Thinking critically about information on the Web

by Betsy Anagnostelis, Alison Cooke, and Alison McNab

This article first appeared in Vine. Originally published as Anagnostelis, B., Cooke, A. and McNab, A. Thinking critically about information on the Web Vine 1997, 104 p21-28, it is reproduced here with permission. The references were checked and found to be correct at the time of submission of the paper.

In addition to the obvious problems of precision and recall in relation to information retrieval via the World Wide Web (WWW), the data retrieved may be inaccurate or out-of-date, even if the sites or resources located initially seem relevant. In response, a number of services have been developed over the Internet which attempt to assist Web searchers in locating relevant, quality resources. Different approaches have been adopted, including the development of gateway services, reviewing resources and the allocation of numerical ratings. However, the effectiveness of the approach, and the quality of the "annotations" used by these sites, can vary enormously.

Using a specific medical WWW resource, a number of services are compared to illustrate some of these issues, to introduce readers to the need to evaluate Web resources, and to provide some guidelines for evaluating the evaluative services. This article forms part of the work by the Advisory Group on Evaluation Criteria for OMNI, the eLib-funded subject gateway for biomedicine. An earlier article compares and contrasts the facilities offered by a wider range of evaluative services (1).

Please note that the article has been mounted unchanged, for historical/legacy reasons, and therefore some of the hyperlinks to remote resources may not work as originally intended

Introduction: a typical scenario?

The Internet provides users of information with the opportunity to access and disseminate a vast quantity and variety of resources, for example databases and datasets, refereed online journals, software, newsgroups and discussion groups. However many users find that their initial attempts at searching the Internet for resources of interest to them can be a bewildering experience. A search engine may return thousands of sites with very little descriptive information to guide the user, who must consequently expend time and energy to follow the links suggested, and sift through the results. Following the links alone can lead the user down numerous blind alleys and dead ends as sites move, resources turn out to be of little interest or relevance, or may contain large graphics which take so long to download that the user loses interest. Recent interviews indicate that medical information users in particular, often do not have the time or perhaps the inclination to explore the Internet in pursuit of resources of interest to them (2). Information seeking is further compounded by the questiononable accuracy of medical information on the Internet today (3), (4), (5).

How might a hypothetical Internet search session run? All too often, someone passing in the corridor may say, "Have you seen such and such resource on the Web? It's on the Home Page of such and such institution - check it out!" Off goes the user to their computer and they begin by opening Netscape and clicking on the Net Search button. Take the example of searching for the AIDS Daily Summary, a news service from the Centers for Disease Control and Prevention (CDC), an agency of the US Department of Health and Human Services. The user may begin by selecting Lycos and searching for "aids" - 7763 hits are returned, displayed ten at a time, with no sign of the required site.

A more discerning user might think about trying to narrow their search. What was that resource again? An AIDS news service? This time, a search on a different tool, AltaVista, for "aids news" retrieves 200,000 hits. The first ten matches include "Study Aids and Reference Material" and "Navigation Aids and Interesting Sites". A number of hits within the first twenty
sound interesting and relevant, such as the "HIV/AIDS Information: Jump Page" and "Global HIV/AIDS Resources". However, clicking on a selection of these links takes the user off on a wild goose chase - including an unhelpful "404 Not Found; the requested URL was not found on this server" message among the first ten hits - and there is still no mention of an AIDS news service.

The problems of recall and precision are not new to information professionals involved with the provision of services to networked information users. As a result of some of these issues, a number of services have been developed which seek to select, evaluate and review Internet sites. These services aim to provide more effective access by using a variety of methods to indicate the quality or usefulness of different sites. Examples include Excite, Magellan (McKinley), Top 5% Sites section of Lycos (formerly known as Point), Infofilter, OCLC NetFirst, OMNI, Six Senses, HealthWeb and Medical Matrix. The different approaches include qualitative reviews, numerical or star rating systems, as well as gateway services which provide access to descriptions of selected, "quality" sites. Some of the services take a subject-specific approach while others are general.

