

Olga Lavrenova
Russian State Library, Moscow

Knowledge Organization Levels for Computerized Information Systems



Dr.(Philology) Olga Lavrenova specializes in "structural, applied, and mathematical linguistics". Since 1970 work in developing IR systems, indexing methods and structuring of thesauri. Since 1986 design of an automated information-library system as Chief of the Automation Department, Russian State Library.

Lavrenova, O.: Knowledge organization levels for computerized information systems.

Knowl.Org. 20(1993)No.4, p. 207

The work represents an attempt to organize concepts used in the field of the development of linguistic software for information retrieval systems, including OPACs. Four levels of knowledge organization are suggested. Each level requires a specific model of knowledge and a specific terminology system. The problems under review are connected with the explicit presentation of knowledge contained in speech statements (texts) and with the preliminary system of knowledge, which helps their comprehension by man or by a system. (Author)

Four levels of concepts associated with knowledge representation are suggested. The first level represents the semantic space, i.e. the abstract multitude of concepts which reflect the world of reality in human consciousness. Here we deal with the conceptual model of the world. The second level is formed with the help of natural languages in the form of texts, which in contemporary linguistics are viewed as the main units of expression of meaning. The meaning of the text corresponds with the multitude of points in the semantic space. These points are coordinates of the meaning of the text contained in it.

For the development of linguistic means for an Information Retrieval Language (IRL), the aspect of organization of knowledge in the process of communication during information searching in the "man - man" system, is significant.

In the research of processes of speech communication, the following 4 aspects are distinguished: communication intention of the speaker, the general base of knowledge of communicants, the correlation of the new and the old (known) elements in the semantic structure of the text, and methods of sign expression of the communicated text.

All these aspects should be taken into consideration when developing IRL and structures of the "man - system" dialog. The artificial models of the semantic structure of texts, created for the solution of pragmatic problems, comprise the third level. The semantic models are formed with the help of IRL and other artificial languages.

The structure of every IRL is based on some hypothesis in the organization of knowledge in the course of information retrieval in a multitude of texts. In a number of papers the model of knowledge based on the idea of the structure of human activity have been described (1-4). The model

makes it possible to design standard structures of knowledge. The model is used for the elaboration of different IRL.

To the fourth level belong means of description of elements of artificial languages that present the meaning of the texts, for example the normative notation. Formats of the UNIMARC Authorities type are essentially descriptions of metalanguages of this level of knowledge organization for IRS.

It should be stressed here that characteristics of use of IRL elements in concrete IRL do not belong to this 4th level. Indexing terms and lead-in terms may form the 5th level - the level of the use of knowledge in a concrete system.

The machine-readable version of the Library-Bibliographical Classification, now being created at the Russian State Library in the form of an authority file can be cited as an example of such an IRL.

On the suggestion of the Russian State Library, a complex database of descriptions of all the main classification systems, codifiers and thesauri is being created (see the papers by Fedosimov and Beloozerov). Unfortunately, for reason of lack of means the work on this project has been discontinued for the moment. This conception could well become a theme of an international program.

Along with this we suggest the following: a) when making up terminological systems and preparing guidelines and standards on terminology, we should aim at describing in a document those terms belonging to level one, and determine only the initial ones through terms of the other levels; b) design the descriptions of IRL and the training of students in the IRL course as a movement along the levels of representation of knowledge.

References:

- (1) Lavrenova, O.A.: Problems of representation of semantic texts in automated information retrieval systems. (In Russian). Nauch.-tekhn.Inf., Ser.2(1980)11, p.26-31
- (2) Lavrenova, O.A.: Model of the knowledge system used in automated information systems. (In Russian) Nauch.-tekhn.Inform.,Ser.2(1989)No.6, p.5-9
- (3) Vostrikov, A.N., Lavrenova, O.A.: The elaboration of the linguistic aspect of the electronic catalog of the Lenin State Library. In: Access to Library Resources through Technology and Preservation. Proc.1988 US - USSR Seminar, July 5-8, 1988. Washington, DC: Chicago: Amer.Lib.Assoc.1989. p.73-76
- (4) Lavrenova, O.A.: Thematical search in the automated library system on the basis of the Library Bibliographical Classification (LBC) with subject entry. In: 57th IFLA General Conf. Moscow, 18-24 Aug. 1991. Paper 4-31-36.