BORNO LIBRARY, ARCHIVAL AND *Information Science Journal*

ISSN: 1596 - 4337

JOURNAL OF THE NIGERIAN LIBRARY ASSOCIATION
BORNO STATE CHAPTER

VOL. 2 NO 2 2003

EDITORIAL BOARD

EDITOR-IN-CHIEF

Dr. Emmanuel M.K. Dawha
Department of Library Science,
University of Maiduguri, Nigeria.

EDITORS

Dr. Emmanuel Camble
Department of Library Science
University of Maiduguri
Maiduguri, Nigeria

James A. Agaja
Ramat Library
University of Maiduguri
Maiduguri, Nigeria

Dr. Ibrahim Haruna
Department of Library Science
University of Maiduguri
Maiduguri, Nigeria

Mr. E. N. Hameen
Ramat Library
University of Maiduguri
Maiduguri, Nigeria

Idris M. Yalwa
Borno State Library Board
Maiduguri, Nigeria

Mrs. Fatima L. Ibrahim
Department of Library Science
University of Maiduguri, Nigeria

BUSINESS MANAGER

Yusuf Makinta
Department of Library Science
University of Maiduguri

CONSULTING EDITORS

Professor C. C. Aguolu
Department of Library Science
University of Maiduguri
Maiduguri, Nigeria

Professor M.I Ajibero
University Librarian
University of Ilorin
Ilorin, Nigeria

Professor M.G. Ochogwu
Department of Library Science
Bayero University Kano
Kano, Nigeria

Dr. (Mrs.) M.I Atinmo
Department Library
Archive and Information Studies
Ibadan, Nigeria

Dr. S.O. Olanlokun
University of Librarian
University of Lagos, Lagos

Dr. John Agada
Emporia State University
K.S. USA.

Borno Library, Archival and Information Science Journal is the Official Learned Journal of the Borno State Chapter of the Nigerian Library Association. It is published twice in a year (January and July). Each year's issues makes one volume. All editorial and general communications should be addressed to the Editor in Chief.

Borno Library, Archival and Information Science Journal, (Vol. 2, No. 2, 2003)

CD-ROM DATABASE SERVICES AND VARIABLES OF SIGNIFICANCE

BY

Obaje Alfred Michael

University of Jos Library, Jos Nigeria

Abstract

The computer is a tool that can greatly assist and speed up human activities but it is not a substitute for intellectual activities. In information retrieval the intellectual aspects of searching remain the prerogative of the searcher while the repetitive, routine and non-intellectual tasks are handled quickly and effectively by the computer. Libraries that operate computer-based information services provide an intermediary between a user and the computer-based system. The intermediary may be an information specialist, information scientist, or reference librarian. He or she handles the intellectual tasks of: selecting appropriate system and database (s) for user's question; negotiating the search question with the user; developing the query or profile with an effective search strategy; conducting the search sometimes; and possibly evaluating the output. In order to maximize the benefits provided by the use of CD-ROM databases, it is of importance that the variables of significance in the use be properly known by information seekers and intermediary..

Introduction

We are in the midst of the Information Age where we are dependent on quality information in our daily and professional lives. However, information is becoming increasingly voluminous and fragmented into different formats and media, as well as duplicated in multiple physical locations. In order to access and use these myriad sources effectively, libraries must be properly equipped and library staff and patrons must be computer literate. As defined by the American Library Association (ALA) in its mission statement for the global information society, 21st century information literacy is the ability to seek and effectively utilize information resources which includes, knowledge of how to use technologies and the forms in which information is stored (ALA, 1998). This means that asking a good question, as well as accessing, locating,

© Nigerian Library Association, Borno State Chapter, 2003

evaluating, and using information is critical not only in scholarly activities but in making daily decisions

In experimental, observational, or survey studies the researcher tries to understand the relationships between the variables involved. Particularly in experiments, the researcher looks for changes in the dependent variables that occur as a consequence of change in the independent variables. For example, one may want to test the relationships between speed (the dependent variable) and the familiarity with the database searched (the independent variable). Assuming that other factors are either equal to or of no consequence, one can set up an experiment to measure the speed with which a certain request is searched in one CD-ROM database by searchers with different degrees of familiarity with the database. The results of such an experiment show the way in which speed of searching varies with the degree of familiarity with the database researched.

