Journal Literature on Digital Libraries: Publishing and Indexing Patterns, 1992–1997

Susan Davis Herring

This study examines publishing and indexing patterns in the area of digital libraries to identify the core journals and the most effective descriptor terms for online searching. A group of 232 citations on digital libraries from 1992 to 1998 was analyzed. Results show that none of the 105 journals listed has emerged as the leading title in the field and that, although two-thirds of the citations were identified by one of two phrases, indexing is inconsistent and subject searching unreliable for full retrieval. The study confirms that skill, persistence, and wide reading are required to keep up with developments in this evolving field.

The number of journal articles published on the topic of digital libraries has proliferated in the past several years. Unfortunately for those attempting to keep up with the subject, the literature is scattered across a number of disciplines, primarily library and information science and computer science. Although it is difficult for specialists in any one field to be familiar with the literature in other disciplines, such shared knowledge is necessary for the successful development of new information technology. Using the same terminology and understanding the same mental models are necessary if computer scientists, information scientists/librarians, and subject specialists are to cooperate in the advancement of digital libraries. One of the best ways to build this understanding and common knowledge is through joint familiarity with the major journals that contribute to the literature in the digital libraries field.

The current research study was designed to explore publishing and indexing patterns in the area of digital libraries. The major purpose of the study was to identify the core journals, those that publish the majority of articles on the topic, in order to help professionals focus their time and attention on the most prolific journals. A second purpose was to examine the descriptor terms used by the organizations that index the articles on digital libraries, to both identify the major terms and reveal inconsistencies in indexing. The study was limited to substantive journal articles, including empirical, theoretical, and practical studies and excluding book reviews, product reviews, and directory lists.

Literature Review

A review of the literature revealed that no study of publishing patterns in digital library literature similar to the current study has been published to date. How-

Susan Davis Herring is an Associate Professor of Bibliography and an Engineering Reference Librarian in the M. Louis Salmon Library at the University of Alabama in Huntsville; e-mail: herrings@email.uah.edu.

ever, the methodology used here was detailed by F. W. Lancaster and Ja-Lih Lee in their discussion of the use of online databases to track emerging issues.[1] They demonstrated the usefulness of conducting online searches in groups of subject-specific databases as a means of tracking the growth of journal literature on particular topics and identifying the core journals in the field. A similar methodology was used by Terrence A. Brooks, who conducted an analysis of the literature in superconductivity through fifteen years of INSPEC files to identify the core journals in that field.[2]

Chun Wei Choo noted that the successful development of information systems requires coordination among domain experts, information technology experts, and information experts; this could be applied to digital library development as well.[3] In a 1997 paper, Lisa Dallape Matson and David J. Bonski stressed the importance of communication between library and computer specialists working on digital library development.[4] They also pointed out the dichotomy between the vocabularies used in library science and computer science literatures, which can easily lead to misunderstanding and a breakdown in cooperation.

An earlier citation analysis by this researcher demonstrated the lack of cross-fertilization in the literature on the development of Internet search engines.[5] Over 40 percent of the articles analyzed appeared in computer science journals, whereas only 20 percent of the articles studied appeared in the traditional library literature. Over 80 percent of the computer scientists writing on the subject published only in computer science journals. Patrick Wilson has documented the common practice of ignoring relevant literature when it does not fall within the defined discipline.[6] However, in the burgeoning and highly interdisciplinary field of electronic information access, limiting oneself to the literature of one's own field could lead to the neglect of a great deal of applicable and appropriate literature.

The Study

The current study focuses on a body of citations on the subject of digital libraries that were identified through computerized searches in a variety of electronic indexes. The study was limited to the period from 1992 through 1997. Earlier research in the field indicated that, prior to 1992, little had been written other than theoretical essays. Because the data were gathered very early in 1999, 1997 was the last year in which complete indexing might be expected.

Two research questions were posed for this study:

1. What are the leading journals in the area of digital libraries? For the purpose of this research, "leading journals" are defined as those that publish the highest number of articles on the topic.

2. Which indexing terms return the largest number of and most relevant articles on digital libraries in electronic indexes?

