REPUBLIC OF AZERBAIJAN

In the right of manuscript

ABSTRACT

of the dissertation for the degree of Doctor of Philosophy

DEVELOPMENT OF METHODS AND ALGORITHMS FOR SYNTHESIS OF NATIONAL TERMINOLOGY INFORMATION SYSTEM

Speciality: 3338.01 – "System analysis, control and information processing" (information technology)

Field of science: Technical sciences

Applicant: **Afruz Mukhtar gizi Gurbanova**

Baku – 2024

The work was performed at the Institute of Information Technology of the Ministry of Science and Education of the Republic of Azerbaijan.

Scientific supervisor: Full member of ANAS, Doctor in Technical Sciences, Prof. **Rasim Mahammad oglu Alguliyev**

Official opponents: Doctor of Technical Sciences, Prof. **Nadir Bafadin oglu Agayev**

Doctor of Technical Sciences, Prof. **Ramin Rza oglu Rzayev**

Doctor of Philosophy in Technical sciences **Elviz Anvar oglu İsmayilov**

Dissertation Council ED 1.35 of Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at the Institute of Information Technology of the Ministry of Science and Education.

Chairman of the Dissertation Council:

Full member of ANAS, Doctor of Technical Sciences, Prof.
______________ **Rasim Mahammad oglu Alguliyev**

Scientific secretary of Dissertation Council:

Doctor of Philosophy in Technical Sciences, Assoc. Prof.
______________ **Fargana Jabbar gizi Abdullayeva**

Chairman of the scientific seminar:

Doctor of Technical Sciences, Assoc. Prof.
______________ **Zarifa Gasim gizi Jabrayilova**

GENERAL CHARACTERISTICS OF THE WORK

Relevance of the topic. Currently, information and communication technologies (ICT) are becoming a way of life for people and are penetrating all areas of activity. On the other hand, the modern era is characterized by a high pace of globalization. The acceleration of the globalization process and intercultural communication also causes certain changes in natural languages. In such a period, it is important to take necessary measures to protect, develop and expand the possibilities of application of languages. Determining the place and role of national languages in the e-government platform in the direction of protecting and developing national languages, which are facing various threats in the conditions of globalization, is one of the urgent issues. The systematization and development of national terminology is of particular importance in the development of the social function of language. As a result of the rapid penetration of the Internet into various spheres of society in our country and the free use of information related to various fields of knowledge by people, it has become a very difficult issue to control the national terminological environment. In recent years, the enrichment of the language with borrowed words is also a natural process. It is impossible to put a barrier to this process, and at the same time, it is impossible to allow the language to be unnecessarily burdened with borrowed words. The reason for the emergence of the real situation is the lack of a single database or terminological register that is accessible to everyone for the use of terminological dictionaries. As a result of Azerbaijanis living in different countries being under the influence of relevant foreign languages, the expansion of education of young people in Azerbaijan in English and Russian, and the Internet becoming a part of people's daily activities, different terminological environments have been formed. There is a need to bring terminological bases into a single, common environment, and to have a national terminological information system accessible to everyone as an integral part of the Azerbaijani language ecosystem [2].

Compared to previous periods, the rapid development of information, including scientific and innovative knowledge, as a result of the widespread application of ICT, on the one hand, makes the terminological process more dynamic, and on the other hand, the very easy access, dissemination and popularization of information regardless of space and time accelerates the de-terminological process.

The above makes research in the field of terminological informatics (computational terminology) necessary. Terminological informatics covers important aspects in the fields of natural language processing - intellectual analysis of text, information search, information extraction, text analysis, ontology construction, and so on. For all these reasons, the creation of an open, accessible National Terminological Information System (NTIS) as an integral part of the Azerbaijani language ecosystem in the e-government environment that is being formed in our country has become a requirement of the time. Taking these into account in the dissertation work, the scientific-theoretical and technological problems of the formation of the NTIS were investigated, and new approaches, methods and algorithms were proposed for the synthesis and intellectualization of the system.

The aim of the work is to develop a model, method and algorithms for the synthesis of the NTIS as an integral part of the Azerbaijani language ecosystem on the e-government platform based on international experience and standards. For this purpose, the following issues were set in the dissertation work:

- Analysis of the current situation in the field of terminological informatics and identification of scientific-theoretical and technological problems;
- Comparative analysis of existing approaches in the field of formation of the NTIS and identification of scientific-technological problems;
- Development of a conceptual model and architectural-technological principles of the NTIS;
- Development of a decision support system (DSS) based on OLAP in the NTIS;
- Development of a model, method and algorithm for modeling, analysis and evaluation of dynamic processes within the framework of the NTIS;
- Development of a method for assessing the level of popularity

of individuals' speech based on the NTIS;

- Development of proposals for the realization of NTIS.

Object and subject of the research. The object of the research is the synthesis of NTIS. The subject of the research is the models and methods of managing the processes of synthesis of NTIS.

Research methods. Natural language processing, OLAP, data mining, text mining, decision-making theory, population model, and statistical analysis methods were used to solve the problems considered in the dissertation.

The main provisions of the defense. The following are the important results of the dissertation work that are scientifically innovative:

- Conceptual and architectural-technological model of the NTIS;
- A system supporting decision-making based on OLAP and DW technologies in the NTIS;
- Polycubic OLAP model for assessing the NTIS;
- A method based on the population growth model for modeling dynamic processes within the national terminological information system;
- A method based on the k-means clustering algorithm and the Naive Bayes classification model for intellectualizing the national terminological information system;
- A method based on the statistical frequency model for assessing the popularity level of individuals' speeches based on the NTIS.

Scientific novelty of the research. The scientific novelty of the research is determined by the following results:

- The conceptual and architectural-technological model of the NTIS has been developed;
- A system supporting decision-making based on OLAP technologies and a polycubic OLAP-model have been developed for the operational analysis of data and metadata collected in the NTIS;
- A method based on the Malthusian population growth model for modeling dynamic processes within the NTIS has been developed;
- A method based on the k-means clustering algorithm and

Naive Bayes classification has been developed within the framework of NTIS for the automatic distribution of new terms across appropriate disciplines;

- A method based on the statistical frequency model for assessing the popularity level of individuals' speeches based on the NTIS.

Practical significance of the work. The approaches proposed in the dissertation can be used in the intellectualization of the NTIS, evaluation and monitoring of terminological processes within the system, language security, protection from external influences and adaptation of new terms to national characteristics, increasing the efficiency of term creation, protecting the copyrights of scientists and specialists who have developed terminological dictionaries, and intellectualization of e-terminological services.

