

Information search as a document heuristic factor

O. M. Zbanatska

National Academy of Managerial Staff of Culture and Arts

Paper received 07.06.18; Accepted for publication 16.06.18.

<https://doi.org/10.31174/SEND-HS2018-178VI29-20>

Abstract. It is investigated the essence of the concept of "information search", which is a component of information activity. The main types of information search are considered: documentary and factographic. The stages and strategy of information retrieval are characterized. It is identified the dependence of information retrieval and documented heuristics. Documentary heuristics is presented as a complex system of knowledge, a scientific direction, studying the theory, history and practice of documentary search, its intellectual and technological foundations.

Keywords: *document heuristics, documentary search, information search, strategy of information search.*

Introduction. Information search for many years is one of the most important areas of information activity. The problem of theoretical comprehension of finding information process, defining its essence, outlining the basic concepts and their interrelations were key since the first collection of documents and the creation of document and information institutions.

The active introduction of information technology in the 1990's led to the transformation many types of information activities, including information search. The creation for new forms of information storage on electronic media, the development of search methods in machine readable arrays, the introduction of data transmission means through telecommunication channels could not but affect the technology of information retrieval. This led to the emergence for new forms of information search organization, different from traditional content characteristics.

Nowadays there is an acute shortage of generalized research on the conceptualization of documentary heuristics. Many aspects in the scientific and methodological plan are not developed. In particular, disclosure of the features of information search in the conditions of the electronic environment. Until now, the essence of documentary heuristics is not understood and its theoretical and methodological basis has not been developed. This greatly complicates the implementation of a documentary search.

Literature review. The analysis of publications revealed certain achievements of the national science on particular issues of heuristics and information retrieval. In general, the term "heuristics" was considered by scientists in their writings P. N. Berkov, E. Bernheim, M. H. Hulishevskaya, N. V. Karamysheva, M. P. Kovalskyi, O. V. Melnyk, H. V. Nezabytovskiy, V. V. Farsobin. На думку N. V. Karamysheva, depending on the direction of creative thinking research, are distinguished scientific, legal, engineering, pedagogical heuristics, etc.

Comprehensive developments in the field of information retrieval were carried out until the 1970's (in the library sector – D. I. Bliumenau, K. A. Beskyi, M. L. Kolchynskiy, V. P. Leonov, P. I. Nikitin, A. V. Sokolov, A. I. Chorniy, E. L. Shapiro; в архівній – V. N. Avtokratov, M. N. Kazanskyy, N. P. Fineshyn). However, they all had a mainly technical descriptive character. Works of the late XX century (A. B. Hlybina, T. I. Kancheli, L. E. Pshenychna, L. V. Sakharnyi) highlighted the local experience of specific information institutions, and the work of the beginning of the XXI century - as features of the search for traditional IPS (V. V. Siedykh, H. P. Terentieva, V. K. Udalova), and features of information search in electronic catalogs and the Internet (V. B. Borshchev, H. B. Varlamov, H. A. Skaruk, O. O. Lavvronova, I. E. Rikun, E. R. Sukiasian, I. V. Olinevych).

Consequently, the information search was considered to be a purely technical practical area of information activity, which had no scientific basis. There were no developments in the issues of documentary search as a kind of informational. A certain gap was formed in the research of the process of finding documents in the document funds of information institutions and scientific direction, which would be aimed at study document heuristics.

The aim and objectives of the study. Aim of research – to review the information search as a factor in documentary heuristics. To achieve this objective, the following tasks were set: 1. to identify the essence of "information search" notion. 2. Identify interrelations of information search and document heuristics.

Materials and methods. The use of methods of scientific analysis and synthesis allowed to reveal the essence of the term "information search". A generalization method was used to penetrate into the essence of the information activity terminology.