However, if our hypothetical user had decided to try any of these services, would they have been better off? The Centers for Disease Control and Prevention (CDC) mentioned above is a key site for academic users of medical information. Two important medical journals are pointed to directly from the CDC home page: the Morbidity and Mortality Weekly Report (MMWR), and Emerging Infectious Diseases (EID). The CDC home page also links to the pages of eleven CDC Centers, Institutes and Offices, among these the pages of the CDC National AIDS Clearinghouse (CDC NAC). In a recent selective guide to HIV/AIDS Internet resources, the CDC NAC is highlighted as a major (US) government server providing HIV/AIDS information: "It serves as the national [US] referral, reference, and distribution center for HIV/AIDS-related information". The AIDS Daily Summary is a key resource of international interest provided by the CDC NAC, and claims a separate entry in the same selective guide: "It summarises news about AIDS from newspapers and journal articles from around the world" and is therefore likely to be referenced widely from non-US as well as US services.

In order to compare a range of evaluative services, each was used to retrieve the CDC site. The breadth and depth of the descriptions are compared by looking for the MMWR, EID and the AIDS Daily Summary.

**General services**

A number of evaluative services are available from the Net Search screen of Netscape Navigator. Examples include Excite, which has a Web Site Reviews section, the Top 5% Sites section of Lycos and Magellan (McKinley).

A search in Excite Web Site Reviews for "aids" returns 495 hits and "centers for disease control" returns a further 3239 hits. Fifth in the list with an 88% "confidence level" is the Centers for Disease Control and Prevention Home Page. The data displayed includes the title, URL, topics ("Politics and Law / Governments / United States / Federal / Cabinet Departments / Health and Human Services") and a description. From the data given, there is no indication of an AIDS news service, or even that the site provides access to health-related information, other than advice for travellers. By selecting the "more like this" option, 3822 documents are retrieved. The first two documents appear to be two separate descriptions of the CDC Home Page. There's still no mention of an AIDS news service and neither the MMWR or EID are mentioned. Searching directly for "morbidity and mortality weekly report" does not point to the resource in the first ten hits returned other than through a link from a personal home page: "To give you some idea of where he's coming from, one of his "cool links" is to the latest Morbidity and Mortality Weekly Report". Searching directly for "emerging infectious diseases" does bring up a review for the journal from the CDC National Center for Infectious Diseases as "only one of the items here that will thrill some and chill others".
Browsing the Top 5% Sites under "Health and Medicine" followed by "Illness and Disorders", retrieves the Centers for Disease Control National AIDS Clearinghouse. The review is fairly informative, particularly on the coverage of the site, and the annotation specifically mentions a "daily summary of AIDS articles in major US news publications". However, the informal language may be unacceptable to a "serious" or academic user. There is no mention of the link between the site and the US Government (a fact less likely to be known outside the US). The Clearinghouse attracts a rating of 40/50 for content, 34/50 for presentation and 33/50 for "experience", though precisely what the scores indicate is unclear. For example, a user may wish to know why the site scored lower on presentation and experience than content - but the annotation makes no reference to them. Furthermore, there is no indication of when the review was written or the knowledge and expertise of the reviewer. Separate reviews are provided for the "Centers for Disease Control" (sic), and the MMWR: "Our favorite chronicle of things that can kill you...". There is no review or mention of EID.

Magellan offers a fairly full and informative description of the CDC home page and awards it a three star rating. Other information includes keywords, audience, language, costs, and producer and contact information. The language of the review borders on the informal, and there is no reference to the rating in the text. (Magellan reviewers rate sites according to depth, ease of exploration and "Net appeal".) A description is offered separately for the EID, which also attracts a three star rating. However, there are no reviews of the AIDS Daily Summary or the MMWR.

The above services are all informal and aimed at a popular audience, although this is not usually stated. Often with support from commercial advertising, the services are not produced by those professionally involved in the provision of information services, but instead can be described as "cyberjournalism". For example, Excite editors have "seen it all. They've done it all. They're bringing it all to you. The best of the net - from the net's best reviewers." However, there are a number of general services which are produced by information professionals. Examples include Infofilter and OCLC NetFirst.

Infofilter is a service developed voluntarily by librarians. The aims of the project are to develop criteria for evaluating Web resources, and to use those criteria to evaluate carefully selected home pages to produce reviews of sites. The intention is not to evaluate all Web pages, but to "select the most useful pages in particular subject areas and evaluate them much as librarians would print reference sources". The Infofilter volunteers have devised a careful process of producing and evaluating the reviews through discussion on a Listserv, in order to achieve and maintain quality standards. The Centers for Disease Control and Prevention is currently the only medical site reviewed by Infofilter. The review includes the full title, URL to the top level page, site contact details, last update of the site and keywords (uncontrolled). The date of the review and name and e-mail address of the reviewer are included. Reference is made to the MMWR series - though not to EID or the AIDS Daily Summary. Infofilter reviews are generally detailed and cover a range of aspects under the criteria of content, authority, organisation, currency, search engine and accessibility. However, there are fewer than fifty reviews available overall, covering a wide variety of subject areas.

