Proceedings of the Contributed Papers Session
Seattle June 11, 1997

Special Libraries Association

?

Biological Sciences Division

INFOMINE: The First Three Years of a Virtual Library for the Biological, Agricultural and Medical Sciences

by
Steve Mitchell
Science Reference Librarian
INFOMINE Co-coordinator
Bio-Agricultural Library
University of California, Riverside
smitch@ucrac1.ucr.edu

ABSTRACT:

The Biological, Agricultural and Medical Sciences INFOMINE is a virtual library founded in January of 1994 by the Library of the University of California at Riverside. It is a unique Web resource of value for its collection of close to 3,000 annotated and LCSH indexed records with links to selected, university-level resources in most major life, medical and agricultural science disciplines. Information in INFOMINE is well organized and easy to find because of the multiplicity of access points provided for browsing and searching. In addition, INFOMINE provides simple and streamlined capabilities for adding resources (e.g., knowledge of HTML - Hyper Text Mark-up Language - is not required) to the hyper-text database system that is at the foundation of our approach. This article describes the Biological, Agricultural and Medical Sciences INFOMINE and what we've learned in over three years of managing and working within one of the first Web virtual libraries. The Biological, Agricultural and Medical Sciences INFOMINE can be accessed at: URL: http://lib-www.ucr.edu/search/ucr_balsearch.html .

I.) INTRODUCTION TO THE BIOLOGICAL, AGRICULTURAL AND MEDICAL SCIENCES INFOMINE:

The Biological, Agricultural and Medical Sciences (BAM) INFOMINE is one of eight INFOMINEs covering most major academic disciplines. INFOMINE has become quite successful and now records over 100,000 accesses each week. It has over 4,000 other Web pages linking to it.

The BAM INFOMINE is noteworthy due to several features:

A.) Content and Disciplinary Focus:

Our collecting goal has been to include items of most use to University of California faculty and students.

Generally this has meant selection of the largest, most comprehensive and highest quality resources. Note that each resource collected tends to link to from dozens to hundreds of other resources in the same or related subjects.

Emphasis is on interdisciplinarity and subject breadth across the basic and applied life sciences. Researchers and students can locate information in their own as well as allied disciplines from within the same tool.

1.) Subject Strengths:

A recent analysis of the content of the Biological, Agricultural and Medical INFOMINE shows that it contains over 1,400 links to resources in the basic life sciences, over 800 links to agricultural resources and over 700 links to medical resources. Our strongest areas are in the biological sciences including biochemistry, molecular biology, genetics, botany/plant sciences, microbiology, ecology and toxicology. The agricultural sciences are next in importance, notably in agronomy, horticulture, pest management and sustainable agriculture. A weak point here is animal science. The medical sciences, though reasonably well covered, with most major sites being represented, are not as strong as the other major areas at this time.

A more detailed breakdown of subject strengths is as follows (included are subject headings representing 15 or more resources):

agricultural sciences; agriculture; agronomy; aids; animal genetics; biochemistry; biocomputing; biodiversity; bioinformatics; biology; biosciences; biotechnology; botany; cell biology; chemistry; climatology; clinical medicine; conservation; culture collections; diseases; ecology; education; entomology; environmental sciences; environmentalism; food; forestry; genetics; genomes; global climate change; health sciences; horticulture; human genetics; immunology; macromolecules; medical sciences; medicine; microbiology; modeling; molecular biology; natural resources; pest management; pharmacology; plant genetics; plant pathology; plant sciences; pollution; proteins; public health; sequences; soils; sustainable agriculture; toxicology; veterinary sciences; water resources; weather; wildlife; zoology.

2.) Major Types of Resources:

One of our major efforts has been to collect other finding aids and virtual libraries in addition to databases, ejournals (if full-text) and other crucial resources in all fields. We are a clearinghouse for such resources as the following indicates (resource type and numbers are presented):

archives - 23; comprehensive resources - 190; databases - 387; directories - 200; ejournals -122; indexes - 207; libraries -122; library catalogs - 24; museums - 32; newsgroups - 43; reference resources - 239; software -125; subject guides - 71; texts (etexts/texts/books/textbooks/primers/manuals) -164; virtual libraries - 121.

B.) Expertise in Selecting Resources:

Seasoned librarians apply their expertise in evaluating Internet resources. Resources are added to INFOMINE utilizing a simplified input form. No knowledge of HTML is necessary thus freeing up contributor time to find and evaluate resources rather than manipulate HTML. INFOMINE was designed to provide ease of use for contributing librarians at most microcomputer skill levels.