Approbation of the work. The main scientific-theoretical and practical results were reported and discussed at the following conferences: The Third International Conference Problems of Cybernetics and Informatics (PCI) (Baku, 2010); "TurkLang" VI International Conference on Computer Processing of Turkic Languages, (Tashkent, 2018); International conference "Development of informatization and the state system of scientific and technical information", RINTI (Minsk, 2019); The 4th International Symposium on Computer Science, Digital Economy and Intelligent Systems (CSDEIS), (Wuhan, 2022); IEEE 17th International Conference on Application of Information and Communication Technologies (AICT), (Baku, 2023); 2nd International Conference on Information Technologies and Their Application (ITTA), (Baku, 2024).

Scientific publications. 20 scientific works have been published on the topic of the dissertation. Of these, 13 articles were published in peer-reviewed journals, 7 theses in conference proceedings. Of these scientific works, 1 article was published in the “Web of Science” and 3 articles in the “Scopus” indexed journal.

The structure and volume of the work. The dissertation consists of an introduction, 4 chapters, a conclusion, a list of 170 references and two appendices, 7 tables and 25 figures. The total volume of the work is 121 pages of typewritten text.

BRIEF OVERVIEW OF THE WORK

In the introduction substantiates the relevance of the dissertation, the purpose of the study and the issues to be resolved. The scientific novelty and practical significance of the results obtained are shown.

In the first chapter ("Research and analysis of scientific-theoretical problems of terminological informatics"), the preservation and development of the language, the role of terminology in the modern era and the history of its formation, foreign experience and international standards in this field are analyzed, and the main scientific-technological problems of terminological informatics are analyzed. In order to implement initiatives related to language preservation in the e-government environment, a conceptual model of the Azerbaijani language ecosystem is proposed on the e-government platform (Figure 1). The main issues that need to be resolved in connection with the development and synthesis of the National Terminological Information System as an integral part of the Azerbaijani language ecosystem are identified [1, 2, 4, 5, 6, 12, 16, 18].

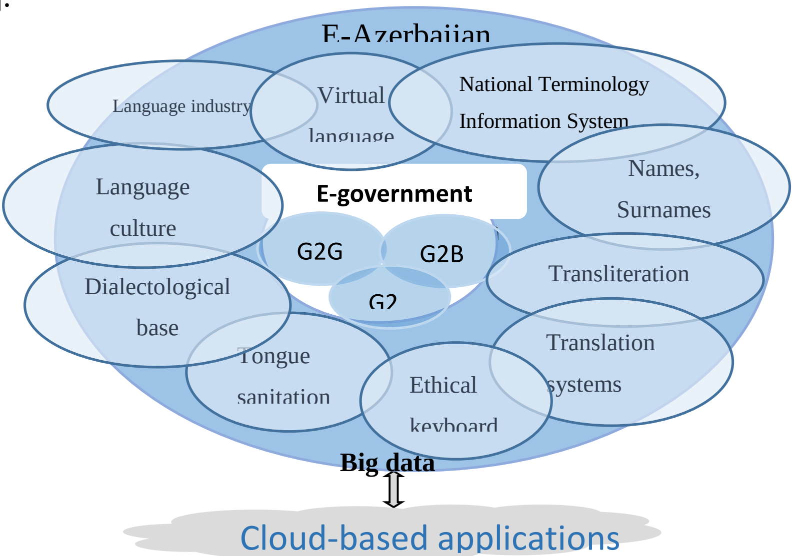

Figure 1. Conceptual model of the Azerbaijani language ecosystem on the e-government platform

In the second chapter ("Development of the conceptual foundations of the National Terminological Information System"), the conceptual model and architectural-technological principles of the NTIS were developed, a DSS based on OLAP and Data Warehouse (DW) technologies was proposed for the evaluation of the NTIS, and the data were analyzed.

In the first section of the second chapter, a conceptual model of the National Terminological Information System was proposed [3, 7, 10]. The functional subsystems of the proposed conceptual model are explained in detail below.

Proposed model. The core of the NTIS is the Terminology Register, which ensures the reliable storage of all terminological dictionaries in a single online environment and is accessible to everyone (Figure 2). Terminological dictionaries published in different years, as well as those available on carriers, are collected, digitized and included in the terminological register. E-terminological services are provided to users via the Internet based on the information collected in the terminological register through the ***terminology web portal***. With the synthesis of the NTIS, the process of term creation will become much easier compared to the previous period.Thus, when creating a new term, each scientist and specialist applies to the Terminology Register of the NTIS to see if this word has been previously registered as a term. If so, then there is no need to adopt a new term (domestic harmonization). If this term is not in the terminological register, then this term is searched for, analyzed and a more efficient decision is made in the terminological information systems of other countries integrated with the NTIS (international harmonization). Within the framework of the terminology expertise subsystem, public discussion and decision-making of dictionaries and other proposals submitted by terminographers is provided both in traditional and online environments. Through the terminographers subsystem, expertise of dictionaries prepared by scientists and specialists (professional terminographers), as well as citizens who want to engage in term creation work (amateur terminographers) is carried out.