NetFirst from OCLC has been designed as a comprehensive index to Internet resources with summaries. Under a CHEST agreement, NetFirst became available free of charge to the UK higher education community in mid-1996. A user ID and password are required to access the service. A keyword search in the summary field for "aids and news" retrieves an entry for the CDC's (sic) AIDS Daily Summaries. The annotation includes the title, publisher, location, type of resource, subject keywords, language, Library of Congress Subject Headings (LCSH) and Dewey Decimal Classification (DDC) and a summary:

"Presents daily summaries from the Centers for Disease Control (CDC) related to AIDS. Notes that CDC releases abstracts of news and journal articles related to HIV/AIDS each business day. Includes formatted summaries in HTML format and in ASCII. Provides access to the Computerized AIDS Ministries (CAM) FTP library, including an FTP to CAM's daily library and an FTP to CAM's AIDSNEWS library."
It is interesting that NetFirst should point to the Computerized AIDS Ministries BBS link to the resource rather than the CDC pages directly. Separate entries are provided for the CDC home page, the MMWR and EID (though following the link for the latter leads, curiously, not to the resource itself, but to a further description of EID, this time from the WWW Virtual Library Electronic Journals List). NetFirst covers a much greater number of sites than Infofilter (approximately 55,000), although the descriptions are comparatively less detailed.

**Subject-based services**

In addition to the general services discussed above, an increasing number of services are available with a subject focus. Services with a medical focus include Six Senses, OMNI (Organising Medical Networked Information), HealthWeb and Medical Matrix.

Six Senses claims to be produced by health care professionals, and appears to receive no input from information professionals. A search for "CDC" retrieves a review of the "Center for Disease Control" (sic). The annotation includes a score of 21 out of a possible 36; the site therefore does not receive a "Six Senses Seal of Approval" awarded to leading sites in health care, even though it scores consistently highly on "content". Three reviews are available by three named reviewers. The language is very informal, parts of it do not actually make sense, e.g. "The site effect to collect" (sic), and the focus is very much on "coolness" and presentation features, and not on the quality of the content. This is reflected in the criteria used to rate the site - content, aesthetics, interactivity, innovation, freshness and character. No discussion of the numerical scores allocated is available in the text. Neither MMWR nor EID are mentioned, and no reference is made to the AIDS Daily Summary.

Six Senses is similar in many ways to the "popular" services already mentioned. However, OMNI, HealthWeb and Medical Matrix are produced by information professionals or subject experts (or a combination of the two) and offer a more formal and focused service.

OMNI is funded by the UK Electronic Libraries Programme (eLib) and aims to provide a gateway service to selected and evaluated resources in health and biomedicine, of interest to the UK higher education and research communities. Each significant resource is represented independently rather than through a description of the site's home page: individual records, including a detailed factual description, keywords (Medical Subject Headings) and the URL, are available separately for the AIDS Daily Summaries, MMWR and EID. The descriptions are informative and highlight the authority, scope, and coverage of each resource. They can be located by direct searching or by browsing an alphabetical list of National Library of Medicine (NLM) section names. Browsing by Universal Decimal Classification (UDC) or NLM classification codes is also possible.

The HealthWeb project was conceived in 1994, with the following expressed goals: "The health science librarians of the Committee for Institutional Cooperation (CIC) schools will work co-operatively to develop an interface which will provide organised access to evaluated non-commercial, health-related, Internet-accessible resources." By browsing under AIDS, the CDC National AIDS Clearinghouse is identified top of the list, with a factual description of the site and an explicit reference to the AIDS Daily Summaries. A brief entry for the CDC home page appears under Microbiology, mentioning both the MMWR and EID. Brief annotations for each of the resources sought can be found by browsing under Epidemiology. A search option is "under construction".

Medical Matrix is a database of annotated pointers to selected medical resources and is aimed at practitioners involved in patient care. The service is produced by a mix of practitioners and information professionals and is a project of the American Medical Informatics Association. The criteria for the selection of resources include utility, clinical application, quality, currency, and unrestricted access; the content-oriented criteria (peer review and application) are weighted favourably. With such a weighting scheme in place, there is no risk that key resources such as the CDC can be "marked down" purely because they may not employ the latest hypermedia features. Like OMNI, Medical Matrix "tends to
reference specific data resources (such as a particular page in a Web site), rather than general locations. Separate entries are available for all the resources of interest and brief descriptions provided.