C.) Collaborative Nature:

While INFOMINE was at first exclusively a UC Riverside effort, the project as a whole now includes librarian participation from all 9 UC campuses and Stanford University (Mooney, Fall/96). One of the major benefits of the INFOMINE project is that it is acting as a model for multi-campus shared Internet collecting

activities and has stimulated important dialogue on the issues and opportunities involved.

D.) Value Added:

Each record included receives an annotation as well as subject and key word indexing. Subject words are Library of Congress Subject Headings (Mitchell, 4/96) while keywords specifically reflect minor concepts, synonyms for major concepts and attributes (e.g., publication types, sponsoring institution, related acronyms) of the resource. Average subjects applied number about 6 in the BAM INFOMINE while key words average about 6. These numbers of terms for retrieval are much greater than those for many library related databases. This alone has been a major contributor to INFOMINE's value.

This level of indexing and annotation effort greatly assist users by allowing them to quickly and easily find and then, by enabling the comparison and contrast of related resources, to evalutate a resource's relevance before investing the time to access it.

E.) Multiplicity of Access Points:

INFOMINE provides for Boolean searching as well as browsing through numerous indexes and access points. Means of access therefore include: a.) Boolean searching; b.) Browsing through our Table of Contents (titles interfiling under their subject terms), Titles, Subjects and Key Words; and, c.) Browsing hyper-linked Subject and Key Word Indexing Terms embedded within a record's long display.

F.) Ease of Use:

All of these ways of finding information, together with a simple and to-the-point presentation where usage options are clearly layed out, mean that INFOMINE is easy to use.

G.) Hypertext Database Management System:

INFOMINE was one of the first Web sites to combine the power of the Web with that of a database manager (Mitchell and Mooney, 3/96). For contributors, this allows us to add, edit, provide access to and generally manage several thousand records easily. For example, all indexes (e.g., The Table of Contents) are automatically generated by the database management system. This saves time for contributors (e.g., instead of going to several HTML index pages in order to change how a link is represented, you simply access the editor utility and make the changes on this form which are then automatically made in those indexes). For users, unique searches yield dynamically created, unique HTML documents customized to their interests in contrast to the more prevalent mode of access via a specific set of already created indexes with preconfigured lists of resources (although we do provide this mode of access as well through our Table of Contents and Subject indexes for example).

H.) Ability to Evolve:

The INFOMINE project is responsive to meeting changing needs, opportunities and challenges.

We think Internet finding and discovery tools like INFOMINE, which are structured around conveying the expertise of information professionals, will be around for a long time (Mitchell, 3/97). Given the complexity and nuance found in most human communication, it will be a long time before our niche is supplanted by the large search engines. Never the less, we too are looking at means to incorporate robotic type full-text indexing into our domain of sites as a complement to our current approach. This would be our "power search" option and a good addition to our current service.

We have also thought about adding more fields. Currently each record includes a URL, title, subject, key

word and annotation field. The challenge here though is that each field added represents more time and money invested in the creation of the record. We have generally been successful in keeping add times down to under 20 minutes per record. While more elaborate approaches to Internet indexing and cataloging abound, we are generally satisfied with what we feel is a balanced, time-efficient and pragmatic approach that has proven itself. One of the approaches that we are interested in and perhaps beginning to move towards is the Dublin Core (see the Dublin Core Metadata Element Set Home Page).

I.) INFOMINE Commendations:

INFOMINE has a long list of honors including: a Point Top 5% of the Web award for agriculture. The resource as a whole has received a Magellan 4 Star Rating, is included in several areas in the Argus Clearinghouse (including for the life sciences and agriculture), was mentioned in a2z (the Lycos-associated index) as one of the top 25 science and technology sites, and is in PC Computing's Map to Navigating the Web (August, 1996.) INFOMINE has also been noted in CyberHound (Gale Research), the InterNIC Scout Report, and the Los Angeles Times (2/3/97), among others. Finally, the BAM INFOMINE was included in a recent College and Research Library News article under the heading of "major guides to agricultural resources" (Clark, 6/96).

II.) What We've Learned:

A.) Value Added:

1.) Quality In:

If you build it, they will come. But only if you've added enough value to what you've built. Though one person's conception of significant value added can differ substantially from another's, all virtual libraries are attempting to add value to the links they collect if only by grouping similar, well-selected titles together.

