

## LINGUISTIC SUPPORT FOR THE FORMATION OF A CONSOLIDATED INFORMATION RESOURCE OF SOCIAL MEMORY INSTITUTIONS

*Halyna Lypak*<sup>1</sup>, *Nataliia Kunanets*<sup>2</sup>,  
*Nataliia Veretennikova*<sup>3</sup>, *Volodymyr Pasichnyk*<sup>4</sup>

**Abstract.** In the article it is analyzed the role of linguistic means as the most important, system-forming elements of any information and library system, including a consolidated information resource of social memory institutions. Metadata of library digital electronic resources, archives and museums is characterized as a fundamental linguistic support of consolidation process of these resources and simultaneously as means of human and machine interaction in the development and operation of a consolidated information resource. It is outlined the criteria for choosing linguistic means of the development of machine-readable descriptions of information resources for the consolidated information resource of social memory institutions in a smart city.

**Key words:** linguistic support, metadata, machine-readable information, human-machine language, consolidated information resource, institutions of social memory.

### Introduction

The stereotypes of relationship between social institutions and society as whole and individual users in particular are changing under the influence of social and cultural and technological changes in the modern world. In new conditions the institutions of social memory (libraries, archives and museums) as the most open communication structures, not just are as objects of information process, rapidly mastering electronic technologies, but also play a significant role as subjects of this process. Using computer and technological tools in the library and information activities, not only substantially accelerates and improves the quality of customer service, but also changes the whole range of existing technologies. Automatization provides a range of new, previously in-

---

<sup>1</sup>Post-graduate student, the Department of Computer Science, Ternopil Ivan Pului National Technical University, Ternopil, Ukraine. [halyna.lypak@gmail.com](mailto:halyna.lypak@gmail.com)

<sup>2</sup>Doctor of Sciences from Social Communications, Professor of the Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine. [nek.lviv@gmail.com](mailto:nek.lviv@gmail.com)

<sup>3</sup>PhD of Sciences from Social Communications, Assistant of the Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine. [nataver19@gmail.com](mailto:nataver19@gmail.com)

<sup>4</sup>Doctor of Technical Science, Professor, Professor of the Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine. [vpasichnyk@gmail.com](mailto:vpasichnyk@gmail.com)

accessible possibilities for information workers and leads to dramatic changes in the information service.

Among the various forms of communication in the institutions of social memory the document communication dominates based on the use of printed and electronic documents. They can be considered as a process of interaction in particular area and suggest the involvement of two or more subjects. These institutions, based on the specifics of their information resources, promote efficient social and communication activity of specialists using special knowledge. Document communication of experts carried out in each of these institutions is not only in their physical space, it forms its contacts with users and in the online mode. At the request of the information society every artifact of cultural heritage, e.g. a book, an archive document or a museum exhibit, should be presented in the form of a digital document that is accompanied by a full comprehensive description of a source that satisfies the deepest information request. Sets of such digital documents form electronic collections and consolidation storages from different origin and content of electronic collections called digital (or electronic) libraries as ways of description and systematization of information resources because they are close to the library cataloging.

## **Main results**

Document and communication direction is leading in the activity of social memory institutions. It is associated with the creation of system directories — the sources of secondary document level, such as descriptions, catalogs and indexes, lists, surveys, registers of displaced values as well as documentary sources, history and culture instructions and other activities, including research products in the form of monographs, collections of scientific papers, articles and surveys [1].

The processes of information transfer that occur between a person and a person in the broadest sense are considered as information distribution between individuals and groups and are defined as a process of message delivery. Today, the communication process in the electronic library imposes new requirements for electronic communication in comparison with normal one, based on natural language, so the important components of communication in the library is a correct combination of natural and artificial languages. They are used to support the knowledge transfer with the difference that an artificial language can be directly used for encoding digital information.

Electronic communication takes the dominant positions with nonlinear structure, because it is possible to solve problems of integrating the whole intellectual potential of society into the documentary communications and ensuring its most efficient further use [2].