The dependent variables in CD-ROM database retrieval studies usually relate to the search process and/or the search outcome; for example, speed of searching and the number of database searched are variables characterizing the search process. Many researchers are interested in identifying the effects of selected independent variables on elements of the search process and its outcome; they choose to investigate certain independent variables because they believe that these variables have some effect on the search process and/or its outcome.

Search outcome variables are always dependent variables while search process variables are used as independent variables, for example the dependence of search outcome variables on search process variables (e.g., the number of descriptors used); the dependence of search process variables (such as the number of databases researched) is the type of request, without measuring search outcome variables. Shaw (1996) analyzed dependence of both search process variables and the search outcome variables on the level of

experience of the searchers, but also suggested dependence of search outcome variables on search process variables. Although there is obviously a relationship between the search process and its results, we know very little about the nature of the relationship and further research is needed (for such studies, search process variables are the independent variables).

Deborah(1996) describes a study that observed undergraduates as they searched bibliographic databases on a CD-ROM local area network. Topics include related research, information needs, evolution of search topics, database selection, search strategies, relevance judgments and CD-ROM interfaces.

Evaluation of a Database

The user of database services will need to evaluate not only the searching methods available but the content of the database itself. The content coverage of data bases may be discipline oriented, mission oriented, problem oriented or multidisciplinary. In evaluating them, a library must know how their subject coverage matches the objectives and the breadth of its own collection. The following questions are necessary in evaluation: Does the database cover material such as government reports, academic journal articles, patents, monographs, theses, reprints and news items? If it does, how complete is this coverage? That is, if it claims to cover a particular journal, will it be covered in its entirety or only for selected issues or articles? Some database producers provide lists of the journals and other items indexed.

Another measure of evaluation is the time lapse between the item's appearance in the primary source, in the secondary source (or index), and finally in the database. In some cases a citation appears on a tape before it is produced in a hard-copy because the hard-copy publication is produced from the tape.

The indexing and coding practices is another important consideration. Does the data base include free-language key word on the tape? Does it include a controlled thesaurus or hierarchical vocabulary terms: are titles given exactly as the author provided them, or are they augmented titles in which additional terminology is added to the author's title? Does it include other kinds of codes to indicate subject matter or any other criteria about the item itself? Are abstracts and extracts available on tape for search and display, or will the library have to go back to the hard copy to obtain them?

The number of citations from one year's accumulation of a file is known from the size and growth of the database. It is crucial to know how the tape version corresponds with the hard-copy version. In some cases, there is a one-to-one correspondence; that is, for each abstract or reference contained on the hard copy there is a tape representation. In many cases, the database itself is a subset of the hard-copy version; or the reverse may be true, that is, the database may contain more citations than the hard-copy. In other cases such as the MARC tapes from the library of congress, these is no corresponding hard-copy publication (except the collection of LC cards). If there is a corresponding hard copy, a library can occasionally do both computer and manual searches of an issue as a cross - check to be sure it is using the right terminology and really getting what is needed.

As a processor, the library must investigate the consistency and quality control exercised by the database supplier, and must also be aware of the frequency with which changes are made in the data. Additions of new data element affects processing time, and possibly require a change in the search program. Adherence to the delivery schedule is another concern of the library. If the database supplier sends his tapes late, then the library will be delayed in providing output to its patrons.

The consideration of subject overlap in data bases coverage is very important that there are costs associated with intellectual processing (indexing

and abstracting) and manual inputting of citations. If the same citation is handled more than once, this can represent wasted time, effort and money. There are a few processing centers that merge several databases to create one common database.

