Connecting Medical Informatics and Bio-Informatics
R. Engelbrecht et al. (Eds.)
ENMI, 2005

Evaluation of Two Search Methods in PubMed; the Regular Search and Search by MeSH Terms

Osman Saka[a], Kemal Hakan Gülkesen[a], Birsen Gülden[b] Oya Deniz Koçgil[b]

[a]*Akdeniz University, Antalya, Turkey*
[b]*Middle East Technical University, Ankara, Turkey*

Abstract

PubMed is the most frequently used medical indexing database by health and biological sciences professionals all over the world. MeSH (Medical Subject Headings) is NLM's (National Library of Medicine) controlled vocabulary used for indexing articles for PubMed. The users usually perform a regular PubMed search, but they can also directly use MeSH terms in their search. Searching by MeSH terms may be more efficient than the regular PubMed search. The aim of this study is to compare the regular PubMed search with search by MeSH terms.
We searched five subjects using the two methods. The results of each search were evaluated by both qualitative and quantitative methods. Results belonging to each subject have been evaluated by five experts to give a satisfaction score and a relevancy score for each method. Mean satisfaction scores that are given by experts were, 3.24 for the regular PubMed search, and 3.72 for the search by MeSH terms. Mean relevancy scores were 3.36 and 4.14 for the regular PubMed and the search by MeSH terms respectively. The quantitative analysis showed that sensitivity was 0.60 in the PubMed search, 0.50 by the MeSH search, whereas precision was 0.70 versus 0.68.
The difficulty of designing such a study is clear, and further studies in this domain needs to be performed. A comparison by qualitative and quantitative analyses showed no dramatic difference for the two search methods. These results may be more easily explained for PubMed searches, but one of the main aims of the MeSH system is to prevent the complications of a routine PubMed search. Our expectation was higher sensitivity and precision as a result of the search by MeSH terms. The results suggest that, there is a problem with indexing procedures. Authors must find appropriate keywords for their article, try to use standard terms and design their abstracts carefully. Reviewing the indexing process in NLM to see if there is a problem may also be helpful. We think that the quality of MeSH indexing must be improved to provide better functionality.

Keywords
PubMed; Medical Subject Headings; Information Storage and Retrieval

1. Introduction

Modern medical knowledge mainly grows with the help of scientific research. The number of research papers increases every year and it is impossible to follow them without an indexing database. A few decades ago, the most widely known medical index was paper based Index Medicus. Computerization of this database has resulted in Medline, which has

dramatically decreased the time needed for searching. The real revolution was PubMed, which is the web counterpart of Medline. PubMed evolves with additional features, and it is free and reachable from every access point on the Internet all over the world. Today, PubMed is the most frequently used medical indexing database by health and biological sciences professionals.

MeSH (Medical Subject Headings) is NLM's (National Library of Medicine) controlled vocabulary used for indexing articles for PubMed. In other words, each article in the PubMed database is indexed according to the MeSH terms. The MeSH terminology provides a consistent way to retrieve information that may use different terminology for the same concepts. When the user performs a search in PubMed, synonyms and the MeSH headings are automatically added to original search terms, and the user is usually not aware of this process.

A user can also directly use the MeSH terms in his/her search. This type of search is a bit more complicated and most of the searchers do not know about searching directly by the MeSH terms. However, searching by the MeSH terms may be more efficient than the regular PubMed search.

The aim of this study is to compare the regular PubMed search with the search by MeSH terms. We tried to get information about the advantages and disadvantages of each search type, to guide researchers about their search strategy.

2. Methods

We randomly selected ten researchers from our medical faculty. They were asked the subject of the last research that they have done using the PubMed, and the exact method and phrases used in their search. As a result, we obtained ten subjects to search. By the help of the original search terms that were used by researchers, two different authors separately determined the optimum search terms for the subject for the search by the MeSH terms. If there was an inconsistency between the selected terms, a consensus was reached by the help of a third author. Five of the ten searches were eliminated because of a confusion in search terms, or very similar search subjects. If the regular PubMed search or the search by MeSH terms produced over 60 results, the number of results was reduced to less than 60 by using a time limit. The reason of this operation was to obtain a reasonable number of titles for qualitative evaluation.

