

# Hand-Made in Iowa

## Organizing the Web Along the Lincoln Highway

Gerry McKiernan

Curator

CyberStacks(sm)

Iowa State University, Ames IA 50011

*gerrymck@iastate.edu*

**D-Lib Magazine**, February 1997

ISSN 1082-9873

---

### Introduction

In Fall 1995, [CyberStacks\(sm\)](#), a World Wide Web (WWW) virtual library, was created to investigate the applicability of using an outline of the Library of Congress classification scheme ([LC Classification Outline 1990](#)) as an organizational framework for enhancing identification, access, and use of Internet resources ([McKiernan 1995](#)). Although still in development, CyberStacks(sm) has become a *de facto* Science and Technology Reference collection, with links to several dozen individual, institutional, and organizational homepages on six continents. In addition to using a standard library classification scheme, CyberStacks(sm) has also adopted a variety of conventional and 'neo-conventional' methods for facilitating efficient access to Web resources ([McKiernan 1997c](#)). Its extensive [Title Index](#), a working file of candidate titles for future incorporation within its collection, and its [Cross-Classification Index](#), an alphabetical listing of the subcategories associated with each incorporated resource, irrespective of its specific location within the Library of Congress classification category, are representatives of these two types of features.

Although established to facilitate access to significant Internet resources through the use of these and other established library organizational methods, CyberStacks(sm) has also served as a vehicle for exploring alternative approaches to managing and accessing Net resources. These investigations have included a study of other projects that have applied standard library practices to Web organization, a review of various methods of 'automated categorization' of Net resources, and an extensive survey of Information Visualization technologies and their potential application for enhancing access and use of Web and non-Web documents.

### *Beyond Bookmarks*

While we believed that the adaptation of a well-established library classification system held the potential of enhancing access to Internet resources when we created CyberStacks(sm), skepticism in some quarters concerning the value of using such an approach led us to undertake a review of other projects that had employed conventional library classification systems to organize the Web. In

Spring 1996, we posted queries to several listservs and newsgroups requesting the addresses of sites that used traditional library organizational methods to manage access to Net collections. Although the postings raised further concern about the suitability of this approach, a number of parallel efforts were, indeed, identified in the process.

In addition to documenting the use of such schemes by others, the survey identified a wide variety of different implementations of this approach. Recently, the features and functionalities of selected sites from this study were reviewed and analyzed ([McKiernan 1997b](#)). While each identified site is unique, several have adopted organizational structures similar to those used within CyberStacks(sm): the presentation of the general outline as a base home page, hotlinks to subcategories or subclasses in appropriate subdirectories, and an indication of the associated subject coverage of a class, section, or category.

Some sites have enhanced access to their respective collection of Net resources by the use of unique functions or features, some of which are currently in review as potential enhancements to the CyberStacks(sm) model. One site, notably [EELS](#), the Engineering Electronic Library of the Swedish University of Technology Libraries, not only provides descriptors for every incorporated resource, but also offers access from these subject terms or phrases to other records assigned the same descriptor. This feature allows users to search from within a record to identify similar resources without leaving a relevant item. The EELS site is also distinct in its classification of a selected resource in more than one category, providing access to that resource from more than one perspective.

Projects identified from this study have been reviewed, categorized, and are hotlinked within a special clearinghouse entitled [Beyond Bookmarks: Schemes for Organizing the Web](#). Since its establishment, *Beyond Bookmarks* has become an established gateway to selected and organized Net resources for over one-hundred libraries, individuals and institutions worldwide. In June 1996, it was designated a *Scout Report Selection Network Tool* by the InterNIC Net Scout Project, an effort to identify significant Internet resources of potential value to researchers and educators.

## Cross-Classification

The need to enhance access to resources in CyberStacks(sm) by content has been an interest since its establishment. In recognition of the limits of browsing to identify resources within the CyberStacks(sm) collection and of an expressed desire by users, we have created a separate index that offers users access to incorporated resources through an alphabetical listing of the subcategories associated with each resource. Within this [Cross-Classification Index](#), users can browse topics covered by described resources irrespective of their specific location within the Library of Congress classification scheme. This Cross-Classification Index is considered a prototype for a more advanced index that will allow users to search or browse a structured thesaurus of subject headings. This Fall, we initiated a preliminary review of other efforts that provide resource access through a structured, hyperlinked, controlled vocabulary ([Net Projects 1996](#)). Although we have not yet created this exact functionality or one that is used at such sites as EELS or [INFOMINE](#), we have established enhanced access to incorporated resources through a link from the entry for such items in the CyberStacks(sm) [Title Index](#) to the appropriate classification range within CyberStacks(sm) collection where a specific resource is categorized.

