

ISBN 1-84544-323-3 ISSN 0264-0473

Volume 23 Number 3 2005

The Electronic Library

The international journal for the
application of technology in
information environments

**ICT developments in Nigerian
libraries**

UNIVERSITY OF IBADAN LIBRARY

www.emeraldinsight.com





Analysis of users' searches of CD-ROM databases

Angela E. Okpala

Department of Library, Archival and Information Studies, University of Ibadan, Ibadan, Nigeria, and

Josephine U. Igbeka

Faculty of Social Sciences, University of Ibadan, Ibadan, Nigeria

Abstract

Purpose – This study investigates the behaviour of searchers of CD-ROM databases in the Kenneth Dike Library of the University of Ibadan, Nigeria. The aim of the study was to identify the behaviour of end-users with different amounts of overall experience in searching electronic databases in order to identify the knowledge of their search systems and syntax, and its effect on their search results, for the purpose of improving user education.

Design/methodology/approach – All end-users who conducted searches during the period of the research were the subjects of the research. A well-constructed questionnaire, including interviews and observations, were used to collect the requisite data. The variables included in this study are user variables, search process variables and search outcome variables. Data were analyzed using descriptive statistics.

Findings – The results show that logical operators "AND" and "OR" are significantly used for searching by end-users; that search results are dependent on search strategy employed by the user. The more sophisticated the search strategy, the greater the result in terms of number and relevance obtained by the searcher. The result also shows that there is no relationship between the level of education of users and quality of search strategy.

Practical implications – The implication of the study is mainly the importance and necessity of training for end-users in CD-ROM literature search.

Originality/value – The uniqueness of the study is based on the fact that certain strategic methods are required for a successful CD-ROM literature search and that this does not depend on the level of education of users. This paper is therefore important and applicable to all end-users who are interested in conducting CD-ROM literature search.

Keywords Online databases, User studies, Information retrieval, Nigeria

Paper type Research paper

Introduction

The mass storage capability of CD-ROMs enables librarians and information centers to offer in-house access to large databases formerly available only in print or online. The discs are extremely stable and not susceptible to scratches or knocks. Tropical temperature, high humidity, dusty conditions, and power outages found in many developing countries do not affect CD-ROMs as they do printed materials, though these conditions can affect a microcomputer work station. The fact that CD-ROM databases are available on a set free subscription basis, and CD-ROM systems are normally self-contained workstations, institutions can directly predict and control the costs of online services. The more a CD-ROM is searched, the lower the cost is per search. CD-ROM systems allow end-users to do their own searching, thus freeing time for members.



In spite of the manifold advantages that the use of CD-ROM holds, its implementation is not without problems. Research studies have indicated that many end-users do not understand basic search concepts and consequently, do not employ effective search strategies when using these databases. Learning basic database design and effective search strategies allows end-user to take the often neglected first steps to successful electronic searching, training in the use of new technologies and new resources. This therefore, forms the first area that must be addressed during implementation. There are different kinds of users with different information needs, as well as different levels of knowledge and abilities to satisfy their own information needs. These differences are manifested in various approaches to interaction between users and computers in searching electronic databases.

There is a vast and ever growing field of information, which makes research in interaction between user and computer very important. The purpose of such research is to understand the cognitive searching process, with the aim of designing information systems that will enable simpler and more efficient searches. During searching, once the searcher understands the request well enough to answer it a plan is developed for the search – a search strategy. This strategy specifies which terms (or search keys) to be used. An example of a search strategy is logical operators, which are used to place individual terms or concepts in mutual relationships. Boolean logical operators and proximity operators are widely used by databases. These operators enable precise searching by connecting two or more concepts within one field or one sentence and generally improve retrieval performance.

There are two distinct types of search key: text words which are used in free-text searching and descriptors from a controlled vocabulary. Information retrieval systems generally use a controlled vocabulary to index documents that are found in databases. The thesaurus is one of the most frequently used forms of controlled vocabulary. It can prevent the separation of related material under synonymous terms, distinguish among homographs, and assist the searcher in comprehensive searching of a particular subject area. Searchers may select single words or phrases from any field or fields in the database record: the matching of any term or phrase in any of these fields is referred to as free-text searching (or natural-language retrieval). Natural language retrieval is also more specific than a controlled vocabulary. In electronic databases searching, in addition to the establishment of strategy, concepts and logical operators, a very important indicator of the knowledge about searching systems with the aim to increase efficiency (e.g. greater recall or time saving) is the use of truncation, sets and possibilities offered by user interface, onscreen help and instructions.

Virtually all databases have truncation facilities. In order to avoid unnecessary typing of individual words or concepts in different combinations during searches databases provide equivalents of word concepts. These equivalents are called sets, and they are represented by a symbol, typical for different software. On-screen help and instructions are especially important in using CD-ROM databases. Good software provides help and instructions at all search levels. The actual benefit of instructions for the user depends on a number of factors related to user's cognitive structure and the type and clarity of instructions provided by the system.