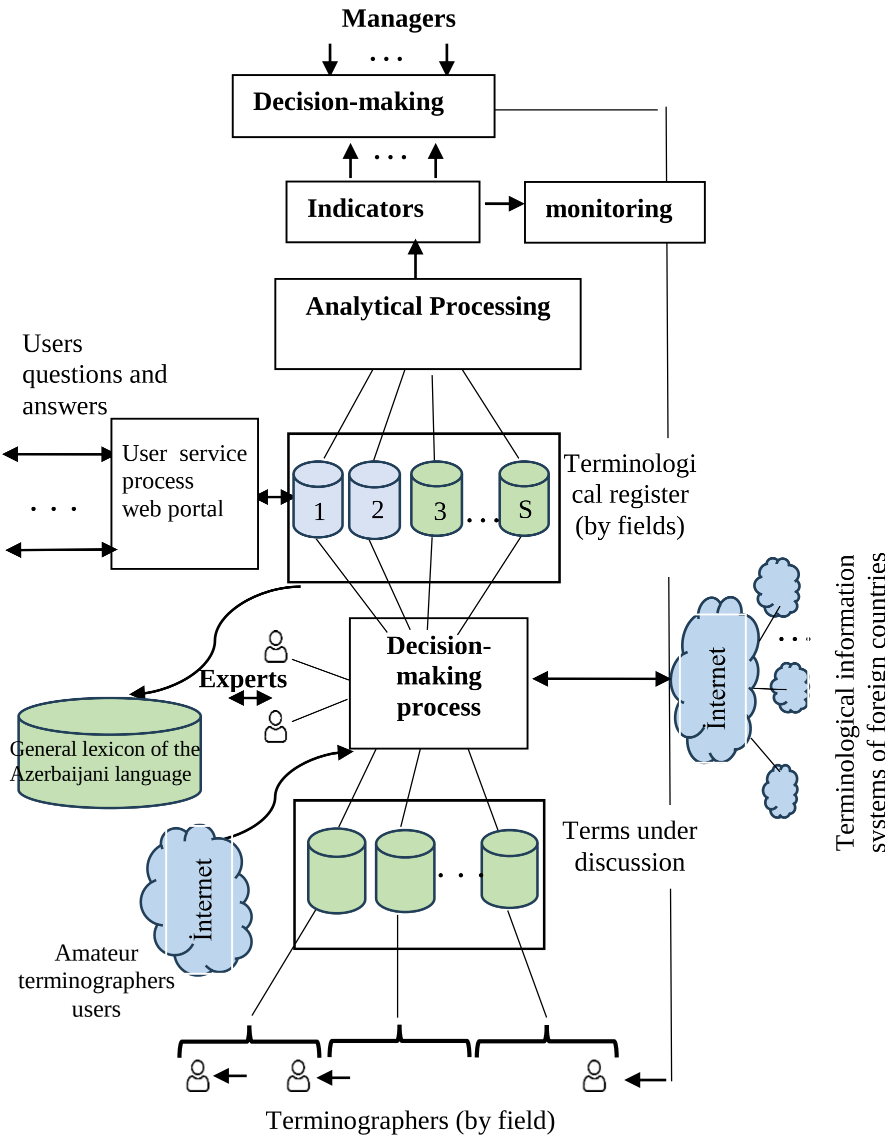

Figure 2. Architecture of the national terminology information system

The analytical processing subsystem performs the function of conducting various monitoring, analysis, and evaluations based on the data collected in the terminological register and web portal.

Thus, valuable knowledge can be obtained by analyzing the dictionaries collected in the Terminology Register of the National Terminology Information System, parameters characterizing terms, usage ratings, etc. indicators.

Thus, the synthesis of the NTIS opens up wide opportunities for conducting research in terminological informatics and other related fields, intellectualizing the system and improving efficiency indicators, expanding its functional capabilities, and applying innovative technologies. The NTIS is a dynamically developing, open national information resource exploited in the Internet environment. When creating the National Terminology Information System, the requirements of international standards were taken into account and e-terminology services will be provided to users. The collection of terminological dictionaries in a single information system will increase the effectiveness of terminological research. The main importance of the NTIS is that thanks to it, the Azerbaijani terminology base will be formed.

In the second paragraph of the second chapter, the structure of the DSS based on OLAP and DW technologies in the NTIS was developed [13, 14]. The main goal is to obtain valuable information by analyzing the terms collected in the Terminological Register of the NTIS and the metadata defining them. The capabilities of the DSS are described in detail below.

Decision Support System. Such systems provide operational processing of information, provision of ready-made reports, and have such an important quality as multifunctionality.

The main tasks of the DSS are data collection, storage, processing and analysis. To solve these problems, DW, a special structure for information storage, was developed.

Data storage. DW consists of a base, history, integrated data and metadata. In DW, homogeneous data from various sources are integrated and become problem-oriented, structured. The structural elements of DW are facts (Fact Table) and dimension tables

(Dimension Tables). In DW, the appearance of a new term will be a fact, and the dimension will be the date, the source of the term, the sphere of activity, the language of the term, etc. The presented DW has the hierarchical levels “Date” - “year - month - day”, and the hierarchical levels “Language” - “language family - language group”. The correct choice of VS architecture is the main factor in the successful operation of the system.

The third paragraph of the second chapter presents a polycubic OLAP-model used for analytical processing of data collected in the Terminological Register of the NTIS [14].

Polycubic OLAP-model. OLAP is a technology that allows you to effectively analyze data from various sources, providing fast access to information for making informed business decisions. The goal is to collect, organize, summarize and analyze data. OLAP allows you to analyze data in several dimensions, providing the end user with the information necessary for more effective decision-making. Typical OLAP operations consist of slices - slice & dice, aggregation (roll-up) and detailing (drill-down). Aggregation and detailing are performed on dimensions with a hierarchical structure.

A DSS architectural-technological model has been proposed for the national terminology information system (Figure 3). It is based on the three-level model of the DW.

Level 1. The first level of the DSS presents various sources of data contained in the Terminological Register of the NTIS. These can be terms obtained as a result of automatic processing of texts, received as a result of discussions held in forums, adopted by scientists and specialists in a particular field, etc. During the transition to the second level, at the intermediate stage, the process of data cleaning is carried out using the technology of *ETL (Extraction, Transformation, Loading)* - data extraction, transformation, loading.

Level 2. Presents the DW, which contains both current and historical data. Historical data related to terms are de-terminated terms, as well as terms that have passed into the category of archaisms. Data kiosks (DK) are separated from the DW. Each DK stores data aimed at solving related issues. In the present case, the DK is aimed at compiling dictionaries - explanatory, translation (az-rus, az-eng., etc.), synonyms, thesauruses, etc. For this, the DW and the DK built on its basis must have all the necessary data.

