

# The Linguistic Support of Electronic Libraries

Nataliia Veretennikova

Information Systems and Networks Department,  
Lviv Polytechnic National University, UKRAINE, Lviv,  
S. Bandery street 12, E-mail: Natasha19061990@bigmir.net

**Abstract – In this article the current state of library in the information technology development is presented, the features of linguistic support of electronic libraries are described, and the characteristics of the major information retrieval languages are detailed.**

Key words – information technology, information retrieval language, electronic library, linguistic support, linguistic tools.

## I. Introduction

The rapid growth of electronic resources has changed the practice of work of library collections, greatly complicating it with various reasons: the regard to the standards of licensing and copyright laws, the necessity to conduct business negotiations concerning the information access, and the need of detailed internal planning for the fund formation. However, the attention level, which is given to a format in an electronic environment, is close to the level of consideration of a substantive component, which is inherent in the document.

The recent studies of library development convincingly demonstrate that an electronic library is the most effective type of information system that stores various data, and is the relevant means of the implementation of integration processes. Electronic libraries are designed to work with a variety of information resources, but it is necessary a common information retrieval language to provide the user interaction with these resources.

## II. The Peculiarities of Electronic Libraries

The rapid development of information technologies, including the incredibly rapid development of the Internet, have recently caused the massive creation of electronic information resources in the all fields of knowledge and in the different activities. The Internet has become an important environment for the creation and the distribution of information resources, and the retrieval realization in them. The class of information technologies, which is aimed at the organizing of electronic resources in the global networks and the retrieval technology in them, is known as electronic or digital libraries [2].

Providing the access to information resources, and the full disclosure of information that remains in libraries and archives, are only possible for using of the modern methods and means of the information transformation, the development of linguistic tools complex that allow to present the document content in a user friendly manner, and thereby to provide the appropriate information service. Unfortunately, the problems of library technologies and, in particular, the technology of semantic processing, on which depends the reliability of the access to information resources, and the quality of the customer service concerning their information needs, were on the periphery of scholar research for many years.

Today, the library is considered and evaluated as a specific producer of information (library) products and services, and their processing is carried out by means of information retrieval languages and indexing techniques.

Modern software and hardware can create a comfortable environment for the user during the work with electronic directory, and provide a wide range of service capabilities. However, the semantic component and, above all, the linguistic support of electronic catalog, which forms the communication interface between the user and the library, are the most important characteristics in assessing the retrieval means in electronic catalogs.

The creation and the use of linguistic support of electronic library are based on information technologies that affect the development and the implementation of language tools. First of all, there are the common technology of operating and the architecture of information systems; it is certainly a theory of information retrieval, the classic and modern solutions in the field of technology databases and, finally, the technology of user interfaces.

## III. The Approaches to the Definition of Linguistic Support

Currently, there are several approaches to the definition of linguistic support. The *classical* approach is the most famous, according to it the complex of information retrieval languages, especially classified and verbal (descriptor) ones, is called the linguistic support [3]. This approach is adopted in the theory of automated library systems with minor modifications. But here, in contrast to the classical approach, the languages of the formation of bibliographic data are usually related to the concept of linguistic support.

There is also an approach that can be called “*linguistic*”, because it naturally follows from the linguistic point of view on the information systems and it is developed mainly by specialists in applied and computational linguistics. According to this approach, the linguistic support is a set of tools that are used for the automated processing of natural language texts (including the processing of queries and retrieval results), which are primarily linguistic processors.

More common is the approach that should be defined as “*semiotic*”, because it comes from the classic semiotic notions of language as a system of signs at various levels, naturally starting with the alphabet. In this approach, the linguistic support is defined as a means of the presenting information in the form of data and the interpretation of this data. In addition, the tools of alphabet coding or the formats of data should be include to the structure of linguistic support, but the instrumental programming language should not be include to it.

We assume that linguistic support of electronic library is a set of language tools and processors for the processing, the presenting and the retrieval of written texts in natural language, mainly on the semantic level.

## IV. The Classification of Tools of Linguistic Support

Languages, that are part of linguistic support of library, are designed specifically for the data presenting in an electronic library. These data are presented in the form of

some isolated and identified information resources that can be called “digital objects” [1].