The last and much more technical-processing concern is that of compatibility between various databases. The variability among the databases complicates handling for those who process them. The standard arrangement of data element tags, data content, and directory information for the records is referred to as the format of the record, and the arrangement of the records on a tape or other media is referred to as the file structure or file format. There are almost as many database formats as there are databases. This leads to confusion and, of course, added expense in processing tapes, because it requires the processor of multiple tapes to either develop multiple search programs or to reformat all in-coming tapes to one standard format. The American National Standards Institute has developed important standard for interchange or transmittal of bibliographic records: the "American National Standard for Bibliographic Information Interchange on Magnetic Tape" The MARC implementation of the standards has been proposed as a Federal Information Processing Standard. This standard deals only with the format for records on tape or the generalized structure, not with the contents of the records.

Database Search Services

The services most often provided are Selective Dissemination of Information and Retrospective searches in either the off line or on-line. An additional database service offered by a few organizations is a private library service. A private library service is one that permits the user to create his own machine readable file either by designating that output from his SDI runs be stored on his own tape or disc file, or by specifying other records (e.g items

selected via his own library searches) he would like to have in his file. He may have its SDI output saved for several weeks or months until he wants to look at it and then he can indicate which of the records should be retained for his subsequent use. The advantage of the system lies in the fact that every record in the user's file represents his own judgment.

All too often a searcher completes a successful search only to be frustrated by the inability to readily obtain needed documents. The process of document delivery includes two major components: the identification of the source location of the document, and acquisition of the document. Delays associated with either or both occur often. Generally, it is left to the user to go to his library to obtain copies or inter-library loan use of a document. Two database producers: ISI (Information Science Institute) and NTIS (National Technical Information Service),simplified the document acquisition process for documents cited in their own databases. In the case of ISI, on-line users of their databases through locked or Systems Development Corporation (SDC) are now able to use the accession number to order full text copies of relevant items through ISI's Original Article Tear Sheet Service (OATS). A specific command is provided for the searcher to order copies of desired items directly on-line through the system. The article order is subsequently transmitted to ISI headquarters in Philadelphia where the orders are filled. Ordered items are mailed to the requester within 24 hours. A similar capacity is available for on-line ordering of NTIS documents.

Possibilities for solving the document delivery problem include:

1. database producer's maintain copies of all documents cited in his database;
2. on-line searching of one or several union lists of serials, and Union lists should be prepared on a national, state, or regional basis;

3. a national serials resources center to function as a central depository. Any one of these solutions would simplify the problem of knowing where to find the document.

The more common solution – the use of copying equipment – appears to be the easiest way of producing copies, though the legality still remains ambiguous. Facsimile transmission is used in cases where fast delivery is mandatory but this technique is still very expensive.

SDI is one of the most successful information services developed in the past decade. A number of factors have led to the development of SDI. Increased availability of computers; the automatic generation of databases through computer typesetting; expansion of the world's literature; and the increasing cost of labour-intensive information services. SDI has also been the primary use to which databases were put in the early years of database development.

Database search services may be used directly by the end-user or indirectly through centers or brokers. The use of the term "center" refers to organizations that acquire and process databases themselves and provide services to users who may be within their own organization or outside of it. The term "broker" refers to a person or organization that searches databases on-line at another location for its own customers. The broker does not process the database but does provide search services from them. Obviously, the use of Centre or broker involves some additional cost. The added cost for the middleman must then be offset by the added value provided by the middleman. The added value may consist of:

(1) augmentation, analysis, screening or interpretation of output;

(2) know-how in effectively using other search services.

(3) Knowing where to go to find the appropriate service or data base;

(4) document location and delivery;

(5) reduction of the purchaser's need for additional personnel with specific skills where the need for such skills may be sporadic; and

Database processing centers may be independent commercial organizations, they may exists in computer centers, libraries or information centers of various sorts. More often than not, the processing of databases has been done outside of the library setting. The brokerage situation though is different, because the brokerage organization needs little investment for equipment and has no need for programmers and computer specialist. Reference librarians or information specialist hired by libraries or information scientist operating independently can establish a search service for users with little capital investment. They can effectively function as intermediaries between users and the systems.