From within this classification range, users can browse through adjoining profiles to identify other resources of potential interest.

### ***Project Aristotle(sm)***

The issue of scalability of the CyberStacks(sm) model has been a general concern and one raised by users since its establishment. While there is strong evidence that conventional library classification systems and controlled vocabularies do offer an organizational framework for effectively identifying and using Web resources, their application in most current environments requires intensive effort to create and maintain. With an interest in expediting the incorporation of selected resources within an organized scheme, a review of efforts that offered automated Web resource categorization was initiated during Summer 1996. This study identified a variety of projects, products, and services that provided a form of automated organization of Web resources at the workstation, system, and network levels. At the workstation level, enhancements include expanded bookmark management and fully automated categorization of resources based upon sophisticated algorithms. At the system level, a number of projects that employ intelligent software agents that traverse the Web in search of products, services, or resources that match a user's interests, or which facilitate access to Web resources by the establishment of narrow and broad concept spaces for focused Web searching, were identified. At the network level, projects such as the [Nordic WAIS/World Wide Web Project \(Ardo and Koch 1994\)](#) and the [OCLC Scorpion experiment \(Vizine-Goetz 1996\)](#) offered features and functionalities particularly relevant to the further enhancement of CyberStacks(sm) at the next stage of its development. Each project has not only endeavored to create enhanced access to Net resources by automatic categorization or organization of Web resources, but also has sought to extend such categorization to automatic classification, a more sophisticated application of existing and emerging technologies.

Projects, research, products and services identified from this study have been profiled in a clearinghouse entitled [Project Aristotle\(sm\) \(McKiernan 1997a\)](#). *Project Aristotle(sm)* has also been designated a *Scout Report Selection Network Tool*.

### **Seeing is Believing**

While the efforts profiled in *Project Aristotle(sm)* offer a range of methods for Web organization that may be adapted to enhance the CyberStacks(sm) model, many of them also provided alternative methods for displaying and accessing resources within a defined collection. Among the most novel in both organization and presentation are the self-organizing semantic maps created through the use of a neural network technique for a [selected collection of Web resources applied by Chen and colleagues \(Chen, Schuffels and Orwig 1996\)](#) at the University of Arizona for the [Digital Libraries Initiative project at the University of Illinois](#). Its display format, as well as those of several others incorporated within *Project Aristotle(sm)*, stimulated an interest in the broader issue of Information Visualization and its potential application for enhancing identification and use of Net resources.

Due to an interest in identifying alternative approaches for navigating Internet resources, a query was posted in Fall 1996 requesting nominations for incorporation within a clearinghouse devoted to visualization of Web collections and documents. From this survey, and from an extensive literature review, a variety of experimental as well as commercial visualization technologies and applications

were identified ([McKiernan 1996a](#)).

Among the better-known products is [HotSauce](#). Originally developed by Apple as Project X, HotSauce makes use of the [Meta Content Format \(MCF\)](#) to present the content structure of a Web site within a three-dimensional interface. By providing a complete overview, users are given an ability to 'fly-through' primary, secondary, and associated links, on demand. In those sites using HotSauce, topics are represented by round-cornered rectangles. By double clicking on a node, the resources linked to a topic are displayed. Color is used to represent a difference in value or relationship. In a more familiar context, [Yahoo!3D](#), provides a VRML-based virtual playground in which a user can walk, fly, slide or spin to identify and select favorite Yahoo! categories represented as appropriate three-dimensional icons. Both approaches hold promise for improving resource identification and access within sites that employ categorization or standard classification schemes to organize Web resources.

Such applications, however, do little in revealing the full semantic content of selected Web resources and the breadth of the relationship of a resource to others within a collection. Although significant, these techniques currently only facilitate access to Web resources through enhancement of established and conventional organizational frameworks. Net resources are treated only as document-like objects to be selected and classified in hand-made categories, no different from physical books and journals.

A number of experimental systems, however, seek to extend access to the underlying content of Web collections by representing them visually. The [SPIRE](#) system developed at the Pacific Northwest National Laboratory and the [HyperSpace/Narcissus](#) project at the University of Birmingham provide the user with an overview of an entire Web corpus and an indication of specific relationships among associated documents through visual images and metaphors. In fully utilizing the electronic format of Web resources, such Information Visualization technologies can be expected to significantly enhance use of the resources identified for CyberStacks(sm) and other digital libraries.

