

¹ Gulmira Bekmanova² Altynbek Sharipbay³ Dinara Kabdylova⁴ Nurzada Amangeldy⁵ Ayaulym
Sairanbekova⁶ Assel Omarbekova⁷ Altanbek
Zulkhazhav

Development of an electronic dictionary of terminology for school textbooks



Abstract: - This research, funded under the BR11765535 “Development of Scientific and Linguistic Foundations and IT Resources to Expand the Functions and Improve the Culture of the Kazakh Language” project, aimed at advancing the scientific and linguistic foundations of the Kazakh language and enhancing its role in interethnic communication through digital resources. Focusing on the urgent need for specialized digital terminological dictionaries for school subjects, the project addresses the limitations of existing resources, particularly for school students. The article discusses the criteria for inclusion in the terminological dictionary, emphasizing the significance, frequency, and thematic relevance of terms across various subjects. The project has undertaken the compilation of a comprehensive database covering key school subjects in the Kazakh language.

Keywords: Electronic dictionary, textbooks, electronic textbooks.

I. INTRODUCTION

This work was funded in the framework of the research project BR11765535 “Development of Scientific and Linguistic Foundations and IT Resources to Expand the Functions and Improve the Culture of the Kazakh Language”. The goal of the project is to develop scientific and linguistic foundations and IT resources to expand the functions and improve the culture of the Kazakh language as a language of interethnic communication in a digital format, which is an urgent and important problem in the strategic direction of the development of our state.

To date, the field of development of terminological dictionaries is one of the rapidly developing areas. However, there is a growing need to develop digital terminological dictionaries, since terminological dictionaries that are widely used today cannot meet all the requirements, especially with regard to school terms. For example, if we take the popular dictionary termincom.kz, despite the fact that it contains 379,465 terms, the term search does not take into account the possibility that the end user is a school student, and the term itself, Russian translation and category are returned in response to a query about the term. That is, this term does not contain information about in which class, in which subject the school textbook was considered, respectively, for younger school students it is inconvenient and useless to use such dictionaries. The dictionary presented in this project can become an indispensable tool to help students, parents and subject teachers, since digitized textbooks, a base of terms collected for each subject are available during the project.

II. DEVELOPMENT OF A DATABASE OF TERMINOLOGY OF SCHOOL TEXTBOOKS

The criteria for determining the inclusion of lexical units in the terminological dictionary were taken as: significance of the term (or semantic value), use (frequency), thematic relationship of the term.

^{1,2,3,4,5,6,7} L. N. Gumilyov Eurasian National University

1gulmira-r@yandex.kz, 2sharalt@mail.ru, 3dinara.kabdylova@gmail.com, 4nurzadaamangeldy@gmail.com, 5sairanbekova98@gmail.com, 6omarbekova@mail.ru, 7altinbekpin@gmail.com

Copyright©JES2024on-line:journal.esrgroups.org

When compiling a terminological dictionary for primary and high school students, one term is given two or more different definitions, depending on age characteristics and the peculiarities of information perception (Table 1, 2).

Table 1 – two-level interpretation

Term	Definition	Form
Figurative words	words that arose from imagining the image of various phenomena in nature	
Figurative words	represents the names of the concepts that are born by seeing the known phenomena that occur in nature with the naked eye	7
Imitative words	words formed from imitation of sound	4
Imitative words	words formed from the perception of sounds arising from the collision of various phenomena and objects that occur in nature, as well as various sounds emanating from the vocal organs of animals, as well as concepts known by the ability to hear	7

Table 2 – three-level interpretation

Term	Definition	Form
Compound verb	Compound verb	
Compound verb	it consists of two or more words and means the same thing	6
Compound verb	verbs that consist of two or more words, have only one meaning, answer one question, and serve as only one member of the sentence	7
Homonym	the sound is the same, the meaning is different	2
Homonym	words that sound the same and have different meanings	4
Homonym	words with the same pronunciation and spelling, but different meanings	5

Taking into account these features, the structures of databases were analyzed and the structure of the database of terms of school subjects in the Kazakh language was determined. In the base structure, the term, definition, and grade in which that term occurs will be specified. In addition, among the subjects of secondary school, a list of the main subjects for which the collection of terms is compiled was determined.

Within the framework of the project, the terms of textbooks of secondary school subjects were summarized due to the lack of an appropriate database in the process of compiling terms from school subjects.