Data were gathered using a selected group of electronic indexes to get a broad and representative selection of the literature. Searches were run in *Library Literature*, *Humanities Abstracts*, *Social Sciences Abstracts*, *Wilson Business Abstracts*, *General Science Abstracts*, *Applied Science and Technology Abstracts*, *Readers' Guide Abstracts*, and *PAIS International*. It is recognized that results from searches in these databases are not comprehensive; however, they will identify articles from the major professional and popular journals that are most likely to be available to readers and researchers in the United States. The search words "digital" or "virtual" or "electronic" adjacent to "library" or "libraries" were used in order to cover all possible terminology. Book reviews, product reviews, and directory listings of Web sites were deleted, and duplicate records were combined, leaving a total of 232 articles. The data were entered into an Excel spreadsheet for ease of analysis.

Data Analysis

The initial question concerned which journals were the major publishers of ar-

Table 1. Publications Requiring... n

Journal	1 2	1 3	1 4	1 5 n	1 6	1 7	t 1
C ... ACM	–	–	–	23	–	1	24
I ... Tech	3	2	1	2	1	1	10
D ... s	–	2	–	3	1	3	9
A ... l	–	–	–	3	1	4	8
L ... e	1	1	2	1	1	3	7
S ... r	–	2	2	2	–	1	7
C ... A	–	1	1	1	3	–	6
J ... p n	–	–	–	–	4	2	6
C ... Tech	–	–	2	2	1	–	5
J ... , t	–	–	–	–	5	–	5
Ti ... y	–	–	1	1	1	2	5

ticles on digital libraries. Of 105 publications listed, only forty-five (42.8%) published more than one article on the topic during the six-year period studied. The top eleven, each of which published five or more articles during that period, are listed in table 1. Twenty-two of the twenty-four articles published by the top-ranking journal, *Communications of the ACM*, appeared in one special issue in 1995. Discounting that special issue, *In-*

formation Technology and Libraries becomes the leader, with ten articles in the six-year period. The steady increase in the number of articles on digital libraries being published is obvious from the data, as shown in table 2. The higher number for 1995 reflects the special issue of *Communications of the ACM* mentioned above.

An additional twelve journals each published three or four articles indexed during this period, including *Journal of the American Society for Information Science* (with four articles), *D-Lib*, *College and Research Libraries*, *Information Processing and Management*, *Library Trends*, *Online*, *Program*, and *Scientific American* (each with three articles). Some of the journals coming in with two articles are *ACM Transactions on Information Systems*, *Computers in Libraries*, *Electronic Library*, *Library Hi Tech*, *IBM Systems Journal*, and *Searcher*, as well as *Forbes*, *Fortune*, and *Science News*.

Inconsistencies in indexing became obvious as the data were analyzed. A summary of the most frequently used subject headings/descriptors appears in table 3. Despite the fact that all but one of the databases searched are produced by the same provider, indexing terms varied remarkably across the databases. Even the primary descriptor for digital libraries differed. *Library Literature* uses "Virtual Library," *Wilson Business Abstracts* and *Applied Science and Technology Abstracts* use both "Digital Libraries" and "Virtual Library," and *Readers' Guide Abstracts*, *General Science Abstracts*, and *Social Sciences Abstracts* use "Virtual Libraries," as does *PAIS*. Of the 232 articles analyzed, 158 were identified with one or more of these three terms; 74 were indexed using other terms entirely.

The number of terms assigned to articles varied from a low of one to a high of eight, with a mean of 2.82 and a mode of two. Thirty articles received only one term; the majority of these were either "virtual library" or "digital libraries."

TABLE 2
Articles Published per Year, 1992–1997

1992	1993	1994	1995	1996	1997	total
10	20	19	62	56	65	232

Other single terms included "libraries," "Internet," and, perhaps most surprisingly, "intellectual capital."

Discussion

The data clearly show the increasing rate of publication of articles on digital libraries. More than six times as many articles were published in 1997 than were published in 1992. The simple proliferation of articles makes it more difficult for the professional to maintain awareness of research in the field. Moreover, this difficulty is compounded by the broad range of publications in which the literature appears.

Analysis of the data gives no clear answer to the first research question, What are the leading journals in the area of digital libraries? It is obvious that no single journal can be considered the leader in publishing these articles. At the least, the professional wishing to remain knowl-edgeable about current issues would be wise to scan *Information Technology and Libraries, Database, Computer, Science,* and the two standard library science journals, *American Libraries* and *Library Journal*, on a regular basis.

