INFORMATION RETRIEVAL BY USING ELECTRONIC DATABASES

Ionela M. BÂRSAN[1]

Abstract: *The retrieval of scientific information can be carried out by searching, browsing or both combined. Selecting the right searching variant to obtain information specific to a particular field of research depends, among other factors, on the type of information needed by a particular user. This article will explore the three different approaches for obtaining information, along with their features and applications: simple and advanced searching, browsing, searching and browsing.*

Key words: *Information Literacy, information retrieval, searching, browsing.*

1. Preamble

The information retrieval (IR) operation is performed through information retrieval systems. As early as the 1950s, various types of information retrieval systems were developed in order to meet various needs. Online systems, CD-ROM systems (*Compact Disc Read-Only Memory*), online public access catalogues (OPAC) and Internet information search systems are the four main categories of systems which have helped to meet the need for information retrieval tools.

2. Online information retrieval systems

Online systems allow the user, by using the computer and the single request is not processed immediately, but later, when a certain number of requests is gathered), and it was only later that real-time interaction between the user and the system was introduced. As early as 1996, Professor Trudi Bellardo Hahn made a

detailed presentation of the pioneers involved in the development of online systems. (Hahn, 1996) Afterwards, in 2003, along with engineer Charles Percy Bourne, head of the development team of Dialog Information Services Inc. wrote several articles on the development of electronic systems between 1963 and 1976. (Bourne, 2003) After the progress and improvements made in the 1950s, electronic systems have become, in the digital age, a well-defined group of IR systems.

2.1. OPAC systems

Online public access catalogues (OPACs) are a form of traditional catalogues telecommunications technology, to research remote databases. Initially, systems only operated in batch mode (i.e. a, but achieved in a different way. (Malinconico, 1984) Prior to 1980, there were only a few attempts, such as the Ohio State University's Library Control

---

[1] *Transilvania* University of Brasov.

System. Since 1980, several prototypes and even operational OPAC systems have been installed in an increasing number of libraries. Few libraries lack a computer-based system today.

This new generation of OPACs appeared by adapting to the digital environment and by embedding already tested information retrieval technologies (Antelman, Lynema, & Pace, 2006). Of course, the web becomes the platform for this generation, enabling access to various features that the web makes available. For example, OPACs of this type allow users to interact with the system through the interface and the Internet browser. Consequently, users do not need special training to get used to the operation mode of this information retrieval system since they are already acquainted therewith. Links to other resources may be created (such as electronic articles and books) starting from the OPAC records, in addition to the links created within the OPAC.

OPACs have helped users locate various materials in library collections and benefit from facilities not provided by their predecessors, catalogues based on individual sheets. In addition, they serve as a gateway for other search systems by creating links thereto. As monographs are usually a major part of library materials, the OPACs are the first, if not the only, means to access them. Given the researchers' expectations in the digital age, without the OPACs, library collections would have become very little used.

The advent of Google Book Search changed and even threatened the dominant role that the OPACs had in obtaining information on books. Google Book Search, originally called Google Print, started in 2004 with a group of publishers (Blackwell, Cambridge University Press, University of Chicago Press) and, two months later, with the libraries of several institutions (Harvard University, University of Michigan, New York Public Library, Oxford and Stanford) called Library Project. The number of participants in this project has continuously increased since then. The books included in this project can be searched after they are made available in digital format. In addition, Google Book Search provides certain facilities, such as links to the books' reviews and the sources where they are mentioned (e.g. web pages). Undoubtedly, Google Book Search provides information research through an approach different from the OPACs'. Google Book Search, along with its web search system, also determines the re-examination of the position, functionality and performance of OPACs in this digital age (Deanna Marcum, 2006).

2.2. Internet search systems

As an Internet Research- IR system, the Internet search system provides virtually all search facilities available in other similar environments. But there are specific elements of the Internet systems, which will be further discussed.

Boolean search, present in non-internet search systems, could not be included in all web search systems in the past. On the other hand, the plus sign (+) was widely used as a surrogate for the AND operator in the early years of Internet search, but this can be misleading because the + sign represents the notation for the OR operator or logic sum. Although this practice was interrupted and the + sign has become the symbol of weighted search, using the + sign as AND operator creates confusion for the user, especially for one who has not used the Boolean search before.

Proximity searching is essential in finding phrases (if the information in the database are not indexed by phrases) and in specifying the corresponding function of the research terms. Online systems support

very well proximity searches. Instead, proximity search has not become a universal feature of the Internet search systems. Out of these systems that support proximity search, few provide the entire range of proximity operators that are normally available in online systems and OPAC systems.