The rapid growth of electronic resources has changed the practice of fund management of libraries, archives, museums, greatly complicating it for different reasons, such as accounting standards of licensing and copyright, the need for conducting business negotiations on access to information, the need for internal planning while forming funds. In the electronic environment, the level of attention given to the format is similar to a semantic component of accounting documents.

Providing access to information resources, the most complete disclosure of information that is stored in libraries, archives, museums is only possible with the use of modern methods and means of information transfer, mastering the complex of linguistic tools to submit the document content in a user friendly manner and thereby ensure appropriate information service. Unfortunately, the problems of library technology and, in particular, the technology of semantic document processing, from which it is depended the access reliability to information resources and the quality of customer service to ensure their information needs, have been on the periphery of scientific interest of researchers for many years.

Electronic libraries are specific enterprises of providing information services and they are simultaneously repositories of information resources, processing them by means of information retrieval languages and indexing methods.

Modern software and hardware allow creating a comfortable environment for a user while working with an electronic catalogue, providing a wide range of service capabilities such as establishing quickly the existence of a specific document in the funds and its location, providing search of the required data in the workplace or at home, getting access to its digital copy. The most important characteristic when evaluating search capabilities of electronic catalogues is their semantic component and linguistic support of electronic catalogue, which forms the communication interface between a user and a digital library.

Actually, user communication with an electronic library comes down to human-machine interaction through a dialog system, search queries, a selection of appropriate properties, and introduction of specifying characteristics. This interaction is supported by linguistic means.

In the conditions of absence of a single research, scientific, methodological and information center for the development and maintenance of linguistic means, electronic libraries face with a number of challenges in practice, unresolved at the theoretical level, the largest of which is the problem of communication between a man and a computer [3]. Live, that is written or oral communication with a machine is a dream of every programmer. Some people believe that any attempt to mechanize language is impossible. Language as a resource and as a means of expression is something only appropriate for a human, or at least for the process of reality representation by a human. The language can be contrasted with one linguistic phenomenon retaining a lot of its properties at the same time is prepared to become an object of mechanization, while remaining a natural means of human expressions. This is a phenomenon of business prose that is a language carrier of industrial relations of people [4].

Speaking about a human and computer interaction using natural language, it is important to note a need and an opportunity to teach the computer to understand and accept fully a business prose. It is unreal to extend computers in all areas of industrial relations and ensure their appropriate interfaces in the form of formal regulations, standard forms and other means of information preparation. Time for processing or document synthesis is always compared to its time performance. However, the number of human and computer interfaces in dialog systems continues to grow [4].

The most important feature of business prose is that it always operates within the simulated situation unlike “live language”, it is always internally

formalized, and no matter whether a user is aware of this. Another specific but important feature is clarity of functions of each message. When a user sends an order for execution, makes an information request, perceives the received information, they should firmly know what a sender or a recipient wants.

Metadata is one of the linguistic means providing clear and unambiguous description and searching “visibility” of information resources. There is a wide variety of interpretations of metadata meaning, but we have chosen the following one: “Metadata is structured information about a resource of any type, which is used to identify a resource, to describe a resource, to manage a resource and provide an access to it” [5].

Interpreting the term metadata as “data about data”, researchers often draw an analogy with the methods of library cataloging that is creating records for information resources. S. Gradman offered to avoid this identification in 1998 at the IFLA General Conference in Amsterdam [6], assuming that metadata belongs to a specific type of materials, namely electronic resources and metadata relation to the source document strongly differs from the relation of catalog records to library books, because they are technologically stronger related to the original than the catalog descriptions. Around the same time, the founder of WWW Tim Berners-Lee pointed out even deeper difference which is crucial for metadata of digital information resources [7]: “Metadata is machine-readable information on web resources on the Internet. The machine-readable word is a key ...”

The various standards for metadata presentation are developed for different types of information resources. They enable the unique presentation of information, increase the reliability and the quality of object description, enhance the quality of documentation, opportunities, completeness and accuracy of information retrieval, and ensure the data compatibility.

Standard metadata can be focused mainly on the description and documentation of objects to search for information or exchange and storage of information. But essentially any standard metadata serves each of these functions to a greater or lesser extent [8].