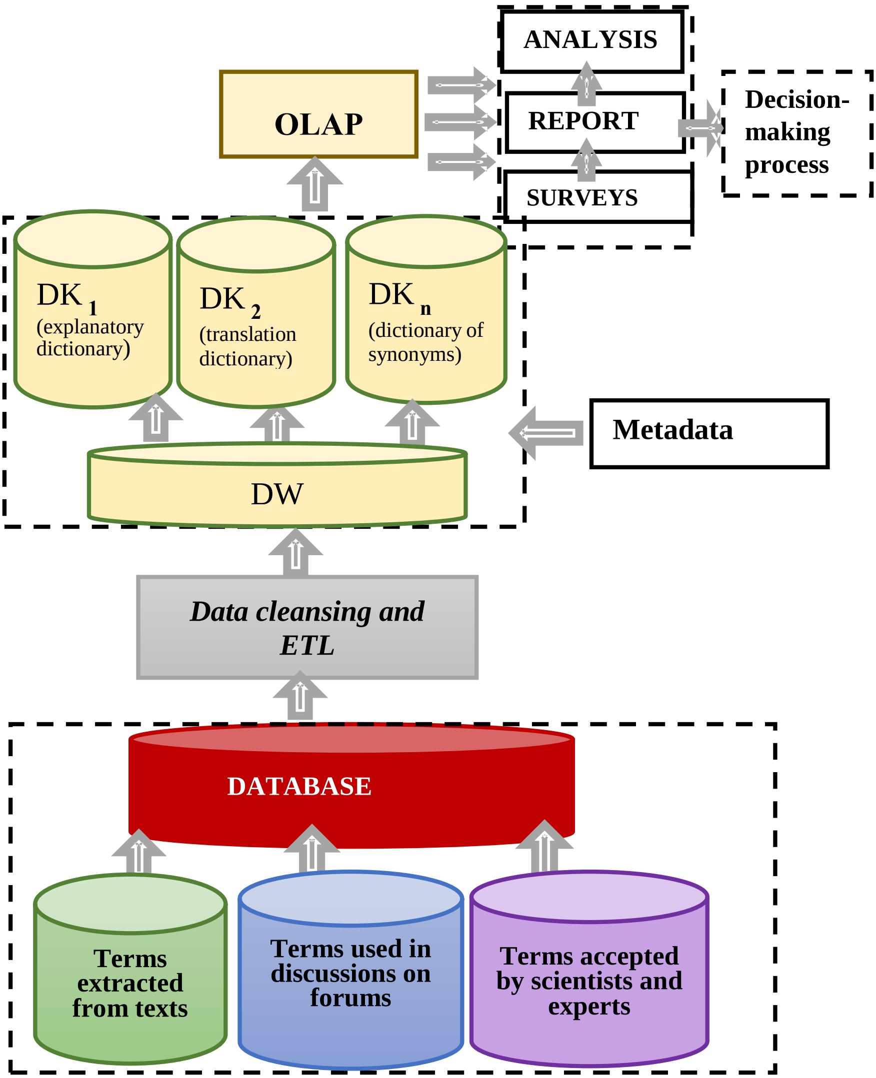

Figure 3. Architectural-technological model of the DSS for the national terminology information system

level 3. The main component of the DW in DSS is OLAP technology. Data is analyzed based on the OLAP cube.

An OLAP cube can have any number of dimensions. Here, three dimensions were selected for analysis, which are attributes of the DW: date (Date), field of activity (Field of activity), language (Language). In the presented example of an OLAP cube, its cells contain numbers reflecting the number of terms entered in the corresponding date, related to the corresponding field of activity, and created in the corresponding language. Since two of these dimensions - language and date - are hierarchical, the results of the SUM() aggregate function performed on them are placed. The OLAP cube helps to detail these values. As a result of these operations, data consisting of the corresponding indicators presented in the form of a two-dimensional table were obtained.

The proposed DSS, developed on the basis of OLAP and DW technologies, will make NTIS more advanced and sustainable. It will provide significant support to DSS, allow improving analytical activities in the field of terminology, and will play an important role in conducting scientific research in this area.

In the third chapter (Methods and algorithms for intellectual analysis of the national terminological information environment), methods and algorithms for automating term creation processes, modeling and intellectual analysis of dynamic processes, and assessing the level of popularity of speech within the framework of NTIS are developed.

The first paragraph of the third chapter considers the issue of automating the processes of term creation. Term creation is the process of creating, systematizing, standardizing and using terms in various fields of knowledge and practice.

Terminological processes are a set of changes occurring in the terminological system of a language. The main terminological processes are terminization, de-terminization, semantic changes of terms, and homonymization of terms. An indicator for the analysis and assessment of terminological processes within the framework of NTIS has been proposed. The proposed approach is explained in detail below:

An indicator for evaluating terminological processes. Let us provide each term (accepted for discussion and approved) in the

Terminological Register (TR) of NTIS with metadata accompanying it. These metadata include the name and address of the organization and authors (terminographers) that submitted the term, the time the term was submitted for discussion, the time the term was approved, an indication of which scientific field the term belongs to, information about the language from which the term was derived, the name (spelling) of the term in that language, the etymology of the term, etc. Formally, let us denote the metadata related to the *i*-th term from the *s*-th field of knowledge by the following vector:

$$T_i^{(s)} = \left\langle t_{ij}^{(s)} \middle| j \equiv \overline{1, m_s} \right\rangle, s = \overline{1, S}, i = \overline{1, n_s} \qquad (1)$$

Here $t_{ij}^{(s)}$ is the jth attribute (metadata) of the i-th term belonging to the j-th knowledge domain. TR consists of the terms (t_i) and their accompanying metadata bases:

$$TR = \bigcup_{s=1}^{S} \bigcup_{i=1}^{n_s} (t_i, T_i^{(s)}) = \bigcup_{s=1}^{S} TR_s \qquad (2)$$

$$TR_s = \bigcup_{i=1}^{n_s} (t_i^{(s)}, T_i^{(s)}), s = \overline{1, S} \qquad (3)$$

Here, the variable n_s is the number of terms belonging to the *s*-th knowledge area, and m_s is the number of metadata characterizing these terms.

For the purpose of analyzing, evaluating, monitoring terminological processes, making comparisons between different fields of knowledge, and other purposes, let us introduce the $n_s(t)$ indicator. Here, $n_s(t)$ is the number of terms in the Terminological Register for the *t*-th time, *s*-th field of knowledge. Naturally, the total number of terms in the Terminological Register for the *t*-th time, $N_s(t)$, will be as follows:

$$N_s(t) = \sum_{s=1}^{S} n_s(t) \qquad (4)$$

To characterize the dynamics of the development of the term creation process in the *s*-th field, let us introduce the indicator Δn_S (t):

$$\Delta n_s(t) = n_s(t) - n_s(t-1),\ t=1,2,...,n \tag{5}$$

This indicator can be used to determine the progress (decline) of the term creation process in knowledge areas. Let's look at the dynamics of the intersection of two areas ($U_s(t)$ and $U_{s'}(t), s \neq s'$):

$$\left|U_s(t) \cap U_{s'}(t)\right| = \Delta U_{ss'}(t) \tag{6}$$

If $\Delta U_{ss'}(t) > \Delta U_{ss'}(t-1)$ conditions are met, the integration of 2 fields has occurred; if $\Delta U_{ss'}(t) < \Delta U_{ss'}(t-1)$ conditions are met, the differentiation of 2 fields has occurred. This can also be done between 3 fields, 4 fields, and so on. This is one of the main functions performed by the NTIS.

In the first paragraph of the third chapter, the technologies for automatic construction of the semantic network of terms of the subject area and automatic extraction of terms from texts are also analyzed, and an applied approach is given [8, 9, 11, 15].

Thus, as a result of collecting terminological dictionaries prepared in various fields in Azerbaijan in a single information system, the efficiency of terminological creation will increase. Within the framework of the concept of NTIS, a National Terminological Web Portal has been created, which is an important step towards automating terminological creation in Azerbaijan.