Result. Inductive term for the documentary heuristic is the term "information search". This term was introduced in computer science by the American mathematician Kelvin Muers in 1947. In the library science, the information search is "the search, selection and provision of certain criteria for information that meets the predefined requirements in the information request" [4, p. 10]; in the archive – "the search for documents, information about them or the facts reflected in the documents" [10, p. 461]. The term "information search" is the main in the information sphere and has several definitions: 1) "the search for certain information or information in the machine memory" [3, p. 10]; 2) "the search for certain information that meets the user's information requests" [5, p. 4]; 3) "process of finding and selecting (delivery) information... from a separate text, a document, a set of documents or, in general, from a storage device of any physical nature" [1, p. 16]. Information search also found a reflection in the field of social communications. Therefore, V. A. Ilhanaieva offers a more complex definition and characterizes the information search as a "set of actions for the selection (location) of information according to the given features; a set of operations, methods and procedures, the result of which is the selection of data stored on a storage device, according to a given topic; one of the elements of the information process" [7, p. 250]. The motive reason for an information search is the information need, expressed in the form of an information request.

According to Yu. A. Shreidera, the provision that the user has a clear need for documents that can be adequately reflected in an information request was formulated in the forced manner, under the influence of the need to create an algorithm for selecting documents based on their search patterns. Traditional bibliographic systems do not provide

the appropriate logical division of the array of disclosed documents and, therefore, do not allow for the final algorithmization of the search. Informatics considered it a defect in systems, an irresistible obstacle to their automation [6, p. 18]. Today it turned out that information search as such is a poorly structured task: the notion of the right document often occurs in the process of its search. The catalog search strategy includes a large percentage of random opacities. The logic of bibliographic systematization corresponds to the logic of intuitive methods of document searching. Too clear system of knowledge structuring stored in library funds would deprive the user of a number of search capabilities.

Explaining the essence of the information search, H. H. Vorobiov noted that of all information processes, the search - the most fascinating and most insidious process. It is fascinating because it is always romantic to look, even when you are looking for what you have lost. Insidious—because it takes revenge those who ignore the science of search [11, p. 163].

An increase in the number of databases has aggravated the problem of obtaining the user the information he needs. Absolute search completeness can not be obtained by accessing only one database because according to the law of dispersion of information documents of related topics can be part of different databases. In this case, the difficulty of finding a complex interdisciplinary problem is gaining ground. According to this connection, the problem of selecting databases and rational addressing the flow of requests arriving to the system becomes the main issue. Needed systems for efficient search, organization and access to information in these databases. Even automation of information processes not only did not solve the problem of search, but rather intensified the need to improve the methods of scientific processing of documents.

Despite the rather high level of development of modern information technology, at the moment there is still no universal method of building a data storage system that would be acceptable in most cases. In each case, such a task is solved individually. Meanwhile, there is a list of requirements that should be taken into account when storing information:

- independantly from programs that use stored data;
- ensuring the completeness and minimum data redundancy;
- the possibility of actualization of data (i.e. replenishment or change of values of data recorded in the database);
- the possibility of obtaining data, as well as sorting and searching according to specified criteria [9, . 30].

It has to be underlined that the important requirement for any system of data is providing of the backup, archiving, structured saving and data restore in the proper term [9, p. 30].

Let us review the main operations, which are needed for effective functioning of the search service. First and the main operation linked to the receiving of the proper information is the selecting documents related to the topic is an information retrieval. Information search may be defined as a sequence of operations that are processed with the aim of finding the documents, that have some data (with further documents issuance or copying) or with the target of issuance of factual data, which has a reply to the set question. Documents search is reduced to comparison of two texts, from which one reflects the topic, which interests the user, and the second - the content of the document. The first text is called a query, the second – a search abstract [12, p. 93].