Metadata standards can be seen as a human-machine language. A human describes the properties of an object in accordance with the standard requirements and they are given in machine-readable form suitable for computer operations, including processing, search, visualization, data exchange with other computer systems. Such machine-readable representation is called a format.

Brakker N. V. highlights in a standard as in any language, vocabulary, syntax and semantics. The human vocabulary includes a name of metadata elements (fields, attributes), lists of terms, thesaurus, authority files in natural language. The designed vocabulary for computer contains names of metadata elements (tags) and other elements of programming languages (e.g., XML).

Syntax defines the role of metadata elements and their relation to each other, defines the logical structure of metadata and corresponding rules that are called metadata scheme for a computer.

Semantics is a semantic metadata content, a pointer to the specific properties of an object. It is determined primarily by a human, but a computer can enrich semantically metadata, for example, adding concepts to lower levels of thesaurus [8].

Today information workers of social institutions that form integrated electronic libraries, the task of creating a unified electronic catalog containing descriptions of library, archival and museum resources, see a definition of “a nuclear set of data elements” based on the standards of all kinds of resources. Thus, this data set should be able to identify the resource, get an access to it by means of electronic library catalog, and it should be implemented a transition to more information, to owners (holders) of originals of resources or to a complete original description in text form [9].

Each of the existing formats offers the set of data elements, depending on the origin of the information resource and the purpose of this description. There are different approaches to the classification of metadata standards. But the biggest problem for professionals working with integrated digital libraries is selection / development of an optimal unique standard that can provide a comprehensive description and quick and quality access to resources of historical and cultural heritage, stored in different social memory institutions (libraries, archives, museums). Studies of consolidation of information resources from heterogeneous sources through the use of metadata standards are found in the works of Western scholars such as Erik T. Mitchell [10], Seth van Hooland [11], Richard Gartner [12], M. Woodley [13], Carolyn McCallum [14], C. Willis, J. Greenberg, H. White [15].

As a means of basic metadata to describe a large class of digital objects is commonly referred to DCMI (Dublin Core Metadata Initiative). This standard is used as a “minimum standard” of metadata to search resources and as a basis for other standards [16]. The prevailing operating standards of metadata description contain elements of Dublin Core or are its derivatives. From a linguistic point of view it is simple in implementation, so it can be recommended as a unique human-machine language to describe heterogeneous information resources to their consolidation within an integrated digital library, which, in particular, may be a consolidated information resource in institutions of social memory [17].

The task of linguistic study of human-machine language is especially relevant in connection with the development of artificial intelligence systems. For those systems it is typical to use adapted natural language in input, output and inside, and its functioning within the system could serve as a prototype of the language, serving as a mediator, an intermediate language between a natural intelligence and a natural language, which is an important element in making the keyword list in generated metadata to a document and information requests [18].

Consequently, the functional aspect of the XXI century library returns to its roots and becomes not only a library, but a repository of heterogeneous elements of humanity accumulated experience, knowledge, cultural and scientific values. Funds of modern electronic libraries are not connected to a single physical location, and consolidate information about available information resources using linguistic means. Along with the performance of traditional tasks of document and communication center, a digital library must ensure the implementation of full technological cycle of intelligent information technologies from bibliographication to the allocation of new knowledge from data warehouse and provide them to users. Linguistic means play a significant role linguistic tools

that are among the most important systemic elements of any information and library system. Their quality determines the effectiveness of the work of the whole system.

The rapid introduction of the new information technologies into the practice of libraries influenced on the development of linguistic software in recent decades. Understanding and using the effects of this exposure are considered to be appropriate, including the formation of consolidated information resources in institutions of social memory. Implementation of such information resources at the level of each city as a part social and communication environment of sociopolis will promote the principles of quality information service of residents and visitors in modern smart cities.

The formation of consolidated information resource of a smart city is based on the use of modern information and communication technologies, a certain order of development and operation of facilities, means of submission and processing poly-model information. An obtained information product as a result of systematically informed actions usually contains unified information of different types and use for its processing and presentation of diverse special technical devices, software and algorithmic and linguistic tools. This ensures the preservation of positive characteristics of traditional methods and forms of information transfer and reproduction and a maximum use of new features of modern information technologies, such as interactive communication, search and sort by set criteria, mobile access to information resources, and integration with the Internet.

