citation. BMC Informatics. 2008 Feb 19; 9(108). doi: 10.1186/1471-2105-9-108. PMID: 18284683
10. Saka O, Gulkesen KH, Gulden, B, & Koçgil OD. Evaluation of Two Search Methods in PubMed; the Regular Search and Search by MeSH Terms. Acta Informatica Medica . 2005; 13 (4):180-3.
11. Hersh WR, Hickam DH. A Comparison of Retrieval Effectiveness for Three Methods of Indexing Medical Literature. Am J Med Sci. 1992 May;303(5):292-300. PMID:1580316
12. Kostoff RN, Block JA, Stump JA, & Pfeil KM. Information content in Medline record fields. Int J Med Inform. 2004 Jun 30; 73(6):515-27. PMID:15171980