An Assessment of the Visibility of MeSH-Indexed Medical Web Catalogs through Search Engines

P. Zweigenbaum, Ph.D.,[1] **S.J. Darmoni, M.D., Ph.D.,**[2,4]
N. Grabar, M.Sc.,[1] **M. Douyère,**[2] **J. Benichou, M.D., Ph.D.**[3,4]

[1] STIM, DSI, Assistance Publique – Paris Hospitals &
Département de Biomathématiques, Université Paris 6, Paris, France

{pz,ngr}@biomath.jussieu.fr   http://www.biomath.jussieu.fr/

[2] Computer and Networks Department, Rouen University Hospital, Rouen, France

[3] Biostatistics Unit, Rouen University Hospital, Rouen, France

[4] Rouen Medical School, Rouen, France

{stefan.darmoni,magaly.douyere,jacques.benichou}@chu-rouen.fr

Manually indexed Internet health catalogs such as CliniWeb or CISMeF provide resources for retrieving high-quality health information. Users of these quality-controlled subject gateways are most often referred to them by general search engines such as Google, AltaVista, etc. This raises several questions, among which the following: what is the relative visibility of medical Internet catalogs through search engines? This study addresses this issue by measuring and comparing the visibility of six major, MeSH-indexed health catalogs through four different search engines (AltaVista, Google, Lycos, Northern Light) in two languages (English and French). Over half a million queries were sent to the search engines; for most of these search engines, according to our measures at the time the queries were sent, the most visible catalog for English MeSH terms was CliniWeb and the most visible one for French MeSH terms was CISMeF.

INTRODUCTION

Finding accurate health information on the Internet is not so easy; therefore, there is a profusion of catalogs (*e.g.*, Yahoo, www.yahoo.com) and search engines (*e.g.*, Altavista, www.altavista.com) on this new media.[1] In a previous study, we demonstrated that manually indexed catalogs are less sensitive but far more specific than search engines.[2] Using a search engine is however the first recourse for most users; as a matter of fact, in a catalog such as CISMeF[3] (www.chu-rouen.fr/cismef), more than half the visitors are addressed from a search engine (Google, Yahoo, AltaVista, Lycos, etc.). This raises several questions, among which the following: what is the relative visibility of medical Internet catalogs through search engines? In other words, when a user queries a search engine for medical information, which catalog is s/he most liable to be addressed to?

Several tools have been distinguished for accessing health information on the Internet,[3] among which quality-controlled subject gateways with a thesaurus, *e.g.*, the UMLS (Unified Medical Language System) metathesaurus or the MeSH (Medical Subject Headings) thesaurus. The MeSH is used to organize information in the following health catalogs:

- CISMeF (Catalog and Index of French-language Health Resources), Rouen University Hospital, France (www.chu-rouen.fr, www.cismef.org or doccismef.chu-rouen.fr);[3]

- CliniWeb, Oregon Health Sciences University, USA (www.ohsu.edu/cliniweb);[4]

- DDRT (Diseases, Disorders and Related Topics, www.mic.ki.se/Diseases), Medical Library and Medical Information Center, Karolinska Institute, Stockholm, Sweden;

- HON (Health on the Net Foundation, Switzerland) (www.hon.ch);[5]

- MedWebPlus USA (www.medwebplus.com);

- OMNI (Organizing Medical Networked Information UK) (omni.ac.uk).[6]

The objectives of this study are to measure and compare the relative visibilities of these six catalogs through different search engines. To characterize the search for medical information related to these MeSH-indexed catalogs, we used MeSH terms in English and in French as search queries. In this purpose, we designed and ran a web robot that sends queries for every MeSH term to designated search engines and analyzes the results. More than half a million queries were sent, mainly in early August 2001.

We first specify the search engines we queried and the terms that were used. We then describe the design of the web robot, the experimental setting for running it and the principles of analysis of its results. We present a summary of this analysis, then discuss its implications and limitations.

