SYMPOSIUM: PSYCHOLOGY AND THE WORLD-WIDE WEB

Organized by Christopher Wolfe, *Miami University, Oxford, Ohio*

Compilation and use of a World-Wide Web index of cognitive and psychological science resources

SCOTT D. MAINWARING

Interval Research Corporation, Palo Alto, California

What started out as an exercise in learning to write World-Wide Web (WWW) documents has evolved into an index (<http://www-psych.stanford.edu/cogsci/>) of over 750 pointers to WWW sites pertaining to cognitive and psychological science. This paper describes the structure of this index, how it was assembled, and how it is maintained. Information about the amount and type of usage has been gathered from automated logging of accesses to the index, as well as from e-mail solicited from users.

In April 1994, I was assisting the Stanford Psychology Department in developing a site on the World-Wide Web (WWW), a globally distributed hypertext system (for an overview, see Vetter, Spell, & Ward, 1994). In part to learn how to write and locate Web documents, and in part to encourage use of Web resources by the department, I created a small index of pointers to various WWW sites of interest from the perspective of cognitive science, broadly construed. In slightly over a year, this "cogsci" index (<http://www-psych.stanford.edu/cogsci/>) has grown to over 750 pointers and receives over 20,000 accesses per month. In the course of this expansion, the overall structure of the index has remained relatively unchanged, although the way it is maintained changed dramatically from being exploration driven to being request driven. Quantitative and qualitative data regarding amount and type of usage of the index have been gathered through automatic logging of file accesses and e-mail messages solicited from users.

Index Structure

The index is conceptually divided into a top-level table of contents and sections devoted to academic programs, organizations and conferences, journals and magazines, Usenet newsgroups, discussion lists, announcement/distribution lists, publishers, software, and miscellany

(Figure 1). Adjunct documents list additions to the index chronologically and acknowledge individuals who have contributed material to or spotted errors in the index.

There have been a few changes to this structure over time. By August 1994, links to sufficiently many on-line sources of research-related software had been collected to move these out of the miscellany section and create the software section. In July 1995, the top-level index was split into a fast-loading concise version including only primary headings (e.g., Academic Programs) and a longer verbose version including both primary and secondary headings (e.g., Academic Programs together with subsidiary headings for each university listed). The concise version replaced the verbose version as the initial entry point for visitors to the site.

The sections are cross-referenced when appropriate; for example, the entry for the Society for Computers in Psychology (SCiP) in the organizations and conferences section points to entries for a sponsored mailing list and for its associated software archive, under the headings Discussion Lists and Software, respectively. In turn, entries for the mailing list and archive point back to the SCiP entry, allowing users to navigate freely between sections of the index. Being able to provide such cross-references at will was a major motivation for incorporating existing references into the index rather than just pointing to them.

Although conceptually partitioned into one set of sections, the cogsci index is physically realized in several different ways. Externally (as visible to the WWW), it is available as a single large file or as a set of interlinked files. The single large file format is supported to allow users to search the index employing the Find option supported by

I am grateful to the Stanford Psychology Department for providing computer resources for the Web index, and to Interval Research Corporation for supporting the presentation of this article. Correspondence should be addressed to S. D. Mainwaring, Interval Research Corp., 1801-C Page Mill Road, Palo Alto, CA 94304 (e-mail: sdm@psych.stanford.edu).

Figure 1. Structure of the index, showing a few entries (ovals) as examples. Both top-level documents (left column) point to each section of the index (middle column); the verbose version points to the first level of entries within most sections as well. Entries can also point to one another. For example, the entries for SCiP, MacPsych (an e-mail discussion list that it sponsors), and the MacPsych software archives are interlinked.

most browser applications with respect to the currently loaded document. Internally, the various external files are derived from a single text file (using the Unix utilities *make*, *sed*, *awk*, and *cpp*). This allows a single change to the internal file to be propagated automatically to several external files.

