<table>
<thead>
<tr>
<th>Institution</th>
<th>IF</th>
<th>CIF</th>
<th>NIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Madurai Kamraj University</td>
<td>0.74</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>National Geophysical Research Institute</td>
<td>0.27</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>National Chemical Laboratory</td>
<td>0.27</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Department of Science &amp; Technology</td>
<td>0.66</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Indian Institute of Technology (Bombay)</td>
<td>0.66</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Bose Institute</td>
<td>0.28</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>Jawaharlal Nehru University</td>
<td>0.27</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>Sree Chitra Tirunal Institute of Medical Sc.</td>
<td>0.27</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Panjab University</td>
<td>0.27</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>Delhi University</td>
<td>0.26</td>
<td>17</td>
<td>20</td>
</tr>
</tbody>
</table>

Impact Factor of a journal is the ratio of the citations received in two years to the number of articles published in the preceding year. IF = Impact Factor; AIF = Average Impact Factor per paper; CIF = Cumulative Impact Factor; NIF = National Average Impact Factor.

INTERNET GRANTHALAYA
DEVELOPING AS NEW SEARCH ENGINE & BROWSER FOR LIBRARIES TO SEARCH AND ORGANIZE THE WORLD WIDE WEB LIBRARY RESOURCES

V Sreenivasulu

1. INTERNET GRANTHALAYA

In India, Nuclear Science Centre (NSC) has developed a Web program which displays the Web-held electronic journals in a systematic manner and it facilitates to browse the Web-held e-Journal user-friendly. This is one of the finest developments and path-finding break-through in the direction towards the organization of World Wide Web Library resources on the Internet. The Concept "INTERNET GRANTHALAYA" was conceived when the Director, INSDOC, who delivered a motivative speech on the occasion of CSIR Foundation Day, September 25, 1999. INTERNET GRANTHALAYA is being originally intended to be developed as a library software which can put the library catalogue on the Internet. The very next day, I submitted, a two-page note on how to develop INTERNET GRANTHALAYA and he made comment on the note and I quote, "This note appears to contain some useful ideas and may be considered while developing Internet version of..."
Granthalaya", and this note was sent to Head of Data, Computer & Software Group, to consider the idea of developing INTERNET GRANTHALAYA. Efforts are under way to develop it at INSDOC.

2 USING FACET ANALYSIS TO SEARCH AND ORGANIZE THE WWW RESOURCES

If Dr Ranganathan was alive today, he would be aware of the potential of his ideas for searching and organizing WWW-held Library materials. Indeed, the genius of Ranganathan is attested to by the very portability of his ideas across time, technology and cultures, simply because they address the very foundations of the business of effective information storage and retrieval.

Facet Analysis can be employed to alleviate some of the problems in searching the WWW using either subject directories or search engines. In some respects, facet analysis merges together word and concept based approaches, in that the concept terms in the facets are selected on the basis of literary warrant. It can also be employed to organize WWW Library resources, not as a whole, but in relation to organizing the results of WWW searches and in relation to Web page indexing.

Facet Analysis can assist the developer or searcher in retrieval to overcome some of the problems in indexing or searching the WWW in a reasonably effective and efficient way. In terms of searching, facet analysis can be used to suggest, arrange and organise the terms, which would be used in WWW search. The initial selection of terms of the facet analysis can proceed in a conventional manner by the use of reference works but equally a broad search on the WWW itself is likely to generate a reasonable set of terms for a facet analysis. The terms of the facet classification can be used in the search as well as to organize the search page. The search page can also be made 'live' by the provision of links to sites can be linked to the terms of the classification on the screen. This makes the facet display live. In subsequent searches this would enable the searcher to go straight from the chosen search terms to the sites identified or searches.

3 WORD AND CONCEPT INDEXING ON THE WORLD WIDE WEB

The different approaches used by subject directories and search engines in terms of indexing are not new. Subject directories use a conceptual approach, which is at the basic of the classification schemes used in many fields. There are two main types of methods for indexing or searching for documents on the World Wide Web-word based and concept based. (Ellis & Vasconcelos, 1999)

Word based methods concern the automatic input of words without any further input or consideration of the exact meaning of each of the terms
extracted from the source of document. The search engines on the WWW work through automatic word indexing of Web sites. (O’Connor, 1996)

Concept based methods require the intervention of the human mind and imply the analysis of the concepts present on the text and the choice of terms best representing these documents. The search directories in WWW have a concept based organization and use human indexing of the different Web sites. Concept based indexes are often structured in order to represent the relationship between the different concepts.