The common term “information retrieval languages” should be correctly used for this class of language tools. These languages are quite naturally classified according to the level of information display that is available in digital objects. There are four levels [1]: 1. The level of display of digital objects completely, including its formal characteristics; 2. The level of display of topics or content of a digital object; 3. The level of display of semantics units of natural language that are in a digital object; 4. The level of display of expression that are in a digital object.

The languages of document describing that are developed in the traditional fields of information activities: librarianship, archives, office work, mapping and so on, are used for describing the digital objects of documents at the first level display. *The presentation languages of bibliographic data* are formed the most famous type of these languages, and is based on the rules of bibliographic description and formats of bibliographic record. These languages appeared in XIX century.

Today, there is an active integration of bibliographic data of traditional documents with languages that are used to describe different types of digital objects. Particularly, this process develops actively online. *Metadata systems* are the common name of languages which are intended to describe digital objects.

At the second level of language display, it is used languages of classified or upper-coordinate type, which also have an old historical tradition. The principal feature of these languages is the distribution of a set of digital objects into classes, which are described by means of a priori connection (upper-coordination) of retrieval features of these classes, often in a hierarchical tree. Languages of this type with regard to the prospects of global information networks cause heated debates, due to their inherent drawbacks, the main one among them is the need for intelligent indexing. These classified languages have significant advantages between other types of retrieval languages and are a compulsory component of practically all electronic libraries.

The newest type of language tools, that appeared only in the automated systems in the 1950s of the XX century, are *verbal languages*. However, the most common name is *descriptor languages*, according to the name of a common form of a representation of lexical units of these languages (descriptors). Sometimes the language data is also called upper-coordinated, emphasizing the contrast with classified languages according to the basic function: the way of display of text information. If the priori connection of retrieval features is used in classified languages, then the features are connected directly to the digital object in descriptor languages.

Sometimes you can come across studies where all problems of linguistic support come to the investigation of verbal languages, especially during the searching for the full texts. It is usually too narrow an idea, but there is

no doubt that verbal languages are a central component of linguistic support of electronic library.

All of these types of language tools can be called languages with more or less convention. However, defining an object as a language, we should be able to distinguish compulsory components for any language in its structure. The units of three levels are distinguished in any language: Alphabet is a set of acceptable signs; Vocabulary is a set of semantically interpreted signs; Texts (discourse) are semantically interpreted significant units of speech [1].

It is also distinguished two classes of rules (grammar) in any language: morphology is the rules of formation and change of lexical units; syntax is the rules of text formation.

Semantically interpreted significant units of language (vocabulary and texts) have three types of relations (properties): Syntactics is the relation between the signs; Semantics is the relation between the sign and the marked (referent); Pragmatics is the relation between the sign and the discourse participant.

This scheme is commonly modified in the linguistic support theory and practice of electronic library.

So, the linguistic support of electronic library includes the following language tools:

1. Information retrieval languages: metadata, classified languages, verbal languages, factual languages.
2. Linguistic processors: automatic text processing, linguistic databases.

In general, the main functions of linguistic support of libraries are associated with providing the documents and queries indexing, and the efficient search in the electronic catalog.

## Conclusion

The concept of linguistic support of library technologies opens up the broad prospects of research on the most pressing issues, which are related to semantic analysis of documentary sources and diversification of information products and services, ensuring compatibility of linguistic tools that serve the functioning of library systems.

## References

- [1] A. B. Antopol'skij and K. V. Vigurskij, “Jelektronnye biblioteki” [Electronic libraries], *Informacionnye resursy Rossii – Information resources of Russia*, vol. 4, 1999.
- [2] V. Kopanieva, *Biblioteka yak tsentr zberezhennia informatsiinukh resyrsiv Internety* [The library as a center for preservation of information resources of The Internet]. Ukraine: Academy of Sciences Publ., 2009, 198 p.
- [3] I Szyh, „Lingvisticheskoe obespechenie jelektronnogo kataloga“ [Linguistic support of electronic catalogue], *Biblioteka – Library*, vol. 6, pp. 35-38, 2009.