Gerry McKiernan
Iowa State University Library

Parallel Universe: The Organization of Information Elements and Access in a World Wide Web (WWW) Virtual Library

Abstract: For generations, libraries have facilitated access to information sources by the development and use of a wide range of appropriate organizational processes. Within a Web-based demonstration prototype, we have applied several established library procedures, principles and practices to enhance access to selected Internet resources in science and technology. In seeking to manage these sources, we have established defined collection, adopted an established library classification scheme as an organizational framework, and sought to simulate the features and functions of a physical library collection and conventional Reference sources. This paper describes the key components of this prototype, reviews research which supports its approach, and profiles suggested enhancements which could further facilitate identification, access and use of significant Internet and World Wide Web (WWW) resources.

"The future means both collections and access" (Crawford and Gorman, 1995, 181).

1. Introduction

A common critique of the World Wide Web (WWW), and the Internet in general, is that while there are numerous and useful information sources that are available, identification is not as efficient nor effective as many desire. To facilitate access to significant sources, CyberStacks(sm), a centralized, integrated, and unified collection of World Wide Web (WWW) and other Internet resources, that uses the Library of Congress classification scheme as an organizational framework, was formally established on the home page server at Iowa State University in Fall 1995 (McKiernan, 1995).

The idea of creating an organized collection of selected Internet resources using an established classification system originated in a series of listserv and newsgroups postings that began in summer 1994 (McKiernan, 1996a, 2). Many of the issues raised in the associated discussions were coincidentally considered by Britten in his excellent review paper on building Internet collections prepared for the 1994 Charleston Conference (Britten, 1995). His articulation of the professional and institutional role that libraries and librarians have historically played in organizing information provides clear support for the benefit that a 'library-organized' virtual library could provide users seeking authoritative, significant and well-described Net resources:

"[T]he long-term institutional perspectives of libraries, librarians' expertise in information organization, and the library profession's organizational structure . . . would add both collection management expertise and permanence [to such a project]." (Britten, 1995, 247).

Britten's profile of the characteristics of networked information, namely the potential of a...
network 'copy' serving the needs of many libraries simultaneously, and the transient and ephemeral nature of some sites and resources, highlights both the value that a centralized collection such as CyberStacks(sm) might generally offer to research and university libraries, and the potential role that librarians could play in the maintenance of access to these electronic sources (Britten, 1995, 249).

The establishment of a central Virtual Library Collection, coordinated collection development of Web and other Net resources using the Internet itself (Britten, 1995, 247) and the application of a widely-used library classification scheme are among the features derived from established library practices and procedures that have been incorporated within the CyberStacks(sm) model of a virtual library.

2. Selection

CyberStacks(sm) seeks not to be a digital library per se, but a true 'virtual' collection. The majority of resources selected for incorporation within its collection are monographic or serial works, files, databases or search services, and all of its selected resources are full-text, hypertext or, hypermedia, and of a research or scholarly nature. CyberStacks (sm) may indeed be considered the service envisioned by Tyckoson in his outstanding review article on the history and future of the catalog:

"The catalog of the future may not be merely an index to the world's collected information, but may actually provide a summary of the information for the user. A catalog that includes full-text works [emphasis added] on a variety of subjects as well as abstracts and citations to materials may serve as something more than an encyclopedia or index. Whereas the scholar of the eighteenth century used two tools to find out about his or her world, the scholar of the twenty-first century may need only one [emphasis added].‖ (Tyckoson, 1991, 25).

In its selection of resources for incorporation, CyberStacks(sm) has adopted many of the same philosophies and guidelines used in the selection of non-Internet resources by most librarians in the development of local collections, including authority of the source, accuracy of information, clarity of presentation, uniqueness within the total collection, recency or timeliness, favorable reviews, and community needs (American Library Association, Reference Collection Development and Evaluation Committee, 1992).

The foundation of the CyberStacks(sm) collection is a selection practice that seeks to identify, isolate, and select Internet resources that are discrete document-like objects. The philosophy of selection of resources for inclusion within the CyberStacks(sm) scheme has been well-articulated by Demas, McDonald and Lawrence in a recent article:

"Title by title selection of high quality resources is one of the most important values librarians can add in providing access to information resources, including those accessible via the Internet. A careful selection of resources is the touchstone of the electronic library" (Demas, McDonald and Lawrence, 1995, 280).

To facilitate the selection of such resources, in the initial phase of the project, we have intentionally sought to locate Internet resources that are the electronic equivalent or analog of standard Reference works, as well as resources that may be so considered. Abstracting and indexing services, bibliographies, biographical sources, dictionaries, directories, handbooks, guides
to the literature, maps, and standards, are but a few of these publication types. We believe that our intentional selection of sources of a Reference nature that serve to index, abstract, guide or review significant and relevant Internet resources, will mitigate some of the frustrating and distressing experiences of users who seek specific data and information via the Net. In general, through CyberStacks(sm) we have sought to provide the value-added service offered by libraries for generations in the careful selection and description of resources. In particular, we have made a systematic effort to identify and incorporate significant, yet elusive sources from significant Web sites (McKiernan and Ames, 1996).

