

Information retrieval features: The relevance and significance of information resources in digital libraries

A. V. Strungar

V. I. Vernadsky National Library of Ukraine, Kiev, Ukraine
Corresponding author. E-mail:19870208@ukr.net

Paper received 26.02.2016; Accepted for publication 10.03.2016.

Abstract. Urgent problems of library interaction with user, that is user's library-information services appear regardless of whether the documents, which contain the necessary information, are within the building of specific library, but depending on how good and how fast the access to them is granted, applying modern high-efficiency technologies for creation of library products. The basic concepts of information search, which could be implemented in formation of scientific digital library, are resulted in the paper. The main information constituents of scientific digital library are considered. The indexes of search pertinence and relevance performance levels, concept of plenitude and exactness of search, as well as «losses of information» and «information noise» are considered. The chart of dichotomy classification of digital library search is presented. Description of user's relation to scientific information and digital documents is presented.

Keywords: *information technologies, digital library, digital document, Internet, information search, pertinence, relevance.*

Introduction. Modern library is a complicated information system, which is formed of both traditional and new, untraditional information subsystems. Among the newest information subsystems of library an important role is played by digital catalogue (DC), and its creation is the primary task and essence of libraries automation. Digital catalogue opens fast and high-quality access to the library information resources. Quality and efficiency of EC is stipulated with the complex of methods and facilities, which determine technology of its creation and use [1]. Digital catalogues are one of major library resources. This resource is especially valuable, when it is linked to the world network, which allows providing round-the-clock access to information about the presence of documents in the library funds. This paper considers the information aspects of library processes, digital library concept determination, digital catalogue of library and principles of its creation, pertinence and relevance of information resources. the structure of information storage and retrieval system is also considered briefly. Bibliographic search is a complex type of activity, performed during the dialog between a reader and information storage and retrieval system of a library [2]. To solve the problem of rational and multifunctional use of information resources accumulated in a library, the ramified system of reference aids (RA) is created, every part of which performs specific function, being a constituent of the unique information potential. High quality of RA and services for readers based on it provide an efficient access to the information and, as a result, the most complete use of library funds by readers. Originality of reference aids system for readers is not only in its content, but to a great extent in the specific format of documents presentation, taking into account social-psychological features of readers of various age-related groups, according to which content and forms of RA organization, method of forming abilities and skills for independent use of information generators by visitors of digital knowledge libraries are analyzed at the choice of literature.

Recent researches and publications analysis. The theory of bibliographic information (BI) was developed by O.P. Korshunov [3] and presented in his textbook. The term «bibliographic document» in relation to documents, which contain bibliographic information, was introduced by E.K. Bespalova [4]. I.G. Morgenstern determines bibliographic resources as «a type of information resources,

which represent the organized ranges of bibliographic information accessible for the use as a bibliographic source» [5]. One of the mechanisms for acquisition of reliable information sufficient for decision-making was offered by A.M. Yanovsky, according to which «characteristics of information and requirements to its content (meaning its specifications) are formulated, and then the sequence of procedures of analysis and its use are determined». G. Selton, professor of Cornwell university, was the author of the SMART system (System of Mechanical Analysis and of Retrieval of Text), based on which many modern algorithms of the automatic indexing and information retrieval have been developed.

Aim of the paper is consideration of information retrieval features in a digital library and familiarization with the methods of digital catalogue usage, definition of the essence of information resources pertinence and relevance from the position of information retrieval in digital libraries.

Research results presentation. Digital reference aid system (DER) is a developing system. The constituent of library ER important for readers is the digital catalogue (DC), which is reliable mediator between the library fund and its users. The necessity to provide readers with the most meaningful, useful information formulates the task to develop and implement methods of high-quality selection of literature and ability to form document searching query into the work of digital libraries by the relevant specialists. An administrator-librarian does not only answer the request of reader, but also makes him an active participant of searching the answer to the question raised in request. The main factors, necessary for the success of digital certificate bibliographic works, are systematic character, consistency and differentiated approach to different reader groups. Digital catalogues provides many entrances to the library funds, and reader must know, which of them it is more appropriate to use in every specific case. The purpose of any information storage and retrieval system is to provide user with the ability to search information from known data, necessary subject or facts. The process of forming queries is performed with multi-optional approach and is an art in a way. In addition, searching procedure has clearly expressed sequence of stages – from determination of information necessity and search area to the analysis of results and choice of pertinent (corresponding to the user's query) objects.

The system «document-user of information» is part or

subsystem of documentary communication system. It could be considered an independent system, because the document, created by communicator for the information transfer, is separated from him and exists by itself. For the user this document is an information generator and a system of relations is set between them. The system «document-user» consists of two basic elements: document and user of information. Except them, it contains different relations between documents and users of information and, if necessary, mediators, helping users to get information from documents. Based on this the concepts of information necessity and information consumptive activity of man are formed. These concepts should be considered separately.