Evaluative services: some general comparisons

While review services, such as those discussed here, have the potential to assist users in locating resources, they may themselves be of varying quality and usefulness. When evaluating review services, a number of issues are worth considering:

Aims and intended audience

Two primary considerations in the evaluation of any service must be its aims and its intended audience. Because any consideration of quality is highly subjective, services with a subject and audience focus are more likely to serve the needs of their users by facilitating access to "quality" materials than a service which attempts to be general and all-embracing. Top 5% Sites, Excite and Six Senses may achieve their aims and serve the needs of some users, especially those who may want to know what new, exciting or "cool" features a particular site has to offer, but would be inappropriate for a medical researcher looking for information on AIDS. Conversely, OMNI, HealthWeb or Medical Matrix would assist a user of medical information, but be of little interest to other users. Likewise, Infofilter could not be used in the same way as the medical gateway services because it has its own aims and purposes i.e. to provide in-depth reviews of just a few key resources.

Resource selection

The services vary in the different approaches used for selecting and evaluating sites. HealthWeb and OMNI are fairly specific in their intention to select sites, and only include "quality" resources in their respective databases. Likewise, Medical Matrix seeks only to include resources which assist in "patient care". However, Top 5% Sites, Magellan, Excite and Six Senses are vague in describing the process of resource selection. For example, Magellan aims "for a lively mix of sites, from familiar Internet favorites to the newest of the new, in all our subject areas". Alternatively, Infofilter takes an entirely different approach by reviewing a very select set of resources within a variety of disciplines.

Formality of the services

As has been highlighted, a number of the services are aimed at a popular audience and produced informally. While this may be appealing to a general audience, the informal nature of such services may be unacceptable to a "serious" or academic information user. The nature of services is most clearly identified by characteristics such as the vocabulary and evaluation criteria used. In addition, formal services tend to rely on the involvement of information professionals, not just Internet enthusiasts. The use of numerical rating systems are a feature of many informal or popular services. While the allocation of a numerical value to a site may be a useful indication of its value, it can only be meaningful if an explanation of its allocation is given. This is not commonly the case.

Choice of criteria

An important distinction can be drawn between two types of criteria used to evaluate Internet resources. Services such as Infofilter and Medical Matrix rely on criteria which refer to the content of sources, e.g. authority, accuracy, and currency. The OMNI evaluation guidelines also state clearly: "In general, we are primarily interested in the value of a resource in terms of information content; quality of design or appearance are of secondary interest, even though they may affect the overall usefulness of a resource." Access issues are important as a corollary to content evaluation, and many of the services did mention the need for a particular piece of software to access MMWR and EID. Services such as Six Senses and the general
popular review sites, on the other hand, tend to focus on the "coolness" of sites, as indicated by presentation features.

The level of detail provided in the annotation and the number of sites included

Some sites offer very detailed reviews or descriptions of the sites included, while others offer the minimum of detail. Where less detail is included, a greater number of sites can often be covered. Infofilter provides very detailed reviews; however, with only fifty sites currently covered, it is unlikely that a user could be directed towards this service as a place to search for quality resources within a particular discipline. Sites such as Top 5% Sites and Excite cover many more sites but offer much less evaluative detail. Some of the medical services (OMNI, HealthWeb and Medical Matrix) have achieved a balance - key resources are covered, but informative descriptions rather than evaluative reviews of sites are provided; inclusion in these services alone indicates the quality of a site.

Evaluative services: a way to finding quality on the Internet?

Many of the differences between the services discussed above can be attributed to the involvement of information professionals. A central role of librarians and information professionals is to facilitate access to information. Skills such as evaluating resources, cataloguing and classifying material, and providing meaningful descriptions of resources, have been developed and updated over many years, and for use with a wide variety of resources.

It is no accident, then, that evaluative services involving information professionals can offer effective access to selected and described "quality" materials. Such services tend to be more formal, make use of subject descriptors or keywords, rely on formal and specified evaluation criteria for assessing the quality of material, and provide fairly detailed and informative annotations, descriptions or reviews to enable users to evaluate the relevance of a resource. Examples include OMNI, HealthWeb, Medical Matrix and OCLC NetFirst. Alternatively, services such as Excite and Top 5% Sites, as well as Six Senses, tend to be informal, either do not specify evaluation criteria or rely heavily on "coolness" criteria, and may result in annotations which are unhelpful for the end-user.