Table 3 – List of school subjects

School subject	Code
Kazakh	1

Literacy	2
Mathematics	3
Natural sciences	5
Self-knowledge	6
World studies	7
Physical training	8
Music	13
History of Kazakhstan	14
Computer science	15
World history	16
Local history	17
Russian literature	18
Algebra	19
Geometry	20
Geography	21
Biology	22
Physics	23
Chemistry	24
Geography of Kazakhstan	25
Fundamentals of law	26
Algebra and analysis initiatives	27
Initial military and technological training	28

Continuation of Table 7

School subject	Code
Graphics and design	29
Fundamentals of entrepreneurship and business	30

Since the beginning of the project, the interpretation of terms according to the plan has been completed by 50%. In particular, the base of interpretations of terms was developed in the following school subjects:

Kazakh;
 Kazakh literature;
 Literature reading;
 Information and communication technology;
 Mathematics;
 Self-knowledge;
 Computer science;
 Algebra;
 Geometry;
 Biology;
 Chemistry;
 Algebra and analysis initiatives;
 Fundamentals of law;
 Graphics and design.

In the process of developing terms, it was found that the definition of the term is given depending on the level of understanding of the school student in each grade. For example, a 1st grade student and an 11th grade student are given two different definitions of the same term. In this regard, the user interface of the electronic terminological dictionary of school subjects in the Kazakh language has been designed.

As a result, a complete database of school terms with digitized textbooks will be formed, consisting of the source code, a database of school textbook terms, a user interface in the Kazakh language with the function of correcting terms and adding new ones.

III. SPECIFIC CHARACTERISTICS AND DESIGN OF THE TERMINOLOGICAL DICTIONARY

On the main page of the terminological dictionary of school subjects under development (Figure 1), it is planned to place the “search” button, additional information about the project, and there will also be a registration window.

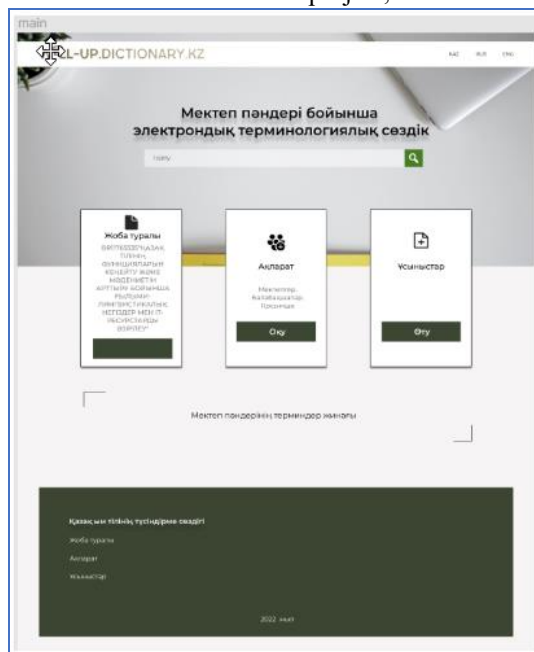


Figure 1 - Front page of the terminological dictionary of school subjects

The “project information” section contains additional information about the authors, on the basis of which programs the project was financed, as well as other information. The “Information” section displays a list of future users, schools using the terminological dictionary. In the “Recommendations” section, educational institutions and individuals can leave their proposals for improving this product.

There will be two modes of using the terminological dictionary: the main one is for grades 5-11, the second is for adapted grades 1-4. In addition, the dictionary users are divided into 3 categories:

- Guest (guests, students, parents);
- Expert (special specialists who have access to the dictionary entry to make correction in the database);

Admin (program administrators).

The “main window” in the mode for grades 5-11 (Figure) consists of the following elements:

- 1) the “search” button provides a dynamic floating list when typing a word;
- 2) information about the found term, information about the “term”, “description”, “grade”, “school subject”;
- 3) links to textbooks from which the term is derived at the bottom;
- 4) you can leave a comment on each term, experts can agree with it or reverse it.