With respect to truncation, field search and search by using capital or small letters are present within the Internet search systems, but on a smaller scale and with limited capabilities. Other IR systems can be very precise by limiting the number of characters which will be reduced. But most Internet systems apply unlimited truncation. However, automatic truncation is default in some cases, generating more irrelevant results.

3. Using electronic databases

The retrieval of scientific information can be carried out by searching, browsing or both combined. Selecting the right searching variant to obtain information specific to a particular field of research depends, among other factors, on the type of information needed by a particular user.

3.1 Information retrieval by simple or advanced searching

Searching is an information retrieval method used for a very long time. It is also a highly debated topic in the discussion on information retrieval (IR), where it is also referred to as database searching, online searching, web searching or the like. Searching involves a process of defining the user's information requirements and a set of techniques whereby, based on these requirements, the user will reach a set of results that will complete at least partially the information needed, in a particular business field..

3.2 Searching and browsing integrated in the information retrieval activity

Searching and browsing are two different information retrieval methods. Half a century ago, the German researcher settled in the United States, Hans Peter Luhn (Lunch, 1958), creator of the Luhn algorithm (still used today in the allocation of numbers on the bank cards) and of the KWIC (Key Words In Context) index, classified information retrieval methods as:
• information retrieval by look-up in an ordered array of stored records;
• information retrieval by search in a non-ordered array of stored records;
• a combination of the two options above.

The terminology and structure of scientific information databases have undoubtedly changed over the years. However, we may ascertain that the first category described by Lunch represents browsing, while the second denotes searching.

Searching is a structured endeavor that requires people to be trained before they can perform it. Browsing, however, is a retrieval approach natural to the user, who does not need to spend time learning it. Moreover, in the browsing process, no query representation is involved, which can relieve the user from the highly sophisticated task and let the user simply focus on browsing. The comparison of the above table shows that both searching and browsing have their own strengths and limitations. Each approach works well in certain conditions and circumstances. Although there are occasions on which one approach seems more appropriate than the other, the integration of the two can result in better retrieval performance in general. Searching may not be needed for every information retrieval activity. But browsing is always necessary in judging the relevance of retrieved results.

4. Quality assessment that relate to user satisfaction regarding information resources. Case study at Transilvania University of Brasov

The internet and Google services radically changed the way we evaluate ourselves with respect to information. Now the libraries are competing for a place in the information market, are based upon a managerial type of culture that is centered upon the library user, also called "customer" in this context. Nowadays info-documentary institutions are centered upon the quality of the offered service. We aim at the access of a high number of users to as many documents as possible, in a very short time. The companies in the private area are looking to satisfy the customers' needs and this is the model that should also be followed by the library. In this competition between the library and the other information providers, the library needs a strategic direction and a good resources management. The library superiority with respect to other information providers is given by the impact of the library in the learning and research process followed by the users. So, the library value can be testified only by its users. LibQUAL+™ represents a range of services that the libraries use to follow and understand the users' opinions regarding the quality of the services offered by the library. Also in this respect, the improvement of library services according to the immediate requirements and needs of the users are taken into account.

4.1. Research methodology

Considering the LIBQUAL evaluation model we accomplished a quality marketing research in Transilvania University of Brasov, for final study years, engineering, sociology and literature students. We selected these areas because

they are provided with information literacy classes starting with first year according to their syllabus.

The research has been developed in the first semester of 2011-2012.

We used the electronic questionnaire generated by www.surveymonkey.com, sent by students' groups email.

We will present only the results relating with students preferences about information resources.

4.2. Research data – data analysis

Concerning the information control, the easy access to information and the sources made available, the students were questioned by a series of questions in order to be able to assess the access to informational resources and their quality. By this quality research we determined the students' preferences. The most used information by the students are the online catalogue of the library: 75,7% literature, 69,6% sociology and 45,9% engineering and in a smaller proportion the online catalogues of other libraries 37,8% literature, 34,8% sociology and 13,5% engineering (Figure 1)
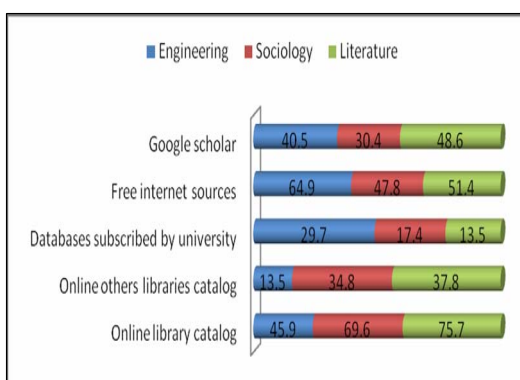

Fig. 1. *Information sources preferred by students*

The free internet sources represent an important information source for 64,9% of

the engineering students, 47,8% of the sociology students and 51,4% of the literature students.