The hypertext nature of the World Wide Web (WWW) and the various browsers used to navigate through it enable and facilitate the inclusion of an entire collection from a site, or only those selected resources considered most relevant. We recognize, as did Britten (Britten, 1995, 248), that the hypertext functionality of the Web permits one to incorporate selected appropriate portions of a site, or its entire collection, into a virtual library. While CyberStacks(sm) does include entire Web sites, only those that are considered well-organized collections of significant Internet resources are selected for incorporation. Such sites are viewed as either directories or encyclopedias, depending on their structure, organization and content, and are categorized and classified as such. Unlike most other media, an individual resource or site may be classified in more than one broad or specific category within a particular organizational scheme, presenting conventional as well as novel opportunities for enhanced access.

3. Library of Congress Classification System

The Library of Congress classification system is a well-established scheme that has been used by libraries worldwide for organizing a variety of publications and media for generations. Within its schedules, this classification system not only denotes subject coverage and content, but information format and conceptual relationships as well. Underlying the CyberStacks(sm) approach is the use of a hypertext outline of the Library of Congress classification scheme (LC Classification Outline, 1990) that allows users to scan broader or narrower subject categories at will and then, if desired, to connect directly to a selected resource.

While an increasing number of sites are adopting the Library of Congress classification scheme as a system for organizing WWW and other Internet resources, most notably the WWW Virtual Library, specific relationships to other resources are typically not indicated. Generally, selected resources are classified only at a broad level, and only listed alphabetically within particular classified groups. While a listing within a broad category does offer the user some assistance in identifying relevant resources, such an arrangement requires a user to review an entire list to be assured that an appropriate resource has not been overlooked.

Unlike most Web sites that have adopted the Library of Congress classification system, CyberStacks(sm) has applied the scheme at a deep level of specificity and description, incorporating resources within a hypertext framework that allows users to browse a collection of resources in order by an assigned classification number. In establishing CyberStacks(sm), we believed that a more comprehensive application of the Library of Congress classification system could offer the value-added structure, organization, context, and an appropriate level of specificity and description for enhancing the efficiency and effectiveness in identifying and using relevant and related WWW and other Internet resources. Indeed, the structure, organization and access provided by CyberStacks(sm) can be viewed as the convergence, merger and fusion of a subject-ordered catalog and an associated collection, where like materials metaphorically 'stand together' in browsable stacks.
4. Browsing

We believe that the use of a standard organizational scheme provides a framework that facilitates the predictable location of resources within an overall system and substantially lessens the cognitive load and psychological burden often associated with information overload and anxiety.

Within CyberStacks(sm) resources are categorized first within a broad classification, then within narrower subclasses, and finally listed under a specific classification range and associated subject description that best characterize the content and coverage of the resource. While we recognize the need to provide alternative methods of accessing relevant resources to suit the variety of information-seeking strategies of a diverse user population, browsing has been adopted as the primary means of identification and access in the initial implementation phase.

We agree with Chang and Rice that:

"Browsing is common... [W]e all browse in various contexts to make sense of the world around us, such as when we read newspapers, scan television channels, go window shopping, or seek information in libraries" (Chang and Rice, 1993, 232).

In citing the work of Herner, Bates, and others, they recognize the importance of browsing in human information seeking and highlight the need for information scientists to incorporate this approach in their system designs (Chang and Rice, 1993, 232-233).

Their summary of critiques of 'specific, direct searching' supports the need for browsable interfaces:

"[There are] some unrealistic assumptions about users and the nature of information seeking—e.g., that users have unbounded rationality, have static and well-defined information needs, know what they want and are output oriented..." (Chang and Rice, 1993, 233).

The clear articulation of the uncertainty that some users experience in the search process further supports this option:

"Users are often in an anomalous state of knowledge as they initiate a search... do not have predefined search criteria, and may alter their interests during a search..." (Chang and Rice, 1993, 233).

Their observations of end-user searching is particularly relevant to the decision to organize selected resources within a browsable, categorized scheme:

"Many end-user systems are difficult to use because they require training, knowledge of mechanical and conceptual aspects of searching, and a high cognitive load from end users... [A]ll of these may be overcome through browsing..." (Chang and Rice, 1993, 233).

Within CyberStacks(sm), users are presented with a variety of broad, narrow and related categories to browse. From a main menu, users may select from an array of primary Library of Congress classes (i.e., Q, R, S, T, U or V) or their associated subjects (i.e., Science, Medicine, Agriculture, Technology, Military Science or Naval Science). As a service to users, only classes which have subclasses with incorporated resources are hotlinked. Within a class (e.g., Science (Q)), users may select from a table of associated subclasses (e.g., Q, QA, QB, QC, QD), each with
an associated subject or topic (e.g., Science (General), Mathematics, Astronomy, Physics, Chemistry). For any subclass, users are presented with a table containing the range of associated class numbers (e.g., QB 1-139, QB 140-237, QB 275-343, etc.) and their associated subjects (e.g., Astronomy (General), Practical & Spherical Astronomy, Geodesy, etc.). At this latter level, users are presented with an option of browsing any (or all) of the related classes with a main class (e.g., Q, QA, QB, QC, QD) at the bottom of the table of class numbers.