MATERIAL AND METHODS

Material
Four search engines were used to assess this visibility: AltaVista (www.altavista.com), Google (www.google.fr), Lycos (fr.lycos.com/vachercher) and NorthernLight (www.northernlight.com). These tools are among the most widely used search engines.[7] Queries can specify the language of the target pages (*e.g.*, English, French) or leave it unconstrained. For some engines, the number of 'hits' by page can also be specified. We used their 'advanced' interface in order to perform this parameterization. Among the six catalogs, MedWebPlus is the only commercial catalog. The other five are academic.

The MeSH thesaurus[8] contained 19,771 terms in its 2000 version. A total of 40,891 English MeSH terms (including 21,120 synonyms) and 28,922 French MeSH terms (including 9,151 synonyms) were thus used as queries.

Web Robot Design
The ranking of Internet medical catalogs was obtained by designing and running a robot that sends the selected search engines a query for each MeSH term and analyzes the results returned. The query is the exact MeSH term, surrounded by double quotes, so that it is handled as one expression by the search engine (when applicable) rather than as independent words. As is common knowledge now for Internet users, the search engine returns a page of results, which contains the first N 'hits' (generally, $N = 10$; each hit is an URL (Universal Resource Locator)) as well as a follow-up URL that will retrieve the page with the next N results. The robot collects these first N hits, together with their rank (from 1 to N), and then accesses the follow-up URL and its batch of additional hits until the target total T is reached. Silverstein et al.[9] showed that on a million queries submitted to AltaVista, 95.7% of the users did not look beyond 30 results, so that we chose $T = 30$.

Analysis of the Hits
We identified the 6 catalogs of our study by matching these hits to their 'root URLs', as listed in the Introduction section. The output for each MeSH term is a list of the catalogs found in the hits, associated with their best ranks for this search engine. We counted the number of terms for which a given catalog is ranked first (1), in the top ten (1–10) or anywhere within the 30 hits. To aggregate the ranks $r(c_i, e_j, t_k)$ of a catalog c_i, obtained with a given search engine e_j, over all MeSH terms t_k, we defined a linear score $lin(c_i, e_j) = \sum_{t_k} (31 - r(c_i, e_j, t_k))$ and a logarithmic score $log(c_i, e_j) = \sum_{t_k} (\ln 31 - \ln r(c_i, e_j, t_k))$.

We compared the linear and logarithmic scores with another criterion, the Web Impact Factor (WIF). WIF was defined in 1998 by Ingwersen.[10] Absolute WIF (*aWIF*) is defined as the number of external pages that have at least one hyperlink to the given site, after the exclusion of internal links. To estimate the absolute WIF, we sent the following type of query to AltaVista: q_1 =`+link:domainname.countrycode -url:domainname.countrycode`; *e.g.*, the formula used to calculate the HON WIF is `+link:hon.ch -url:hon.ch`. The relative WIF (rWIF) is a normalized value defined as the abolute WIF (aWIF) divided by the number of pages indexed by the search engine. Using AltaVista, the two queries needed to measure the relative WIF are q_1 and q_2 =`+url:domainname.countrycode`, the result being $rWIF = \frac{q_1}{q_2}$. The existence of a correlation between the catalogs and the search engines use was assessed by using the Pearson and Spearman rank correlation coefficients.

Implementation
The web robot was implemented in Perl5 by adapting a former program[11] used to collect text corpora on the Web. Queries are composed and sent in parallel with the help of the LWP::Parallel::UserAgent package (search.cpan.org). URLs are extracted from the result pages with regular expressions, of which Perl provides a powerful set. The formats of queries and regular expressions are declared for each search engine (incidentally, they must also be updated several times a year to follow search engine updates). The programs have been run on Linux. Because of the large number of queries sent (over 600,000), the queries were spread over several days: on 7–9 August 2001 (most of the queries), and then on 6–11 October 2001 (after some technical problems were solved, English queries to Google and NorthernLight).

Analyses were performed using BMDP New System for Windows, Version 1.1 (BMDP Statistical Software Inc) and StatXact software version 3.0.2 (Cytel Software Corporation).