Google Scholar also represents an important information source for 48,6% of the literature students, 40,5% of the engineering students and 30,4% of the sociology students**.**

The databases subscribed by the university are accessed by the students at a lower extent: 13,5% literature, 17,4% sociology and 29,7% engineering.

It is normal there are similarities between sociology and literature, so that the correlation coefficient between the obtained data strings containing the answers concerning the preferred information sources is 0,936037, meaning a strong connection. The situation is different with respect to the data string obtained by the engineering students, the value 0,486141 with respect to the sociology students and the value 0,498655 by comparison to the literature students, which indicates a medium connection. (Table 1).

Table 1

Pearson correlation regarding information sources for students

	Engineering	Sociology	Literature
Engineering	1		
Sociology	0.486141	1	
Literature	0.498655	0.936037	1

Regarding the way they were informed about the information sources, the engineering students found out from the teaching staff in about 70,3%, from the colleagues 51,4%, from the faculty courses 40,5%, 27% by chance and did not take part in training classes at the library.

The sociology students found out in a 73,9% percentage from the colleagues, 47,8% from the teaching staff, 30,4% by chance, 13% from the faculty courses and only 4,3% attended training classes. The literature students found out from colleagues 56,8%, from the teaching staff 45,9%, by chance 37,8%, from the faculty courses 37,8% while 13,5% attended training courses at the library. (Figure 2).
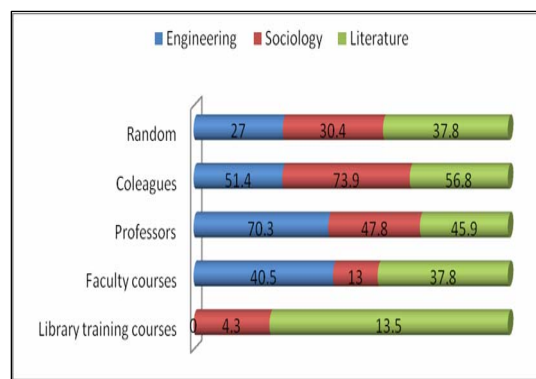

Fig.2. *Sources of information for students*

5. Conclusions

Users benefit from this integrated approach not only because they have the right tool for the right retrieval task, but they also get more out of the same IR system. For example, Science Direct allows searching within its browsable categories (e.g. Mathematics). Searching within a browsable category is therefore similar to searching a database which is specialized in a particular field represented by the browsable category. Browsing and searching become integrated activities in this environment. Thus, we can metaphorically say that one plus one could be more than two if the combination is done wisely. This is true for the integrated information retrieval approach presented above (searching plus browsing).

The evaluation culture represents a component of the library management that must be continuously developed. In Information Literacy courses have to be implemented different chapters with information research resources and new studies about searching and narrowing behavior have to follow. The promotion of the information resources should also be a continuous action.

References

1. Antelman, K., Lynema, E. Pace, A.K., et al.: *Toward a twenty-first century library catalog.* In: Information Technology and Libraries (2006), vol. 25, no. 3, p. 128-139.
2. Bourne, Ch.P., Hahn, T.B.: *History of online information services 1963-1976.* Cambridge, MA. MIT Press, 2003.
3. Fenichel, C.H., Hogan, Th.H.: *Online searching: A primer.* Marlton, NJ. Learned Information, 1981.
4. Hahn, T.B.: *Pioneers of the online age.* In: Information Processing & Management. (1996), vol. 32, no. 1, p. 33-48.
5. Herner, S.: *Browsing.* In: *Encyclopedia of Library and Information Science*, Kent, A., Lancour, H. *et al.* (Eds.). Marcel Dekker, New York, 1970, vol. 3, p. 408-415.
6. Cook, C., HEATH, F., et al.: „*LibQUAL+(2001): Service quality assessment in research libraries*". In IFLA Journal 27 (2001), no. 4, p. 265 Available at: http://ifla.queenslibrary.org/V/iflaj/art2704.pdf. Accessed: 09 September 2013
7. Luhn, H. P.: *Review of information retrieval methods. Report No. RC-59.* Yorktown Heights, NY. IBM Research Center, 1958.
8. Malinconico, S. M.: *Catalogs & cataloging: Innocent pleasures and enduring controversies.* In: Library Journal. (1984), vol. 109, no. 11, p. 1210-1213.
9. Marchionini, G..: *Information seeking in electronic environments*. New York. Cambridge University Press, (Cambridge Series on Human-Computer Interaction 9), 1995.
10. Marcum, D.B.: *The future of cataloging.* In: Library Resources & Technical Services. (2006), vol. 50, no. 1, p. 5-9.