STRUCTURAL MODELING OF ANALYSIS AND SYNTHESIS OF TECHNICAL TEXTS PROCESSES

Vasyl Andrunyk [1], Rostyslav Bekesh[2], Liliya Chyrun [2]
Information Systems and Networks Department, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, UKRAINE, E-mail: [1]phottt@ukr.net, [2]rostyslav.bekesh@gmail.com, [3]lchirun@mail.ru

The peculiarity of modern Ukrainian scientific and technical terminology development is an increased interest in its authenticity, as historically this terminology has become unavailable to users. The thing is such terminology has been removed out of official dictionaries and textbooks. Prohibited dictionaries were sent to the special storages of libraries and they were given only by special permission. Until today the dictionaries dated by 1920-1930 years exist in the form of single copies, or are not reached at all – they were lost or destroyed. Even the very existence of many terminological dictionaries is now known only to a narrow audience of industry specialists.

It is estimated that approximately 90 % of the new words, which appear in each language, are terms. Modern Ukrainian terminology is also actively being updated with the new units – in most cases, they are borrowings from English or linguistic calques from Russian. Despite the fact that the Ukrainian language partially assimilates words of others, it is still a large number of borrowed words which constitute a threat to the national terminological system clarity and often negatively affect the rate of the learning process. Flattering is the fact that Ukrainian equivalents arose before the emergence of specific new borrowings in Ukrainian terminology, for example: трастове товариство – довірче товариство (trust company), апроксимація – наближення (approximation), детектор – виявляч (detector) etc. In case such tendency continues, the majority of "modern" borrowings will become passive; the subject matter terms will remain. The terms are not translated as the regular words from one language to another. The optimal way of terms translation should be as follows: "concept -> Ukrainian term" rather than "foreign-language term -> Ukrainian term"; in this case, source language doesn't play a role (V. Morhunyuk). In other words, equivalent word searching begins with the analysis of new concept properties. Unfortunately, in most cases, the translation of terms into Ukrainian is done through "calques".

The purpose of this work is intelligent system development for modeling analysis and synthesis processes of texts of a technical nature, such as: checking the correctness of terms usage in articles according to the generally known rules and the possibility of these articles ontology's constructing. The established system is based on morphological analysis algorithms, namely: it is based on the modified morphological analyzer. The difference of the established system from already existing morphological analyzers is its narrow specialization while searching terms in technical texts, especially in articles. The object of this work is to study morphological systems of text analysis. The object of study is the algorithms of morphological analysis, stemming and automated constructing of ontology's. Theoretical significance of the work lies in the analysis of already known algorithms of morphological analysis, stemming and ontology's constructing methods. Practical significance of obtained results lies in the realization of methods composition of morphological analysis and stemming in order to improve the efficiency of searching incorrect words and phrases in the text, and the possibility of ontology's constructing.

Structural intelligent system of process analysis modeling and technical texts synthesis was developed. In order to provide the design and implementation of system structural analysis the environment All Fusion Process Modeler 7 was used; with the help of which a system functional diagram as well as data flows diagram and operations sequence logical diagram were created. All components and functional parts of the system were developed by means of visual programming environment C ++ Builder 6.0, using SQL-queries. Database designing was carried out in My SQL Workbench environment. Literature data analysis and research in the field of automated processing of texts and morphological analysis were conducted; the existing systems (similar in functionality) were considered. The developed system of modeling analysis and synthesis processes of texts of a technical nature performs the search of incorrect words and phrases in the text of a technical nature, especially in articles. Keyword searching function, which is based on Zipf's law, is additionally provided in the system. The found keywords can be used for constructing an article ontology in the form of XML-documents, which is a very important fact because this format has become a standard for data exchange between applications. In order to automate the validation stage it would be more efficient to

use third-party dictionaries and thesauruses, which may lead to the development of WordNet Ukrainian equivalents.

The result of system operation is finding out an incorrect word or phrase as well as a list of words or phrases that can replace it. The system also displays a list of keywords mentioned by the author of the article as well as a list of keywords founded by the system, and the result of checking their matches. Despite the functionality of the system, it also has some drawbacks. Since the whole operation of the system depends on the dictionary content, first of all, the replenishment of the terminological base should be performed. Deprivation of the developed system drawbacks could be the first step in its further development. In particular, the work of the system should be continued in the direction of improving algorithms of irregular words and phrases searching, e. g. searching algorithm implementation, which is based on neural networks. In order to increase the efficiency of keyword searching algorithm, it is necessary to implement the possibility to perform keyword searching not in the single file, but in multiple files related to the similar subject for the purpose of discarding words, which are typical for the texts with a similar subject, but not crucial. The distinctive feature between created and existing systems at the current stage of its development is narrow specialization of the system in specific subject texts (especially in technical articles). So, the developed system of modeling analysis and synthesis processes of texts of a technical nature is a simple tool for searching incorrect words and phrases as well as for ontology's constructing, and may be used for non-commercial purposes.

Keywords – generative grammar, structured scheme sentences, computer linguistic system.