Word indexing, although fast and economic, has problems associated with it, as it does not consider the meaning of words, and it does not consider the concept and bibliographic access points indexing to retrieve the content. Concept based indexing, used by subject directories, tries to address these problems basing indexing, not on extracting words from the text, but on identifying the meaning of the different words in the context they are used and choosing one keyword or expression for each concept. These keywords or expressions will form the subject index of subject directories.

4 INDEXING AND SEARCH TOOLS ON THE WWW

The hypertext structure of the web means that retrieval is done through the links between different web pages, through browsing and navigation. Because each web page will have multiple links, there will be multiple paths to the same piece of information. Although, the WWW does not have a central catalogue or index, different search tools have been devised in order to assist people in finding information on the largest repository of documents in the world. These search tools can be divided into two main categories: subject directories and search engines.

4.1 SUBJECT DIRECTORIES

Subject directories (also called subject trees or subject: guides), allow people to browse information by subject. The searcher can browse through the index in search of relevant subjects and navigate to the relevant web sites, by clicking on hotspots, which represent those sites. Examples of subject directories include Galaxy (http://www.einet.net/galaxy.html); Magellan (http://www.mckinley.com) and, yahoo! (http://www.yahoo.com) probably the latest directory.

4.2 SEARCH ENGINES

Search engines, rather than searching documents through an index, are based on allowing users to enter keywords that are matched against a database. Search engines are composed of three different parts: a program called a spider (or robot or crawler), a database with an index and search software (Sullivan, 1998, & Haskin, 1997). Spider wander through web, crawling
from site to site. Different search engines use different types of spiders. The first type of spider finds huge volume of information short space of time, the type latter generates a smaller number of pages, with perhaps more relevant results. Every page found by a spider is stored in a database and an index of its contents is built, by extracting automatically words from the Web pages and ranking them alphabetically, listing the same principles which inverted files use. The index is, therefore, a list of every word found (except irrelevant words such as and, the, an, called stop words), other search engines index only the title and top level phrases of a web site. The third element of the search engine is the software, which is a program which compare search queries keyed in by people.

The different approaches which search engines use in crawling the web, finding new pages and indexing them will produce totally different results. This is why when searching the same topic in different search engines there can be very striking differences in the results. Examples of search engines include: Altavista (http://altavista.digital.cim/); Excite (http://www.excite.com); Hotbot (http://www.hotbot.com/); Infoseek (http://www.infoseek.com/); and Lycos (http://www.lycos.com/).

5. SALIENT FEATURES COVERED IN THE PRESENT SEARCH ENGINES

The Web search engines have broad popular appeal. They are the tools that have finally made end-user searching a reality. And while they cater to the general public, the Web search engines are increasingly important tools for the information professional as well (Garmaman, 1997). No ones to disdain free searching or free information, we make good use of these search tools, even while we sometimes wish for the precision and advanced capabilities of our traditional command line systems with accurate Boolean processing, extensive field searching capabilities, advanced output and display options, truncation, search saves, controlled vocabulary, current awareness services and a targeted subject focus (Hock, 1999, Notess, 1999, & Sherman, 1999). The following features of search engines are briefly described:

—Search Engine Size: The "size" stated by search engine producers conventionally refers to the number of unique Web pages (unique URLs), rather than "sites" (which may contain numerous "pages"). In Lycos, for example, the 35 million, does not include the personal home page or its pictures and sounds databases.

—Boolean Operators (And, Or, Not) and Parentheses: In general, there are two levels of Boolean capability among these search engines. The "Simplified" from uses a plus sign in front of a term to indicate that records should only be retrieved, if that term is present. At this simplified level, a "NOT" is achieved by
means of a minus sign in front of a term. Boolean capabilities familiar in traditional online services, the engine must provide the equivalents of AND, OR, and NOT, plus the capability of nesting (The use of parentheses). In all engines that use these Boolean connectors, the capitalized form will work.

—*Phrase Searching, Exact Phrase Match:* In almost all cases, a phrase can be indicated by putting the phrase in quotes (" ") in the query box. In some cases, a phrase can be designed by choosing the phrase option from a pull-down window.

—*Proximity Searching:* Phrase searching which can be as sophisticated as that found in commercial online searching (Candy, 1998). It is, of course, one form of proximity searching. The next most common proximity option is NEAR, which specifies "within 10 words" in Altavista and "within 25 words" in Lycos Pro. The latter also allows NEAR/n, where "n" is a user-specified maximum distance, e.g., NEAR/5.

—*Truncation:* Truncation (terminal and internal) or conversely, inhibition of automatic stemming, character strings, allows for initial truncation. If a truncation or stemming feature is available, appropriate symbol is shown.

