ED 403 900   .   IR 056 274

AUTHOR   Mitchell, Steve
TITLE   Library of Congress Subject Headings as Subject
                Terminology in a Virtual Library: The INFOMINE
                Example.
PUB DATE   96
NOTE   7p.; A downloaded version of one of 29 conference
                papers (for those in ERIC, see IR 056 258-274) placed
                on the Internet under the following overall title:
                "Untangling the Web. Proceedings of the Conference
                Sponsored by the Librarians Association of the
                University of California, Santa Barbara, and Friends
                of the UCSB Library" (Santa Barbara, CA, April 26,
                1996).
AVAILABLE FROM Electronic version:
                http://www.library.ucsb.edu/untangle/
PUB TYPE   Reports - Descriptive (141) -- Speeches/Conference
                Papers (150)

EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   Academic Libraries; *Authority Control (Information);
                *Electronic Libraries; Higher Education; Indexing;
                Information Technology; *Internet; Research Tools;
                *Subject Index Terms
IDENTIFIERS   *Library of Congress Subject Headings; Links
                (Indexing)

ABSTRACT
                INFOMINE is an academic virtual library containing
close to 5,000 records describing and providing hypertext links to
academically useful Internet resources. Creating a virtual library
helps make new territories of the Internet available to others.
Librarians have traditionally identified printed-form scholarly
resources of value to students and researchers, and they should
continue to do so in the new information environment, utilizing new
technologies and resources. INFOMINE is among an important class of
tool that is helping define a useful scholarly information landscape
on the Internet. Separate virtual collections or INFOMINEs exist in
most major areas of university level research and educational
interests. Library of Congress Subject Headings (LCSH) are used as
subject terms in INFOMINE for the following reasons: LCSH indexing is
very quick relative to other methods; LCSH is a standardized
descriptive language that provides a common thread or a set of
controlled access points across the disciplines, and that is familiar
to librarians in all disciplines; LCSH is useful for both general and
high-level subject description; and in the future, LCSH can be used
in conjunction with other controlled vocabularies. The majority of
the paper looks at the INFOMINE technique for applying LCSH subject
headings. (Author/SWC)

Untangling the Web

Library of Congress Subject Headings as Subject Terminology in a Virtual Library: The INFOMINE Example

Steve Mitchell
INFOMINE Co-coordinator
University of California, Riverside

Abstract:

INFOMINE (http://lib-www.ucr.edu) is an academic virtual library containing close to five thousand "linked" records describing and providing access to academically useful Internet resources (Mitchell and Mooney). These records are organized via a hypertext database and presented in HTML format. INFOMINE began in January of 1994 as a project of the Library of the University of California, Riverside. It now includes participants from most other UC campuses. We initiated INFOMINE because we believe that by creating a virtual library we help make the new territories of the Internet available to others. By supplying links to our own documents and those of others we further shape and develop our conceptions of what's important and situate that in the web as new vistas to be discovered and entered by others. For hundreds of years this is what we as librarians have done with print resources and the concept of the well-selected "collection". Few groups are as qualified as we to continue, in a new information environment, the identification of scholarly resources of value to students and researchers. INFOMINE is among an important class of tool that is helping to define a useful scholarly information landscape in the Internet. Separate virtual collections or INFOMINEs exist in most major areas of university level research and educational interests.

We currently use Library of Congress Subject Headings (LCSH) as subject terms. In meeting the challenge of employing this major standard in library subject description, we use a time-efficient means of selecting and applying LCSH terms which is based on the use of the wealth of LCSH cataloging to be found in TEN, the current 10 year subset of Melvyl, the University of California's online public access catalog (among other bibliographic databases). Our method and technique allow us to use this important subject vocabulary without having to catalog (or re-catalog) Internet resources thus saving much time.

Why We Use LCSH in INFOMINE Indexing:

A.) We have developed a method and technique that makes the use of LCSH reasonably simple, straightforward and efficient for organizations where cost and personnel limitations are commonplace. An INFOMINE record typically takes under fifteen minutes to create (fields utilized include URL, title, subject, key word, and annotation) as compared with one to one and one-half hours for MARC cataloged resources (Morris).