In the second paragraph of the third chapter, the processes of termination and de-termination are analyzed, dynamic processes within the framework of NTIS are formulated on the basis of mathematical regularities, and a method based on a population model is proposed for the general analysis of the system [17]. Each stage of the proposed approach is explained in detail below:

De-termination threshold. The acquisition of new meanings of old terms – de-termination and the emergence of new terms are the result of the dynamics of terminology. The de-termination threshold of terms depends on the number of people using that term, that is, it is determined by the frequency of use of this term in the general lexicon of that people. In other words, if the number of uses of the term T_i from the term base T on the Internet and other resources is f_i, and the number of people under consideration is P, then

$$\frac{f_i}{P} \le d, i \in T, d \in [0,1] \tag{7}$$

those that correspond to the condition are called terminology, and those that do not correspond are called determinologized.
This indicator depends on the literacy and education of that people. By analyzing this limit, one can obtain certain knowledge about the popularization of knowledge in society for different periods, as well as the literacy of the people.

Populyasiya modeli. Terms are created by people who have lived a certain life, so an analogy can be drawn between the size of the human population and the process of forming new terms. Similar to the population dynamics model, for the dynamics of the number of terms

$$T'(t) = kT(t) \tag{8}$$

where k is the growth rate. $T'(t)$ is the number of terms included in DB at time t, $T(t)$ is the number of terms in DB up to time t. As can be seen, the growth rate of terms at time t is proportional to the number of terms in that period. Let us define the number of terms at the initial time t_0 as T_0:

$$T(t_0) = T_0 \tag{9}$$

The general solution of (8) and (9) is an exponential function:

$$T(t) = T_0 e^{kt} \tag{10}$$

e - forms the base of the natural logarithm. If the specific growth rate $k>0$, then the number of terms will increase infinitely with increasing time. If $k<0$, the number of terms will decrease towards zero with increasing time. When $k=0$, the number of terms will remain unchanged over time.

By comparing the average number of terms in the terminology base with the average number of words in the general lexicon, it is possible to analyze the processes taking place in the language and make decisions. For this, a special indicator has been introduced and analyzed.

The average number of terms in the TB. Let the TB be the filling rate $T'(t) = \eta(t)$. The average number of terms in the TB up to the moment *t* is determined as follows:

$$N_T(t) = \frac{1}{t - t_0} \int_0^t \eta(t) dt \tag{11}$$

Average number of words in the lexicon. Let the rate of change of the number of words in the lexicon be $l(t) = L'(t)$, and the average number of words in the general lexicon of the language up to time *t* is defined as follows:

$$N_L(t) = \frac{1}{t - t_0} \int_0^T l(t) dt \tag{12}$$

Obviously, $N_T(t) << N_L(t)$

The processes occurring in language can be analyzed and evaluated using the relationship between formulas *(11)* and *(12)*. For this purpose, the *n(t)* indicator has been introduced.

$$n(t) = \frac{N_T(t)}{N_L(t)} \tag{13}$$

$N_L(t)$ is a monotonically increasing function. Thus, the increase or decrease of the function *n(t)* can occur due to the introduction of new terms into the language and the regular entry of some terms into the general lexicon by exceeding the limit value of their frequency of use. By analyzing this, one can gain knowledge about whether new terms are widely used in education, enlightenment, popularization, whether they are de-terminated through social media or not, on the contrary, they are not de-terminated, etc.

In the second paragraph of the third chapter, a general model of dynamic processes within the framework of NTIS (Figure 4) and methods based on machine learning for the analysis and evaluation of these processes are proposed. K-means and Naive Bayes methods are proposed to automate the distribution of new terms across relevant disciplines [20].

Proposed method: Suppose that a database has been created for the environment under study on the topic under consideration. Each stage of the proposed approach is explained in detail below:

Step 1. First, texts in various fields are collected from the NTIS web browser. These texts are viewed as documents: $D = \{d_1, d_2, ..., d_n\}$ where n is the number of documents.

Step 2. The collected documents are pre-processed. Terms are extracted from the documents and semantic analysis is performed.

Step 3. Terms are extracted from the texts, $T=\{t_1, t_2, ..., t_m\}$**.** Then, the set of documents is described as a vector using the "Term Frequency-Inverse Document (TF-IDF)" scheme. Euclidean distance is used to calculate the distance between documents, $d_i = [w_{i1}, w_{i2}, ..., w_{im}], i = 1, ..., n$, where w_{ij} is the weight of the j-th term in the i-th document and is calculated as follows:

$$w_{ij} = tf_{ij} \log \frac{n}{n_j}; i = 1, ..., n; \qquad (14)$$

where tf_{ij} is the frequency of the i-th word in the j-th document, n_j is the number of documents in which the word t_j occurs, n is the total number of documents, m is the number of terms.

Step 4. Using the Cosine metric, the proximity between two vectors is calculated as follows:

$$sim(\mathrm{d}_i, \mathrm{d}_l) = \frac{\sum_{j=1}^{m} w_{ij} w_{lj}}{\sqrt{\sum_{j=1}^{m} w_{ij}^2 . \sum_{j=1}^{m} w_{lj}^2}}, \quad i, l = 1,2 ..., n \qquad (15)$$

where $\mathrm{d}_i = [w_{i1}, w_{i2}, ... w_{im}]$ and $\mathrm{d}_l = [w_{l1}, w_{l2}, ... w_{lm}]$ are vectors corresponding to documents d_i and d_l, w_{ij} and w_{lj} are the k-th elements of the frequencies of occurrence of the vectors being compared.