—*Title, Date, and URL Fields:* Where these fields are searchable, the approximate prefix is shown, or an indication is given that the searcher uses pull-down window or text boxes. For prefixes, the searcher should enter the prefix shown followed by the term to be searched - for example, title: lupus.

*"Links to" a URL:* This refers to the capability of identifying which pages in the search engine’s database contain a link to a particular URL. This is somewhat analogous to "Citation search" and enables the searcher to identify sites that have some interest in the site referred to.

—*Language:* Entries here indicate whether one can search by the language in which the Web page is written.

—*Media Searching:* This refers to the capability for searching by type of media - images, audio files, and video files. The implementation is quite different among the search engines that provide this.

—*Name:* The entries here refer to whether the search engine claims to be able to identify proper names, persons. This actually boils down to either limiting retrieval to instances in a page where each word appears in its capitalised form and/or automatically allowing for the inverted form of a word pair.

—*Case Sensitivity:* Some engines can identify upper and lower case
letters. This is an important instances when "AIDS" needs to be distinguished from "aids." In general, when a query is entered using all lower case, the search engine will retrieve both lower and upper case. When upper case is entered by the searcher, the engines will return only those records with an exact case match. For examples, "next" will retrieve "next" and "neXt," whereas "neXt" will only retrieve "neXt".

---Searches All Common Words: This refers to whether literally words are indexed and searchable. Critical not only when one wants to search for "The Who," but when one needs to search for any phrase containing a very common word.

Directory Attached: This is an indication of whether a Web directory is included as a part of the search engine's search page. (In some cases, the directory may be embedded in a "Channel" option.)

Gives Count for Answer: Being told the number of items retrieved might seem to be something we could take for granted, but not so for Lycos.

Gives Term Count: Some engines tell not only the overall answer, but also the retrieval count for the individual terms searched (as with traditional online services.)

Output Option: This line indicates which format options are available, whether the user can specify the number of records on each results page, and also if results can be "grouped or ungrouped" by Web site.

"More Like This": For some search engines, you are given the option, when you see a record you like, of having the engine find other records that are similar to that record.

5.1 SPECIAL FEATURES

What is listed here are additional features provided by the search engine that should be of interest to the serious searcher. The choice of what is included in this line is admittedly somewhat subjective. Features listed here are ones directly related to performing a search on the search engine's Web database. Additional features (or "add-ons"), such as company directories, free email, weather reports, etc are ignored.

6 IG: A SEARCH ENGINE & BROWSER EXCLUSIVELY FOR EVERY LIBRARY USE ON INTERNET

Common usage has evolved, and now "search engine" is a broadly-used, loosely-defined term for any kind of software or service that offers searchable access to electronic files. Search engines may search Internet or Intranet files, locally-mounted databases, files on your own computer desktop, image
databases, or many other kinds of computer-readable files.

INTERNET GRANTHALAYA is a search engine that offers searchable access to Internet-held OPACs, displays the library holdings World Wide through Internet. It may be first in its kind in the World and exclusively for every library use on the Internet.

6.1 DEVELOPMENT OF INTERNET GRANTHALAYA

INTERNET GRANTHALAYA can be developed initially in three models/phases.

*Model-I*: It can be developed as a Browser, such as Netscape, Internet Explorer, which displays Internet-held OPACs, Library holdings, Library databases in INTERNET GRANTHALAYA Format.

*Model-II*: It can be developed as a search engine, such as Yahoo, Lycos, AltaVista, which does searching, indexing and displaying of Internet-held OPACs, digital library holdings in INTERNET GRANTHALAYA Format.

*Model-III*: A combination of these two. INTERNET GRANTHALAYA functions as browser as well as search engine exclusively for every library use. This model may be absolutely suitable for INTERNET GRANTHALAYA.

6.2 SALIENT FEATURES OF INTERNET GRANTHALAYA

- Displays Internet-held OPACs of every document on Internet
- Provides "Universal Access" and "Access to All" to the Global Library User to the Library
- holdings of World Wide Library Resources on the Internet
- Displays OPACs in an INTERNET GRANTHALAYA Format
- Provides "An Integrated OPAC Format", with different browsing icons, irrespective of cataloguing code and classification scheme and the location of the library. For example, OPAC of INFLIBNET, Library of Congress, The British Library, OCLC, to INSDOC
- Facilitates "Auto-classification" and "Auto-cataloguing" in an Integrated INTERNET GRANTHALAYA Format
- Works as "Knowledge Mining Tool, "Knowledge Discovery Search Engine" with reference to library holdings World Wide. It can be a new tool and technique exclusively of digital libraries and electronic libraries for knowledge discovery for every library use