B.) LCSH is a vocabulary that provides a common thread or a set of controlled access points across the disciplines and, therefore, across all INFOMINEs.

C.) LCSH is useful for both general and high-level subject description. It provides an often broadly focused, controlled vocabulary to supply useful subject and meta-subject terms (i.e., terms that often describe whole disciplines; think of them as subject category terms similar to the data in the SC field in the Current Contents data base).

D.) LCSH is a standardized descriptive language familiar to librarians in all disciplines. As a profession, we are very familiar with this vocabulary. This familiarity can be said to exist as well for many academic library users; as flawed as their employment of LCSH in searching may be.

E.) In the future, INFOMINE will be able to supplement LCSH as needed with other, more specialized controlled vocabularies (e.g., MeSH headings) as well as with appropriate key words.

F.) Note that we can convert the information in our records into the MARC record structure if at some point this makes better sense in terms of economics and improved user access.

The INFOMINE Technique for Applying LCSH:

Ingredients: a.) The Melvyl System suite of bibliographic databases which employ LCSH, especially TEN. These databases are rich and varied. The experience and wisdom that has gone into applying LCSH in these bases is both immense and immensely useful. b.) The Windows environment allows multi-tasking to occur in multiple windows and enables the easy and rapid transfer of subject and other text from one window to another. c.) The Internet/web environment allows multiple, integrated access to the various tools/resources/support documents we need, all on one screen.

Result: We often are able to simply and quickly copy relevant subject headings to INFOMINE from these databases. Importantly, for many INFOMINE records, the subject indexing and title words present in the bibliographic database being searched have often amounted to one-stop shopping for our subject and key word application work.

A.) Finding LCSH - The Whole LCSH Domain:

Basics: Cutting-and-pasting from Internet/web accessible resources in a multiple windows environment is what makes our technique possible. Our common-sense approach simply applies the cut-and-paste technique to LCSH focused bibliographic databases as we copy their subject indexing over to INFOMINE. We search bibliographic databases which utilize LCSH, via a telnet session, for both title words and subjects that reflect the content of the Internet site we are describing/indexing and, once found, cut-and-paste the subject terms over to the subject field in the open INFOMINE record. Significantly, we often use title words and some (minor) subject terms we discover in doing this as additional key words as well.

Melvyl System Settings: Once in the database of choice (TEN is most commonly used; see below) we set the timeout period to be long (typically, "set timeout 30") and the subject display to just show titles and subjects ("set display ti,su"). Sometimes, it's useful to use "set display ti,su,date"; date is added to shed light on the age of the indexing (relatedly, you can "set date current"). Doing a "set loc all" command will return all records for all campuses. Most often, depending on the results returned, "loc" and "date" are handled interactively from the keyboard as the need arises.

Searching for LCSH via Title Words: Usually, searching by title word ("f tw xxxx" or "f xt xxxx") will return a large enough set of diverse results to indicate the full scope of how a concept was subject-described, in LCSH, by the catalogers. Usually, the INFOMINE contributor will scan a dozen or two records and copy those LCSH terms which most closely describe the Internet resource being looked at and which have repeated most frequently. Sometimes, the LCSH terms which don't repeat, but which are still appropriate, ARE used in the INFOMINE subject field though more typically these will find their way into the INFOMINE key word field.

Searching/Browsing for LCSH via Subjects: When the concept(s) being searched via a title word approach is yielding too many records, especially if they are too diverse in subject application, we switch modes to subject searching ("f su xxxx" or "f xs xxxx") and/or we browse subjects ("bro su xxxx" or "bro xsu xxxx"; usually we employ "d bro counts" to show us the subjects and number of titles to which any individual heading has been applied). This can give us more focus.

Number of LCSH Terms Applied: Note that our method can be accomplished rapidly enough so as to encourage the application of more rather than fewer subject headings (as compared with most typical LCSH focused databases). This alone is an IMPORTANT boon for INFOMINE users. Minimum LCSH applied in this fashion are 2-3. Overwhelmingly, though, most INFOMINE records have many more than this (est. 4-8 headings or more in addition to several keywords). There is no maximum limit. The INFOMINE contributor stops when the record has been adequately described.