Table 1: Ranks of six catalogs according to English MeSH term hits through four search engines.

catalog	AltaVista					Google				
score	1	1–10	1–30	*lin*	*log*	1	1–10	1–30	*lin*	*log*
CISMeF	4	23	48	844	61	7	56	137	2379	152
CliniWeb	1295	5241	7438	168445	13328	2873	6962	9495	224015	19232
HON	165	1150	1922	39725	2837	424	1469	2260	49059	3821
MedWebPlus	19	237	406	8199	563	101	419	775	15074	1103
DDRT	54	410	763	14819	1029	85	305	573	11349	826
OMNI	227	729	1083	24163	1971	347	901	1311	29310	2453
catalog	Lycos					NorthernLight				
score	1	1–10	1–30	*lin*	*log*	1	1–10	1–30	*lin*	*log*
CISMeF	3	27	54	979	66	2	64	200	2993	174
CliniWeb	1911	4217	5762	136498	11982	537	1471	2028	47883	4098
HON	73	710	1226	24812	1728	0	0	0	0	0
MedWebPlus	17	91	153	3131	231	40	187	293	6265	484
DDRT	112	587	1013	20583	1482	11	192	431	7785	496
OMNI	225	828	1131	26490	2225	814	1960	2328	60310	5591

Table 2: Ranks of three catalogs according to French MeSH term hits through three search engines.

score	1	1–10	1–30	*lin*	*log*
catalog	AltaVista				
CISMeF	35	161	298	5855	423
CliniWeb	0	0	0	0	0
HON	279	1292	1634	40208	3412
	Google				
CISMeF	2764	4549	5304	139635	13578
CliniWeb	167	301	343	9222	872
HON	873	1607	1981	50175	4703
	Lycos				
CISMeF	1240	2290	2780	71243	6753
CliniWeb	0	0	0	0	0
HON	949	1506	1727	46254	4577

RESULTS

Each search engine was sent each MeSH term in English (AltaVista, Google, Lycos, NorthernLight) and in French (the same minus NorthernLight). Two queries were sent for each term: one requesting that the target language be the same as that of the term (*e.g.*, English for an English term) and one leaving the target language unspecified. As a general rule, we observed that constraining the target language results in better scores for the catalogs. We thus only present results obtained with this strategy. Table 1 (viz. table 2) shows the aggregated ranks of each catalog obtained with each search engine for English (viz. French) MeSH terms.

A total of 40,673 English MeSH terms (including synonyms) reached at least one of the four search engines; 40,025 terms (98.4%) obtained at least one answer from one search engine; for 14,441 (35.5%) of the terms, at least one of the catalogs was found in the top 30 hits of at least one search engine. Among them, 11,115 obtained a hit in one catalog, 2,407 in two catalogs, 630 in three, 218 in four, 68 in five, and only 3 terms (*Brain Diseases, Metabolic*; *Deglutition Disorders*; *Hemic and Lymphatic Diseases*) obtained a hit in the six catalogs of this study.

A total of 28,922 French MeSH terms reached at least one of the three search engines, 23,193 (80,6%) obtained at least one answer, and 7,372 (25.6%) were found in a catalog, among which 1,167 in the two French-language catalogs (CISMeF and HON) and none in more. Indeed, if more than the top 30 hits of each search engine had been explored, these numbers would probably have been higher.

We have tested the pairwise correlation between the four search engines, comparing the linear score *lin* (and logarithmic score *log*) of the six catalogs (see table 1). Significant correlations were obtained for all pairs of search engines except those including NorthernLight. As an example, the observed Pearson correlation coefficient and the Spearman rank correlation coefficient between Google and AltaVista were significant: $r_{pearson} = 0.9987$ ($p < 0.0001$) and $r_{spearman} = 0.9429$ ($p = 0.0167$). Furthermore, we have tested the internal correlations within each search engine, comparing respectively the number of first positions in the results with the number of position between 2–10 and with the number of position between 2–30 (see table 1). We found a significant internal cor-

Table 3: Scores cumulated over search engines (AltaVista, Google, Lycos, NorthernLight).

