

SUMMARY

Of the PhD Thesis on the specialty 6D070300 – «Information Systems»

Aislu Kassekeyeva

«The model of semantic library containing information in English, Russian and Kazakh»

Digital literacy is the basis of security in the information society, one of the most important knowledge of the 21st century, one of our main themes in today's society. Digital literacy is the readiness and ability to reliably and effectively use digital technologies in all spheres of human life. The use of this technology improves the quality of life of the population.

In his Address dated January 31, 2017, "The third revival of Kazakhstan: global competitiveness," President Nursultan Nazarbayev emphasized the need to develop a state program "Digital Kazakhstan": "... We must develop new industries created using digital technologies. This is an important and complex task. The country needs to develop areas such as 3D printing, online shopping, mobile banking, digital services, healthcare, education and other promising areas. These industries have already restructured the economies of developed countries and have given a new quality to traditional industries. In this regard, I instruct the Government to develop and adopt an individual program "Digital Kazakhstan". ... The development of the digital industry will give impetus to all other areas. Therefore, the Government should pay special attention to the development of the IT industry. "

Each country chooses certain factors that affect economic development and integration into the world space. The Republic of Kazakhstan makes one of the main factors education that meets the requirements of the world economy, increases the standard of living and the general well-being of the population. The governments of many countries consider increasing the competitiveness of the economy through the development of the quality of educational services as one of the main tasks. This is due to the fact that the level of education of society and scientific potential is an important condition for economic growth. In the context of modern globalization processes, an international system of higher education is being formed, which is a set of national interconnected systems.

Education is one of the most important criteria of the "Kazakhstan - 2030" Strategy. The main goal of educational programs is to adapt the educational system to the new economic environment. The President of the Republic of Kazakhstan set the task of joining the republic among the 50 competitive countries of the world. Also, according to the Law of the Republic of Kazakhstan "On National Security of the Republic of Kazakhstan", one of the threats is the deterioration of the quality of education and the intellectual potential of the country, which proves the enormous importance of creating and improving a quality education system.

Until recently, digital libraries were perceived by ordinary users as electronic versions of traditional library catalogs that contain descriptions of physical library objects (usually books or other printed publications). The definition of the subject matter, content and structure of objects are considered and perceived as additional, but optional functions of such libraries. The development of the Internet and

semantic technologies makes its own adjustments and allows you to take a broader look at the concept of digital libraries and generalize the accumulated experience in the implementation of information systems in different areas of knowledge to form a new type of libraries.

The very concept of a library in the context of the rapid development of the Internet acquires a completely different meaning and denotes the active involvement of the user in the processes offered by the libraries. Such a library assumes the participation of users in the process of creating, searching and classifying the library content that is necessary for this particular user.

Topicality of the research. Due to the rapid growth in the volume of textual information, research in the field of computational linguistics in natural language remains relevant. Today, the amount of information that humans and machines reproduce in natural language has greatly increased. The creation of algorithms and the creation of data collection systems, the model of semantic libraries containing information in different languages, the classification and clustering of libraries with text documents are still complex tasks.

Many researchers are inclined to the need to carry out a deep semantic analysis of texts to create their semantic images, on the basis of which it would be possible to carry out a fine ranking of documents. This approach is undoubtedly the most reasonable, but it requires careful and long work to create suitable tools for automatic word processing. In particular, a detailed description of different areas of expertise may be required. Therefore, it also makes sense to search for partial solutions, one of which is presented in this work.

The continuous increase in the intensity of the flow of textual information makes the task of the semantic model of a library containing information in different languages more and more important.

Semantics is a section of linguistics that studies the semantic meaning of language units: individual words, phrases, sentences, text fragments. Currently, there are several machine-oriented methods for representing operator values.

For example, I.A. Melchuk introduced the concept of lexical function, developed the concept of syntactic and semantic valence and considered it within the framework of an explanatory combinatorial dictionary. V.Sh. Rubashkin and D.G. Lahuti introduced a hierarchy of syntactic links for the efficient functioning of the semantic analyzer.

The continuous increase in the intensity of the flow of textual information makes the task of the semantic model of a library containing information in different languages more and more important.

The theoretical basis was scientific works containing research on grammars of links, syntactic analyzers of texts in natural language, methods for comparing sentences and determining topics of texts, algorithms on graphs and mathematical logic.

The aim of the research and scientific outcomes.

The purpose of the dissertation research is to develop a method and model of the semantic library containing information in English, Russian and Kazakh

languages, based on research and results obtained in the analysis of space-time designs in a natural language, and the developed Semantics application.

The object of the research is a semantic library to support scientific and educational activities.

The subject of research is models, algorithms and methods of semantic libraries, as well as methods of classification and clustering of resources in the electronic library.

The scientific novelty. The following questions are of the greatest scientific interest: The classification of spatiotemporal constructions in the sentences of the Kazakh, Russian and English languages is proposed. The method and algorithm of semantic equivalence of sentences containing spatiotemporal constructions and simple sentences of Kazakh, Russian and English languages are proposed. A model and prototype of a semantic library using semantically equivalent sentences of the Kazakh, Russian and English languages are developed.

The algorithm of equivalence of spatiotemporal relations is implemented using the Link Grammar Parser and Dealing analyzers, the server part of PHP, the client part of HTML, CSS, and MySQL databases. The information system of the semantic library containing information in English, Russian and Kazakh languages is created, taking into account the model of formal representations.

Research objectives:

In accordance with this goal, the following tasks are solved in the dissertation work:

1. Analyze and classify spatiotemporal structures;
2. Create a mathematical model of the set of formal representations of sentences in the language of equivalence;
3. Develop an information system for the semantic library Semantics.