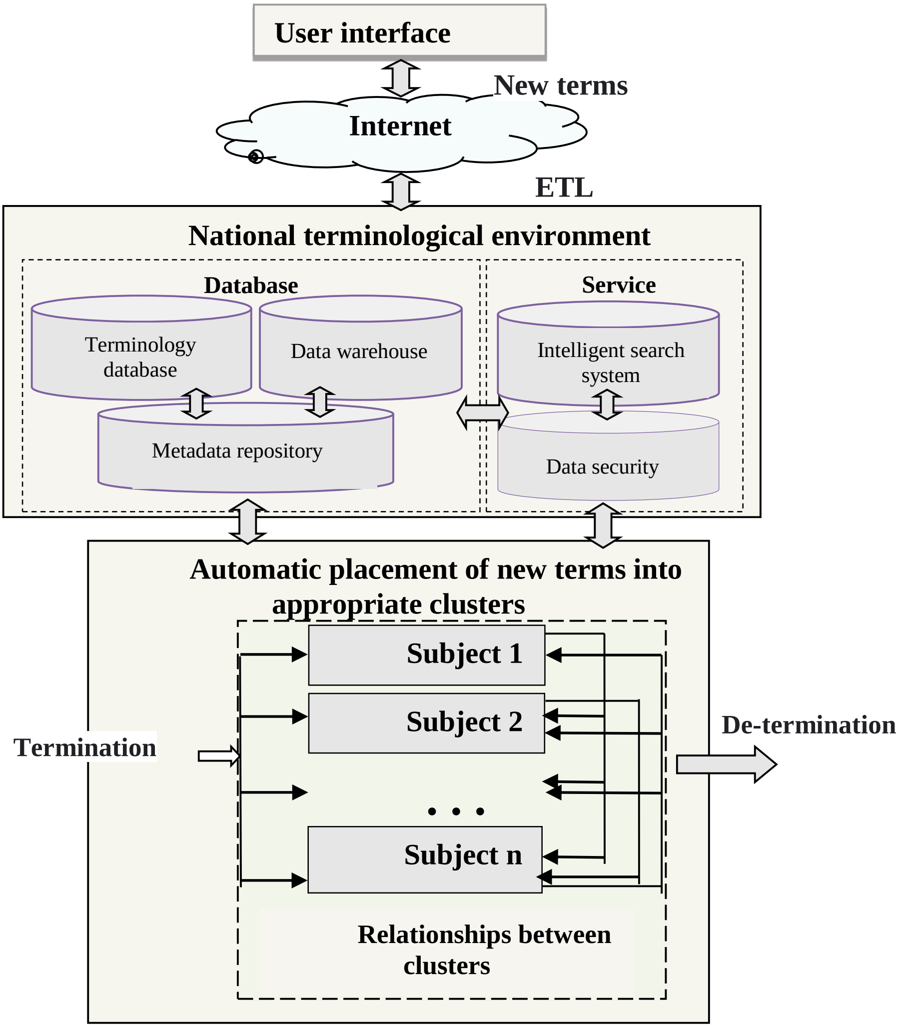

Figure 4. General model of dynamic processes in the national terminology system

Here, the vectors are ordered by the words used, their lengths overlap and are equal to the maximum length of one of the two vectors.

Step 5. After the documents are represented in vector form, they are clustered. It is proposed to use the k-means method to cluster the documents. k-means is considered one of the popular algorithms in big data analysis due to its low execution time and ease of application.

The documents included in the database must be grouped into $C=\{C_1,C_2,\ldots,C_p,\ldots,C_k\}$ clusters. To solve the clustering problem, some notes need to be clarified:

- $x_{ip}=\begin{cases}1, \text{ if } d_i \in C_p\\ 0, \text{ other wise}\end{cases}$

$R=[R^1,R^2,\ldots,R^g]$ – is the central repository of documents.
where the R -th coordinate of l is defined as follows:

$$R^l=\frac{1}{n}\sum_{i=1}^{n}w_{ij};l=1,2,\ldots,g \qquad (16)$$

- $R_p=[R_p^1,R_p^2,\ldots,R_p^g]$ – C_p is the center of the cluster.

The *l*-th coordinate of R_p is calculated as follows:

$$R_p^l=\frac{1}{\left|C_p\right|}\sum_{i=1}^{n}w_{il}x_{ip};l=1,\ldots,m \qquad (17)$$

where $\left|C_p\right|=\sum_{i=1}^{n}x_{ip}$, $p=1,\ldots,k$, C_p – is the number of documents in the cluster. During clustering, it is necessary to minimize the degree of proximity between the centers of the clusters and the centers of all document sets.

$$\max f(x)=\sum_{p=1}^{k}\frac{1}{\left|C_p\right|}\sum_{i=1}^{n}sim(\mathrm{d}_i,R_p)\,x_{ip} \qquad (18)$$

The "purity" coefficient was used to evaluate the results of clustering:

$$Purity=\frac{1}{N}\sum_{k=1}^{K}\max_j\left|C_p\cap T_j\right| \qquad (19)$$

where, *N* is the total number of terms, *K* is the total number of clusters, C_p is the set of terms belonging to the p-th cluster, and T_j is the set of terms belonging to the *j* -th real class.

For the experiment, we clustered terms with different Purity Coefficient values (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7) using the k-means clustering method. We also increase the number of resources (150, 300, 450, 600, 750 terms) to estimate the total volume of data in the result.

As can be seen from the table, the results of clustering depend on the number of resources. As the number of resources increases, the result improves. Increasing purity increases the correspondence between clusters and true classes.

Step 6. The next step is to use a Naive Bayes classifier to automatically assign a new term to an already defined cluster (subject). Naive Bayes is a simple and fast classification method that is very suitable for analyzing large amounts of data. The main idea is to use the features of new terms to estimate the probability of which class they belong to.

Table. Purity coefficient of clustering at different values of *a*

a	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Number of resources	**Purity coefficient**						
150	0.54	0.62	0.68	0.73	0.78	0.82	0.85
300	0.57	0.65	0.71	0.76	0.81	0.84	0.87
450	0.58	0.66	0.72	0.77	0.83	0.86	0.88
600	0.60	0.68	0.73	0.79	0.84	0.87	0.89
750	0.61	0.69	0.75	0.80	0.85	0.88	0.90

Let $C = \{c_1, c_2, \ldots, c_m\}$ be classes. We take the class of terms as w_i: $\{w_i, i = 1, \ldots, m\}$, *m* - is the number of terms. Let us define as $p(w_i)$ the probability that a term w_i belongs to the set of signs *T(x)*. Signs refer to the semantic, grammatical, and morphological characteristics of words.

The condition is as follows:

$$L(w_j, c_i) = \begin{cases} 0, w_j \notin c_i \\ 1, w_j \in c_i \end{cases} \quad (20)$$

The solution ϕ_i is assumed to have losses if the input sample belongs to class w_j. The conditional risk is determined by the loss function within the condition $x \in w_j$:

$$R(\phi_i / x) = \sum_{j=1}^{m} L(\phi_i / w_j) p(w_j / x) \qquad (21)$$

ehtimalı () Bayes teoremi vasitəsilə aşağıdakı kimi hesablanır The probability $p(w_j / x)$ ($x \in w_j$) is calculated using Bayes' theorem as follows:

$$p(w_j / x) = \frac{p(w_j) p(x / w_j)}{\sum_{k=1}^{m} p(w_k) p_k (x / w_k)} \qquad (22)$$

Thus, using the Naive Bayes classifier, we can create our own individual classification and solve the problem of automatic distribution of terms by features into appropriate disciplines. Such a solution to the problem helps to effectively solve the Big data problem, as well as improve the quality of search in the system.

The third paragraph of the third chapter considers the issue of automating the assessment of the level of popularity of speech within the framework of the national terminology information system. For this purpose, a method for assessing the level of popularity of individual speeches was proposed [19].

Method for assessing the level of popularity of speech.

The goal is to assess the level of understanding of speeches, texts by the public or target audience. For this, a method based on a statistical frequency model was proposed. The speech popularity index for unigrams and bigrams was developed.