The information necessity arises up in humans in the process of activity, when the need to obtain new knowledge or find out certain facts appears. Activity based on which information necessity is formed is named «base activity». It could be any type of activity, which became the «basis» of information necessity origin: production, scientific, educational activity and others [6].

The information necessity takes man to information consumptive activity (ICA) – activity, aimed at information retrieval and consumption. User’s ICA allows getting necessary information and going back to base activity. This process is illustrated at Fig.1, and it should be noted that this process works not in circle, but in spiral, because the return to the base activity is done to already new, higher level, enriched with new knowledge [7].

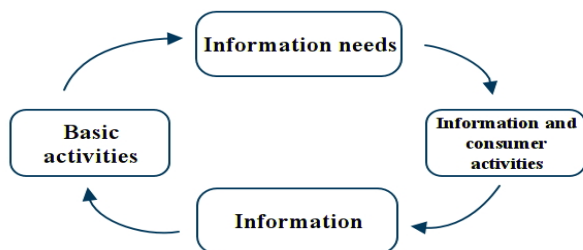


Fig.1. Origin and satisfaction of information necessity

Extremely important role in the process of information search is searching sign or set of searching signs, which are the basis for information retrieval conduction. A sign is a property of object, which stipulates its distinguishing or common features with other objects. Document search profile (DSP) is the set of searching signs, which characterize a document and are necessary for search and authentication of document in accordance with a query. For example, it is possible to consider a bibliographic record as a search profile of document.

Set of searching signs, which characterize an information query and are necessary for an information retrieval in accordance with a query, is the query searching method (QSM). The query searching method can be formulated in the terms of human language or translated into an artificial information retrieval language (IRL). Possible examples of information retrieval language are tables of library-bibliographic classifications, used for drafting library systematic catalogue. The query searching method, expressed with the help of such information retrieval language, will represent the list of classification indexes of those sections, in which it is possible to find documents (in other words searching profile of documents) corresponding to a query [8].

Translating query searching profile into information retrieval language is a searching order in obedience to which search is conducted in information storage and retrieval system.

Bibliographic search as well as information search in general has certain characteristics or, more exactly to say, requirements to the search: pertinence and relevance.

Pertinence is correspondence of the found (or given) information to the consumer need. Pertinence is obviously the greatest aim of information search, but its achievement becomes complicated with difficulties of information necessities study, with inadequacy of their reflection in information queries, with imperfection of information storage and retrieval system. Pertinence is the subjectively evaluated accordance of the obtained information to the information necessity of user. Consequently it is attitude of volume of information useful for the user to the volume of information, received on request, in other words this is search efficiency ratio. It could be improved, taking into account past interests of this user, user’s conduction in digital search engines, clarification of queries formulation, ranging with gravimetric criteria, and limitation of number of documents given out as a result of search and others like that. Earlier Google has realized the new algorithms of informal pertinence (relevance) achievement and as a result it became the most popular searching system in the internet.

Other requirement – relevance – means accordance of the found (or given) information to the user’s query. Relevance of bibliographic search appears by comparison of search image of query with searching offenses of documents, which found out as a result of search, but finally it can appear only at comparing of information query to maintenance of documents. There is a great number of determinations of relevance. For example, GOST 7.73–96 states: «relevant: accordance of the obtained information to an information query». Thus, relevance is determined only by the search algorithms of the specific searching system. It is also stated in the same GOST, that «pertinence, pertinent: accordance of the obtained information to the information necessity», in other words pertinence determines the degree of correspondence between expectations of user and results of search. On the whole, relevance may strongly differ from pertinence; however these concepts are constantly mixed up at interpretation [9].

The indexes characterizing the degree of implementation of search relevance requirement are plenitude and exactness of search, and also «losses of information» and «information noise».

The indexes of information delivery plenitude (or its losses) and exactness of information delivery (or information noise) characterize quality of work of IRL, which can be expressed in number.

For description of practical bibliographic activity the concept of bibliographic search have an important value. Some bibliographers consider that they belong only (or mainly) to the processes of bibliographic service, especially – to one of these processes (reference aids). Sure, the concept of bibliographic search is directly connected with reference aids, but they can be also treated as fundamental concepts, which are applied to various processes of practical bibliographic activity. Therefore they should be considered here to represent general description of bibliographic practical activity [7].

A bibliographic search is a type of documentary and information search. Its correlation is resulted at Fig.2.

Search on the whole is an activity, aimed at finding of that or other object, phenomenon, thing among others and its comparison with the need, which induced the search. Applying dichotomy classification (logical procedure of dividing the volume of concept into opposite to each other), search is divided into information and non-information search depending on the object of search.