CD-ROMs were introduced in developed countries at least two decades ago and much research was done at the time into their use. For example, Siegfried *et al.* (1993) researched the behavior of a group of 27 scientists from the humanities during online database searching. The search techniques of these scientists were analysed, the features they used and their learning curves were also studied. Similarly, Hsieh-Yee

(1993) studied the influence of domain knowledge experience in searching of novice and experienced users to search strategies and found that search experience influenced search strategies and more experienced users apply more tactics. On the other hand, the use and subsequent study of CD-ROM databases in areas like Africa or India by end-users in libraries are still relatively new.

Thus the primary purpose of this study is to investigate the searching behaviour of users in the Kenneth Dike Library of the University of Ibadan, Nigeria. In the study the search behaviour will be followed up on the basis of qualitative and quantitative analyses of strategies and search results of users who differ in the level of independence (experience) in searching. The strategies and search results of a professional searcher will be used for comparison. The aim of the research is to determine individual differences between users in using database resources and searching styles in relation to search success. It is expected that the results of this study will indicate the necessity for the education of users to enable them to use this type of database more efficiently.

Objectives and methodology

The aim of the study was to identify the behaviour of end-users in the Kenneth Dike Library with different amounts of overall experience in searching electronic databases, in order to identify the knowledge of their search systems and syntax, and its effect on their search results, for the purpose of improving user education. The study thus sought to provide answers to the following questions:

- What are the prevalent search systems and syntax used by CD-ROM searchers in Kenneth Dike Library?
- Is there any significant relationship between users quality of search strategy and search results?
- Is there any significant relationship between users quality of search strategy and level of education?

The study was carried out in November 2003, involving end-users with different level of searching experience. All the users who came for a CD-ROM search during the period of the research were included. There were 15 users in all and for the purpose of comparison, each search was also performed by a professional searcher. The relevance of the obtained references was assessed by the user. A well-constructed questionnaire was used, including interviews and observations. Descriptive statistics was used to analyse data collected in the study. In addition to descriptive statistics data were presented by means of correlations. The variables included in this study were: user variables; search process variables; and search outcome variables.

The user variables included:

- Computer and databases searching experience.
- Characteristics of user groups in relation to the usage of on-screen instructions and tools.
- Data on characteristics of the searcher (user variables) were collected by means of a comprehensive questionnaire.

The search process variables included three types:

- (1) variables relating to the knowledge of searcher system and syntax;
- (2) variables of words, concepts; cycles and a new area – steps; and

- (3) variables of qualitative search analysis clarity, specificity and complexity of strategies in relation to the title of search topic, and errors made during searching (see Appendix).

Within this a number of variables of search strategies were included:

- Variables relating to the knowledge of search system and syntax.
- Concepts, steps and cycles.
- Variables of qualitative search analysis.
- Clarity of strategies.
- Specificity of strategies.
- Complexity of strategies.
- Errors made during searching.

The search outcome variables were:

- Number of retrieved references.
- Number of relevant references.
- Search precision.
- Result with zero hits.

Results and discussion

The results are presented according to the three research questions.

Search processes

The first question asked what are the prevalent search systems and syntax among users of CD-ROM in the library? Table I displays at a glance, the frequency of usage of the available search strategies.

Variable	Frequency	Percentages (%)
Truncation	4	32
Index	4	32
Thesaurus	-	-
<i>Database fields</i>		
Title	6	48
Author	5	40
Language of text	1	8
Abstract	4	32
<i>Boolean logical operators</i>		
And	10	80
Or	8	64
Not	6	48
With	7	56
Near	5	40
Syntagmas	5	40
Sets	9	72

Table I.
Variables of search
systems and syntax

In the investigation the logical operator "AND" – employed in 80 per cent (10) of cases – is the most commonly used search strategy followed by "OR" used in 64 per cent (8) of cases. The logical operator "with" is used by 56 per cent (7) of cases, while the logical operator "NOT" was used in six cases (48 per cent), and "Near" was used in five cases (40 per cent). These findings corroborate those of Jokic (1997) who noted that the logical operator "AND"; and "OR" are significantly used in searching by end-users.

Searching by individual fields most frequently results in decreased recall and increased relevance. In the present study, title field is the most commonly used field with six cases (48 per cent). This is closely followed by the author field with five cases (40 per cent). Abstract and descriptors are equally used in four cases, each of which is 32 per cent. The least used is the language of test field in just one case (8 per cent).

Word truncation is a practical technique in searching databases, especially in cases when users are not sure whether to use singular or plural, how to spell a word, or when they want to increase recall by using only word stems. Truncation in this study was used in four cases (32 per cent). Syntagmas or Natural Language Phrase was used by five searchers only – which is 40 per cent. Sets are used exclusively to speed up the search and avoid possible mistakes in subsequent typing of a concept. A set in the present study was used in nine cases (72 per cent).