2. 1 Quantitative evaluation

Each article title in the list was classified as relevant or non-relevant, based on the question of “If you would want to see the studies related to the subject, is this article may be accepted as relevant to the subject?” The classifications were made by three authors/experts for each article and the majority of opinions were accepted. Some additional searches with different phrases were performed to find more titles that were missed by our searches. The results of all of the searches were examined, and the number of total relevant articles was determined.

By these processes, we obtained some measures about features of each search subject; (1) Number of total relevant articles. (2) Number of articles which were not indexed (the articles cannot be found by the MeSH search because indexing procedure needs time after the article is entered to database, so some new articles are unindexed yet). (3) Number of total articles which were obtained by the regular PubMed search. (4) Number of relevant articles which were obtained by the regular PubMed search. (5) Number of missed articles by the regular PubMed search. (6) Number of total articles which were obtained by the search by MeSH terms. (7) Number of relevant articles which were obtained by the search

by MeSH terms. (8) Number of missed articles by the search by MeSH terms. As a result, some indexes, namely *sensitivity* and *precision* were obtained from these figures (Table 1).

Table 1 - Definition of "sensitivity" and "precision".

	PubMed	**MeSH**
Sensitivity	Number of relevant articles which were obtained by the regular PubMed search/Number of total relevant articles	Number of relevant articles which were obtained by the search by MeSH terms/(Number of total relevant articles-number of articles that are not indexed)
Precision	number of relevant articles by the regular PubMed search/number of total articles obtained by the regular PubMed search	number of relevant articles by the search by MeSH terms/number of total articles by the search by MeSH terms

2.2 Qualitative evaluation

The results of the regular PubMed and MeSH term searches for each subject were printed. Results belonging to each subject have been evaluated by five experts. The experts were not the same five people for every subject, and a total of 12 experts evaluated the results. They were given a questionnaire to give a satisfaction score and a relevancy score of one to five for both the regular PubMed search and the MeSH terms search results.

2.3 Statistical analysis

The reliability coefficients were calculated by an established method [1] for qualitative scores of each subject. If the reliability was under 0.7, the expert with lowest part-whole correlation was excluded from the study, and an additional expert was asked to evaluate the results.

3. Results

Details of selected five couple of searches are presented in Table 2.

3.1 Quantitative analysis

The results of quantitative analysis are presented in Table 3. Using the data in this table, some indexes were produced (Table 4).

3.2 Qualitative analysis

The mean satisfaction scores that were given by experts were, 3.24 for the regular PubMed search, and 3.72 for the search by the MeSH terms. The mean relevancy scores were 3.36 and 4.14 for the regular PubMed and the search by the MeSH terms respectively (Table 5).

Table 2 - Details of each PubMed and MeSH search

No	**The Subject**	**Regular PubMed search**		**Search by MeSH terms**	**Extra search(es)**
		Original search	**Details**		
1	Use of laparoscopy in prostatectomy	"laparoscopic radical prostatectomy" "prostate cancer"	(laparoscopic[All Fields] AND radical[All Fields] AND ("prostatectomy"[MeSH Terms] OR prostatectomy[Text Word])) AND "prostate cancer"[All Fields] AND ("2003"[PDAT] : "3000"[PDAT])	"Prostatic Neoplasms"[MeSH] AND "Prostatectomy"[MeSH] AND "Laparoscopy"[MeSH] AND ("2003"[PDAT] : "3000"[PDAT])	laparoscop* radical prostatectomy

