JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY

JASIST | VOLUME 53 | NUMBER 14 | 2002

CONTENTS

Papers published online in Wiley InterScience 5 November 2002 through 15 November 2002

Letters to the Editor

Engineering a Search Engine (WebLib) and Browser (Knowledge Navigator) for Digital Libraries: Global Knowledge Discovery Tools Exclusively for Librarians and Libraries on the Web

Sir:

The ultimate objective of the concept of "engineering a Search engine for digital libraries" is to provide theoretical framework and guidelines to build new browser and search engine–WebLib–The Ultimate Search Engine for digital libraries with Browser–KNOWLEDGE NAVIGATOR exclusively for library use on the Internet and offers broadly universal enthusiasm for the future among the architects, engineers, librarians and information scientists who are engineering and building the next generation of Web search tools and this conceptual idea provides food for thought for digital librarianship. Such search engine and browser can be used as knowledge mining tools and knowledge discovery Web tools for discovering the knowledge from digital libraries and Electronic Knowledge Houses. The importance of this letter can be stated as: "we need to engineer a search engine, WebLib with Browser-KNOWLEDGE NAVIGATOR (KN), for digital libraries founded by the library and information profession, supported by the information professional, for the global library community for every library use on the Web, for the dissemination of digital information globally, and enjoyed by the end-user global community."

Engineering the WebLib to meet the challenges of digital librarianship. Librarians and information scientists need smart tools for better display of filtered digital library content, the improved search options such as title, author, publisher, date, ISBN, including Online, CD-ROM databases and full-text databases to navigate on the Web with the help of engineering such search engine with browser for mining the knowledge from huge knowledge marts, knowledge houses, global digital libraries, global electronic libraries, online libraries and information centers, virtual information markets, from end-users, online users, virtual user community, online communities to the library and information user.

Engineering WebLib: A search engine and browser exclusively for digital libraries on the Web. WebLib is a search engine that offers searchable access to Web-held OPACs, and displays Web-held digital library resources and the library holdings Worldwide through Internet. It may be first in its kind in the world and exclusively for every library use on the Internet. WebLib can be developed initially in three models/phases: In model/phase-I, it can be developed as a browser called KN, such as Netscape, Internet Explorer, which can displays Internet-held OPACs, Library holdings, & Library databases in WebLib format (set of bibliographic elements). In model/phase-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 WebLib format. In model/phase-III, a combination of these two, and WebLib should work as browser as well as search engine exclusively for every library use. This model may be absolutely suitable for digital libraries.

Salient features of WebLib. It should display Internet-held OPACs of every document on the Internet; provide "universal access" and "access to all" from the local library user to the global user, end-user, virtual communities, online users; from libraries to the global digital libraries, virtual information markets, the electronic libraries to the global knowledge warehouses; from library holdings to worldwide library resources on the Web; provide "An integrated OPAC format on the Web", with different browsing icons and mining fields, irrespective of cataloging code and classification-scheme; facilitates "auto-classification" and "auto-cataloguing" in an integrated WebLib format on the Web; should work as "knowledge mining tool", "knowledge discovery search engine" with reference to digital library holdings worldwide. It can be a new tool and technique exclusively digital libraries and electronic libraries for knowledge discovery for every library use on the WWW for the global digital library centres. WebLib should have the special features such as metadata indexing, bibliographic access points indexing, subject access, subject indexing, multimedia indexing and documentation, concept indexing, field searching, keyword indexing and the use of facet analysis to search and organize the worldwide Web library resources and digital information systems. WebLib should overcome the demerits of present search engines (Sullivan, 1998) such as word-by-word searching and indexing. It should index only controlled vocabulary rather indexing every word or word-by-word.

Engineering the browser of WebLib as Knowledge Navigator (KN): How it works. Browser-KN provides a uniform framework for communicating, structuring, and sharing information on the Web. A software prototype browser called Knowledge Navigator (KN) is to be developed. This prototype should facilitate the specification of text, image, video and sound data to access and retrieve on the Web. KN should also provide natural language interface for navigation and browsing over the Web. The present Web browsers such as Internet Explorer and Netscape Navigator are not designed for the librarians and information professionals for every digital library use on the Web. The WebLib KN should be engineered and developed to facilitate the bibliographic access points such as title, author, keyword, ISBN, ISSN, publisher, place, and date. These access points are able to provide the concept indexing, and subject indexing rather than word-by-word indexing on the Web. This KN should also facilitate to catalogue and classify the digital documents, and digital content on the Web. This KN should aim to find the content from the digital libraries and electronic knowledge houses; navigate digital knowledge glo-

