Indexing the Web? A Comparative Study of Three Medical Web Servers on the Internet: CliniWeb, 'Diseases, Disorders and Related Topics', OMNI

S. J. Darmoni, MD, PhD,   **B. Thirion,** Medical Librarian,

Rouen University Hospital, F 76031 Rouen Cedex, FRANCE
Tel: +33.232.88.88.29 Fax: +33.232.88.88.32
Email: Stefan.Darmoni@chu-rouen.fr & Benoit.Thirion@chu-rouen.fr

Last update : May, 28 1997

In September 1995, G. Taubes [1] published an article about 'indexing the Internet', regarding the subject of search engines such as Lycos, Yahoo, Webcrawler, Alta Vista and Hotbot. For medical practitioners and librarians, one of the main problems of the Internet is to find organised medical resources because (1) the content is highly distributed therefore difficult to find (2) there are a lot of biomedical servers (N=7000 - Medweb) (3) it is difficult to separate professional-oriented from consumer-oriented in healthcare, (4) the source and quality of the information is of vital importance in medicine.

There is a need for classified resources and furthermore for constructive criticism. In our particular case as the Web server of the Rouen University Hospital (http://www.chu-rouen.fr/), we have organized, like C. Hukins [2], access to medical resources on the Internet for our medical practitioners, specially francophone servers (http://www.chu-rouen.fr/ssf/ssf.html).

Methods

In this study we compare three medical Web servers which index the biomedical Internet resources.

1 **CliniWeb** [3] (http://www.ohsu.edu/cliniweb/), from the Oregon Health Sciences University (USA), is an index for medical clinical resources. It is organised using MeSH terms (Medical Subject Headings from the National Library of Medicine, Bethesda, USA), and has been in existence since 17 October 1995. To conduct a search, one may either browse through the MeSH hierarchy or use the MeSH terms directly.

2 **Diseases, Disorders and Related Topics (DDRT)** (http://www.mic.ki.se/Diseases/index.html), from the Library and Medical Information Centre at the Karolinska Institute (Stockholm, Sweden), which also use MeSH terms, but not only for clinical information. This service also supplies genetics, imaging, etc. This list was a Gopher in 1993 and a Web since May 1995. To conduct a search, the end-user may either browse through the MESH hierarchy or use the alphabetical list of MESH terms.

3 **OMNI** (http://omni.ac.uk/), the Organising Medical Networked Information (UK), is a catalogue of biomedical resources with NLM classification and MeSH terms indexing and it has been in existence since 29 November 1995. It is the work of a UK consortium of seven members: National Institute for Medical Research Library (main institution), British Medical Association, Library, Cambridge University, Medical Library, King Edward's Hospital Fund, Library, Nottingham University, Greenfield Medical Library, Royal Free Hospital School of Medicine, Medical Library, Wellcome Centre for Medical Science. There are two ways to search, either directly using MeSH terms, or by browsing through the NLM classification. There are also three list servers (OMNI, OMNI-all and OMNI-collaborators, accessible by Mailbase, http://www.mailbase.ac.uk/lists-k-o/) which are created in such a way as to facilitate collaboration with OMNI or to provide information about it.

We have not included Medical Matrix (http://www.slackinc.com/matrix/), neither Medweb,

(http://www.cc.emory.edu/WHSCL/medweb.html), Emory University Health Sciences Center Library, which are very important Guides to Internet Clinical Medicine Resources, because they are not indexed with MESH terms.

Results

We introduced the same requests into the three Web servers on October 1996. Seven MeSH terms were used which were selected as being representative of general needs in clinical information research on the Internet:

Alzheimer's disease - Dyslexia - Melanoma - Osteitis - Pain - Pancreatitis - Sciatica

Table 1 - Results

	CliniWeb	Karolinska	OMNI	Alta Vista
	Servers	Servers	Servers	Pages
Alzheimer's disease	29	19	10	20000
Dyslexia	4	6	3	5000
Melanoma	60	5	1	7000
Osteitis	1	0	2	187
Pain	50	12	15	200000
Pancreatitis	48	1	0	2000
Sciatica	0	0	1	800

Discussion

Although we were thinking of ordering these three Web servers, we did not do so for two reasons. Firstly, the creation dates of these Web servers are different from DDRT having begun much earlier. Secondly, comparing the quantity of resources available on these servers, the service is really inadequate.