Search strategy and results

The second question queried whether there was any significant relationship between users quality of search strategy and search results? The result in Table II also shows that search results are dependent upon the search strategy employed by the user. The more sophisticated the search strategy (higher number), the more results in terms of number and relevance obtained by the searcher. For example, searcher no. 8's search strategy is 35, while his grade/search result is 80 per cent; the search strategy for search 14 is 11, and the search result is 45 per cent.

Search quality and education

The third question asked if there was there any significant relationship between users' quality of search strategy and level of education? The result shows (Table III) that there is no relationship between the level of education and the quality of the search

Searchers	Grade (%)	Search strategy
1	60	31
2	30	10
3	80	33
4	55	23
5	40	10
6	60	28
7	40	11
8	80	35
9	60	30
10	40	9
11	50	12
12	70	33
13	55	12
14	45	11
15	70	32

Table II.
Users' search results
(grade) and quality of
search strategy

Searchers	Level of education	Quality of search strategy
1	8	31
2	4	10
3	8	33
4	8	11
5	7	10
6	3	28
7	4	11
8	8	35
9	8	30
10	3	9
11	4	12
12	7	33
13	8	12
14	7	23
15	7	32

Table III.
Relation between
education and search
quality

strategy. For example, searcher 12's level of education is 7, and his quality of search is 33, searcher 14's level of education is also 7, while his quality of search is 23. Also, searcher 10's level of education is 3, and the quality of search is 9; while searcher 6's level of education is 3 but the quality of search strategy is 28. Searcher 1's level of education is 8, while his quality of search strategy is 31, on the other hand searcher 4 with an education level of 8, has a quality of search strategy of 11. This therefore shows that performance or search result does not depend on level of education.

Conclusion

The objective of this study has been to identify prevalent search strategies among users of CD-ROM in the Kenneth Dike Library of the University of Ibadan, as well as any relationship existing among the users' quality of search strategy, search results and level of education. The study has shown that there is a significant relationship between users' quality of search strategy and search results. However, there was no significant relationship between users' quality of search strategy and level of education. In other words, the level of education does not determine the quality of search results. It was also ascertained that Boolean logical operators and sets are the most commonly used search methods and syntax. The results of this study has implications for the training of CD-ROM end-users in Nigerian libraries, as well as elsewhere.

References

- Hsieh-Yee, I. (1993), "Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers", *Journal of American Society for Information Sciences*, Vol. 44 No. 3, pp. 161-74.
- Jokic, M. (1997), "Analysis of users' searches of CD-ROM databases in the National and University Library in Zagreb", *Information Processing & Management*, Vol. 33 No. 6, pp. 785-802.
- Siegfried, S., Bates, M.I. and Wilde, D.N. (1993), "A profile of end-user searching behavior by humanities scholars: the Getty Online Searching Project Report, No. 2", *Journal of the American Society for Information Science*, Vol. 44 No. 5, pp. 273-91.

Appendix. Testing for variables

- (1) Relating to the knowledge of search system and syntax.
- (2) On typical search strategies: number of words, concepts, and a new one-steps.
- (3) Of qualitative search analysis.
 - i. What are the typographical errors noticed?
 - ii. Did the error(s) affect the search results?
 - iii. What are the search strategies used?:
 - Topic title.
 - Picking essential concept from the topic title.
 - Picking all the concepts from the title.
 - Use of several strategies.
 - Synonyms.
 - Related concepts.
 - Descriptors.
 - Narrow terms.
 - Broad terms.
 - Natural Language (i.e. personal syntactic & semantic construction).
 - iv. What are the number of words on the search title?
 - v. Did the searcher use commands such as select; combine; and/or expand terms?
 - vi. How many steps did the searcher take to reach the desired result? (i.e. a step is one or more commands terminated by enter).
 - vii. Did the searcher make use of Boolean logic (AND; OR and NOT).
 - viii. Did the searcher make use of PROXIMTY operators? (NEAR; WITH)
 - ix. Were the searches made through database fields (title; author; language of text; Abstracts; descriptors); index
 - x. Did the searcher engage in word truncation? (e.g. toxic instead of toxicants; toxicology, toxicity etc).
 - xi. Computer operator's grading of research result compared to his own percentage.

(Angela E. Okpala is a registered librarian and a PhD student in the Department of Library, Archival and Information Studies, University of Ibadan, Nigeria. She has a Bachelor's degree in Communication and Language Arts and a Master's Degree in Library Science and has written several papers. She can be contacted at E-mail: umejei2001@yahoo.com

Josephine Umejei Igbeka is a Senior Librarian in the University of Ibadan, Nigeria and currently the Faculty Librarian in the Faculty of the Social Sciences. She has a Bachelor of Science degree (Ohio State), a Master's Degree in Library Science (MLS) and a PhD from the Department of Library Archival and Information Studies, University of Ibadan. She has 12 publications in reputable journals and has also published two textbooks. She can be contacted at E-mail: umejei2001@yahoo.com)