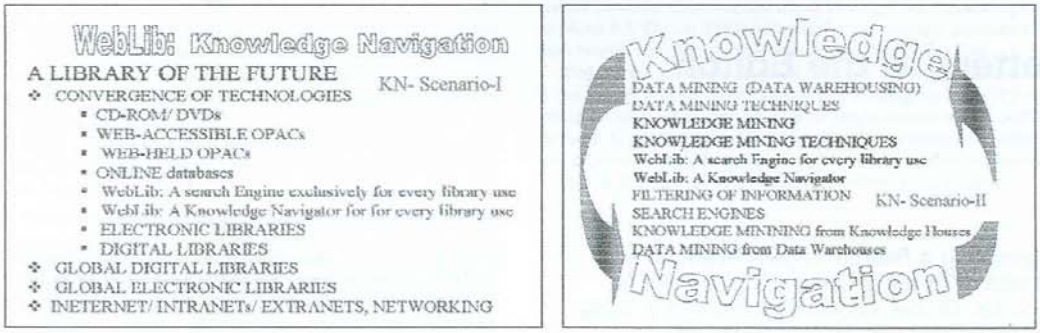

FIG. 1. The concept of Knowledge Navigator (KN) with WebLib in Scenario-I and Scenario-II.

bally; display and access to Web-held OPACs; find global bibliographical information; access to full-text databases from the WWW; retrieve Web-held information and CD-ROM databases; discover the digital content and knowledge from the global electronic libraries and global digital libraries.

Sreenivasulu (2000) depicts the search engine's role in the context of digital libraries, and describes scenario-I (Figure 1, left-hand side) where the convergence of information tools, and information technologies combined with WebLib for digital libraries for exclusively for every library use on the Web to organize the Web-held OPACs, several sets of digital information systems (Sreenivasulu, 1998, 2000) such as CD-ROM databases, Online databases, electronic libraries, several knowledge networks and digital libraries accessible instantly to the global library and information community. The WebLib search engine should be able to organize these information systems on the Web and facilitate the universal access to digital information systems to the global library community. Scenario-II denotes about the browser's (KN) role to discover knowledge from the digital libraries. It denotes the use of browser WebLib: Knowledge Navigator (KN) to mine the knowledge in scenario-II (Figure 1, right-hand side), where global knowledge mining is possible. It should provide a comprehensive KN for the global library community to navigate the global knowledge for digital libraries for exclusively for every library use on the Web to display, retrieve, find, and to discover the Web-held OPACs and digital information. Several sets of digital information systems are accessible instantly to the global library and information community through this browser. The KN has to be an alternative to the existing browsers namely Internet Explorer, and Netscape Navigator, which are designed by the computer scientists, of the computer scientists and for the computer scientists, but not for the exclusive library use of the global library society. KN should provide the features such as improved search options; better display system with bibliographic access points; concept indexing, subject indexing, and multimedia indexing and documentation; content filtering, better filtering options; and keeping up with trends to meet the challenges of modern digital information retrieval.

Organization of digital library content on the Web. By applying and using these concepts such as WebLib and KN theoretical frame work, guidelines; facet analysis and the Dublin core metadata set of fifteen elements, viz. (http://www.lub.lu.se/cgi-bin/nmde.pl or http://purl.org/meta-data/dublincore, the digital library content can be organized, and eventually lead to provide universal access to digital library content to the global digital library society. Additionally, there are some other related concepts can be used and applied for organizing the digital library content. These concepts include, i.e., 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 Cyberdeway, 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 cataloging 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://orc.rsch.oclc.org: 6109/bintro.html). One has to follow an integrated approach of using all these concepts while designing, engineering and development of WebLib and KN to organize and mine the digital library knowledge globally.

Concluding note. By applying all these concepts with an integrated approach towards WebLib and KN, an attempt has to be made to develop and engineer such a search engine exclusively for every library use on the Web and this idea offers broadly universal enthusiasm among the architects, engineers, and information scientists, who are building the next generation of Web search tools and it is an opportunity to library associations and institutions and information science organizations to take lead in doing research on engineering the WebLib and KN with 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 engineering and developing WebLib and KN for every library use on World Wide Web.

This letter finally advocates and targets at the task of engineering and architecting a new search engine—WebLib with Browser-Knowledge Navigator (KN)—for every library use on the Web by applying and using this conceptual theoretical framework and guidelines and this may be considered as a major great contribution to the digital librarianship and information science profession from any country, who ever can build and develop it. The author

ultimately aims to communicate this concept to the global library information community for which I am thankful to Prof. Donald H. Kraft, Editor, *JASIST*.