The present structure does not easily allow multiple categorization of entries. For example, it might be desirable to treat an entry as a discussion list at one time but as a linguistics resource at another. Another possible structural change that has been suggested would be to break the index into smaller, more easily loaded files. Support for users with slow links and text-only terminals has been a design guideline from the start (e.g., images have so far been entirely avoided).

Assembly and Maintenance

For the initial push in building the index, I searched for WWW resources in a variety of ways. I posted queries to the Usenet newsgroups *sci.cognitive* and *sci.psychology*, research to see what existing resources were known to those communities. The search engines Archie (<http://hooohoo.ncsa.uiuc.edu/archie.html>), Veronica ([gopher://veronica.scs.unr.edu/11/veronica](http://veronica.scs.unr.edu/11/veronica)), and, especially, Web-

Crawler (<http://webcrawler.com/>) were very useful. The National Center for Supercomputing Applications (NCSA) "What's New with NCSA Mosaic and the WWW" page (<http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/whats-new.html>) was valuable, as were the existing topical resource lists, particularly Fehrmann's guide to scholarly conferences in psychology and psychiatry (<http://www.mid.net/KOVACS/CD/S0055S.html>) and Cormicle and Bonario's *Neurosciences Internet Resource Guide* (<http://http2.sils.umich.edu/Public/nirg/nirg1.html>). I also did some general link-following around the Web, a strategy that became increasingly useful as interlinked, relevant Web sites started to achieve a sort of critical mass.

As the index grew in size (through many small increments), it began to be referenced in turn by the resources it was referencing and by some popular Web indices, including Yahoo (<http://www.yahoo.com>) and Global Network Navigator's *Whole Internet Catalog* (<http://www.gnn.com/gnn/wic/>). This publicity, combined with the huge increase in the number of organizations and individuals using the Net, allowed maintenance to change from being an activity of active exploration and testing to one of more passive responding to requests for various Web addresses to be included in the index and for cor-

rections to existing ones to be made. Thus, in a sense the index became self-maintaining.

A literally self-maintaining index would be updated directly by its users themselves. However, in addition to the technical problems of implementing such a scheme, the issue of (lack of) editorial selectivity would be a considerable problem. Psychological and cognitive sciences is already perhaps too broad a topic for a useful index, and I have started to make an effort to avoid including self-help and predominantly clinical resources in order to focus on materials most relevant to basic cognitive research. Striking a balance between openness to alternative conceptions of cognitive science (for example) and avoiding information sprawl is difficult, and a user-driven solution to this problem is difficult to envision.

Usage

Whenever a user accesses a file in the index, the server creates an entry in its log recording which file was accessed and the address of the remote computer being used. The total number of file accesses from May 1994 through October 1995 is shown in Figure 2.

Growth in the number of accesses has come in spurts, with fairly long periods of little or no growth. Usage increased by a factor of 6 over the previous month with the announcement of the page on NCSA's "What's New" list in June 1994. (At that time, NCSA's browser Mosaic was used by most Web users, and its What's New list was a focus of interest.) Following this initial spurt, usage remained fairly level through October 1994. A period of rapid growth followed through March 1995, probably due in part to the index being included in GNN's on-line *Whole Internet Catalog*. Five months of negligible or negative growth followed, ending with large increases in the number of accesses for September and especially October 1995. A seasonal slowdown in academic activity may account in part for the lack of growth in the summers of 1994 and 1995. The reason for the most recent jump in activity is unclear.

A large proportion of all file accesses were to the top-level files, as shown in Figure 3. On average, for every ac-

Figure 2. Total number of file retrievals per month.

Figure 3. Number of file retrievals, by month and file category. The top line represents section files, the middle line top-level files, and the bottom line the index in its single-file format.

cess of a top-level file there were 1.5 accesses to the files constituting the sections of the index proper. This suggests that the average user after accessing the top level looked at only one or two of the nine sections of the index.

Use of the single large file format for the index was negligible until April 1995, and increased rapidly in September and October 1995. The reasons for this increase are unclear, although it may in part be due to the reorganization in July 1995 of the default top-level page that users first see. Prior to this, information about the single-file format appeared after several screen-fulls of the verbose top-level file; with the new concise top-level file, this information appeared on the first screen. The increasing popularity of the larger single-file format may also reflect the use of higher bandwidth Internet connections. It is also possible that this popularity is due to a link to the single-file version having appeared at a popular Web site somewhere.