LCSH in Print: In our work there will still be rare times when a look at the print LCSH (and subject trees) will be needed and very useful.

B.) Finding LCSH Meta-subjects - The Controlled LCSH Domain of "Disciplines - Subjects"

Meta-subject terms provide large handles for things. They serve to counteract for LCSH's tendency to feel boundless in scope to users (and indexers). They provide a useful subject "backdrop" against which to search more precise terms (e.g., searching for "icons" and the meta-subject "agriculture").1-2 Terms are applied at the level of meta-subjects or subject categories as you would find, again, in the subject category (SC) field of the Current Contents database or in many web subject lists (which are often organized via broad headings).

Our meta-subject term lists are somewhat lengthier than the 90 subject categories found, for all subjects, in Current Contents. A controlled domain of 50-200 broad LCSH terms exists for each INFOMINE (e.g., the Life, Agricultural and Medical Sciences INFOMINE). These will be found in our "Disciplines - Subjects" document. Since this group of controlled, broad LCSH terms tends to be short enough to commit to memory for any one set of disciplines, they often can be applied without consulting the support document itself. Usually, though, these terms are suggested in the course of the basic LCSH term searching and discovery process as described in A.), above. Experienced INFOMINErs only occasionally need to consult the list that contains these terms and this saves time.

Above and beyond creating a more controlled domain within the LCSH universe, these terms are an attempt to ensure that this controlled domain of LCSH is one that is relevant to the greater UC campus community. As such, all terms were taken from the indexes of several UC campus catalogs. In many ways the campus catalog is the primary document by which the greater UC campus community describes itself. Terms were specifically derived from names for Academic Departments, Programs, Institutes and Colleges. Roughly, 80% of the terms exist as is in LCSH. For the other 20%, we used the closest possible LCSH derivatives.

C.) Configuring Your Session or "The Integrating Environment: Windows on the Internet" --

There are a number of ways to configure your session as you pursue the finding activities mentioned in Sections A. and B. above. All of us have slightly different styles. These are often dictated by the screen size of our monitors, machine RAM and the power of our operating system. I.e., the bigger the screen and RAM and the more robust the operating system, the greater the number of open windows and multi-tasks you can do sensibly. Small monitors and weaker machines mean that you'll: have fewer windows open (and more of your tools/resources as bookmarks or bottom screen icons); be doing more "back and forth"; and, be taking more time to do it. Generally, though, with a 15" or greater monitor and a fast 486 or a pentium with 8 plus megabytes of RAM, we have done the following:

Windows layout: The following assumes overlap among the open windows. The individual window needs to be as small as possible within the limits of being large enough so that all scroll buttons are visible when the window is active and in the "forefront". The three primary windows mentioned below will be open at the same time. The Netscape window in which the resource is being reviewed is generally on the left and the INFOMINE Record Adder form is on the right. The telnet session is also on the right, usually behind the Adder, though when in use it is brought to the fore and takes the place of most of the Adder form. The telnet session is usually called from the left Netscape session.

Primary Open Windows (generally open at the same time):
Resource being viewed/described (via a Netscape session)
INFOMINE Adder Page: The forms-based page in which the description is placed (via a Netscape session to INFOMINE)

Telnet Melvyl System Session I: Our means of accessing one of the databases mentioned below (via access through INFOMINE/Netscape)

Optional - Telnet Melvyl System II: Advanced INFOMINErs can efficiently employ a second window with which to use a second Melvyl System accessible base (via access through INFOMINE/Netscape)

Support Documents Available From the Adder Page (or can be bookmarked):
The "Disciplines - Subjects" or LCSH Selected Domain of Meta-subjects List
    (accessible from the INFOMINE Adder Page; can be bookmarked)
"Shared Subjects/Key Words" (accessible form the Adder Page) (=our protocols/
    conventions concerning resource/site attributes; e.g, a world wide web document
    we use "web")
"Key Word Type Checklist" (to help beginning INFOMINE participants)
"Help: INFOMINE for Indexers" (a lengthy document detailing indexing choices and styles;
    includes our rules for handling LCSH)

Specific Windows Microcomputer Environments: a.) MS Windows (Windows 3.x) is commonly used by contributors and generally is quite sufficient. Sometimes the cutting and pasting of lists (e.g., if several successive LCSH in a TEN record were relevant) requires that each item in the list be transferred individually. This has not been a major problem. b.) Windows NT - not Yet tried c.) MACs - fine. d.) Linux: This is probably the best and quickest bet if you are an experienced micro user. Powerful and efficient. To move a single LCSH from TEN to the Adder takes up to six clicks in Windows 3.x and only two in Linux.