What represents a high value virtual library? I have found that:

"Those VLs that work best seem to generally share the following traits: They add more rather than less value in regard to the value added equation; they usually are hybrid systems bundling more than one [organizing approach, e.g., database and static pages]; the access points they provide are accessible via both search and browse modes; regardless of the specific approach or mode, they provide numerous access points through which the information contained can be accessed/discovered; they are intended to handle larger rather than fewer numbers of resources and cover more rather than fewer numbers of disciplines; they link conveniently at numerous points to other VLs and search engines; and, maybe crucially, because this makes all the above possible, individual entries or records can be added and be well described in a minimum of time"; (Mitchell, 3/97).

2.) Quality Maintained:

Unlike the "good" book which, once added to the collection, can be generally forgotten about for a long time, the "good" link will require recurrent involvement. If anything, link maintenance is somewhat more akin to serials maintenance.

Vanishing sites (also discussed in section II.C.3) can be a challenge: Though perhaps a less common occurrence in academically oriented and sponsored sites than in other realms (e.g., popular culture), "link rot" or non-working links can be a problem. Many sites, including ours, have link checkers that periodically report back on the viability of links (our link checker sporadically works and thus we have more link rot than we find acceptable and are working to correct this).

More challenging in the long run perhaps than link rot, given its labor intensiveness, is major site content redirection / changes / augmentation. Coping requires occasional systematic re-examination of the sites you've collected. Not only are there ever more useful sites proliferating on the Web but, in addition, those that are already there are continuing to expand and/or evolve, many quite rapidly. A site that offered a single interactive database 2 years ago may now offer access to several databases, an ejournal, newsletter and directory of societies as well as a subject index to related sites. Our goal is to try and verify site content at least once every year and we haven't been able to do this as well as we've wanted. This is a MAJOR job since annotations may need to be re-written and indexing modified.

B.) Creating Publicity and Visibility and Targeting Your Audience:

INFOMINE was created as a service both to the University of California community and the global academic community. As a result, our efforts in creating awareness of the tool have been focused on a number of levels and types of audiences. One factor that has helped INFOMINE a lot has been that we began our project, relatively speaking, early on. It was easier to be visible then. This alone helped us to develop a large following among other relatively early Web users. Those seriously using the Web then tended to be predisposed towards information technology and using the Internet as a serious information medium, they tended to be explorers and open to change and they tended to be good communicators. This early word of mouth provided quite a lot of visibility for us.

1.) The General Academic Community Audience:

One of the best and most focused ways of getting to your target audience is by simply identifying good finding and/or subject tools in your interest areas and letting their Web manager know what you're about. I keep a canned email message that does this on hand which I send to such resources as I come across them. Specifically, you'll want to notify other virtual libraries, subject guides and finding tools that you exist.

Get a message into the newsletters of important related scholarly and/or information societies and groups. Listservs (e.g., Web4Lib) are another useful way of getting the word out. The same is true for newsgroups. They want to hear about you.

Presenting to or publishing for librarian, professional or scholarly groups with similar subject interests is also effective. Interestingly enough, my feeling is that formal presenting and publishing, while useful, is less effective and takes much more time than simply using the many Web tools, listservs, and other new and/or non-traditional ways of announcing. In other words, virtual conferences, virtual meeting places and virtual publishing forums are increasingly doing much of what traditional conferences and publishing have done in terms of creating awareness.

2.) The Global Web Community:

You can submit a site description directly and/or you can use a number of Web services which specialize in taking the description you provide and delivering this to the input forms of a great number of search engines or virtual libraries. Submit-it (http://www.submit-it.com/default.shtml) and PostMaster (http://www.netcreations.com/postmaster/index.html) are examples.

HTML "meta-tags" or phrases that describe the mission and subjects covered by your site should be supplied so that the robots for the large search engines will provide access to your tool via the most appropriate terms. Embed these terms in your top page(s). They will be visible to the robots but remain invisible to your users.

3.) The Local Academic Community:

For us, early on, contacting Web savvy faculty and staff one-on-one was a key in getting local awareness generated. In addition, because our Library was the first campus organization to take the Web seriously, we were in a position to train, between 1994 to date, over 1,400 people regarding general Web usage and the use of Mosaic and then Netscape. Because virtual libraries tend to be, among other things, ideal finding tools for beginners, a few minutes of INFOMINE have been taught in each session. We also have made efforts to do departmental, campus and inter-campus presentations whenever possible.

C.) Collecting Issues for Virtual Libraries:

1.) Collecting Tools or Sources for Links:

There are numerous ways to go about finding resources that fit your subject profile. A number of sources can be used (Mitchell, 3/96). These generally can be classed into collecting tools which provide either relatively unfiltered or unprocessed information about sites or, conversely, relatively filtered information. I generally refer to the former as primary collecting tools and the latter as secondary collecting tools.