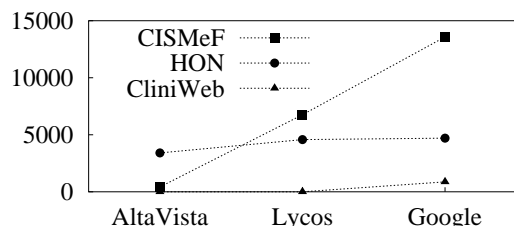| catalog	*lin*	*log*	*#res*	*aWIF*	*rWIF*
CISMeF (En)	7195	453	10190	8675	1.56
CliniWeb (En)	576841	48640	8000	8303	1.37
HON (En)	113596	8386	40000	171971	89.66
MedWebPlus	32669	2381	25000	3136	0.03
DDRT	54536	3833	11000	11081	99.83
OMNI	140273	12240	4894	11559	2.48
CISMeF (Fr)	216733	20754	10190	8675	1.56
CliniWeb (Fr)	9222	872	8000	8303	1.37
HON (Fr)	136637	12692	40000	171971	89.66

Figure 1: Cumulated logarithmic scores (French).

relation for each search engine.

Table 3 cumulates for each catalog the results of the linear (*lin*) and logarithmic (*log*) scores, the number of included resources (*res*), the absolute (*aWIF*) and relative (*rWIF*) Web Impact Factors, for all four search engines (English) and all but NorthernLight (French). Figure 1 shows logarithmic scores for French MeSH terms. No significant correlations of linear score (or logarithmic score) were observed with number of resources included in the various catalogs, nor with absolute and relative WIF.

DISCUSSION

Nearly all English MeSH terms (98.4%) were found by one of the four major search engines examined (AltaVista, Google, Lycos and NorthernLight); however, close to two thirds of the English MeSH terms (and three quarters of the French ones) did not return any of the six major health catalogs studied in the first 30 results.

Variability over search engines. There is a general agreement among three of the four search engines tested, NorthernLight alone having really different results, with the surprizingly complete omission of HON for which we found no explanation.

Language. Results are quite different depending on the chosen language. CISMeF, with pages all in French (although it also mentions the English MeSH term in addition to the French ones), is very low in the English queries. Conversely, all English-language catalogs are low or absent in the French queries. Only HON, being multilingual, is present at a good level in both.

Visibility vs *coverage.* As an important caveat, the visibility of one of these catalogs through a search engine is linked to many factors, among which the coverage of the search engine, its ranking strategy, and the organization of the catalog itself. It does not necessarily entail catalog coverage. For instance, the prominence of CliniWeb for English may be related to the fact that it provides links to PubMed from every MeSH term, whether or not it itself has links to other relevant sites.

Redundancy. Some web sites have several addresses that are aliases or redirections. Some of the engines seem to resolve this redundancy, whereas others (*e.g.*, NorthernLight) consider them as different sites. This can change hit counts dramatically. However, it is hardly surprising that some general MeSH terms such as *accidents* or geographical MeSH terms such as *Quebec* do not generate a hit for health catalogs. A more detailed analysis of the relative visibility of these catalogs for specific MeSH subtrees should thus be performed.

Outside these catalogs. Beyond the six catalogs, some web sites were found to consistently obtain a high rank. They include sites that have the MeSH online (*e.g.*, INSERM, who translated the MeSH into French), or results of MeSH queries for various studies. Again, the fact that a site is often ranked high with MeSH-term queries does not entail by itself its relevance for retrieving high-quality health information.

There is a difference between the visibility as defined here and the reality of usage statistics as measured through an access log. As an illustration, in January 2002, among the visits to the CISMeF site, 102,720 (69%) were referred by Google, 19,849 (13%) by Yahoo, 17,465 (12%) by AltaVista, 5,446 (4%) by Lycos, and 38 (0.02%) by NorthernLight, which is quite different from the CISMeF visibility. This study did not intend to assess the overall quality of main health catalogs available on the Internet but to assess their respective visibility. Nevertheless, coverage should be considered an indirect indicator of quality as the Web Impact Factor, although studies show that the correlation between quality of the site and quality of the medical content itself was still disputed in 2001.[12,13]

This work measures the relative visibilities of health catalogs when accessed through search engines by formulating queries composed of exact MeSH terms. This is consistent with the fact that these catalogs are all indexed with MeSH terms. One can see our experimental setting as modeling a sort of virtual meta-search-engine for MeSH-mediated access to the Web. We assessed the relative positions of six health catalogs in this virtual meta-engine. We made no provision, however, for synonymy beyond that explicit in the MeSH. Therefore, the queries examined can be expected, on average, to be natural for medical librarians, a little less for health care professionals, and less again for the general public. In another study though, examining actual user queries to the CISMeF search engine Doc'CISMeF,[14] we showed that 50% of the words in these queries were present in the MeSH (this figure increases to 85% if frequency of occurrence is taken into account). That study dealt with words rather than full MeSH terms, but it suggests however that our present work on MeSH-term queries might also have some relevance for a large panel of user queries.

