Information Retrieval in Libraries and Information Centres: Concepts, Challenges and Search Strategies

Chimah, J.N.[1], Unagha, A. O.[2] and Nwokocha, U.[3]

National Centre for Energy Research & Development Library, University of Nigeria Nsukka [1,2], Dept. of Library & Information Science, Abia State University, Uturu[3]

Abstract

Purpose: *To cater for these classes of library users, libraries and librarians provide easy and efficient ways of accessing and retrieving information in the library. This paper examines concepts of Information Retrieval (IR), identifies its challenges and search strategies including Boolean logic.*

Design/Methodology/Approach: *Exploratory approach was applied in this study. By this method the problems of quick, accurate and useful information can be possible.*

Findings: *Many library users have been known to leave the library frustrated. This is because these users have not been adequately schooled in the art of making the best out of the library. The study reveals some search processes which help to build reliable, effective and efficient Information Retrieval System in libraries.*

Practical Implication: *Recommendations are offered for effectiveness and efficiency of Information Retrieval in libraries and information centres. When applied the problem of retrieving irrelevant information from our databases will be highly reduced.*

Originality/Value: *Explanations of terms association with IR are given for fuller understanding of the concept of Information Retrieval in Libraries and Information Centres. The value of this lies in fact that the information manager and seekers are brought to a level of satisfaction of services given and received.*

Keywords: *Information Retrieval, Search Strategies, Boolean logic, Recall, Precision, Relevance and Fall-out*

Paper Type: *Conceptual*

Introduction

Information Retrieval, in computer science, is term used to describe the organization, location and retrieval of encoded information in computer systems. Important factors in retrieving information are the type of media, or storage device, used to store information; the media's storage capacity; the speed of access and information transfer to and from the storage media, the number of times new information can be written to the media; and how the media interacts with the computer (Unagha, 2010).

The set of decisions and actions taken during a search is known as search strategy. Some searchers are more methodical in the construction of search strategies than others, but every searcher aims to retrieve sufficient relevant records and avoid:

1. retrieving irrelevant records
2. retrieving too many records
3. retrieving too few records

Searching for and locating relevant information requires careful thought and strategy. User can often find answers to their questions by first looking through general reference sources, such as encyclopedias, dictionaries, atlases, and other materials that are usually located near the library's reference desk. These sources can provide overviews of the subject that may lead to more detailed source of information. Users looking for a wide range of literature on a particular subject can search through various other indexes, abstracts, and databases. These sources provide references to relevant magazine and journal articles (Rowley & Farrow, 2000). The Internet can also be a useful source of information.

Information Retrieval Challenges

Information retrieval involves the retrieval of information from a collection or database in response to an information problem. Imagine a student who a student who has an assignment that requires him or her to acquire a certain piece

of information. How does he/she get that piece of information with the least trouble? If there are only a few information sources (say ten or less books) available and the information is contained in one or two of them, it will be easy in that the student could easily page through the books to locate the one that has the piece of information required. However, if the number of information is huge (say ten thousand books), then it becomes unpractical to go through all the sources to locate the required piece of information; hence, the user is faced with an information retrieval problem. The latter case reflects what information seekers almost always face due to the multitudes of information sources that are available in any given subject area (Aina, Mutula and Tiamiyu, 2008).

Search concepts and strategies

Users approach and interact with an information retrieval system because they want to get documents that can satisfy their information need. The whole process that leads to the final state of actual keying in a query into the system is important.

The route to getting a solution to an information problem varies in many ways, such as time, mental effort and skill required, and the satisfactory level with the result obtained. When faced with a problem of any kind, people will always look for the shortest possible route to get to the solution. The process of retrieving information (IR process) is by nature interactive, and involves an interaction between the user and the IRS (Xie, 2003). The process can be divided into several steps. It should be noted that these steps may not be followed in the order given below.

- Recognizing and accepting an information problem
- Understanding the information problem
- Deciding on and selecting the IRS
- Formulating a query

Steps in Online search strategy

Certain steps are to be followed during on-line search. Steps are taken to guide a searcher so that he/she does not confuse his/her search response with output. The steps as identified by Unagha (2010:233) are as follows:

1. Clarifying and negotiating the information need and search objective: In this regard, the search has to interview the requestor so as to clarify his story of need as he narrates with the aim of ensuring retrieval of relevant items (high recall); retrieval of only relevant items (high precision); and retrieval of some relevant items (brief search).
2. Identifying retrieval databases: The essence of this step is to avoid waste of time and labour or duplication of efforts. Here, the searcher tries to identify which online database is useful for first consultation, second, and so on.
3. Formulating basic search logic (planning search strategies): In this step, the searcher tries to analyze the search topic into parts called facets or concept groups. He plans his approach to search strategy to combine concepts of the topic.
4. Compiling search terms: Choose indexing terms from database's thesaurus or other printed word list. Select terms for text search of the subject-conveying fields (title, author, abstract, etc). Decide to use thesaurus and alphabetic word lists online.
5. Making choices: This has to do with limiting and printing output (offline or online). Selecting and approach to search strategy which best satisfies the search objectives expressed by the requestor.
6. Conceptualizing the search as input to the retrieval system: This involves arranging the search terms; for example, truncation, and word proximity. Nothing most important and less important concept groups and deciding on sequence input to access these concept groups efficiently. Restricting or limiting output based on search objectives.
7. Evaluating preliminary results: This step involves review search results step by step considering alternative search strategies to meet search objective.
8. Evaluating final results: This is determining requestor's satisfaction with search results.

Heery (1996) outlines the four possible outcomes of a search for documents in a database as follows:

1. Some relevant items are successfully retrieved. These we call Hits.
2. Some items that are not relevant are retrieved. This is known as Noise.
3. The search fails to retrieve some relevant items. These are Misses.
4. Some irrelevant items are not retrieved. These have been successfully. Dodged.

Boolean Logic

This a system of logic developed by the English mathematician George Boole (1815-1864) that allows the user to combine words or phrases representing significant concepts in a key word's search of an online catalogue or bibliographic database. Three logical commands (sometimes called "*operators*") are available in most search software (a computer program designed to execute a search for information when queried by a user. User-friendly search software provides both a menu-driven interface for novices and a command-driven interface for experienced searchers. Sophisticated search software permits the use of Boolean logic, nesting, truncation, wildcard, and proximity operators in search statements and allows the user to limit search results by various parameters (Reitz, 2004).

The OR command is used to *expand* retrieval by including synonymous and related terms in the query. This is also called *logical sum* (In Boolean logic, the result obtained when the OR command is used to find all the members common to two or more sets of entities). Logical sum is the search strategy used to determine which records in a library catalogue or bibliographic database contain term A or term B, or both term A and term B. It is synonymous with *logical addition* and *union*. Search statement: **violence** or **conflict** or **aggression**

The AND command is used to *narrow* search results. Each time another concept is added using "and, the search becomes more specific. In some online catalogues and databases, the "and" command is *implicit* (no need to type it between terms). In other interfaces, keywords will be searched as phrase if not separated by "and". It is also called *logical product* (in Boolean logic, the result obtained when the AND command is used to find all the members common to two or more sets of entities from those of another). It is the database strategy used to determine which records in a library catalogue or bibliographic database contain term both term A and B. It is synonymous with *logical multiplication* and *conjunction*. Search statement: **violence** and **television** and **children**

The **NOT** command is used to *exclude* unwanted records from search results. It is also called *logical difference* (in Boolean logic, the result obtained when the NOT command is used to separate members of a set of entities from those of another). It is the database strategy used to determine which records in a library catalogue or bibliographic database contain term A but B. It is synonymous with *logical subtraction* and *negation*. *Search statement*: **television** not **news**.

When two different Boolean commands are used in the same search statement, parentheses must be included to indicate the sequence in which they are to be executed (syntax). This technique is called *nesting*. *Search statement*: **television** and (**violence** or **aggression**) and **children**

Search statement in information retrieval is an information need or query entered as input in a form acceptable to the search software used by the retrieval system. Most online catalogues, bibliographic databases, and search engines allow Boolean logic, nesting, truncation, wildcard, and proximity operators to be used in keyword(s)' search statements and permit the user to limit search results (Tseng, 2003).

Recall, Precision and Relevance and Fall-out

Recall - is the ability of the retrieval system to reproduce a good number of documents that will satisfy the need of the user. The primary concern of every library user is getting the documents that will satisfy his need. In some cases, one document can suffice and in some other cases, more documents are wanted. Therefore the recall system is to ensure the retrieval of more documents when further search is done to satisfy the need of the user. The more we search the more materials we recall.

Recall is an important requirement for a user since his main purpose of approaching an information retrieval system is to obtain one or more document relevant to his information needs. The ability of the retrieval system to uncover relevant documents is known as the *Recall Power* of system. According to Vickery (1991), the recall power can be expressed quantitatively for each indexing term by a simple ratio defined as the *Recall Ratio*. This is given as:

$$\frac{R}{C} \times \frac{100}{1}$$

Where R is the number of relevant documents that is retrieved in the conduct of a search using an indexing term to interrogate the system and C is the total number of documents in the system that is established as being relevant to a particular request. The *recall ratio* is the number of *hits* (relevant documents that have been retrieved) as a percentage of the total number of relevant documents in the system or database.

Precision is the probability that an item would be relevant given that it is retrieved. Precision relates to the system's ability to filter out unwanted items.