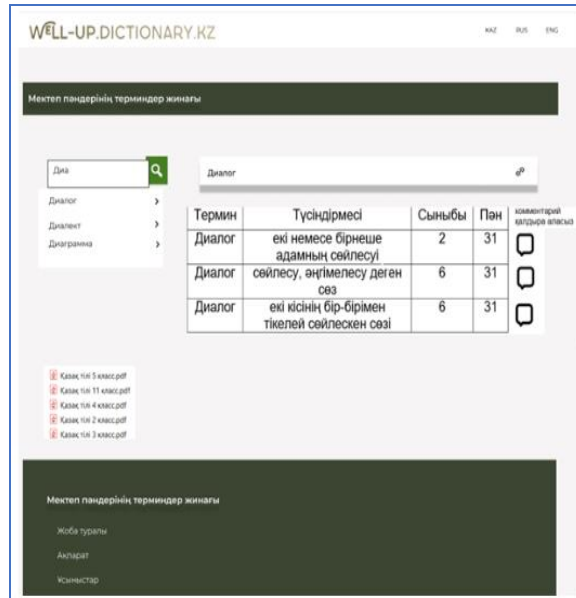


Figure 2 – “Main window” in the mode for grades 5-11

If you enter the main window as an expert (Figure), you will have the opportunity to enter a change, delete or add a new term to the terms in the terminology dictionary, as well as accept or reverse the comment left by ordinary users. As the terminology dictionary is developed and introduced into use, changes will be made and further improved.

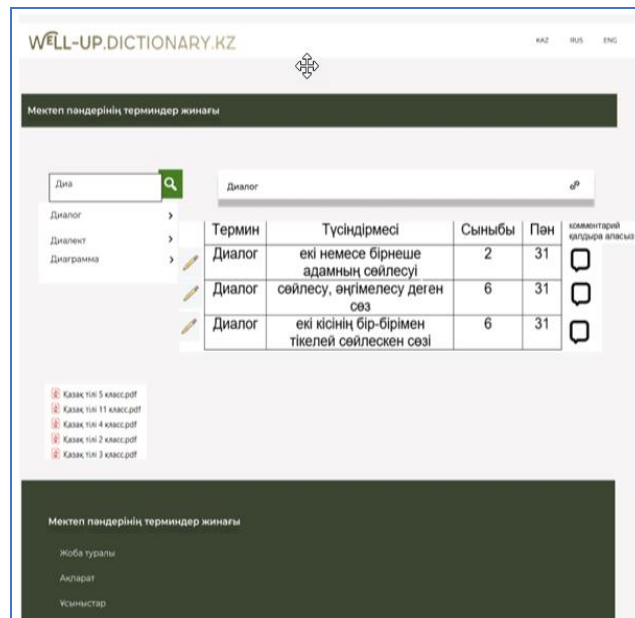


Figure 3 – The main window as an expert

When developing the terminological dictionary, the age of users is especially taken into account, an individual interface is developed for 1-4 grade students (Figure). Moreover, 1-2 grade students are considered to convey the term with sound due to the imperfection of reading skills. The interface will also be equipped with animated elements for children.

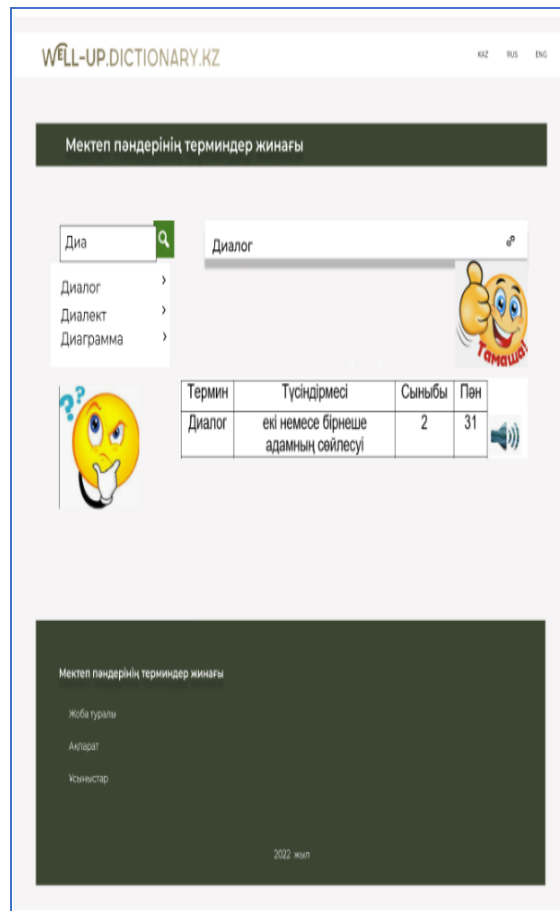


Figure 4 - Individual interface for 1-4 grade students

Brief conclusions based on the results of this stage of the research

In the course of the results obtained at this stage of the research, a database of textbooks on school subjects was compiled as the initial information for creating terminological words of school subjects in the Kazakh language, a database of terms was developed, and a technical task for a software product of the terminological dictionary was developed, namely:

A database of terms was collected from school textbooks on the Kazakh language, Kazakh literature, Literary reading, Information and communication technologies, Mathematics, Self-knowledge, Computer Science, Algebra, Geometry, Biology, Chemistry, Algebra and analysis initiatives, Fundamentals of law, Graphics and design. This is about 50% of the planned base of terms;

School terms are defined in the database, key fields are fixed, and the relationship between those fields is established;

Based on the created database of about 50%, the terms of reference of the terminological dictionary were developed and the functionality was determined;

The design of a software product has been developed, which can serve as the basis for planning a terminological dictionary for school subjects and compiling code.

Assessment of the completeness of solving the tasks set:

All the planned tasks of this stage were fully solved.

Recommendations and initial data on the actual use of the research results:

The database of about 50% of school terms created during the reporting period allows creating a software product of the terminological dictionary, that is, the collected data will serve as the basis for the full functionality of the program, and it is enough to test the software product.

Assessment of technical and economic efficiency of implementation:

The assessment of the technical and economic efficiency of the implementation was not carried out, since the implementation of the results obtained at this stage was not provided for.

CONCLUSION

The completed research work has a high scientific and technical level and will become one of the dictionaries of school subjects in the Kazakh language, which currently has no analogues, because many modern terminological dictionaries, including the provided terminology dictionary of school terms termincom.kz does not cover all of the terms, and does not provide for an opinion on any terms for the end user. For example, in the subject Information and Communication Technologies of the termincom.kz dictionary, 39 school subject terms are found, and in the dictionary created during this project, there are 479 terms considered.

REFERENCES

- [1] Baitursynuly A. Til – kural. - Almaty: Sardar, 2009. - 348 bet. (In Kaz.)
- [2] Baitursynul A. Til tagylymy. - Almaty: Ana tili, 1992. - 448 b. (In Kaz.)
- [3] Gruber T. R. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2):199-220. – 1993. (In Eng.)
- [4] Protégé. – URL: <https://protege.stanford.edu/>. (karalghan kuni: 10.008.2023). (In Eng.)

AUTHORS PROFILE



Gulmira Bekmanova received the B.S. degree in information systems from D. Serikbayev East Kazakhstan State Technical University, Ust-Kamenogorsk, Kazakhstan in 2004 and the M.S. degree in information systems from the same University in 2006, candidate of technical sciences on specialty: 05.13.11 – Mathematical software for computers, complexes and computer network and the Ph.D. degree in Computer science, computer engineering and control from L.N. Gumilyov Eurasian National University, Kazakhstan in 2010.

From 2004 to 2006, she was Assistant of Professor with the D. Serikbayev East Kazakhstan State Technical University, Kazakhstan. Since 2006 to 2019 she was senior lecturer, associate professor, Researcher of Artificial Intelligence Research Institute, head of the Computer Science and Information Security Department, L.N. Gumilyov Eurasian National University, Kazakhstan. Since 2019 she was the director of the Distance learning center, L.N. Gumilyov Eurasian National University, Kazakhstan. Since 2020 she is the director of the Digital Development and Distance learning Department, L.N. Gumilyov Eurasian National University, Kazakhstan. She is the author of 5 monographs, 4 manuals and more than 100 articles, and 2 inventions.



Altynbek Sharipbay received the B.S. degree in Applied Mathematics from the from S. M. Kirov Kazakh State University, Kazakhstan in 1974, the candidate of science degree in Mathematical software for computers, complexes and computer network from Lomonosov Moscow State University in 1981 and doctor of technical sciences since in Mathematical software for computers, complexes and computer network in 1997.