6.3 SPECIAL FEATURES OF INTERNET GRANTHALAYA

- Metadata Indexing and Documentation about documents, objects with the help
of metadata schemes such as DDC, LC, MARC formats, LCSH, AACR-II

— Bibliographic Access Points Indexing and Retrieval rather than word by word indexing and retrieval by the search engines
— Subject Access via hierarchical guides (hotlists) on the Web search engine
— Subject Indexing through classification and controlled vocabularies for library resources on the Web
— Concept Indexing rather than word by word indexing
— Multimedia Indexing and Documentation

6.4 ONLINE ACCESS TO CD-ROM DATABASES

Many CD-ROM database producers are now offering access to their databases through Internet. The Internet option provides a number of benefits to users (Silver Platter information):

a) Any fully Internet connected workstation can get access. This means that users can work from their desks wherever they are located;

b) There is no need to purchase a CD-ROM server and therefore the budget can be spent more on information databases and less on hardware;

c) Internet subscriptions offer a budgetable fixed cost for database searching;

d) The SPIRS retrieval clients are the same for other forms of access which means that there is no need to learn new interface;

e) Databases are automatically updated at the server and therefore there is no need to receive discs and there is no problem of damaged or lost discs. Updates are also available sooner.

6.5 BROWSER OF INTERNET GRANTHALAYA

When systems become very complex, it is necessary to break them down into subsystems. The interaction between the subsystems needs to be defined very carefully. (Anil Kumar, 1998) Browser provides a uniform framework for communicating, structuring and sharing information. A software prototype called "Browser" has to be developed. This prototype should facilitate the specification of text, image, video and sound data to access and retrieval on the Web. Browser also provides natural language interface for navigation and browsing over the Web Library.

7 ORGANIZATION OF INFORMATION ON THE WEB

The recent developments in the internet and WWW have brought new challenges to information professional. Mandal and Wolven considered the relevance of Charles Ammi Cutter's principles of bibliographic access to internet resources. They further examined identifying authors' roles, collocating works and versions, and providing subject access through classification and controlled vocabularies for resources on the Web. (Choudhary, 1999)
A core set of metadata elements needs to be identified and each creator of electronic documents should be able to implement it in the record that he or she creates. With this objective, a simple resource description set of data has emerged which is known as the Dublin core (http://purl.org/metadata/dublin-core). The Dublin core metadata set prescribes fifteen elements, viz (http://www.lub1.u.se/cgi-bin/nmde.pl):

- Title;
- Creator;
- Subject (keywords, controlled vocabulary; and
- Description (abstract and content description);
- Publisher;
- Contributor (Other than the creator);
- Date;
- Type (Category of the resource);
- Format (HTML, postscript, etc);
- Identifier (URL, string or number used to identify the resources);
- Source (from which this resource is derived);
- Language;
- Relation (with other resources);
- Coverage (spatial and/or temporal characteristics of the resources);
- Rights (link to a copyright notice, etc).

Woodward mentioned the various strategies used by librarians for cataloging and classifying items on the internet, such as: automatic classification projects and classified subject trees, like the BUBL Subject Tree, Cyberdewey, and the World Wide Web Virtual Library (Woodward, 1996). The cyberstack project that uses Library of Congress (LC classification to organize Internet resources is a significant example of the use of classification on the WWW (http://www.public.iastate.edu/~CYBERSTAXCS). Similarly, an example of the use of Dewey Decimal Classification (DDC) for organizing information on the Internet can be found in the Bulletin Board of Libraries (BUBL) that allows searching materials classified according to DDC. (http://link.bubl.ac.uk/ISCL)

Scorpion is another project (http://www.oclc.org/oclc/research/publications/review96/scorpion.htm) that has been built around this concept. The scorpion system is a research project at OCLC that explores indexing and cataloguing electronic resources. Since subject indexing is a key to information retrieval, browsing and clustering, the primary focus of scorpion is building tools for automatic subject recognition based on classification schemes like the Dewey
8 CONCLUDING NOTE

The subject approach to information, concept indexing, keyword indexing facet analysis, which have been traditionally controlled by library and information professionals with the help of classification and cataloguing and vocabulary control tools, are an important areas of research in the Internet environment. By applying all these concepts; an attempt is being made to develop a search engine so called INTERNET GRANTHALAYA exclusively for every library use on the Internet and this idea offers broadly universal enthusiasm among the architects, engineers and information scientists who are building the next-generation - Web search tools. Their collective ideas, plans and projects epitomize the most prescient vision of the future possible. Information professionals with their rich experience in organizing and searching information, as well as in identifying user needs, should work hand in hand with software professionals in organizing and developing search engines for every library use on World Wide Web.