These issues raise a further question: what is "information quality"? Information professionals may be comfortable with seeking out information that can be relied on as being accurate, factual or objective, but are still faced with identifying what the user wants and needs to know in order to assess quality. The explosion of Internet resources has resulted in a need for Internet users to think ever more critically in relation to the information they retrieve.

If critical thinking can be defined as: "A readiness to question all assumptions / An ability to recognise when it is necessary to question / An ability to evaluate and analyse"(7) then it is supremely applicable in the networked environment, and extends beyond "the more mechanistic skills of evaluation of resources and mastery of search tools"(8). Direct access to potentially valuable information resources, far from the presence of a librarian, requires that the information seeker is encouraged in developing such skills so that they may be exercised independently.

The development of evaluative services, together with an understanding of their context, can go some way to ensuring that the task of identifying quality networked information resources is less a case of finding a needle in a haystack and more a case of choosing the right haystacks to search!
References


2. Alison Cooke is currently investigating the information seeking and information use behaviour of medical research users of the Internet. The reference is taken from preliminary analysis of recent interviews conducted with thirty users of the Internet.

3. Uretsky, S. Beware of useless or dangerous medical advice online *Internet World* February 1996
   URL: http://www.interne tworld.com/1996/02/medicine.html

   URL: http://www.bmj.com/bmj/arch ive/7054n1.htm

   URL: http://www.mednet-i.com/advil.html


8. Jones, D. Critical thinking in an online world *Untangling the Web: Proceedings of the Conference Sponsored by the Librarians Association of the University of California, Santa Barbara and Friends of the UCSB Library*, April 26, 1996, University Center, University of California, Santa Barbara
   URL: http://www.library.uc sb.edu/untangle/jones.html

URLs of Internet sites discussed

CDC http://www.cdc.gov/
MMWR http://www.cdc.gov/epo/mmwr/mmwr.html
EID http://www.cdc.gov/ncidod/EID/e id.htm
CDC National AIDS Clearinghouse http://www.cdcnac.org/
AIDS Daily Summary http://198.7 7.70.131/cdcnac/cgi/databases/news/adsdb.htm
Lycos http://www.lycos.com/
Alta Vista http://www.altavista.digital.com/
Excite http://www.excite.com/Subject/
HealthWeb http://hsinfo.ghsl.nwu.edu/heal thweb/
Infofilter http://www.usc.edu/us ers/help/flick/Infofilter/
Magellan (McKinley) http://www.mckinley.com/
Medical Matrix http://www.slackinc.com/matrix/
OCLC NetFirst http://www.oclc .org/oclc/netfirst/9352nf/netflyer.htm
http://www.netfirst.ac.uk/
OMNI  
http://omni.ac.uk/

Top 5% Sites  
http://point.lycos.com/categories/index.html

Six Senses  
http://www.sixsenses.com/

Contact details

Betsy Anagnostelis  
Medical Library Royal Free Hospital School of Medicine

Alison Cooke  
Dept of Information & Library Studies University of Wales, Aberystwyth

Alison McNab  
Pilkington Library Loughborough University

Disclaimer

BIOME is a filter-based service that presents a catalogue of descriptions of carefully selected life sciences, health and medical Internet resources.

The information provided should not be used for diagnosis or treatment of any medical condition.

Please note that we can only provide help on BIOME-related services. The BIOME Service cannot give actual medical or health advice or advice on the practices or products that exist within the life sciences sector.

BIOME provides links to Internet resources. We do not guarantee that the information found in an Internet resource found using BIOME is always accurate and complete. This is because we have no control over these resources and their contents. Therefore neither BIOME or any of it's contributors are responsible for any loss, damage or expense that might be caused by any action, or lack of action, that a user of this service might take as a result of reading material on a site found using BIOME. Responsibility for such actions, or lack of actions, remains with the reader concerned.

Likewise, whilst BIOME endeavours to provide accurate information on all the BIOME website pages, we do not guarantee that these pages will not contain any errors, incomplete or out-of-date information. Therefore, neither BIOME or any of it's contributors are responsible for any loss, damage or expense that might be caused by any action, or lack of action, that a user of this service might take as a result of reading material on any of the BIOME web pages. Responsibility for such actions, or lack of actions, remains with the reader concerned.

© University of Nottingham