A. A. Hrechykhin made a classification of types of information search, pointing out nine criteria of selection: purpose, object, method, genre, chronological coverage, geo-

graphical coverage, completeness, intensity, category of information consumer [2]. Each selection criterion corresponds to certain types of search. In our opinion, to the named selection criteria you can add two more: search tools and search engine. To the criterion of the selection of "search tools" in the varieties of search indicate: 1. Search in traditional information retrieval systems; 2. Search in automated information retrieval systems; 3. Search the Internet. To the criterion of the selection of the "performer of search" in the types of search indicate: 1. Search, carried out by the information worker; 2. Search by the user.

Information search is carried out according to certain rules that define the search strategy, that is, the means to achieve the optimal result. The strategy of information retrieval depends on the type of search query, the criteria for issuing and the nature of the dialogue between users of information and the information retrieval system.

There are two different approaches to solve the issue with the information search – empirical and semantic. Basis of the first is the assumption that information search is essentially a simple process, the modeling and automation of which requires the solution of only technical tasks. Automation consists mainly of creating a vocabulary of terms from a certain field of knowledge (dictionary of descriptors) and appropriate equipment for storing and finding information. The second approach suggests that information retrieval is a complex creative process which subject matter is content. Automation of information search includes modeling of intellectual activity of a person, in particular understanding of its content of texts [12, p. 93].

Traditionally, two main types of information search are distinguished: documentary and factual. Accordingly, the documentary information retrieval system in response to a request formulating the requirements for the requested information (for example, the characteristics of a specific site or device that are of interest to the user) are indicated on the documents containing the necessary information - a description of the site or device. The factual information retrieval system provides the user directly with the information sought - technical data of the device, etc. In the process of information retrieval, two tasks are successively solved. First, it is necessary to find out in which the actual documents contain the sought information, and then - to obtain these documents or their copies [12, p. 93].

In general the process of information search can be defined on the following stages:

- formulation of task search;
- development of working search program;
- search implementation;
- search results publication.

The wording of a search task is most often executed in the form of a thematic section (keywords). The formulation of the task allows you to determine the area and peculiarities of used documents, facilitate the compilation of the search program.

The search program should maximally expand and specify the task, defining: object, types and methods, possible directions, necessary search restrictions – thematic, chronological, linguistic, genre, etc., possible objective and subjective complexities, degree of completeness, form submit search results.

In the process of information retrieval, a number of general methodological recommendations should be followed. For sure, the nature of the search is entirely determined by the content of the task. It is important to distinguish here that the search initially seems to be tied to the topic and the purpose of a detailed, comprehensive development of its plan,

when it is particularly necessary to have a wide coverage of information sources.

Therefore, at first the bibliographic and documentary types of search are used to a greater extent. And only then the information search is specified and limited to the direct content of the task being solved. At this stage, a factual search prevails. At the same time, throughout the process of information retrieval, it is necessary to combine different types and methods of search in two main directions: on the one hand, from general bibliographic aids to separate sources and facts contained therein, and on the other hand – from individual facts and sources to general bibliographic manuals.

In each particular case, it is important to clearly identify the starting point of the search engine in the information support system. For example, it is better to search in reverse order, from new manuals and sources, if necessary, consistently going deeper into history. Or start with what is already known or easily accessible, from the directories of one library. And only then to turn to other libraries, document-information institutes.

The most effective method for finding documents containing scientific information is to read each document in a separate library. But this is practically impossible because the number of documents is usually too large so that all of them can be read at each information request. Therefore, it is necessary to use another, less effective method, in which information search is carried out not by the texts of documents, but by short characteristics of the content or certain external features of documents. To do this, each document is provided in a searchable way – a characteristic, in which the basic semantic content of the document is briefly expressed. In the form of the same short characteristic, the search query image should be formulated. Due to this, the information search procedure can be reduced to a simple comparison of the search image of a document with a given search query image. Therefore, when an information search query, as a rule, the query is compared not with the full text of the document, but with its search pattern. If the search document of the document is sufficient and corresponds to the search query, then this document is considered to correspond to the information request. Such a comparison is justified only if the search image and the search query are formulated in terms of the same language, and in the same way in which each phrase permits one and the same interpretation.