Primary collecting tools are generally sites where other sites are either being announced to the world and/or are part of the domain of a large search engine that doesn't provide subject classification. Among these primary sources are: announcement or what's new sites; mentions in listservs, newsgroups and newsources; mentions in popular and scholarly journals; and, general search engines. In addition, INFOMINE has several sites announced to it weekly by Web site coordinators who want us to be aware of their work.

Just as important are the secondary tools. In these, some evaluation of site quality and, usually, some subject classification and/or ordering of links has occurred. These secondary tools include virtual libraries and subject guides. Generally, of course, given the focus of INFOMINE, we're concerned only with those secondary sources with academic or serious research interests.

There is no one tool or even type of tool to which collectors should restrict themselves. A suite of different tools is needed to be effective in finding useful Internet resources. What is in the particular suite of finding tools used will vary depending on the type of search being conducted.

a.) Primary Collecting Resources:

Use primary tools when you are getting into specialized subjects or subjects that just cannot be found through other means. They are also used when you have comprehensive coverage in more general subjects and are looking to further expand and deepen your collection. The use of these tools is sometimes the only way to find the desired content. Their use, however, is generally more time consuming than finding links through secondary tools. You often, as in the case of the search engines, have to spend more time searching for (and creatively limiting) results. You also have to spend more time doing initial, first-cut analysis of content and quality.

b.) Secondary Collecting Resources:

For resources found in secondary tools, certain decisions about the resource's substance have been made. Your decisions are at least partially reinforced by those others who have already examined the resource and who have made judgements about overall quality and stability. Secondary sources often vet one another. This usually saves time for the collector.

Still, serious analysis is required of a resource found anywhere. With more frequency than would seem probable, secondary resources can make mistakes in adding to their collections and such mistakes can often be echoed by other secondary resources. Of the total number of resources we find at secondary as well as primary sites, only a small percentage are added to INFOMINE.

Use secondary collecting tools when new to a subject and/or in need of an overview of the overall landscape of a subject. Generally, these are of value when looking for crucial, comprehensive and/or reference level sites for a subject. You will find such sites through the primary tools as well but often you will find them faster via the secondary tools.

There is a trend for secondary resources to become increasingly inter-linked with each other through more ways than just providing a hyperlink. Most of us, in fact, do bounce off of one another in discovering sites. This can be good and, indeed, is part of what INFOMINE is for: to help not only regular users but librarians, information specialists and site managers as well in finding useful resources. We are a public domain finding tool working within an atmosphere of informal cooperation in a community of other non-proprietary finding tools.

c.) Subject Resource Hubs:

Most people who spend significant time collecting resources on the Internet do so by constructing groupings of most valuable site discovery tools. These groupings might be called subject resource hubs. They consist of productive primary and secondary tools.

As collecting goals vary, the mix of tools in a subject hub will vary as well. For instance, a collector in a very specialized subject may rely on mostly primary tools, including: a combination of faculty word of mouth and subject specialized serials, listservs or newsletters as well as the general search engines. Those maintaining a brief, representative reference level collection will probably more strongly rely on secondary tools such as the the broader more general subject guides and virtual libraries.

I have numerous subject hubs of "best" finding tools for the many disciplines in the life and other sciences in which I work: one each, for example, for biochemistry, genetics, ecology, and so on. With some experience, I've been able to determine the most productive sites and weave them into my collecting routines. Most of my subject hubs include announcement sites as well as other virtual libraries and subject guides with backup from two search engines, Alta Vista and HotBot. With the exception of the search engines, the specific sites I've gone to have varied quite a lot over time within each subject. And, they vary quite a lot among subjects. Other virtual libraries rely on INFOMINE's efforts in the same manner.

2.) Collecting Issues in Adding Links:

a.) Making Good Choices:

Since most quality criteria that apply to print resources also apply to electronic, the transition into Web collection development has been generally quite smooth. Librarians and information professionals know good quality when they see it. This is a given. As far as format goes, our assumption has generally been that Web format is at least as useful as print. If needed, Web documents can of course usually be easily printed.

Among others things I look for, though, which are more specific to this information environment than perhaps others, are those relating to site stability, currency of information and access to archival information. For example: Is the site supported by a large and/or apparently stable corporate sponsor (e.g., research program, college, corporation)? If part of a program or college, is the site a personal or corporate presence? Is the content of the type that will need updating and, if so, is this occurring frequently (though realize that Web maintainers often update site content without remembering to revise the date)? You need to note that the more energetic and well-funded the Web presence, the more the site will be growing and the more frequently you will need to revise your information.

