
ORIENTAL MANUSCRIPTS AND NEW INFORMATION TECHNOLOGIES

***G. V. Lezin, K. K. Boyarsky,
E. A. Kanevsky, A. I. Popova***

PROGRAMMING OF TEXTS CONCEPTUAL TREATMENT

Introduction

Computer representation of knowledge is an actual but not yet solved problem of historical investigations methodology. The huge volume of knowledge in the humanities is presented, as a rule, in the form of texts. In historical investigations the texts of the literary monuments written in ancient languages as well as pieces of ancient scripts serve as the principal source of information. The problem of scientific knowledge formalisation is thus connected with the problem of computer representation of semantics of natural language texts and their conceptual modelling.

Within a wide scope of the problems relevant to the modelling of communications employing a natural language (NL), we are taking interest, first of all, in a special case of the NL communications, which aims at the accumulation, systematisation, and the conceptual modelling of the texts in a concrete objects area. This process is meant to create a sublanguage with a formalised semantics, providing computer "understanding" of the texts in a given sphere of communications. It is the process of creating a problem-oriented sublanguage where a user of the knowledge base is to determine main practical and communicative objects of information accumulation.

The studied approach to the use of the model of understanding of a NL-text requires realisation and investigation of two level-interrelations of a user with the knowledge base created by him. One of them, conventionally denoted as the "level of formation of communication sphere", is used in those cases when the base should be upgraded by a set of primary objects, in order to ascertain their properties

or initial states, to complete a conventional list of semantic linkages, to introduce certain scripts or initial rules for the performing of plausible reasoning necessary for the text understanding. This is the level for communications employing formal means of interaction with the knowledge base. The next one (the NL-level) assumes exchange with the base using a problem-oriented sublanguage as well as automatic — that is, performed by the base analyser, extraction of conceptual information from the text.

In this paper we consider the formal means of interaction with the knowledge base on the level of creation a communication sphere. At the base of determining of the means under consideration lies a view on the knowledge base of a humanitarian investigation as on the system of hypertext formatted as an electronic card catalogue. In a number of works [1] were published the backgrounds of the general approach to creation of such a system and were considered certain elements of the technique of semantic encoding of NL-texts (with the example of employing the V. V. Martynov's semantic primitives [2]). The means of exchange with the knowledge base considered in this work are determined as a language for recording the texts content and for programming their conceptual treatment (a formal language). Using a definition "conceptual treatment" we follow R. Shank [3] and assume the procedures of the text analysis and synthesis, which are based on the scope of definitions and their interrelations created beforehand and upgraded in the process of the analysis.

1. The Formal Communication Language

In our approach to the definition of the formal language we were guided by the following requirements to the practical representation of the means of interaction with the knowledge base:

a) the formal language is a part of the general set of means for the description of a linguistic processor, the "carrier" for the problem-oriented language;

b) in the framework of this language, the means of hypertext (formation and requests at the reading), the philological toolkit of a humanitarian investigation (creation of thesauruses and lexical pointers, determination of semiotic juxtapositions and synonyms series), the means of control over factographical data bases, and the means of definition of a problem-oriented sublanguage (description of infor-