Let's introduce the following notations:

- $D = \{d_1, \ldots, d_n\}$ – the set of texts (documents) generated by an individual; where *n* is the total number of documents.
- $T = \{t_1, \ldots, t_m\}$ – the set of words found in the set of documents D; where *m* is the number of words;

- $F^{1-gram}=\left\{f^{1-gram},\dots,f_m^{1-gram}\right\}$ – the frequency of occurrence of words $(1-gram)$ in the frequency dictionary of the language; where f_j^{1-gram} is the frequency of occurrence of the word t_j;
- $s_j=\left\{t_{(j,1)},\dots,t_{(j,m_j)}\right\}$ – the word t_j has a lot of synonyms; where m_j is the number of synonyms; $m_j \geq 1$; $j=1,\dots,m$. It is believed that the first element (synonym) $t_{(j,1)}$ of this set is the word t_j itself.
- $F_j^{1-gram}=\left\{f_{(j,1)}^{1-gram},\dots,f_{(j,m_j)}^{1-gram}\right\}$– the frequency of use of synonyms $t_j=\left\{t_{(j,1)},\dots,t_{(j,m_j)}\right\}$ of the word t_j in the frequency dictionary of the language; where $f_{(j,k)}^{1-gram}$ is the frequency of use of the *k*-th synonym $t_{(j,k)}$, $k=1,\dots,m_j; j=1,\dots,m$.
- $F^{2-gram}=\left\{f_{ij}^{2-gram}\middle| i,j=1,\dots,m; i\neq j\right\}$ – the frequency of occurrence of bigrams $(2-gram)$ in the frequency dictionary of a language; where f_{ij}^{2-gram} is the frequency of processing of the bigram $(t_i t_j)$.
- $F_{ij}^{2-gram}=\left\{f_{(i,k)j}^{2-gram}, f_{i(j,p)}^{2-gram}\middle| \begin{matrix} k=1,\dots,m_i; i=1,\dots,m \\ p=1,\dots,m_j; j=1,\dots,m; i\neq j\end{matrix}\right\}$;
- $f_{(i,k)j}^{2-gram}$- is the frequency of occurrence of the bigram $(t_{(i,k)}t_j)$ in the frequency dictionary of the language.
- $f_{i(j,p)}^{2-gram}$- is the frequency of occurrence of the bigram $(t_i t_{(j,p)})$ in the frequency dictionary of the language.
- $(t_{(i,k)}t_j)$ - it is a bigram formed by the *k*-th synonym of the word t_i and the word t_j.
- $(t_i t_{(j,p)})$ - it is a bigram formed by the *p*-th synonym of the

word t_i and the word t_j .

It is assumed that the database of the NTIS contains complete information about terms, their synonyms and usage frequencies.

Uniqramlar üçün nitqin populyarləıq indeksi. In the frequency dictionary of a language, the lowest and highest frequency values of each word (taking into account synonyms) are determined as follows:

$$f_{j\_\min}^{1-gram} = \min_{k=1,\ldots,m_j}\left\{f_{(j,k}^{1-gram}\right\}, \; j = 1,\ldots,m \tag{23}$$

$$f_{j\_\max}^{1-gram} = \max_{k=1,\ldots,m_j}\left\{f_{(j,k)}^{1-gram}\right\}; \; j = 1,\ldots,m \tag{24}$$

If the set of synonyms s_j consists of only one element, in other words, if the word has no synonyms, then $f_{j\_\min}^{1-gram} = f_{j\_\max}^{1-gram}$.
From (23) and (24) it follows that:

$$f_j^{1-gram} \in \left[f_{j\_\min}^{1-gram}, f_{j\_\max}^{1-gram}\right], \; j - 1,\ldots,m \tag{25}$$

If we sum up all the words in (23)-(25) and calculate the average value, we obtain the average value of the frequency of use of words for the set *D*:

$$f_{\min}^{1-gram} = \frac{1}{m}\sum_{j=1}^{m} f_{j\_\min}^{1-gram} \tag{26}$$

$$f_{\max}^{1-gram} = \frac{1}{m}\sum_{j=1}^{m} f_{j\_\max}^{1-gram} \tag{27}$$

$$f^{1-gram} = \frac{1}{m}\sum_{j=1}^{m} f_j^{1-gram} \tag{28}$$

Using formulas (25)-(28), the popularity index of a speech is determined as follows:

$$p^{1-gram} = \frac{f^{1-gram} - f_{\min}^{1-gram}}{f_{\max}^{1-gram} - f_{\min}^{1-gram}} \tag{29}$$

From (29) it follows that $p^{1-gram} \in [0,1]$. It can be seen from the definition that as the value of *p* approaches 0, the popularity level of the speech will be low, and vice versa, as the value of *p* approaches 1, the popularity level of the speech will be high. Thus, by choosing synonyms, the popularity level of the speech can be controlled. It should be noted that the frequency of word combinations (two-two, three-three, four-four, etc.) can also be looked at. This can be done using the simplest model, N-grams (one word, two words, three words, etc.), as well as on the basis of a frequency dictionary.

Speech popularity index for bigrams. The smallest and largest values of the frequency of use of each of the $(t_i\ t_j)$ bigrams (taking into account synonyms) in the frequency dictionary of the language are determined as follows:

$$f_{i\_\min,j}^{2-gram} = \min_{k=1,\dots,m_i}\left\{f_{(i,k)j}^{2-gram}\right\}, i, j = 1,\dots,m, i \neq j \tag{30}$$

$$f_{i\_\max,j}^{2-gram} = \max_{k=1,\dots,m_i}\left\{f_{(i,k)j}^{2-gram}\right\} \tag{31}$$

$$f_{i,j\_\min}^{2-gram} = \min_{p=1,\dots,m_j}\left\{f_{i,(j,p)}^{2-gram}\right\} \tag{32}$$

$$f_{i,j\_\max}^{2-gram} = \max_{p=1,\dots,m_j}\left\{f_{i(j,p)}^{2-gram}\right\} \tag{33}$$

From (30)-(33) it follows that,

$$f_{ij\_\min}^{2-gram} = \min\left\{f_{i\_\min,j}^{2-gram}, f_{i,j\_\min}^{2-gram}\right\} \tag{34}$$

$$f_{ij\_\max}^{2-gram} = \max\left\{f_{i\_\max,j}^{2-gram}, f_{i,j\_\max}^{2-gram}\right\} \tag{35}$$

From (34) and (35) it follows that,

$$f_{ij}^{2-gram} \in \left[f_{ij\_\min}^{2-gram}, f_{ij\_\max}^{2-gram}\right] \tag{36}$$

If we sum over all words in (30)-(36) and calculate the average value, we will obtain the average value of the frequency of processing of bigrams for the set *D*:

$$f_{\min}^{2-gram} = \frac{1}{m(m-1)}\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} f_{ij\_\min}^{2-gram} \tag{37}$$

$$f_{\max}^{2-gram} = \frac{1}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} f_{ij\_\max}^{2-gram} \tag{38}$$

$$f^{2-gram} = \frac{1}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} f_{ij}^{2-gram} \tag{39}$$

Using formulas (36)-(39), the speech intelligibility index for bigrams is determined as follows:

$$p^{2-gram} = \frac{f^{2-gram} - f_{\min}^{2-gram}}{f_{\max}^{2-gram} - f_{\min}^{2-gram}} \tag{40}$$

Written and oral speech materials related to an individual can be evaluated separately by periodizing them over time and then looking at their dynamics, etc.