This study has some limits and potential biases. One limit is the dependency on the coverage of search engines. In recent studies, the coverage of search engines was estimated at about only sixteen percent for US AltaVista[15] and for Northern Light[7] of the whole set of documents available on the Internet. To limit this drawback, we used four search engines to calculate the visibility of health catalogs. Furthermore, a relatively small set of search engines is responsible for the majority of online searches,[7] among them Google, AltaVista and Lycos, as can be exemplified from the above figures for CISMeF. A second limit is the stability of Web search engines, Selberg and Etzioni[7] showed a difference greater than 40% over one month using 8 search engines, with the exception being AltaVista which changed by slightly over 20%. This means that this study should be repeated over time if one wants to obtain up-to-date figures.

This method also provides insights into medical web catalog referencing by search engines. We have made more recently a similar experiment to compare the relative visibilities of a 'primary resource' (the Orphanet site on rare diseases, orphanet.infobiogen.fr) and various secondary resources including CISMeF, which helped Orphanet make their web site more visible.

REFERENCES

1. Flannery MR. Cataloging Internet resources. *Bull Med Libr Assoc* 1995;83(2):211–5.

2. Darmoni SJ and Thirion B. Indexing the Web? a comparative study of three medical Web servers on the Internet: Cliniweb, "Diseases, Disorders and Related Topics," Omni. In: Proc 1st European Congress of the Internet in Medicine, 1996.

3. Darmoni SJ, Leroy JP, Thirion B, et al. CISMeF: a structured health resource guide. *Methods Inf Med* 2000;39(1):30–5.

4. Hersh W, Ball A, Day B, et al. Maintaining a catalog of manually-indexed, clinically-oriented World Wide Web content. *J Am Med Inform Assoc* 1999;6(suppl):790–4.

5. Boyer C, Baujard O, Baujard V, et al. Health On the Net automated database of health and medical information. *International Journal of Medical Informatics* 1997;47(1–2):27–9.

6. Norman F. Organising medical networks' information (OMNI). *Med. Inf* 1998;23(43–51).

7. Selberg E and Etzioni O. On the instability of web search engines. In: Proceedings of RIAO 2000, Paris, France. 2000.

8. National Library of Medicine, Bethesda, Maryland. Medical Subject Headings, 2001. http://www.nlm.nih.gov/mesh/meshhome.html.

9. Silverstein C, Heizinger M, Marais H, and Morcz M. Analysis of a very large AltaVista query log. Technical report 1998–014, Compaq Systems Research Center, Palo Alto, Ca, USA, 1998.

10. Ingwersen P. The calculation of WEB impact factor. *Journal of Documentation* 1998;54(2):236–43.

11. Grabar N and Berland S. Construire un corpus Web pour l'acquisition terminologique. In: Terminologie et intelligence artificielle, Nancy. 2001:44–54.

12. Hernandez-Borges AA, Macias-Cervi P, and Gaspar Guadardo MA. Can examination of WWW usage statistics and other indirect quality indicators help to distinguish the relative quality of medical Web sites. *Journal of Medical Internet Research* 1999;1(1).

13. Cui L. Rating health Web sites using the principles of citation analysis: a bibliometric approach. *Journal of Medical Internet Research* 1999;1(1).

14. Grabar N, Zweigenbaum P, Soualmia L, and Darmoni SJ. A study of the adequacy of user and indexing vocabularies in natural language queries to a MeSH-indexed health gateway. *J Am Med Inform Assoc* 2002;8(suppl). *To appear*.

15. Lawrence S and Lee Giles C. Searching the Web: General and scientific information access. *IEEE Communications* 1999;37(1):116–22.