Projects, research, products and services identified from this study have been profiled in a clearinghouse entitled [The Big Picture: Visual Browsing in Web and non-Web Databases](#).

### ***'Use the Force, Luke!'***

The tedium associated with identifying resources for inclusion in the CyberStacks(sm) also prompted an investigation of other methods for identifying candidate resources for the CyberStacks(sm) collection. During the process of reviewing sites for *Project Aristotle(sm)*, a number of efforts employing some form of intelligent software agents for resource identification and organization were identified. A subsequent study of major agent sites ([Adbu and Bar-Ner 1996](#), [Hermans 1996](#)) provided insight into the potential application of this technology for facilitating the identification of appropriate candidates for a Web collection.

In late summer, a query requesting information about the application of software agents for such library services as acquisitions, cataloging and collection development was posted to several library lists and newsgroups. While less than a handful of relevant projects have been identified to date, the proliferation of Web-based services within libraries and the growing interest within the academic

and research communities in agent technology should stimulate their use for these and similar applications.

## Built-By-Hand

These and other sophisticated technologies will no doubt become common-place by the end of this decade. Until they are well established, CyberStacks(sm) will continue to be developed as it has from its inception. Although its construction continues to require a significant investment of time and energy, it offers the rare satisfaction in our digital world of building the future one brick at a time.

## References

**Adbu, D. And Bar-Ner, O.** 1996?. *Software Agents: A General Overview*. Technion, Israel Institute of Technology, Haifa, Israel. [Available at URL: <http://t2.technion.ac.il/~s3180501/agent.html>] 23 November 1996.

**Ardo, A. and Koch, T.** *Automatic Classification of WAIS Databases*. Technical report, Lund University Library, 1994.[Available at URL: <http://www.ub2.lu.se/autoclass.html>] 22 November 1996.

**Chen, H, Schuffels, C. and Orwig, R.** 1996. Internet Categorization and Search: A Self-Organizing Approach. *Journal of Visual Communication and Image Representation* 7(1):88-102.

**Hermans, B.** 1996. *Intelligent Software Agents on the Internet: An Inventory of Currently Offered Functionality in the Information Society & A Prediction of (Near-)Future Developments*. Tilburg University, Tilburg, The Netherlands, July 9, 1996.[Available at URL: <http://www.hermans.org/agents> ] 23 November 1996.

**LC Classification Outline.** 1990. Library of Congress, Washington, D.C.

**McKiernan, G.** 1995. CyberStacks(sm): A 'Library-Organized' Virtual Science and Technology Reference Collection. In *D-Lib Magazine*. [Available at URL: <http://www.dlib.org/dlib/december95/briefings/12cyber.html>] December 1995. 19 November 1996.

**McKiernan, G.** 1996a. Information Visualisation: The World Wide Web Gets Really Graphical! *Intelligence: The Magazine of the Information Age*, Special Edition (1997 Guide to the Internet), 116-118, December 1996.

**McKiernan, G.** 1997a. Automated Categorization of Web Resources: A Profile of Selected Projects, Research, Products, and Services. *New Review of Information Networking*. In review.

**McKiernan, G.** 1997b. Beyond Bookmarks: A Review of Frameworks, Features, and Functionalities of Schemes for Organizing the Web. *Internet Reference Services Quarterly* 2(1/2) 1997. In final review.

**McKiernan, G.** 1997c. The New/Old World (Wide Web) Order: The Application of 'Neo-Conventional Functionality' to Facilitate Access and Use of a WWW Database of Science and

Technology Internet Resources. *Journal of Internet Cataloging* 1(1). In Press [Abstract available at: <http://jic.libraries.psu.edu/jic1nr1.html>] 12 February 1997.

"Net Projects." [<http://www.public.iastate.edu/~CYBERSTACKS/Projects.htm>] 20 November 1996.

**Vizine-Goetz, D.** 1996. Online Classification: Implications for Classifying and Document[-like Object] Retrieval, in *Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington, DC, USA*, (Washington, D.C: INDEKS Verlag, Frankfurt/Main, 1996), pp. 249- 253.

Copyright © 1997 Gerry McKiernan

*Approved for release, February 14, 1997.*



---

*hdl:cnri.dlib/february97-mckiernan*