2	Effect of melatonin on arteries in coronary bypass surgery	bypass melatonin	bypass[All Fields] AND ("melatonin"[MeSH Terms] OR melatonin[Text Word])	"Coronary Artery Bypass"[MeSH] AND "Melatonin"[MeSH]	1. pass melatonin 2. coronary melatonin
3	The current situation in treatment of pheochromocytoma	pheochromocytoma AND treatment	(pheochromocytoma[Text Word] OR phaeochromocytoma[Text Word] OR "pheochromocytoma"[MeSH Terms]) AND ("therapy"[Subheading] OR "therapeutics"[MeSH Terms] OR treatment[Text Word]) AND Review[ptyp] AND ("2003"[PDAT] : "3000"[PDAT])	"Pheochromocytoma"[MeSH] AND "therapy"[Subheading] AND Review[ptyp] AND ("2003"[PDAT] : "3000"[PDAT])	pheochromocytoma therapy
4	Do cellular telephones cause brain cancer?	"brain cancer" phone	"brain cancer"[All Fields] AND phone[All Fields]	"Brain Neoplasms"[MeSH] AND "Cellular Phone"[MeSH]	1. phone malignancy brain 2. phone brain tumour
5	Telemedicine applications for diabetes mellitus patients.	telemedicine applications for Diabetes	("Telemedicine"[MeSH Terms] OR telemedicine[Text Word]) AND applications[All Fields] AND ("diabetes mellitus"[MeSH Terms] OR "diabetes insipidus"[MeSH Terms] OR Diabetes[Text Word])	"Telemedicine"[MeSH] AND "Diabetes Mellitus, Type II"[MeSH]	1. home care applications for Diabetes 2. Diabetic AND telemedicine NOT "diabetic retinopathy"

Table 3 - The results of quantitative analysis

Subject No	Total number	Not indexed	Regular PubMed search			Search by MeSH terms		
			N of found	N of related titles	N of missed titles	N of found	N of related titles	N of missed titles
1	68	14	33	31	37	54	43	11
2	8	2	8	8	0	2	2	4
3	10	-	35	10	0	25	9	1
4	35	0	8	7	28	12	12	23
5	21	0	17	7	14	17	4	17

Table 4 - Sensitivity and precision of each search

Subject No	Regular PubMed search		Search by MeSH terms	
	Sensitivity	Precision	Sensitivity	Precision
1	0.46	0.94	0.75	0.80
2	1.00	1.00	0.33	1.00
3	1.00	0.29	0.90	0.36
4	0.20	0.88	0.34	1.00
5	0.33	0.41	0.19	0.24
Mean	0.60	0.70	0.50	0.68

Table 5 - The mean scores given by experts on satisfaction and relevancy of the searches

Subject No	Regular PubMed search		Search by MeSH terms		Reliability
	Satisfaction	Relevancy	Satisfaction	Relevancy	
1	3.8	4.2	4.2	4.0	0.74
2	4.2	4.0	2.8	5.0	0.79
3	2.2	2.2	3.4	3.4	0.77

4	3.0	3.8	3.6	4.2	0.74
5	3.0	2.6	4.6	4.2	0.79
Mean	3.24	3.36	3.72	4.16	-

4. Discussion

The number of medical articles rises each year, and by the help of the information technologies, the researchers can find the articles which they are interested in. National Library of Medicine (USA) has a free web service that assists searching the medical article database for several years. This web service is known as the PubMed, and widely known by a large number of researchers all over the world. It is possibly the most frequently used database for medical subject search.

Scientific literature is not disinterested in the PubMed [2], and some reports advising better search methods using the PubMed are published [3, 4]. However, we are aware of only one study that systematically evaluates PubMed searches [5]. We think that it is because of the difficulty of designing a study that evaluates an application that used by a wide spectrum of people for a wide spectrum of purposes. The design of the present study was really troublesome trying to solve a lot of confusions about the study method, and led to long discussions between the authors.

A statistical comparison was not performed because of the low number of subjects. A comparison of qualitative and quantitative analyses showed no dramatic difference for the two search methods (sensitivity: 0.60 versus 0.50, precision: 0.70 versus 0.68). However, these figures are not very satisfactory for researcher. A researcher can reach about half of the articles by a search, and he/she must see one non-relevant article for two relevant articles. In other words, a regular search using the PubMed or a MeSH term search can find 50 of 100 articles about the subject, and the researcher must see an additional 25 article to find this 50 article. In another study search by MeSH terms provided higher precision than the regular PubMed search (0.66 and 0.47, respectively) with a lower sensitivity (0.78 versus 0.88) [5].

These results may be more easily explained for PubMed searches. A well known problem for text based databases is the presence of some synonyms, or use of different terms for the same entity. In some cases of PubMed search, an algorithm adds synonym terms to search spontaneously. For example, in the PubMed search of subject 3 (table II), the original search term is “treatment”, and the subheading of “therapy” was automatically added to the search terms. This function is very valuable for the user, and prevents missing some articles. On the other hand, this function does not work for every term. The search with the subject 4 contains “brain cancer” term, and “brain neoplasms” term is not automatically included in the search.