D.) The LCSH Focused Databases Used:

TEN: As mentioned, TEN has been used most frequently. It contains, roughly, the current TEN years of UC cataloging wisdom and LCSH terms (averaging over 1.6 million titles in each of the last 3 years). That is, it represents the current ten year slice of our electronic union catalog, Melvyl. Its big enough to return significant result sets for most searches and current enough to provide reasonably up-to-date indexing. It is also readily accessible and response time is good.

MAGS and other Melvyl System/IAC Bibliographic Databases: TEN can lag in dealing with new LCSH terms for new concepts and, thus, can be behind in dealing with terminology for fast moving disciplines (e.g., high energy physics or the Internet). MAGS is better in this regard though supposed LC terms used may or may not really be LC...still this and COMP and some of the other IAC databases accessible through the Melvyl System will suggest alternative subject and key word terms. Seldom needed.

RLIN/Eureka BIB Database and OCLC/FirstSearch WorldCat: We're currently exploring these to see if they are of use in augmenting our work done in TEN. Limited ports for connecting is a problem, especially for WorldCat.

References:

Mitchell, S. and Mooney, M. 1996 (in press). "INFOMINE - A Model Web-Based Academic Virtual Library." Information Technology and Libraries.

Morris, D. E. 1992. "Staff Time and Costs for Cataloging." Library Resources and Technical Services 36 (1):79-95.

For her many insightful comments and suggestions, the author would like to thank Heidi Hutchinson (Cataloging Librarian, Rivera Library, UC Riverside). Discussions with Lynne Reasoner (Government Documents Librarian, Rivera Library, UC Riverside) also contributed greatly to this paper. All opinions expressed remain those of the author.

Direct questions or comments about this page to mail to: webmaster@library.ucsb.edu">webmaster@library.ucsb.edu

This is an official University of California, Santa Barbara Library web page.

Author:

Corporate Source:

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant
materials of interest to the educational community, documents announced in
the monthy abstract journal of the ERIC system, Resources in Education
(RIE), are usually made available to users in microfiche, reproduced paper
copy, and electronic/optical media, and sold through the ERIC Document
Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the
source of each document, and if reproduction release is granted, one of
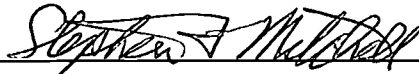the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified
document, please CHECK ONE of the following two options and sign at the
bottom of the page.

____✓____ Permission is granted to the Educational Resources Information Center
(ERIC) to reproduce this material in microfiche, paper copy, electronic,
and other optical media (Level 1).

or

_____ Permission is granted to the Educational Resources Information Center
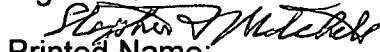(ERIC) to reproduce this material in other than paper copy (Level 2).

Sign Here, Please _____

Documents will be processed as indicated provided reproduction quality
permits. If permission to reproduce is granted, but neither box is
checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC)
nonexclusive permission to reproduce this document as indicated above.
Reproduction from teh ERIC microfiche or electronic/optical media by
persons other than ERIC employees and its system contractors requires
permission fro the copyright holder. Exception is made for non-profit
reproeudction by libraries and other service agencies to satisfy
information needs of educators in response to discrete inquiries.

Signature: Position:
                                    Sciences Reference Librarian
Printed Name: Organization:
Stephen F. Mitchell   University of California, Riverside, Library
Address:
Riverside, CA 92521
Telephone Number:
(909) 787-3316

Date:

$7/3\emptyset/96$

III. DOCUMENT AVAILIBILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of this document from another source, please provide the floowing information regarding the availability of the document.
(ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more strigent for documents which cannot be made available through EDRS).