

Evaluating Search Engines: Task Force Progress Report

by Margaret (Peg) Allen, MLS-AHIP

from NAHRS (Nursing and Allied Health Resources Section of MLA) Newsletter (March 1998) 18(1):4,6-8.

Numerous ways to search MEDLINE, CINAHL, and RNdex are available via the Internet. Do they all give equally satisfactory results? How can you plan and execute your literature searches to get the best possible results? In an attempt to help librarians and health professionals with these decisions, members NAHRS started an evaluation research project in the fall of 1996, under the umbrella of our Research Committee. The Task Force to Evaluate Search Engines for MEDLINE, CINAHL, and RNdex developed an evaluation tool and 20 typical search questions to use in the evaluations. This tool was developed by conducting a three round Delphi survey. Members used an extensive annotated bibliography compiled by Co-Chair Win Sewell. This bibliography is available on the Dr. Felix's Internet MEDLINE Bibliography website, at <http://www.grhlib.demon.co.uk/medrefs.html>. This site is maintained by Helga Perry, another task force member, and contains links to several articles on evaluating MEDLINE search engines. We ended up with both a features checklist and criteria for rating performance on a scale of 1 (poor) to 4 (excellent), with 0 for not applicable. The features checklist and evaluation criteria were organized into the following categories:

- System description/type
- Basic search functions
- Limit functions
- Thesaurus/vocabulary based search functions
- Display/output functions
- User aids
- Management features

Once the evaluation tool was done, our goal was to collect and compile at least three professional evaluations of each database/search engine combination. Task force members volunteered for search engines, and at least three were assigned to each database search engine combination. In addition to librarian evaluations, two classes of University of Michigan nursing students participated in the evaluations of selected search engines. Pat Redman coordinated the student aspect of the project. One class was graduate students and the other undergraduate. Pat presented the project and some basic search concepts to each class, and arranged for compilation of the results. Students used the same evaluation criteria, with minor modifications. We took out some of the criteria from the longer task force tool, and added some clarification. Students were asked to do three searches, not all 20. The student version of the evaluation tool and the 20 questions are reproduced following this article.

Unfortunately, we never reached the goal of at least three professional evaluations for each search engine. The amount of time required was more than some volunteers were able to handle – one member reported that it took her an average of four hours to learn a search engine, do the 20 searches, and complete the evaluation tool. Since many of the search engines have changed (or gone away with the advent of free MEDLINE from NLM), we do not feel comfortable in publishing what results we have.

The evaluation project was shared with three groups. The methodology was presented as part of a four-hour MLA CE class, Evaluating MEDLINE and CINAHL Search Engines: Meeting the Needs of Librarians and End Users, which I presented July 11, 1997 in Oregon. Helga Perry used the criteria for a hands-on class on search engine evaluation in the UK, with librarians working in pairs to try and use the criteria.

Charts comparing librarian and professional evaluations were part of my presentation "Optimizing Your Literature Searches" on December 2, 1997 in Indianapolis, at the Sigma Theta Tau International Tech Expo, the lead event for the 75th anniversary conference of this organization. Results were presented in the context of understanding the databases and the search engines in order to devise effective search strategies. As I prepared the charts for the presentation, the most striking fact was that librarians tended to rate search engines higher than the students. In other words, we can learn a new search engine much faster than the average nursing student. Also, graduate students had less difficulty than undergraduates did – student ratings in the sample comparison chart are the average of the undergraduate and graduate ratings. Students were required to write up their search experience and reactions the project. These papers showed that most had not used MEDLINE or CINAHL, and many were unfamiliar with computers. Many admitted that they were pleasantly surprised with the potential value of these databases. Generally, they preferred search engines that were forms based or had clear instructions. The following table shows professional ratings contrasted to the student MEDLINE ratings in parentheses. Note that the HELIX free MEDLINE was via SilverPlatter's WebSPIRS; this free MEDLINE is no longer available. Internet Grateful Med required a password when tested, and has been updated since this spring 1997 evaluation. Some students chose to do CINAHL instead of MEDLINE; looking at

two telnet search engines and only one web version. Librarians on the task force looked at eleven MEDLINE search engines and five for CINAHL, so the comparisons below are just a small sample of the effort. Also, HealthGate and HELIX were each evaluated by just one librarian – far short of our goal.