Variables of Significance in CD-ROM Services

In this paper, the following variables will be considered: the setting, the user, the request, the database, the searcher, the search process, and the search outcome.

The setting

The type of parent organization is usually described according to one or more of five facets: the organization's orientation (research, education, etc); whether it is for-profit or not-for-profit organization; organization affiliations (government, commercial, etc) of particular importance are the charging policies and sources of funding for searching.

Another aspect is the position of the searching unit within the structure of the organization. Management attitude towards CD-ROM databases searching and the length of time provided for the users are all part of the setting variables.

The user

The user is a person who has an information need and who initiates a search. Whether the user decides to submit his request to a searcher or prefers to conduct the CD-ROM databases search himself, a description of this person is important for a better understanding of request.. User characteristics such as cost behaviors and education, whether perceived by the searcher or not, may have an effect on the user searcher interaction and, in turn, may influence the search process. The user's prior experience with computer-based information retrieval and the users attitude toward computer-based retrieval are factors very important for consideration in CD-ROM databases search.

The Request

The term "request" includes any specific attribute a searcher may take into consideration before and during the search, and any specific characteristic that may aid in a better understanding of the information need. A request is submitted by a user and is processed into an image by the searcher. The distinction between the user's statement of the request and the searcher's image of the request is very important. The query is that part of the request that can be expressed by a formulation.

The search usually begins with initial plan called "search strategy". It is assumed that in planning the search, the user meets with a subject or reference librarian to develop the initial search formulation before the actual search. The degree to which the request is formulated before logging or may differ from one search to another. The query formulation consists of terms to search and the logic used to combine them. In query formulation process, three important variables are: the degree to which the searcher can predict the nature of the resulting set of references, the amount of citations the searcher expects to retrieve, and the degree to which the expected amount of retrieval determines the nature of the query formulation.

The degree to which the query could be translated into a Boolean expression and the extent to which the concepts used could be translated into the system vocabulary should also be taken into consideration. Finally, the information given by the user when submitting the request is important; does this information include a written statement of the query, the names of authors of particular interest, precision-recall requirements, and preferred databases to search?

The Database

CD-ROM database is an organized collection of information in machine-readable form. The collected information may be bibliographic, or bibliographic-related; natural language text; numerical; or representational.

Database attributes include coverage; frequency of updating; availability of thesauri, dictionaries, cross-reference listings, and scope notes for terms; cost; and type of vocabulary. The vocabulary may be controlled, with or without hierarchical structure, or uncontrolled. When the vocabulary is controlled, the index terms may be formulated in natural word order, in inverted word order (important when using word adjacency in free-text searching), or the terms may be formulated to consist of a heading/ sub-heading structure. Two possibilities exist for indexing: either index terms are specifically assigned by human indexers or a computer program, or the terms supplied by the author and/or abstractor in the title, abstract, or full text are used.

The Searcher

Characteristics of searchers are widely described in the literature. They range from personality attributes to detailed analyses of searching experience .Of the attributes that are easily defined, only few have proven to have any effect on databases retrieval. The characteristics are related primarily to personality traits, cognitive factors, demographic variables and education (e.g) academic degree, major field or study).

An "average" searcher should have attributes such as: a logical (analytical) mind, flexibility, adaptability, and imagination. Training and experience are most frequently mentioned in the literature. Training variables include the type of initial training (formal versus informal) and the duration of training. Descriptions of searcher experience may refer to experience within the same setting or to total experience.

The attitude of the searcher towards databases searching seems to play an important role in his searching behaviors. The general attitude of the searcher may be determined by factors such as his perception of the utility or value of the information, interest in and enthusiasm towards searching and a sense of professionalism.

The Search Process

The term "search" as used here includes any activity performed by searchers in order to provide users with retrieved information to satisfy their needs. Searching is a process of accumulating information and the order in which the information is acquired by the searcher is likely to determine many elements in the search process.

Interaction with the user may occur at the beginning of the search, during or at the end of it, when there is an interaction during the search, it is important to mention whether the user is present at the terminal. The main part of the interaction with the user may take place at the reference interview.