V. Sreenivasulu
Library
The Indian Law Institute
Opp. Supreme Court of India
Bhagwandas Road
New Delhi 110 001 India
E-mail: drvs67@hotmail.com;
v_sreenivasulu@hotmail.com;
library@ilidelhi.org

Published online 7 November 2002 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/asi.10188

References

Sreenivasulu, V. (1998). Computer and information literacy: Challenges for modern information professional at the advent of XI century. ONLINE & CD-ROM Review: International Journal on Online, Optical & Networked Information, UK. 22(6), 395–397. Online: http://www.learned.co.uk.orl/target2000.asp

Sreenivasulu, V. (2000). Building vehicles of universal access to information: Information tools, techniques, technologies for librarians in the context of Samudramanthan (Information filtering). A paper presented on the occasion of 2nd CV Subbha Rao Memorial Lecture at Ranganathan Research Circle, New Delhi.

Sreenivasulu, V. (2000). The role of a digital librarian in the management of digital information systems (DIS). The Electronic Library: An International Journal (UK), 18(1), 12–20. Online: http://www.mcb.co.uk/portfolio/el/sample or http://www.mcb.co.uk/el.htm

Sreenivasulu, V. (2000). Building competitive intelligence knowledge: Processes, activities in a corporate organization. FID Review, 1(5), 42–49. Online: http://www.kb.nl/infolev/fid/about/news/4–5.html

Sullivan, D. (1998). Search engines features chart. Online: http://searchenginewatch.internet.com/

Woodward, J. (1996). Cataloguing and classification information resources on the Internet. Annual Review of Information Sciences & Technology, 31, 189–219.

The Special Competency of Information Specialists

Sir:

In a new article published in *Journal of Documentation* (Hjørland, 2002), I claim that the special competency of information specialists and information scientists are related to "domain analysis." Information science grew out of special librarianship and documentation (cf. Williams, 1997), and implicit in its tradition has in my opinion been a focus on subject knowledge. Although domain analysis has earlier been introduced in *JASIST* (Hjørland & Albrechtsen, 1995), the new article introduces 11 specific approaches to domain analysis, which I claim together define the specific competencies of information specialists. The approaches are

(1) Producing and evaluating literature guides and subject gateways,
(2) Producing and evaluating special classifications and thesauri,
(3) Research on and competencies in indexing and retrieving information specialties,
(4) Knowledge about empirical user studies in subject areas,
(5) Producing and interpreting bibliometrical studies,
(6) Historical studies of information structures and services in domains,
(7) Studies of documents and genres in knowledge domains,
(8) Epistemological and critical studies of different paradigms, assumptions, and interests in domains,
(9) Knowledge about terminological studies, LSP (Languages for Special Purposes), and discourse analysis in knowledge fields,
(10) Knowledge about and studies of structures and institutions in scientific and professional communication in a domain,
(11) Knowledge about methods and results from domain analytic studies about professional cognition, knowledge representation in computer science and artificial intelligence.

By bringing these approaches together, the paper advocates a view which may have been implicit in previous literature but which has not before been set out systematically. The approaches presented here are neither exhaustive nor mutually exclusive, but an attempt is made to present the state of the art. Specific examples and selective reviews of literature are provided, and the strength and drawback of each of these approaches are being discussed.

It is my claim that the information specialist who has worked with these 11 approaches in a given domain (e.g., music, sociology, or chemistry) has a special expertise that should not be mixed up with the kind of expertise taught at universities in corresponding subjects. Some of these 11 approaches are today well-known in schools of LIS. Bibliometrics is an example. Other approaches are new and represent a view of what should be introduced in the training of information professionals.

First and foremost does the article advocates the view that these 11 approaches should be seen as supplementary. That the professional identity is best maintained if those methods are applied to the same examples (same domain). Somebody would perhaps feel that this would make the education of information professionals too narrow. The counter argument is that you can only understand and use these methods properly in a new domain, if you already have a deep knowledge of the specific information problems in at least one domain. It is a dangerous illusion to believe that one becomes more competent to work in any field if one does not know anything about any domain.

The special challenge in our science is to provide *general* background for use in specific fields. This is what domain analysis is developed for. Study programs that allow the students to specialize and to work independent in the selected field (such as, for example, the curriculum at the Royal School of LIS in Denmark) should fit well with the intentions in domain analysis.

In this connection it should be emphasized that the 11 approaches are presented as *general* approaches that may be used in about any domain whatsoever. They should, however, be seen in connection. If this is not the case, then their relative strengths and weaknesses cannot be evaluated. The approaches do not have the same status. Some (e.g., empirical user studies) are dependent on others (e.g., epistemological studies).

It is my hope that domain analysis may contribute to the strengthening of the professional and scientific identity of our discipline and provide more coherence and depth in information studies. The paper is an argument about what should be core teachings in our field. It should be both broad enough to cover the important parts of IS and specific enough to maintain a special focus and identity compared to, for example, computer science and the cognitive sciences. It is not a narrow view of information science and on the other hand it does not set forth an unrealistic utopia.