| Criteria | HealthGate | HELIX | Internet Grateful Med |
|--|------------|---------|-----------------------|
| Basic search functions | | | |
| Boolean | 4 (3.0) | 4 (2.7) | 3.0 (3.0) |
| Textword phrase | 4 (3.1) | 0 (2.3) | 1.3 (3.0) |
| Truncation | 0 (2.6) | 4 (2.5) | 0.3 (2.9) |
| Author searching | 4 (2.3) | 4 (2.4) | 3.3 (3.0) |
| Limits | | | |
| Human | 4 (3.4) | 0 (1.3) | 4.0 (3.6) |
| Age groups | 2 (3.3) | 2 (1.1) | 2.7 (3.5) |
| Publication types | 4 (3.3) | 4 (1.5) | 3.0 (3.4) |
| English | 4 (3.5) | 4 (3.5) | 4.0 (3.7) |
| Selected years | 4 (3.7) | 4 (3.5) | 3.7 (3.3) |
| Thesaurus/vocabulary functions | | | |
| Mapping | 3 (1.7) | 4 (2.4) | 3.0 (3.0) |
| Attached subheadings | 0 (2.2) | 4 (2.0) | 1.3 (3.0) |
| Abbreviations/acronyms | 4 (2.7) | 4 (2.5) | 3.0 (2.7) |
| Display/output functions | | | |
| Printing | 4 (3.1) | 4 (3.0) | 2.7 (3.0) |
| Downloading | 4 (1.8) | 0 (2.6) | 3.0 (2.9) |
| User aids & overall performance | | | |
| Front end | 4 (2.6) | 3 (2.4) | 3.7 (3.2) |
| Screens | 4 (2.5) | 3 (2.6) | 3.3 (2.4) |
| Manual/tutorial | 3 (2.2) | 3 (2.3) | 2.7 (2.5) |
| Ease of use | 3 (3.0) | 3 (2.3) | 2.7 (3.0) |
| Speed | 4 (2.9) | 4 (2.4) | 3.3 (3.4) |

Next Phase:

Some of us from the first year (1996-1997) of this project are willing to continue, and I have a few new volunteers. We propose refining the criteria and developing a second set of just ten questions, which might include some from the first phase. We now have a better idea of what types of questions will help the most with the evaluations. However, we need more help! If you would like to help with criteria and question development, please e-mail me at pegallen@tznet.com (e-mail address updated 10/10/01). You should also be willing to do some evaluations. It is my goal to refine the objectives from the July 1997 workshop and turn this process into an online MLA CE class – there is Midwest Chapter interest in sponsoring the idea of online CE, and this might be a good place to start. The concepts are best learned if you can practice applying the criteria, but as Helga learned, a couple of hours is not enough time. Our board meets in March and will discuss a strategy to develop the course.

Finally, I want to thank all the task force members noted below, who contributed to the development of the tool and/or completed evaluations. Your work was appreciated!

Winifred Sewell, Co-chair, MLA Fellow (retired)

Theresa Arndt, Philadelphia College of Pharmacy & Science (moved to University of Nebraska)

Allan R. Barclay, Indiana University School of Medicine Ruth Lilly Medical Library

Judy F. Burnham, University of South Alabama Medical Center Library

June Levy, Cinahl Information Systems & Glendale Adventist Medical Center

Sandra O'Neill, Health Information Service, Bathurst Campus Library Charles Sturt University, Bathurst New South Wales, Australia

Connie Ostrove,

Ohio Department of Health, Bureau of Radiation Protection

Helga Perry, Medical Library, Gloucestershire Royal NHS Trust, UK

Pat Redman, Taubman Medical Library, The University of Michigan

Vislava Tylman, University of Miami, Calder Library

Susan Warner, Library of the Health Sciences, Texas Tech University Health Sciences Center, Lubbock TX 79430