For example, the Web server 'Overview of Pain & Pain Management Modalities - TALARIA, University of Washington (US)' quoted by DDRT has a very large clinical practice guideline with an index, and very numerous hypertext links to different chapters. The server 'Harvard: Therapy for Bone Pain' quoted by CliniWeb is an article with some figures, but makes no mention of a peer reviewed publication.

Also 'Candid-dementia (Mailbase, UK)', a list server quoted by OMNI for Alzheimer's disease, as well as, the 'Chromosome 14 - OMIM-GDB, John Hopkins University (US)' quoted by DDRT are very different and deal with different types of data.

We can however observe some disturbing differences. In March 96, when we look at main medical Internet resources, such as 'AHCPR Clinical Practice Guidelines (NLM)', we find two guidelines about pain : 'Management of Cancer Pain' and 'Acute Pain Management'. Curiously, these two do

not appear in CliniWeb. 'Cancer pain' and 'Acute pain' are both in OMNI and DDRT. (Talaria is an hypermedia version of the AHCPR Guidelines on Cancer pain). In order to obtain the maximum amount of information, it would seem necessary to look at all three Web servers.

Table 1 illustrates the results of human indexing vs machine indexing (search engines). Human indexing is labor intensive and includes some inconsistensies but with a very low fallout (proportion of documents found among the non relevant ones). Machine indexing has a good recall (proportion of documents found among the relevant ones) but with a very high fallout. In the case of 'Pain', we advise to look at the servers indexed by Cliniweb, DDRT and OMNI and not Alta Vista. In the contrary, in the case of 'Osteitis' or 'Sciatica', as there are very few servers indexed by Cliniweb, DDRT and OMNI , the end-user has to use the search engines.

Although OMNI was the last of these three servers to appear on the Internet, and is therefore not very highly developed, it provides very high quality information: each indexed document contains a title, an OMNI description, MeSH terms and the URL (Uniform Resource Locator). An example is displayed in the Table 2. When one conducts a search in OMNI, one can be confident that all documents chosen for insertion into the database have been validated as being interesting.

Table 2: Description of a Web server (OMNI)

o **Title**: Management of Cancer Pain, (AHCPR) Clinical Practice Guideline Number 9

o **Description**: The Agency for Health Care Policy and Research (AHCPR) is a United States government agency established in December 1989 to enhance the quality, appropriateness, and effectiveness of health care services and access to these services. The AHCPR guidelines are systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical conditions. This guideline is designed to help clinicians understand the assessment and treatment of cancer pain and associated symptoms, and contains a section on HIV positive/AIDS- related pain.

o **Keywords**: pain; neoplasms; pain measurement; acquired immunodeficiency syndrome; HIV

o **Available from**: http:11/10/96/text.nlm.nih.gov/ftrs/dbaccess/capc/

Although our objective was not to try and find the ideal biomedical Web server, we want to emphasize the considerable diversity of medical Web servers organised with a similar classification. We suggest that the following underlines the benefits of each server:

* CliniWeb is an excellent tool for browsing among MeSH tree structures from level to level using only a mouse click. The explode command is not yet available.

* DDRT is extremely impressive with its great range of medical resources as far as we are aware, it was the first MeSH organised Web Server to appear on the Internet.

* OMNI and its description of resources is perhaps the type of Web server that physicians and medical librarians will require in the future: an information base, containing indexed, described and validated resources.

In the near future we plan to compare OMNI, DDRT, CliniWeb, Medical Matrix and any other Web server which will index servers using MeSH terms. In conclusion, we hope that a meta-list or even better a meta-index based on the best lists of biomedical Internet resources which use the same classification may become available.

References

[1] G. Taubes. Indexing the Internet. Science, 269:1354-6, 1995.

[2] C. Hukins and M. Pitcher. Medical library innovates on the Internet. British Medical Journal, 311:1305, 1995.

[3] W. R. Hersh, K. E. Brown, L. C. Donohoe, E. M. Campbell and A. E. Horacek. CliniWeb: Managing Clinical Information on the World Wide Web. JAMIA, 3:273-280, 1996.