Communications of the ACM, an important scholarly journal in computer science, only occasionally includes articles dealing with the topic. However, when they appear, the articles are important, numerous, and in-depth, and well worth reading. *Science* is another well-respected and scholarly journal that is worth regular perusal by the librarian, information specialist, or computer scientist. The *Journal of Academic Librarianship* is the only library science title on this list that specializes in research articles. The *Journal of Information, Law, and Technology* also is potentially useful because of its legal focus.

The alert reader will immediately notice that this list does not include some journals that might be expected to appear. The outstanding example is *D-Lib*, the electronic journal that focuses specifically on advances in digital libraries. However,

TABLE 3
Indexing Terms and Occurrence Rate

Index Term	1992	1993	1994	1995	1996	1997	total
Subject-specific Terms							
Virtual Library	1	8	8	15	27	24	83
Virtual Libraries	-	1	2	3	4	5	15
Digital Libraries	1	4	3	27	10	23	68
Other Terms							
Libraries	1	-	1	5	6	15	28
Information Systems	1	2	2	5	6	11	27
Internet	-	1	1	8	5	9	24
Electronic Publishing	-	-	-	4	8	6	18
Automation of Library Processes	2	4	-	2	5	-	13
Information Retrieval	2	-	2	5	1	2	12
World Wide Web	-	-	-	2	4	5	11

of the databases searched, this journal was indexed only by *PAIS International. PAIS* began selective indexing of *D-Lib* in 1997, although the journal started publication in 1995. Only three articles from *D-Lib* were identified in this study.

The simple difference between "library" and "libraries" could make the difference between an adequate search and one that is barely successful.

The second research question concerned the most effective indexing terms. The results of this analysis indicate that two phrases, at a minimum, are required for reasonable results. Using only "Digital Libraries" would have returned less than one-third of the total; and "Virtual Libraries" alone would have returned less than 10 percent. Even broadening the search and truncating the descriptor terms to "Virtual Librar* OR Digital Librar*" would miss almost one-third of the final results. This inconsistency in indexing terminology could present a serious problem for the naive searcher who is unaware of the various possibilities or the specialist who thinks in strictly defined terms. Not all users understand either boolean operators or truncation. The simple difference

between "library" and "libraries" could make the difference between an adequate search and one that is barely successful.

Perhaps more surprising was the number of articles that received only one indexing term. Although these comprised only 13 percent of the total, it is still disappointing that the indexers found no other, perhaps more useful, terms.

Conclusions

The digital library is a new phenomenon, and the present study has confirmed that the literature on this topic is still evolving. As is normal in other advancing fields, indexing terminology is still evolving and common terms are only beginning to be established. Clearly, the person wishing to remain current on the subject must cast a wide net in an attempt to retrieve relevant publications.

The study described here focused on only one part of the literature in this field. A great deal of the discussion and research on digital libraries is appearing outside the traditional journal format, primarily in conferences and online publications. It might be illuminating to conduct a study similar to this one examining the publication patterns and means of access to conference proceedings and online papers.

Notes

1. F. W. Lancaster and Ja-Lih Lee, "Bibliometric Techniques Applied to Issues Management: A Case Study," *Journal of the American Society for Information Science* 36, no. 6 (Nov. 1985): 389–97.

2. Terrence A. Brooks, "Core Journals of the Rapidly Changing Research Front of 'Superconductivity,'" *Communication Research* 16, no. 5 (Oct. 1989): 682–94.

3. Chun Wei Choo, *Information Management for the Intelligent Organization: The Art of Scanning the Environment*. (Medford, N.J.: Information Today, 1995).

4. Lisa Dallape Matson and David J. Bonski, "Do Digital Libraries Need Librarians?" *Online* 21, no. 6 (Nov.–Dec. 1997): 87–92.

5. Susan Davis Herring, "The Value of Interdisciplinarity: A Study Based on the Design of Internet Search Engines," *Journal of the American Society for Information Science* 50, no. 4 (Apr. 1999): 358–65.

6. Patrick Wilson, "Unused Relevant Information in Research and Development," *Journal of the American Society for Information Science* 46, no. 1 (Jan. 1995): 45–51.