NAHRS/MLA SEARCH ENGINE EVALUATION CRITERIA (student version):

Directions: for each criterion, rate the search engine on a scale of 0-4:

- 0 not applicable (search engine does not offer this option)
- 1 Poor
- 2 Fair
- 3 Good
- 4 Excellent

A. Search engine and database tested:

B. Basic search functions

- ___1. Ease of using basic Boolean operators (and, or, not)
- ___2. Ease of searching for a textword phrase: Are phrases easy to ask for? Or must I follow rules that aren't intuitive?
- ___3. Ease of using truncation feature - can you determine how to search on the root/stem of a word using this system? I.e. diabet: to retrieve diabetes/diabetic
- ___4. Ease of conducting an author search (use your own test question - try your name)

B. Limit functions

- ___1. Ease of limiting search results to humans or animals (MEDLINE)
- ___2. Ease of limiting search results to desired age groups
- ___3. Ease of limiting to selected publication types, for example review articles.
- ___4. Ease of limiting to English
- ___5. Ease of limiting to selected years

D. Thesaurus/vocabulary based search functions

- ___1. Does the software guide you to the right terms and suggest additional terms for full retrieval? If so, how effective is this feature?
- ___2. Ease of using attached subheadings
- ___3. Handling of search vocabulary: abbreviations, acronyms (CBC, BUN), names of organisms, antibodies, etc.

E. Display/output functions

- ___1. Ease of printing of records
- ___2. Ease of downloading records

F. User aids

- ___1. Is it easy to get started on (i.e. "Hello! Welcome to MEDLINE! Press <F1> to begin a search" as opposed to a blank command line and no screen instructions).
- ___2. Clear, easy-to-read screens: Consistent screen layout; clear Instructions
- ___3. Is online user manual/tutorial clear?
- ___4. Is it online context sensitive help clear?

G. Overall performance:

- ___1. General ease of use.
- ___2. Speed: Is the system reasonably fast in executing searches and returning results?

H. Evaluation of searches:

Search your choice of at least three of the following questions in each of the three search engine/database combinations you choose, review first 25 (or all if less than 25) citations retrieved. In first column, indicate number retrieved (in current file for MEDLINE); in second indicate how many of first 25 (or less) were relevant.

| # Retrieved | Relevant in first 25 | |
|-------------|----------------------|---|
| | | 1. Risk of cancer in men after vasectomy |
| | | 2. What is the incidence of tuberculosis in the inner city. |
| | | 3. Adverse effects associated with tamoxifen use in women with breast cancer. |
| | | 4. Use of eye drops in the diagnosis of Alzheimer's disease. |
| | | 5. Umbilical cord blood for bone marrow transplantation. |
| | | 6. Diet therapy in insulin-dependent diabetes. |
| | | 7. How is TENS being used for treatment of pain in patients suffering from various peripheral nervous system diseases? |
| | | 8. Identify review articles in English on drug therapies for multiple sclerosis. |
| | | 9. What is the role of therapeutic exercise in treating patients with osteoarthritis of the hip? |
| | | 10. What effect does hormone replacement therapy have on the risk of developing colon cancer? |
| | | 11. How are percussion, vibration and postural drainage used in treating cystic fibrosis patients? |
| | | 12. What tests can be used to diagnose depression in patients with rheumatoid arthritis? |
| | | 13. Is fluticasone effective when administered to children (aged 2 to 18) with rhinitis? |
| | | 14. What is the QUEST test? Is it reliable? |
| | | 15. Use of behavioral therapy in the treatment of nausea and vomiting in patients receiving chemotherapy. |
| | | 16. Research articles on balance in stroke patients. |
| | | 17. Care of patients receiving total parental nutrition in the home |
| | | 18. Looking for information on downsizing of hospitals & hospital mergers with respect to nursing staff. |
| | | 19. Is therapeutic touch effective? |
| | | 20. Information on end-of-life issues that need to be discussed with HIV patients, including the importance of Advanced Directives. |