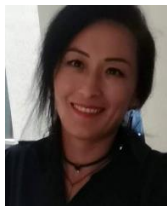
From 1974 to 1975 he was mathematician-programmer with S. M. Kirov Kazakh State University. From 1975 to 1981 he had positions of intern researcher and graduate student of Lomonosov Moscow State University, USSR. From 1981 to 2000 he was Senior Lecturer, Associate Professor, Head of Department with S. M. Kirov Kazakh State University, head of the Laboratory of the Institute of Informatics and Control Problems of the National Academy of Sciences of Kazakhstan, head of the Department at the Kazakh State Academy of Management, director of the Center for Information Technologies. Since 2000 he has been head of the department, professor of the department Technologies of artificial intelligence, L.N. Gumilyov Eurasian National University, Kazakhstan. He is the author of 27 books, more than 400 articles, and more than 30 inventions. His research interests include programming, artificial intelligence, information security and information protection, e-learning.

In 2001 he became laureate of the State Award in the field of science, technology and education and in 2004 he received the State Award of the Republic of Kazakhstan.



Dinara Kabdylova received the B.S. degree in Computer Engineering and Software and M.S. degree in Computer Science from L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, in 2009 and 2014.

Since 2022 Manager of the Information Systems Analytics and Testing Sector, from 2011 to 2022, she has been working as a eLearning Knowledge Engineer of Sector e-learning support and document flow Department of Digital Development and Distance Learning, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.. She participated in scientific projects “Development of scientific and linguistic foundations and IT resources to expand the functions and improve the culture of the Kazakh language” (2022-2023). She is the author of one collective monograph and 3 articles.



Amangeldy Nurzada, a postdoctoral fellow at L.N. Gumilyov Eurasian National University, specializes in object recognition, computer vision, and artificial intelligence. Her leadership in spearheading several key projects, including the creation of the Kazakh sign language recognition system, underscores her innovative approach. Her involvement in the study of detecting personal protective equipment using the YOLOv8 architecture points to her commitment to elevating safety standards through cutting-edge technologies. Set against a backdrop of global trends, Nurzada's work addresses the pressing need for automated safety solutions, making her contributions timely and relevant. Her research endeavors not only foster technical advancements in the field but also address pressing contemporary challenges.



Ayaulym Sairanbekova received the B.S. degree in Information System and M.S. degree in Administration of management and protection of computer systems and networks in enterprises from L.N. Gumilyov Eurasian National University, Astana, Kazakhstan, in 2016 and 2022, respectively, where currently is an active Ph.D. student in Computer Science.

Scientific interests: artificial intelligence, computational linguistics, processing of natural language texts, processing of big data, data mining.

She participated in scientific projects: "Development of scientific and linguistic foundations and IT resources to expand the functions and improve the culture of the Kazakh language" (2022-2023), AP19679847 "Development of methods for the analysis of the Kazakh political discourse" (from 2023). She is the author of 1 collective monograph and 10 articles.



Assel Omarbekova graduated with honors from the S. Toraihyrov Pavlodar State University, Kazakhstan in 2000 and postgraduate from L.N. Gumilyov Eurasian National University, Kazakhstan by specialty Mathematical and software of computers, complexes and computer networks in 2000. In 2008 she was received the degree of Candidate of Science.

Since 2020 she is the deputy director of the Digital Development and Distance learning Department, L.N. Gumilyov Eurasian National University, Kazakhstan

Her research interests related with intelligent tutoring systems, and distance learning technologies. She has more than 250 scientific and scientific-methodical works, of which 24 publications are indexed by Scopus (of which 3 are in peer-reviewed scientific publications with a CiteScore percentile in the Scopus database of at least 35) and 14 - Web of Science, 28 - in publications recommended by CQASES, 10 textbooks, 23 study guides.

She is the author of the following software: technology of automatic create electronic textbooks, system electronic questionnaire of teachers and etc. Also she was involved in projects related to the creation of multimedia learning systems for the Ministry of Education and Science, Ministry of Culture and Information, the Ministry of Finance of Kazakhstan and etc.



Altanbek Zulkhazhav pursued his master's degree from 2014 to 2016 at L.N. Gumilyov Eurasian National University in Astana, Kazakhstan. Currently a Ph.D. candidate.

Since 2014, he has been actively engaged in research within the realms of computer linguistics, nlp, knowledge bases, data processing, and the semantic web. H-index (Scopus):2.

He was involved in the following research projects: "Development of electronic thesauri of Turkic languages to create systems for multilingual search and extraction of knowledge" (2018-2019), "Development of scientific and linguistic foundations and IT resources to expand the functions and improve the culture of the Kazakh language" (2022-2023), AP19679847 "Development of methods for the analysis of the Kazakh political discourse" (from 2023)