The present search engines are not rightly suitable to search and organize the WorldWide Library resources on the Web. Developing suitable search engine exclusively for every library use on the Web is the need of the hour. The information professional should be best served with multiple search tools.

This paper finally urges world wide, advocates and targets at the task of creating a new search engine and dedicated browser exclusively for every library use on the Web. In this direction, INTERNET GRANTHALAYA may be the ultimate search engine exclusively dedicated for every library use to search and organize the World Wide Web Library Resources.

REFERENCES

5. GARMAMAN Nancy: The homepage: the ultimate, original search engine. (ONLINE, May/June, 1999, 6).

INICAE
A core set of metadata elements needs to be identified and each creator of electronic documents should be able to implement it in the record that he or she creates. With this objective, a simple resource description set of data has emerged which is known as the Dublin core (http://purl.org/metadata/dublin-core). The Dublin core metadata set prescribes fifteen elements, viz (http://www.lub.1u.se/cgi-bin/nmde.pl):

- Title;
- Creator;
- Subject (keywords, controlled vocabulary; and
- Description (abstract and content description);
- Publisher;
- Contributor (Other than the creator);
- Date;
- Type (Category of the resource);
- Format (HTML, postscript, etc);
- Identifier (URL, string or number used to identify the resources);
- Source (from which this resource is derived);
- Language;
- Relation (with other resources);
- Coverage (spatial and/or temporal characteristics of the resources);
- Rights (link to a copyright notice, etc).

Woodward mentioned the various strategies used by librarians for cataloguing and classifying items on the internet, such as: automatic classification projects and classified subject trees, like the BUBL Subject Tree, Cyberdewey, and the World Wide Web Virtual Library (Woodward, 1996). The cyberstack project that uses Library of Congress (LC classification to organize Internet resources is a significant example of the use of classification on the WWW (http://www.public.iastate.edu/-CYBERSTACKS). Similarly, an example of the use of Dewey Decimal Classification (DDC) for organizing information on the Internet can be found in the Bulletin Board of Libraries (BUBL) that allows searching materials classified according to DDC. (http://link.bubl.ac.uk/ISCL)

Scorpion is another project (http://www.oclc.org/oclc/research/publications/review96/scorpion.htm) that has been built around this concept. The scorpion system is a research project at OCLC that explores indexing and cataloguing electronic resources. Since subject indexing is a key to information retrieval, browsing and clustering, the primary focus of scorpion is building tools for automatic subject recognition based on classification schemes like the Dewey
Decimal classification (http://Horc.rsch.oclc.org:8109/bintro.html).

8 CONCLUDING NOTE

The subject approach to information, concept indexing, keyword indexing facet analysis, which have been traditionally controlled by library and information professionals with the help of classification and cataloguing and vocabulary control tools, are an important areas of research in the Internet environment. By applying all these concepts; an attempt is being made to develop a search engine so called INTERNET GRANTHALAYA exclusively for every library use on the Internet and this idea offers broadly universal enthusiasm among the architects, engineers and information scientists who are building the next-generation - Web search tools. Their collective ideas, plans and projects epitomize the most prescient vision of the future possible. Information professionals with their rich experience in organizing and searching information, as well as in identifying user needs, should work hand in hand with software professionals in organizing and developing search engines for every library use on World Wide Web.

The present search engines are not rightly suitable to search and organize the WorldWide Library resources on the Web. Developing suitable search engine exclusively for every library use on the Web is the need of the hour. The information professional should be best served with multiple search tools.

This paper finally urges world wide, advocates and targets at the task of creating a new search engine and dedicated browser exclusively for every library use on the Web. In this direction, INTERNET GRANTHALAYA may be the ultimate search engine exclusively dedicated for every library use to search and organize the World Wide Web Library Resources.

REFERENCES

1. ANIL KUMAR Nishta: Going digital... Developing the digital library of PRL, Proceeding of 49th FID Conference held at New Delhi, 1998, FID Publication No 719, 11-28-34.


5. GARMAN Nancy: The homepage: the ultimate, original search engine. (ONLINE, May/June, 1999, 6).

INTRODUCTION

In the present age of information explosion, the libraries and information centres are generally ill-equipped to handle and retrieve information efficiently. Due to increasing cost of reading materials and limited financial resources of the libraries, it is very difficult to fulfill the requirements of the users. The option left with us is to promote the sharing of resources by networking to bridge the gap between the user and information provider.

Anil Singh, Division of Library, Documentation and Information, National Council of Educational Research and Training, Sri Aurobindo Marg, New Delhi-110016.