Statements for the defense:

1. Formal classification of space-time structures.
2. Mathematical model of a set of formal representations in the language of equivalence.
3. Software implementation of the semantic library information system Semantics.

Practical significance of the research outcomes. The proposed models, methods in the form of a prototype of an intelligent system can be used in intelligent information search systems. The results of the study is also of interest to scientists of linguists.

The practical significance of the main provisions of this dissertation was confirmed by the results of the use of developed models, methods and algorithms for developing a semantic library at the Institute of Informatics Systems. A.P. Hershova Siberian branch of the Russian Academy of Sciences.

Personal contribution of the PhD student. Models and methods of the semantic library containing information in English, Russian and Kazakh languages were proposed, described and developed by the author personally. Analysis and sample of the most important concepts related to space-time relations, from an intelligent dictionary S.I. Ozhegova and other sources and their translations,

identifying space-temporary forms in documents for use in the semantic library in three languages was personally developed by the author. Creating an array of paraphrased variants of various proposals and the method of assessing them to also also personal contribution of the author. The experimental assessment of the algorithm and the model was carried out in conjunction with the scientific consultant and the Institute of Informatics Systems. A.P. Yershova of the Siberian branch of the Russian Academy of Sciences.

The main provisions submitted to the defense are the personal contribution of the author to the published work.

Approbation of the thesis outcomes. The validity of the results obtained in the dissertation work is based on the use of proven research methods, the correct application of the mathematical apparatus of graph theory, set theory and vector algebra, statistical methods of data processing, the coordination of the results obtained with the known theoretical provisions in the field of natural language text processing and intellectual decision support.

The adequacy of the proposed methods and algorithms is confirmed by the results of the implemented intelligent methods on the presented text corpus, as well as the results of testing and the act of implementing the prototype of the intelligent system. The reliability of the results obtained is ensured by the use of well-known methods and approaches to the classification of texts, correct statistical data processing.

The main issues and results of the work were announced and discussed at the following scientific conferences:

- Collection of abstracts of the Republican scientific and practical conference " modern problems of mathematical and computer modeling in the context of the development of the digital industry of Kazakhstan "(May 3-5, 2018, Astana);

- International scientific and practical conference "integration of Science, Education and production - the basis for the implementation of the national plan" (Saginov readings No. 10) (June 14-15, 2018 Part 2, Karaganda 2018);

- XIX International Scientific and practical conference" Advances in Science and Technology "15 March 2019 scientific and innovative center" actuality. Russian Federation", (March 15, 2019, Moscow);

- BIG DATA and analysis of the highest level, collection of materials of the V International Conference. (March 13-14, 2019 . Minsk, Belarus);

- Six international scientific and Practical Conference BIG DATA and Advanced Analytics BIG DATA and analysis of the highest level (May 20-21, 2020 Minsk, Belarus);

- Collection of materials of the XV International Scientific Conference of Students and Young Scientists "YLYLYM JÁNE BILIM-2020", Kazakhstan, (April 10, 2020, Nur-Sultan).

- III – IV International Multidisciplinary Conference Prospects and key tendencies of science in contemporary world, Madrid, Spain, February, 2021.

The results of the dissertation work were implemented at the A. P. Yershov Institute of Computer Science Systems of the Siberian Branch of the Russian Academy of Sciences, which is confirmed by the acts on the implementation of

scientific provisions and developments of the dissertation in the practice of educational organizations given in the appendix.

Publications. According to the results of the dissertation research, 13 publications were published, including 4 works in peer-reviewed publications recommended by the CCSON, 1 work in publications indexed in Scopus, and 8 publications in other scientific journals and conference proceedings. 1 copyright certificate on entering information into the state register of rights to objects protected by copyright of computer programs was obtained..

Volume and structure of the thesis. The dissertation work consists of an introduction, three sections, a conclusion and a list of references. The volume of work is 115 pages, including 22 figures, 3 tables, 3 appendices. The list of references contains 107 titles.

The introduction presents the scientific apparatus of the research, substantiates the relevance of the topic, the degree of its development in theory and practice, defines the purpose, object, subject and objectives of the research, reveals the scientific novelty, theoretical and practical significance of the work, defines the research methods, presents the provisions submitted for defense, the personal contribution of the author, the list of publications and approbations of the results of the work.

The first chapter includes an analysis of the current state of logical and philosophical research of semantic libraries. The article analyzes the methods of information search, which are actively developing, relevant in the scientific and practical aspect today. The following models of information search are considered: Boolean model, vector model, probabilistic model. Shortcomings of some models are revealed. The key problems of comparison of modern used semantic libraries are identified.

The second chapter describes the formal analysis of paraphrased natural language sentences by Link Grammar Parser analyzers and Dialing for analyzing dictionary entries and usage examples in fiction related to space-time relations.

The third chapter describes the model and algorithms of the semantic library containing information in English, Russian and Kazakh, as well as the software implementation of the analysis of sentence diagrams. The interface of the system for forming a model of a semantic library and a database for classifying spatiotemporal constructions of three languages is described. A mathematical model of the semantic library containing information in three languages is also provided.

In conclusion, the results of the study are summarized, the main conclusions are formulated, confirming and proving the truth of the provisions submitted for defense.

The **appendix** contains practical research materials.

The author expresses gratitude to his scientific advisor, Doctor of Physical and Mathematical Sciences, Professor of the Department of Information Systems Tusupov Dzhamalbek Aliaskarovich for setting the problem and helping in the course of the dissertation research.

The author is grateful to his foreign consultant, Ph.D. A.P. Ershov SB RAS Murzin Fedor Alexandrovich for consultations during the research.

The author also expresses gratitude for the advice of D. Filol.professor of the Department of General Linguistics and Translation Theory, Saula Zhaksylykbaevna Tazhibayeva.