P (B/A)

Retrieved	AμB	AμB
Not retrieved	-AμB	AμB

A = retrieved

B= relevant

In measuring the ability of performance of an indexing system, there is a ratio complementary to the recall ratio called the **Precision Ratio.** The *precision ratio* is the number of hits (relevant and retrieved documents) as a percentage of the total number of documents retrieved documents) as a percentage of the total number of documents retrieved, whether relevant or not.

The precision ratio is given as

$$\frac{R}{L} \times \frac{100}{1}$$

Where R is again, the number of relevant documents retrieved by using an indexing term in interrogating the system, while L is the total number of documents retrieved in that search.

Relevance – This is the ability of the retrieval system to retrieve documents that will satisfy the information needs of the user. In other words, retrieval documents must be able to match a reader's requirement. Usually, every library user who comes to the library has certain documents in mind either via the title, author, subject etc. When the retrieval system is searched (catalogues, bibliographies, computers etc) it should be able to release these documents that are relevant and ready to match the information needs. Therefore as a feature, every retrieval system must show high level of efficiency in retrieving documents that are relevant to a user's need.

The recall ratio and precision ratio can further be illustrated thus:

$$\text{Re}callration = \frac{(A \cap B)}{A} = \frac{(relevant\ \ document\ \ retrieved)}{(total\ \ of\ \ relevant\ \ documents)}$$

$$\text{Re}callration = \frac{(A \cap B)}{B} = \frac{(relevant\ \ document\ \ retrieved)}{(total\ \ of\ \ relevant\ \ documents)}$$

Fall-out – another term which is sometimes used is fall-out ratio, defined as: is the probability that an item would be retrieved given that it is not relevant.

$$Fall-out\ ratio = \frac{B-A}{A} = \frac{(retrieved\ \ but\ not\ relevant)}{(total\ not\ relevant)}$$

The set B – A consisting of documents retrieved but not relevant, may be regarded as noise, while the set (AUB)', documents neither retrieved nor relevant, may be thought of as dodged, to use Vickery's term.

Conclusion

Researchers and practitioners of information retrieval are grappling with information problems such as the one described above to fine ways which information seekers could be helped to be able to find the information they require from the multitude of sources available with the least "pain" possible. The goal of IR research is to find the best ways of helping information seekers to find information relevant to their information problem by developing information retrieval systems that are "user friendly". Understanding the information behaviours of the communities of users of information retrieval system is necessary for both developing and searching IRS.

For any particular reader who comes to the library with a need for information, there will be certain items in our collection which will be relevant. Among these it be possible to establish some sort of precedence order; some will be definitely relevant, others will be useful, but less so, while others will be only marginally relevant. To find more information we may broaden our search: that is, present to our reader those items which, though they do not deal specifically with

the subject he is interested in, do include it as not deal specifically with the subject he is interested in, do include it as part of a broader subject.

Recommendations

Information retrieval researchers have, however, long maintained those information seekers characteristics, the context and the system in the market place.

It is imperative that those who endeavour to facilitate the retrieval of documents by users should the information retrieval system, including the structure of the database, the file structure and the indexing system used, stand to use it appropriately, formulating queries that will enable them to retrieve the most relevant documents, and should be able to move quickly through the information seeking process, especially at the level of interaction with the system.

More importantly, maximum benefits from IRS require a tripartite relationship between the research in user studies, information technology pundits and information retrieval researchers. Kuhlthau (2005) notes, *We need to take a 'just for me' approach to system design that is based on the user's perspective of information seeking and use*.

References

Heery, R. (1996) Review of metadata formats. Program 30(4), October. Available at http://www.ukoln.ac.uk/metadad/review

Reitz, J. M. (2004) Dictionary of Library and Information Science. West, Connecticut: Libraries Unlimited.

Tseng, G. (2003) Boolean Logic in International Encyclopedia of Information and Library Science. London: Routledge.

Rowley, J. & Farrow, J. (2000) Organizing Knowledge: an introduction to managing access to information. 3rd ed. Hampshire: Ashgate.

Unagha, A. O. (2010) Knowledge Organization and Information Retrieval. Imo. Whytem Publishers

Vickery, B. C. (1991) Retrieval System Theory. London: Butterworth.

Xie, H. I. (2003) Supporting Ease of Use and the User Control: Desired Featuresand Structure of Web-based Inline IR Systems, Information Processing and Management, 39(6) pp.899-922.

Aina, L.O., Mutula, S.M. and Tiamuyu, M.A.(2008) Information and Knowledge Management in the Digital Age: Concepts, Technologies and African Perspectives. Ibadan. The World Information Services Ltd.pp.429-432.

Kuhlthau, C.C. (2005) Towards Collaboration Between Information Seeking and Information Retrieval, Information Research, 10(2) paper 225. Available at: http://InformationR.net/ir/10-2/paper 225.htm .