The PubMed search also produces some undesired titles. This is a common problem for the text based search, because the search is not semantically structured and the use of the same word in different context can produce positive results.

In fact, one of the main aims of the MeSH system is to prevent the complications of the routine PubMed search. Our expectation was higher sensitivity and precision by the MeSH search. The results of the search by the MeSH terms are not completely disappointing, but the presence of a functionality problem is clear. A search by the MeSH terms has no clear advantage over the regular PubMed search according to our results. The logic behind the design of the MeSH system is clear and defensible, but the results suggest that, there is a problem with indexing procedures. In a study, simultaneous use of textword and indexing terms retrieved only 82/107 (77%) papers [6]. Authors do not define their study with correct terms in their abstracts, and sometimes do not give appropriate keywords [6, 7].

Authors need to improve the quality of abstracts to make retrieval and screening of relevant papers more effectively and efficiently [8]. Over one million articles enter the database each year, and because of wide area of medicine and complexity of contemporary scientific research, indexing the articles is difficult for the team of indexers in NLM. Authors must find appropriate keywords for their article, try to use standard terms and design their abstracts carefully. Reviewing the indexing process to see if there is a problem may also be helpful. The workload of the people who perform indexing, and the workflow of indexing process may need to be checked for further improvement. However, because of the large number of articles that must be indexed, a higher quality indexing process may need a substantial investment. The ultimate and best solution may be the development of intelligent algorithms that analyses semantic relation in texts.

5. Conclusion

In this study, we tried to evaluate the sensitivity and precision in the regular PubMed and the MeSH term searches. A significant difference could not be observed by both qualitative and quantitative methods. These results may be more easily explained for PubMed searches, but one of the main aims of the MeSH system is to prevent the complications of the routine PubMed search. The MeSH indexing does not seem to be very reliable, and the situation must be evaluated by further studies in this domain. We think that the quality of MeSH indexing must be improved to have a wider acceptance and better functionality.

6. Acknowledgements

The authors wish to thank Dr. Buket Cinemre and Dr. Uğur Bilge for revising the English of this article. We also wish to thank Dr. Sadi Özdem, Dr. Cahit Nacitarhan, Dr. Coşkun Usta, Dr. Burhan Savaş, Dr. Hakan Bozcuk, Dr. Türker Köksal, Dr. Mustafa Usta, Dr. Erol Güntekin, Dr. Uğur Bilge, Mr. Özgür Tosun, Dr. Özlem Batu, Dr. Neşe Zayim for their assistance in qualitative evaluation. This research is supported by Akdeniz University Research Foundation.

7. References

[1] Friedman CP, Wyatt JC. Computing reliability coefficents. In: *Evauation Methods in Medical Informatics*. New York: Springer-Verlag, 1997; pp. 115-117.

[2] Kostoffa RN, Blockb JA, Stumpa JA, Pfeil KM. Information content in Medline record fields. *Int J Med Inform*. 2004; 73: 515-527.

[3] Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying Diagnostic Studies in MEDLINE: Reducing the Number Needed to Read. *J Am Med Inform Assoc*. 2002; 9: 653–658.

[4] Montori VM, Wilczynski NL, Morgan D, Haynes RB. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ*. 2005; 330: 68.

[5] Jenuwine ES, Floyd JA. Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals. *J Med Libr Assoc*. 2004; 92: 349-53.

[6] Derry S, Kong Loke Y, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. *BMC Med Res Methodol*. 2001; 1: 7.

[7] Murphy LS, Reinsch S, Najm WI, Dickerson VM, Seffinger MA, Adams A, Mishra SI. Searching biomedical databases on complementary medicine: the use of controlled vocabulary among authors, indexers and investigators. *BMC Complement Altern Med*. 2003; 3: 3

[8] Dijkers MP. Searching the literature for information on traumatic spinal cord injury: the usefulness of abstracts. *Spinal Cord*. 2003; 41: 76-84.

Address for correspondence

Dr. Osman Saka, Akdeniz Universitesi Tip Fakultesi, Biyoistatistik AD, Antalya, Turkey.
saka@akdeniz.edu.tr