At the top of every subclass table and resource profile is a horizontal menu of broad subject classes that permits users to access any class currently incorporated with the CyberStacks(sm) scheme (i.e., Q, R, S, T, U, and V).

6. Record Format

For each resource incorporated within CyberStacks(sm), a profile of its coverage, content and special features is provided. For each record, when and where possible, we have excerpted the summary data from the original resource to provide sufficient information about its subject coverage, and other features, to enable users to judge a resource's potential usefulness. Our focus is not to analyze a resource and delineate each of its potentially important data elements, but to characterize it. We believe that the meaning and value of a given resource can be conveyed as well, or better, by structure and organization. We believe that the CyberStacks(sm) scheme provides meaning through context; a context that is appropriate and relevant to users while seeking a resource which may satisfy an information need.

Although we have decided not to standardize the format of data within a record, many of the categories recommended for describing a resource within the CyberStacks(sm) scheme are identical or similar in function to the metadata elements of the Dublin Metadata Core Element Set, the Dublin Core. While the Dublin Core seeks to identify and isolate the data elements of an Internet resource, the summary information provided for a selected resource within CyberStacks(sm) is intended to describe the resource only to the extent that users can judge its potential value for meeting an information need.

While we prefer to describe resources within CyberStacks(sm) more holistically than analytically, we understand the desire to identify and define core data elements. Although we do not believe that an exhaustive, analyzed set is necessary for use within our scheme, we believe that a set such as the Dublin Core can provide the optimal level of descriptive and subject cataloging for World Wide Web (WWW) and other Internet resources that will facilitate their incorporation within emergent as well as conventional information and retrieval systems and services. We also do not believe that the holistic approach taken by CyberStacks(sm) and the more analytical approach represented by the Dublin Core are mutually exclusive. Indeed, many of the planned enhancements for CyberStacks(sm) will require the manipulation of delineated data elements to be realized.

7. Planned Enhancements

We recognize that there are inherent limitations in providing a browse function within CyberStacks(sm) with limited visual context. The addition of images and appropriate graphics could "further reduce cognitive load and memory tasks" as summarized by Chang and Rice (Chang and Rice, 1991, 242). Likewise, a "visual system that models a ... library's physical layout on the screen (including visual displays of book shelves [emphasis added], floor maps, and walking paths)"(Chang and Rice, 1991:242) would offer a familiar environment that could further facilitate
identification and selection of relevant resources by users. The Science Library Catalog Project (Borgman et al., 1995) not only has implications for the design of information retrieval systems for children, but offers an excellent model for the enhancement of other hypertext retrieval systems, including the CyberStacks(sm) scheme.

Although not as sophisticated as Borgman's efforts, we believe the recent incorporation of relevant icons to identify all broad subject categories as well as subclasses within the major CyberStacks(sm) classes provides additional visual information and context that will further facilitate selection and use of relevant Internet resources.

The application of virtual reality browsers (e.g., Pesce, 1995) to the CyberStacks(sm) scheme could literally provide the added dimension that would further facilitate use. The landmark experimental project directed by Arthur W. Rhyno, Systems Librarian, University of Windsor, Canada, that has applied VRML to 'navigate information space' at the Leddy Library, represents a major advance in user-friendly interfaces to the local online library catalog (OPAC) and has significant potential application for enhancing CyberStacks(sm).

8. Conclusion

In a recent review article on the pitfalls and promises of the virtual library, (Rooks, 1993, 22-29), Rooks provides a good general critical review of some of the major ramifications which the emergence of the virtual library may have on conventional library services. In concluding her article, she offers a balanced view on the future of libraries and clearly articulates the overall role that librarians should play in their development:

"The virtual library is not the ultimate answer to everyone's information needs. It is merely another step in a dynamic and evolutionary process. The traditional print library and traditional library services will not disappear. But, as librarians, we must accept and adapt to the introduction of new techniques and systems. We must recognize the enormous potential of the virtual library, address the issues involved in its creation, and take a leadership role in integrating these new systems and services into our libraries, for our own good and for our users" (Rooks, 1993, 28).

In creating the Virtual Library, librarians must look beyond individual and local needs, and begin to give fuller consideration to the ways in which our professional skills can be applied not just to the creation and maintenance of a special collection of Internet resources, but to the larger issue of centralized, integrated and unified national collections of Internet resources built through cooperative collection development projects in a networked environment. Within systems such as CyberStacks(sm), all well-managed local efforts potentially will not only benefit a targeted audience, but a broader, world user community as well.

References