In the search form of the document in a compressed form, only the main semantic content of the document is expressed. Therefore, such a method can not provide for finding in the document-information institutions of all documents containing the necessary information. In addition, among found documents may be those that do not actually correspond to this information request. These documents form the so-called "search noise".

Today the information search is carried out mainly in automated information retrieval systems. At one time Yu. A. Shreider outlined their advantages:

- compactness (the density of recording in automated directories is much higher than in publications);
- the possibility of using at a distance;
- the possibility of simultaneous access of a large number of users;
- efficiency of the management and use of directories;
- multifunctionality;
- the possibility of integration;
- visibility [6, p. 16].

Each document is presented in its search pattern in the information retrieval system, which characterizes the content of the document in the information retrieval language. Such a language can serve, in particular, any subject classification. In this case, the search image of the document coincides with the subject heading. The user's request is formulated as an indication of the subject heading, and in response to it, the system should issue all that belongs to this heading. These can be the actual documents or descriptions that allow them to be identified. Issuing documents (or information about them) from the search array, corresponding to the information request, is the final result of the search. If the result does not satisfy the user, he must make a new request, etc. [6, p. 17].

So, A. V. Sokolov, has outlined three stages of the development of the ideas of information search in bibliographic science:

- familiarization with the problem (by the middle of the 1960's);
- comprehension of the problem (the end of the 60's – early 1970's);
- development of the problem (nowadays) [8, p. 52].

The idea of information retrieval has become a factor in the emergence of a new scientific direction – document heuristics, which is intended to cumulate knowledge about the types and species of documents, their location, peculiarities of searching, taking into account the information needs of users.

Conclusions. The notion of "information search" is available in most national information standards. Information search is a set of actions to find information on a given basis. Since any information is recorded in documents, the search for such information can be called documentary. Although national science and practice have accumulated sufficient experience in the field of documentary search, however, a certain algorithm has not been developed yet.

Information search can be considered as a factor that prompted the emergence of a scientific direction – documentary heuristics, whose task is to study the theoretical and practical issues of documentary search.

REFERENCES

1. Explanatory dictionary on the basics of information activities. Kiev : UkrINTEI, 1995. 249 s.
2. Hrechykhin, A. A. Bibliographic heuristics: History, theory and methodology of information retrieval. Moskva, 1984. 46 s.
3. Information and documentation. Glossary: DSTU ISO 5127:2007. Kyiv: Derzhspozhyvstandart Ukrainy, 2010. 237 s.
4. Library and information activities. Terms and definitions : DSTU 7448:2013. Kyiv: Minekonomrosvitku Ukraini, 2014, III, 41 s.
5. Scientific and informational activity. Terms and definitions: DSTU 5034:2008. Kyiv: Derzhspozhyvstandart Ukrainy, 2009. 38 s.
6. Shreyder, Yu. A. Bibliographic tradition and catalog automation // Sovetskaya bibliografiya, 1986, № 2. P. 14–22.
7. Social communications: (theory, methodology, practice): slovar-spravochnik / [avt.-sost.: V. A. Ilganaeva]. Harkov, 2009. 391 s.
8. Sokolov, A. V. Automation of bibliographic search Moskva : Kniga, 1981. 167 s.
9. Starostina, O. V., Yanovskaya, O. V., Borisenko, D. A. Informational support of the population in the conditions of technogenic catastrophes, socio-ecological and interstate conflicts (Crisis Informology). Vinnitsa : Baranovskaya, 2015. 93 s.
10. Ukrainian Archival Encyclopedia Kyiv, 2008. 880 s.
11. Vorobev, G. G. Your information culture. Moskva : Molodaya gvardiya, 1988. 303 s.
12. Yakovishak A. V., Yevtukhova T. I. Аналіз сучасного стану технологій інформаційного пошуку An analysis of the current state of information search technology // Problemy nauky, 2014, № 4/5. P. 92–95.