Generally speaking, the interaction process concentrates on transformation of the request (e.g., from a broad one to a narrow one).

For a good search process the following are very necessary steps.

1. Formulate the search question in a clear and concise search statement.
2. Analyze the search question into component aspects, which can be represented by keywords or descriptors (search terms).
3. Develop and expand the search statement by means of synonyms, broader terms and terms related to the descriptors, so that all possible aspects are covered.
4. Limit the search to suit the purpose of the search (project, application, thesis etc.), the subject area and the time span required.
5. Search for the relevant information.
6. Evaluate the results of the search.
7. If necessary, modify and alter the search strategy.

The Search Outcome

Search outcome variables (which are always dependent variables) refer usually to the "quality" of the retrieved set of information. The measure of outcome involves judgement of relevance or utility of retrieved materials. User's satisfaction is also a good measure of search outcome. It is very important for subject or reference librarians to strive to discover outcome measures that relate directly to requirements arising out of problem situations of specific user(s). Lack of precision in defining the real information need can lead to poor search results.

Conclusion

Having the right information determines the difference between the rich and the poor in the Information Age. Universities can make a vital contribution toward closing this gap by equipping their libraries with new information technology facilities and let staff and students have access to retrieve and utilize information. Database services are very important part of services provided by libraries world wide. In order to provide effective services and also satisfy users of CD-ROM databases, reference librarians must acquit themselves with different types of databases, evaluation, techniques involved in search strategies and the major variables of significance in the use of CD-ROM databases. This will go along way in improving services provided by libraries in this age of information technology.

References

Adeniran, O.R.& Adigun, T.A. (1994). Availability and use of CD-ROM products in Nigerian libraries and information centres. Electronic library 12(3), 155-168.

American Library Association. (1998). Planning document "New Visions: Beyond ALA Goal 2000."

Awogbami, P. A. (1991). The diffusion of CD-ROM into Nigerian libraries. CD-ROM librarian 7(5), 30-33.

Ercegovac, Z. (1998). Information Literacy: Search Strategies, Tools & Resources. Los Angeles: InfoEN Associates.

Ercegovac, Z. (1997) "The interpretation of library use in the age of digital libraries: Virtualizing the name." Library & Information Science Research, 19 (1), 31-46.

Fallon, H. (1998). CD-ROM: Introduction, problems and evaluation In: An Leabharlaan, 14(1), 3-17. Ireland.

Huang, Samuel. T. (1989). CD-ROM databases: Unique resources for libraries. Laserdisk professional 2(2), 62-67.

Igbeka, Josephine. U. & Okpala. Angela. W. (2004). Analyses and Patterns of CD-ROM Database Use in kenneth Dike Library, University of Ibadan. The Information Technologist 1(1)

John, M. Budd.,& Karen, A. Williams.(1993). CD-ROMs in Academic-libraries: A survey. College and Research libraries 54(6), 529-535.

Karen, Sandlin .S. (1990). CD-ROM in Libraries: Access, Trends and challenges. Serialslibrarians 17(3), 52.

Kuhlthau, C.C. (1993). Seeking meaning: A process Approach to Library and Information Services. Norwood, NJ: Ablex.

Mackellar, M., & Elliott, F. (1999). The importance of training CD-ROM end user. Computers in libraries 19 (2). 63-66.

Meyer, Richard .W. (1993). Locally mounted databases: making information as close to free as possible. Online 16 (1), 17-24.

Mosuro, T. (1996). CD-ROM technology and its application of Kenneth Dike Library, university of Ibadan. Nigerian library, Archives and Information Science Review 14 (1 &2). 46-50.

Oduwole, A. A. (2000). A study of the use of CD-ROM Databases in Nigerian academic libraries. Journal of Information Science 26 (5). 364-369.

Shaw, Deborah (1996). Undergraduate Use of CD-ROM Databases: Observations of Human-Computer Interaction and Relevance Judgments. Library & Information Science Research; 18 (3) 261-274 .