The index has been accessed from hosts in 78 different Internet domains around the world. (Most countries have a single Internet domain; the United States has several.) A coarse categorization is shown in Figure 4. As would be expected given the location and content of the index, the single largest group of users has been in the U.S. educational domain. The U.S. commercial domain represents the second largest number of accesses. The index receives a substantial number of accesses from Europe, particularly from the United Kingdom, Germany, and the Netherlands.

The relative popularity of the various sections of the index differed substantially (Figure 5). Sections for academic programs, journals and magazines, and organizations and conferences accounted for 60% of all section retrievals. In general, the larger sections were accessed more often, though the relatively small journals and magazines section was accessed almost twice as often as the largest section, discussion lists. Interest in on-line journals and magazines appears to be increasing over time; this section replaced academic programs as the most accessed section in July 1995.

Figure 4. Number of file retrievals according to Internet domain of retrieving host.

Figure 5. Number of file retrievals for each section of the index.

To get a sense of how the index was being used, for a number of months I asked users (in a prologue to each file) to "tell me how you use (or plan to use) the information" in the index. Of the 92 messages received, 10 were noninformative. Of the remaining 82 messages, 46 were from individuals seeking specific information: 15 sought information on clinical or self-help topics, 6 sought various psychometric instruments, 3 sought e-mail addresses of particular individuals or types of individuals, and 22 sought information on a wide range of other topics. The remaining 36 messages were categorized as follows: 14 were from prospective students or job hunters, 9 were from people using the index for general reference or browsing, 8 were from people building other Web sites or Internet guides, and 5 were from writers researching a paper, thesis, or book.

This feedback suggests that the index is accessed for many different purposes by users of quite differing back-

grounds. Users choosing to comment do not represent a random sample, and some categories of usage are most likely disproportionately represented (e.g., people building Web indices). Nevertheless, it appears that the index is most often used for specific queries, a large proportion of which and perhaps the majority of which are on clinical or self-help topics outside the scope of the index. Thus many users of the index may be frustrated. This is hardly surprising considering the rapid outnumbering of academic researchers on the Internet by lay persons. To help address this problem, a clarification of the scope of the index was added to the first screen users encounter, along with links to some clinical/self-help Web sites.

Although most often used for specific queries, the main strength of the index may be as a browseable general reference, especially useful for students searching for undergraduate or graduate programs. Some of the most enthusiastic replies were from people in this category. For example, a user from Finland wrote the following (some spelling and phrasing has been corrected):

This is my very first time on the Internet and I was lucky enough to find the right page to start this interesting journey to the world of psychology. I'm studying psychology in the University of Helsinki and first I [contacted] our department and I [picked] this address from someone's homepage. THIS IS GREAT!!!! Unlimited possibility to learn new things and to know what is going on all around the world. As a beginner I don't know yet how I'm going to use the information I'll get from here, but definitely I'm going to spend some time here learning the possibilities.

Conclusion

At least in its early stages, the Web as a large-scale interactive medium allowed individual effort to be amplified by feedback from the community responding to that effort. The Web provided a supportive environment for the incremental growth of the cogsci index, which after an initial phase of focused development was maintained as a background activity. As the Web continues to expand and evolve, the number of resources relevant to any given topic will increase; some topics will become too large (in the context of the Web) to be indexed by an individual. It is possible that index building will also be made obsolete by advances in automated information retrieval technologies such as WebCrawler. However, there will probably continue to be topics of appropriate generality that could serve as the basis for fairly self-maintaining indices. The Web will become what its users collectively make of it.

REFERENCES

- VETTER, R. J., SPELL, C., & WARD, C. (1994). Mosaic and the World-Wide Web. *Computer*, 27(10), 49-57.

(Manuscript received November 13, 1995;
accepted for publication November 20, 1995.)