## Conclusion

Therefore, the choice of linguistic means for the development of machine-readable descriptions of information resources for the consolidated information resource of the institutions of social memory in the city must ensure the possibility of achieving the goal of its functioning, and it is to provide the widest and the most convenient access to information digital resources of libraries, archives, museums. To achieve this goal, it is expedient to use information-oriented metadata created purposefully by information workers, for example, using a minimal set of elements of the Dublin core, and so-called “user metadata” created in the process of social communication in social networks that can provide the necessary depth of linguistic identification of the artifact description of cultural and historical heritage for further convenient search.

## References

- [1] Dubrovina, L., A. Kyrydon and I. Matiash. 2017. Archives, libraries, museums are the source of national memory and cultural heritage of Ukraine. *Library Bulletin*. 1:3–10 (in Ukrainian).
- [2] Kopanieva, V. 2007. Library in the system of scientific electronic communication. *Library Bulletin*. 5:3–9 (in Ukrainian).
- [3] Veretennikova, N. 2013. The Linguistic Support of Electronic Libraries. Computer science & Engineering: Proceedings of the 6th International Conference of Young Scientists CSE-2013. P. 232–233.

- [4] Ershov, A. P. To the methodology of building interactive systems: the phenomenon of business prose. Dialogue-78. <http://ershov.iis.nsk.su/archive/eaindex.asp?lang=1&did=11235> (in Russian).
- [5] Digitization: a landscape of standards for European museums, archives, libraries. 2016. [http://www.minervaplus.ru/publish/standards\\_landscape.pdf](http://www.minervaplus.ru/publish/standards_landscape.pdf) (in Russian).
- [6] Gradman, S. 2010. Digital Library Metadata. In: *Business Planning for Digital Libraries: International Approaches*. Leuven University Press, Leuven, Belgium. P. 157–166.
- [7] Berners-Lee, T. Metadata Architecture. <https://www.w3.org/DesignIssues/Metadata>
- [8] Brekker, N. V. About metadata standards in the museum area. <http://www.minervaplus.ru/publish/publish.htm> (in Russian).
- [9] Elings, Mary W. and Günter Waibel. 2007. Metadata for all: Descriptive standards and metadata sharing across libraries, archives and museums. <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1628/1543>
- [10] Mitchell, Erik T. 2013. Metadata developments in libraries and other cultural heritage institutions. *Library Technology Reports* 49.5:5–10. <https://journals.ala.org/ltr/article/view/4691/5583>
- [11] Van Hooland, Seth and Isabelle Boydens. 2009. Metadata quality in the cultural heritage sector: stakes, problems and solutions.
- [12] Gartner, Richard, Hervé L'Hours and Grant Young. 2008. Metadata for digital libraries: state of the art and future directions. JISC.
- [13] Woodley, M. S. 2008. Crosswalks, metadata harvesting, federated searching, metasearching: Using metadata to connect users and information. Getty Research Institute.
- [14] Mitchell, Erik and Carolyn McCallum. 2012. Old data, new scheme: An exploration of metadata migration using expert?guided computational techniques. *Proceedings of the American Society for Information Science and Technology*. P. 1–10. <http://onlinelibrary.wiley.com/doi/10.1002/meet.14504901091/full>
- [15] Willis, Craig, Jane Greenberg and Hollie White. 2012. Analysis and synthesis of metadata goals for scientific data. *Journal of the American Society for Information Science and Technology*. P. 1505–1520. [http://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=5391&context=faculty\\_scholarship](http://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=5391&context=faculty_scholarship)
- [16] <http://dublincore.org>
- [17] Baca, M. (Ed.) 2008. *Introduction to metadata*. Getty Publications.
- [18] Veretennikova, N. V. 2013. The creation of the linguistic corpus of the Ukrainian terminology texts on library science. *Science and World: international scientific journal*. 2(2):49–51.