The difficulties and challenges have been building over the last year as we begin to need to discriminate

among the increasing number of important subscription based Internet resources. Everyone (publisher, researcher, librarian) is experimenting. For the considerable value added advantages of information in electronic format, what is a reasonable price to pay? What features, ranging from guaranteed archival access to sensible fair use policies to guaranteed smooth access (i.e., minimal traffic problems) to zoomable/high-resolution graphics can we expect and for what price? How decentralized will modes of Internet information distribution become? Will learned societies be subscribing and organizing access, will libraries or consortiums of libraries be subscribing and organizing access and/or will individual researchers subscribe directly to publishers who organize access? We are, in fact, going in all these directions at once so that what was once straightforward isn't anymore.

b.) Facetting A Site:

Often the main entry point for a site is not the most useful part for your users. Sometimes a site will have several major components of which only a few are important. These are situations in which, instead of representing the whole site, you want to represent only an important facet(s) of it instead. A good example is a site that is a compendium of access points to a number of medical databases of which just a couple are pertinent to your users. In this case, we would create a record for each of the two individual databases instead of, or in addition to, the page representing the whole site. The benefit of this is that you can represent important content at more length and get your users directed faster. Facetting can bring vital but buried resources to the forefront making them much more visible to users. The down side is that you're more subject to link rot as changes in servers and directories occur and it does take longer to create multiple records.

c.) Links that are Announcements:

Increasingly, relevant sites are available on a subscription only basis. Given that we're in changing times, the Library response to this can be slow in that money is often not available and/or sensible conventions in access arrangements between publishing house and library are in the process of being defined. Given this situation, we will at times create a record for such an unsubscribed to site in the hopes of making it visible to our researchers and students who may require it and who would be motivated to figure out a way of accessing it on their own. In many of the sciences, for example, access to crucial articles weeks before their availability in print format by itself represents an absolute necessity in ensuring that research efforts remain fully informed and competitive.

d.) Annotations:

Surprisingly, annotations have been easier to do than we had initially thought. In academic sites, over one-third of the time, we find accurate, ready-made short descriptions already present. Either as they are or with slight modification, they can be entered (with quotations) without too much effort or time.

3.) Collecting Issues in Maintaining Links:

Vanishing Sites:

Academic sites generally show a lot more stability than commercial or other types of sites. Disappearance usually signals that the resource has moved rather than terminated service. When they move, there is usually adequate notice given and generally the move is to another directory on the same server rather than to a completely new address. Re-finding is relatively easy. And, when a new server address is involved, it's typically with the same institution. Often, when a move is accomplished, the old address will continue to function and aliases, or alternative URLs, for the resource are supported. If you can't locate it on the same server or within the same institution, its then off to one of the larger search engines for a retrieval attempt

which is often successful: good sites simply don't usually disappear.

Author's address:
Steve Mitchell
Bio-Agricultural Library
University of California, Riverside
Riverside, CA 92521
smitch@ucrac1.ucr.edu

BIBLIOGRAPHY:

Clark, Kathleen. 6/96, "Internet Resources for Agriculture." *College and Research Library News*. 57(5): 359-364.

The Dublin Core Metadata Element Set Home Page.
URL: http://www.oclc.org:5064/research/dublin_core/

INFOMINE. 1/94-.
URL: http://lib-www.ucr.edu/search/ucr_balsearch.html

Mitchell, Steve. 3/96, "General Internet Resource Finding Tools: A Review and List of Those used to Build INFOMINE."
URL: http://lib-www.ucr.edu/pubs/navigato.html

Mitchell, Steve. 4/96, "Library of Congress Subject Headings as Subject Terminology in a Virtual Library: the INFOMINE Example." *Untangling the Web Conference, UC Santa Barbara.*
URL: http://www.library.ucsb.edu/untangle/smitch.html

Mitchell, Steve. 3/97, "INFOMINE." *Ariadne*, issue 8.
URL: http://www.ukoln.ac.uk/ariadne/issue8/infomine/

Mitchell, Steve and Margaret Mooney. 3/96, "INFOMINE - A Model Web-Based Academic Virtual Library." *Information Technology and Libraries*, 15 no.1: 20-25.
URL: http://lib-www.ucr.edu/pubs/italmine.html

Mooney, Margaret. Fall, 1996, "Linking Users to Internet Government Information Resources through INFOMINE." *DLA Bulletin*, 16 no.1: 31-43.
URL: ftp://ftp.dla.ucop.edu/pub/dlabulletin/issue35/infomine.txt

---

Comments should be directed to Tom Turner tom.turner@asu.edu

HTML Document Preparation by Linda Shackle and Tom Turner
May 1997