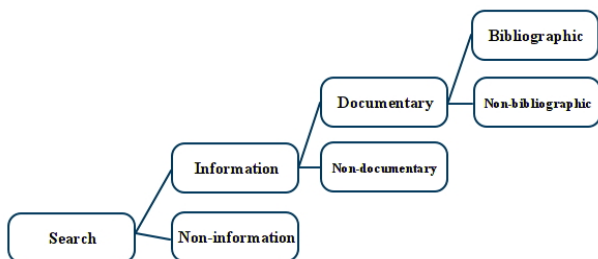


Fig.2. Dichotomy classification of search

An information search is the information, which has certain thematic content or formal description retrieval. Non-information search does not relate to the problems of bibliography. An information search is divided into documentary and non-documentary.

A documentary search is a search for document, which contains necessary information. A non-documentary information search is carried out without addressing to the document, but using other information generators (usually –questioning other people). A documentary search is divided into bibliographic and non-bibliographic.

A bibliographic search is a search of document, which contains necessary information, using bibliographic information generators without direct addressing to the document. In other words, a bibliographic search is a bibliographic information retrieval, meaning information about any document, which corresponds to the information necessity, but without addressing to the document. Non-bibliographic documentary search is carried out in case, when it is conducted among the documents themselves, for example, on a bookshelf or on a counter.

As it is shown on the chart of dichotomy classification of information search – a bibliographic search is always documentary and information by its essence (Fig. 3.). Consequently, all requirements to the documentary and information search and their properties are also inherent to the bibliographic search.

For an information search it is very important to distinguish such concepts, as information necessity, information query and information interest. All these concepts characterize user’s attitudes to the information [10].

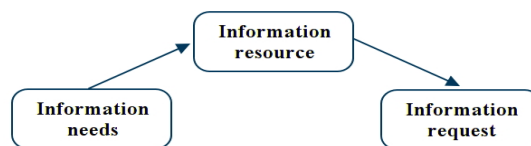


Fig. 3. Description of user relation to information

An information query is formulated and presented in the concepts of human language. The query obviously represents information necessity, but not always sufficiently clear and full.

Conclusions. Description of basic concepts given in the paper is related to bibliographic search, which is initial basis of all bibliographic processes: its interaction with documentary and information search, correlation of information necessity, query and interest concepts; concept of information storage and retrieval system, search profile of the document, searching query, and others like that.

It is obvious, that thematic search is sometimes difficult and labor intensive enough, but due to exact application of special facilities and clear strategy of search in the digital mode we can obtain the desired result. Among the advantages of thematic search in digital catalogues it is possible to mark: operability, multidimensional character, expansion of possibilities of document search image formation, with the simultaneous aggregation of different searching signs; accounting hierarchical connections of subject concepts, that allows specifying, narrowing or extending the limits of search, complex acquisition of various information about geographical object, person, area of knowledge.

Further investigations will be aimed at enhancement of virtual reference aid service, actively using the resources of global network: digital libraries, web-catalogues, on-line dictionaries, encyclopedias and reference books. Considering high rate development of information technologies, which promotes quality of reference aid service, changes the processes of information search and use for maximal satisfaction of users’ queries of the usual, organization of professional interaction of user with the library-information system is necessary, but they are still based on general rules and methods of bibliographic reference provision.

REFERENCES

1. Andreev, V. Digital Libraries: the experience of creating abroad. Libraries and Associations in the Transient World: New technologies and new forms of cooperation: 4th International Conference «Crimea-97». Retrieved from: <http://www.gpntb.ru/win/inter-events/crimea97> [in Russian].
2. Amlinsky, L. (2008). Scientific libraries of information society: organization and technology. St. Petersburg: Professia, 200 p. [in Russian].
3. Korshunov, O. (2001). Bibliography. *Fundamentals of the theory of bibliography*. Part 1. Moscow Reg. University of Culture and Arts. (Parts1-2; Part 1). Moscow [in Russian].
4. Bespalova, E. (1994). Formation of bibliographic thought in Russia (up to 60s. XIX c.). Moscow: Moscow State University of Culture, p. 12 [in Russian].
5. Morgenstern, I. (2001). Bibliography. General Course. (2nd ed.). Chelyabinsk Reg. Acad. Culture and Arts. Chelyabinsk, p. 27 [in Russian].
6. Fokeev, V. (2006). Bibliography: theoretical and methodological bases. St. Petersburg: Professia, 352 p. [in Russian].
7. Shvetsova-Vodka, G. (2011). The introduction in bibliography. Rivne, 231 p. Retrieved from: <http://oa.elibukr.org/handle/1984/1453> [in Ukrainian].
8. Korshunov, O. (2009). Bibliography. Fundamentals of the theory and methodology. Moscow: FAIR, 336 p. [in Russian].
9. Khramtsov, P. About search engines, domain names and search engine optimization. Retrieved from: <http://www.osp.ru/os/2005/11/380556/> [in Russian].
10. Barkov, O. (2001). Experience of creation of scientific electronic library of V. I. Vernadsky National library of Ukraine. *Bibliotchnyi visnyk*, 4, pp. 51-62 [in Ukrainian].