

Searching as a Primary Internet Discovery Paradigm  
Michael F. Schwartz  
University of Colorado - Boulder  
Published in Internet Society News 2(2), Summer 1993

There are two resource discovery paradigms in common use in the Internet: organizing/browsing, and searching. Organizing refers to the human-guided process of deciding how to interrelate information, usually by placing it into some sort of directed graph (e.g., the hierarchy of directories in an FTP file system). Browsing refers to the corresponding human-guided activity of exploring the organization and contents of a resource space. Searching is an automated process, where the user provides some description of the resources being sought, and a discovery system locates some matches.

What is the most appropriate paradigm for supporting resource discovery in the rapidly growing collection of Internet-accessible information? Already the amount of information may conservatively be estimated at one terabyte, and is growing rapidly with the addition of new user communities (such as K12) and applications (such as multimedia mail).

Searching becomes the most feasible approach in large information spaces, because the effectiveness of browsing depends heavily on how well organized the information is. As users generate and collect increasing amounts of data, keeping it all well organized becomes increasingly difficult. In fact, the notion of "well organized" is highly subjective and personal. What one user finds clear and easy to browse may be difficult for users who have different needs or backgrounds. To some extent this problem can be alleviated by systems that support multiple views of information, such as Prospero. Yet, doing so really pushes the problem "up" a level - users must locate appropriate views, which in itself is another discovery problem. Moreover, because there are few barriers to "publishing" information on the Internet, there is a great deal of information that is useful to only very few users, and often for only a short period of time. To other users, this information clutters the "information highway", making browsing difficult.

Beyond the scalability problems of the organizing/browsing paradigm, the importance of searching can be seen in its increasing popular appeal. A number of file system search tools have been introduced in recent years (including NeXT's Digital Librarian, the University of Colorado's Essence system, MIT's Semantic File System, SunSoft's SearchIt tool, and the University of Arizona's Personal Information Retrieval System), and recently the Veronica system was introduced to support searching through Gopher - which is, without a doubt, the world's largest browse-based system. Moreover, it is interesting to note that while the Prospero model focuses on organizing and browsing, the Prospero system is primarily used in its capacity as an interface to thearchie search system.

The above discussion is not meant to imply that organizing information is unimportant. Clearly, a well organized space is easier to use than a poorly organized one. Rather, the point is that once an information space gets large enough, searching is the most feasible option. Ideally, resource discovery systems can combine these approaches into a search-then-browse paradigm, in which users submit search queries and then use the results as starting points to begin browsing. This model is supported, for example, by the Alex file system's "archia" tool, which preforms anarchie search request and returns the results in the form of Alex path names. The user can then change directories through Alex and browse the relevant parts of the FTP information space.