

Comparative Analysis of Internet Search Methodology for Electrical Engineering Technology

*Ronald H. Rockland, Ph.D.
Engineering Technology Department
New Jersey Institute of Technology
Newark, New Jersey 07102*

The Internet has become an invaluable tool for research in a wide variety of topics, both for students and faculty. Researching topics for a term paper or gathering new information for next week's lecture can be very time efficient if done properly on the Internet. However, information overload can result due to an overabundance of potential sources, many of which are not effective. Attempting to narrow the information obtained to the most relevant part is a challenge, and can prove very frustrating and time consuming. There is even more of a challenge when technical topics are chosen.

In conjunction with other members of the Electrical Engineering Technology faculty at New Jersey Institute of Technology, ten topics were identified to be searched on the Internet to evaluate various search methods. These topics relevant to courses we are teaching, topics that are being researched, or topics that were used for student presentations in class. Also, a listing of what type of expected information from this search was developed for each topic, so that criterion for "relevant hits" could be made.

Out of 10 topics that were identified by this questionnaire, 4 were randomly chosen to be used for the search criteria. The importance of these phrases is to develop a comparative analysis between the various search engines, rather than to identify which search engine is the best. Short phrases were developed for each topic that would relate to the desired information to be obtained. In addition, these phrases were chosen not to include any boolean logic (using words such as and, or, near) since these would be used for custom searches.

Most indices in other articles on search engines for general information [1,2] review the first ten terms of each search for the number of relevant hits. From the author's experience, this makes sense, since the relevancy ratings by the search engines decrease dramatically for the second page. It is usually much quicker to alter the keywords in the search rather than to search through multiple pages.

Assigning a weighting factor to each of the results developed the index. The problem with relevancy is that different individuals might consider certain material on a home page more relevant than others might. Therefore, a simple relevancy index was established, where a weighting factor of 0 was assigned if the a match: linked to a page, which did not contain any

relevant material; the link was not valid; relevant information might be obtained from a page which was at least two links away from the original match; the material was relevant but minimal, and no link was provided. A weighting factor of 1 was assigned if the page resulted in exactly the material needed or provided a direct link to a second page, as well as possibly providing additional links for other information.

An average of the indices from the four topics was calculated for each search engine, and expressed as a percentage (having all ten articles as relevant would be 100%).

The search engines that were included in this paper include the most popular ones: AltaVista, Excite, HotBot, Infoseek and Lycos. Another way of searching on the Internet is to utilize metasearch sites, which ferret out information from

multiple search engines. A relevancy index for one of these metasearch sites (ProFusion) was developed.

Other methods, such as newsgroups or home pages for organizations such as the IEEE and search engines for magazines (such as Electronic Design or EDN) were explored.

The metasearch engine ProFusion seems to perform best, and would be the best for a first method of searching for a topic. All other search engines performed approximately the same for technical terms, especially when using advanced features such as Boolean logic.

References

1. J. Munro, R. Kwon, "Searching the Net", PC Magazine, Dec. 2, 1997
2. R. Overton, "Search Engines Get Faster and Faster, But Not Always Better", PC World, Sept. 1996
3. C. B. Leshin, "Internet Investigations in Electronics", Prentice Hall, NJ 1997