

Revenge of the Librarians

December 10, 2009

Originally published May 10, 1996 in Web Review magazine.

I have a confession to make. I'm a librarian. Now, I've never actually worked in a library and I don't wear my hair in a bun, but I do have a Master's degree in *Information and Library Studies* from the University of Michigan's [School of Information](#). In addition, many of my friends are librarians and I'm getting married to a librarian in a couple of months. So you'll understand, when I proclaim that librarians are destined for greatness on the Internet, that I'm a little biased in my opinions.

The greatness of which I'm speaking relates to the development of tools for finding information on the Internet. To date, most of the information retrieval tools on the Net have been designed by people with a computer science background. Don't get me wrong. I don't want to offend any programmers. First of all, the Internet itself wouldn't even exist if it wasn't for the creative talents of programmers from around the world. Second, and perhaps most important, librarians and other non-programmers such as myself depend heavily upon the skills and good will of programmers to carry out our grand schemes in this highly technical environment. We need programmers!

However, anyone who has spent much time looking for information in this chaotic and distributed environment will agree that finding useful information on the Internet is no easy task. Phrases such as *drowning in information* and *finding a needle in a haystack* come to mind. Searching on the Net can be difficult, frustrating, and very time consuming.

You've probably heard the saying that to someone with a hammer in his hand, all problems look like nails. Well, on the Internet, computer science folks tend to hit information retrieval problems over the head with relevance ranking algorithms, intelligent agents and lightning quick processors. [Alta Vista](#) is a perfect example of a fast, powerful and highly automated search tool, and for many types of searches it works wonders. However, the complexity of the query language and the sheer volume of

information can be overwhelming. A search on "*Web Review*" for example, returns 900,000 documents. That's a few too many for me to start sifting through.

Now, there are ways to refine a query, and I'll get to those in a moment, but even the expert searcher can't get around the very real problems of information retrieval that are caused by the ambiguity of language. Without an understanding of the context of a query, it is difficult for a search tool to understand exactly what you're looking for. Do you want shoe polish or Polish shoes? Are you looking for biology resources for the professional researcher or for a class of 6th grade children? Do you only want good information? How do you define "good" information anyway?

I happen to believe that the people best equipped to solve some of these information retrieval problems are librarians. We might not be very strong in the marketing department and we're certainly not very good at turning great ideas into revenue generating products and services, but we have invested a substantial amount of time and energy studying the information seeking behavior of real people in the real world and developing skills and tools to meet those information needs.

To best state my case, I'd like to discuss a loose categorization scheme I've developed for information retrieval tools on the Internet which include automated search tools, Internet directories, and virtual libraries.

Automated Search Tools

Automated search tools comprise the richest and most varied category of tools. These search tools employ software *robots* and *spiders* that crawl the Web indexing everything they find. Examples include [Alta Vista](#) and [Lycos](#). The highly automated nature of these tools allows them to provide access to the most comprehensive indices of Internet resources available. Alta Vista's database, for instance, contains 15 billion words indexed from over 30 million web pages.

However, the weaknesses of search tools also derive from their automated nature. First, since search tools do not employ organization schemes, the onus for sorting the information falls to the user. To refine our search for "Web Review" magazine using [Alta](#)

Vista, we might enter ***title: "web review"*** as our query phrase. This requires a knowledge of the query language and an understanding of the principles of online searching that many novice users lack. Second, because search tools exercise no editorial control over the resources they index, the quality of information varies widely. Search tools make no distinction between "good" and "bad" information. As the volume of information on the Net continues to grow exponentially, both of these problems will become increasingly troublesome. General purpose search tools will not scale well at all.

Internet Directories

Internet directories or collections of resources maintained by the global Internet community are fairly comprehensive and easy to use. [Yahoo](#) is the best example. The idea behind directories is that anyone can add resources to a directory. With Yahoo, for instance, anyone can submit a resource with a brief description. That resource is integrated into a subject hierarchy that is largely designed by the staff at Yahoo. In this way, Internet directories balance central control with distributed independence, thereby melding the efforts of human and machine. With several million potential contributors, the strength of directories clearly lies in their ability to be comprehensive and current.

On the other hand, the weaknesses of directories also derive from their distributed independence. Because anyone can add resources, everyone does. This results in information of varying quality. In addition, the self-submission of resource descriptions and evaluations can be problematic. Who isn't going to hype up their resource, at least a little? A search in [Yahoo](#) on the word "great" returns 523 hits, "excellent" returns 306 hits, and "lousy" only 6 hits. Either Yahoo is packed with amazing high quality resources or we've got a problem with objectivity.

Like search tools, it is difficult to see how these general purpose directories are going to scale over time. Already, Yahoo's subject hierarchy is close to collapsing under its own weight. There are just too many menus and levels in the hierarchy. As Yahoo doubles and quadruples in size, even the search results screens will become unwieldy. With a billion dollars in market capitalization, give or take a few million, I'm sure the folks at Yahoo will come up with a few creative solutions to these problems, but I doubt they'll solve the underlying issues caused by the ambiguity of language. That's where us

librarians come in.

Virtual Libraries

Virtual libraries or value added collections of Internet resources are among the more civilized areas of an otherwise chaotic and unruly cyberspace. Although a far cry from the order and stability of traditional libraries, virtual libraries do provide a taste of the value that librarians can add to the Internet through the application of traditional skills in a vastly non-traditional environment. Through the identification, selection, organization, description, and evaluation of Internet information resources, *digital librarians* or *cybrarians* create virtual libraries which help people to find "good" information.

Examples include the [Argus Clearinghouse](#) and [About.com](#). The topical guides within these virtual libraries can serve as an excellent tool for finding useful information. The human authors have identified, described, evaluated, and organized many of the best resources on the Internet for many specific topics. All the hard work has been done for you. And it's free!

The strength of virtual libraries clearly derives from the process through which real people add value to the raw information environment of the Net. The weaknesses stem from the fact that there are a limited number of guide authors, many of whom have day jobs and maintain their guides as a free service to the Internet community. They simply can't keep up with the volume of information and rapid speed of change. For these reasons, virtual libraries tend to be less comprehensive and less current than directories and automated search tools. Now if we found a way to compensate guide authors for their efforts...perhaps we could dramatically improve the quality of guides and subsequently the effectiveness of virtual libraries as Internet information retrieval tools.

The Rise of Librarians

So what does all this mean? What is the future of searching on the Internet? Why are librarians poised for global domination?

Well, there are a few clear trends that we can see. First, the volume of information on the

Internet is increasing exponentially, doubling every 6 to 8 months. Second, the sheer volume of information is leading people to spend more and more time using information retrieval tools on the Net. They desperately need these tools to find what they're looking for. Third, entrepreneurs and investors alike have recognized these trends. The initial public offerings of companies like Yahoo and Lycos testify to the *gold rush* fever in this area.

In my opinion, the highly automated "one size fits all" approaches to indexing the Internet are doomed to failure as the volume and diversity of resources continue to expand. Instead, people will come to depend upon the audience or subject specific value-added guides to be found in virtual libraries. These guides will cut through the clutter and provide a sense of organization and perspective in a way that no automated tools or intelligent agents ever will.

Since librarians are well suited to the development of these guides, we are positioned to capitalize on the growing demand for value-added information services. Maybe on the Internet, librarians will finally figure out how to make some money while helping to make the world a better place. Then again, maybe not, but at least it's worth a try.