Time-dependent dynamics of the popularity index. The time-dependent dependence of the popularity level of speech materials is determined as follows:

$$\begin{cases} \frac{dp(t)}{dt} = ap(t) \\ p(0) = p_0 \end{cases} \tag{41}$$

It follows from this that,

$$p(t) = p_0 e^{at} \tag{42}$$

Here a is the rate of increase in the popularity level of speech materials, e is the base of the natural logarithm. Thus, the dependence of the popularity level of an individual's speech on time is an exponential function.

As can be seen, if $a > 0$, that is, the popularity level of speech materials related to the individual increases over time, then the individual learns the language well, if $a < 0$, that is, the popularity level of speech materials related to the individual approaches 0 over time, then the individual loses language knowledge, if $a = 0$, that is, the popularity level of speech materials related to the individual remains unchanged over time, then the individual's language knowledge remains stable.

When assessing the popularity level of speech, it is necessary to assess both its general and scientific popularity levels and compare them. Thus, as a result of this assessment, it is possible to determine the level of language and scientific knowledge of the individual and make a decision. The assessment is carried out in the interval $p \in [0,1]$: if the general popularity level is good, it is assessed as 1, and if it is weak, it is assessed as 0. The value of the popularity level of speech can be different. Thus, the general and scientific popularity level of speech can be good or weak at the same time. There may be cases when the general popularity level of speech is good, but the scientific popularity level is weak, and vice versa.

The value of the overall popularity index of a speech is determined as follows, analogous to formula (40):

$$p_L = \frac{f_L - f_L^{\min}}{f_L^{\max} - f_L^{\min}}, p_L \in [0,1] \quad (43)$$

The value of the scientific popularity index of a speech is determined analogously as follows:

$$p_T = \frac{f_T - f_T^{\min}}{f_T^{\max} - f_T^{\min}}, p_T \in [0,1] \quad (44)$$

Let's summarize the general and scientific indicators of the popularity of a speech according to the importance factor:

$$p = \alpha_L p_L + \alpha_T p_T \quad (45)$$

α – is the importance coefficient. Thus, by regularly analyzing the popularity index of individuals' speech (writing), it is possible to determine the level of language and scientific knowledge of an individual and make decisions

In the fourth chapter (“Possibilities and prospects for creating a NTIS”), proposals are made for the efficient collection and organization of data used in the analysis of the NTIS, experimental verification of the methods and algorithms used in the dissertation is carried out (21,22). The expected results from the implementation of the NTIS are given and recommendations are made regarding the prospects of the system.

In the first paragraph of the fourth chapter, reports were formed based on the software modules corresponding to the data warehouse of the DSS in NTIS. OLAP was visualized by setting up PivotTable Service in Microsoft Excel. For analysis, a PivotTable was built on 50 terms entered into VS from various sources and metadata about those terms (selected term of the term, selected language, selected field and authors). The data selected for analysis is relational, that is, although they are stored in separate tables, they can be combined according to common values. Reports were received as a result of surveys conducted in various sections (Figure 5-7).

Thus, the study showed that it is possible to monitor and manage the processes taking place in the national terminology information environment by analyzing terms and their feature vectors. Thus, by determining how many terms have entered our language during a year, mainly from which field, from which language more terms have been added, etc., we get the opportunity to make decisions based on the information we have obtained about the ecology, safety of the language, and what field of activity is developing in the country. The analysis was conducted on three features (language of the term, field of activity, and date of creation). By expanding the feature vector based on the proposed polycubic OLAP model, it is possible to conduct analysis and make decisions on more parameters.

In the second paragraph of the fourth chapter, the functions and capabilities of the national terminology web portal created within the framework of the NTIS are analyzed (4), the possibilities and prospects for the implementation of the NTIS are determined:

National terminology web portal. The latest technologies were used in the software of the national terminology web portal (www.terminologiya.az): Adobe Photoshop CS6, Adobe Illustrator CC17, HTML5, CSS3, JavaScript, jQuery, AJAX, PHP5.6, MySql. The web portal belongs to the Web2.0 and Web3.0 development stages of web technologies (Appendix 2), provides e-terminology services to citizens. The quality indicators of the web portal are evaluated through the Webometrics system, the LiveInternet web portal, the Google-analytics system. Obtaining information such as a term, its explanation, translation into other languages, its source, etc. through the information search system of the web portal improves the quality of term creation work.

	A	B	C	D
1	Field of act	computer science		
2				
3	Количество по полю Author	Названия столбцов		
4		Deutsch	English	Общий итог
5	Названия строк			
6	Algorithm	1		1
7	Anisotropic filtering			
8	AntiCam			
9	Authentication			
10	BASIC		1	1
11	Computer			
12	Differential evolution		1	1
13	Digital electronic computer	1		1
14	Java		1	1
15	Kabel			
16	Modem		1	1
17	Semantic web		1	1
18	World Wide Web		1	1
19	Общий итог	2	6	8

Figure 5. Distribution of the number of terms across computer science for all languages

1	Language	(All)								
2										
3	Quantity across the field Date	Column names								
4	Row names	biochemistry	chemistry	computer science	ethics	history	economy	mathematics	pedagogics	physics
5	Abscissa							1		
6	Algorithm			1						
7	Anisotropic filtering			1						
8	AntiCam			1						
9	Aqua-Lung									1
10	Authentication			1						
11	Aviation									1
12	Barometer	1								
13	BASIC			1						
14	Beurs						1			
15	Bunsen-Rosco Law		1							
16	Capitalism					1				
17	Catalysis		1							
18	Catharsis									
19	Computer			1						
20	Deontology				1					
21	Differential evolution			1						
22	Diffusion		1							
23	Digital electronic computer			1						
24	Ellipse							1		
25	Exponent							1		
26	Golden ratio							1		
27	Imperium					1				
28	Integral							1		
29	Isomerism		1							
30	Java			1						

Figure 6. Distribution of the number of terms in different fields of activity across all languages

Quantity across the field Language	Column names												
	⊞biochemistry	⊟chemistry									chemistry total	⊞computer science	⊞ethics
Row names		1830	1833	1835	1858	1860	1862	1928	1930	1951			
Abscissa													
Algorithm												1	
Anisotropic filtering												1	
AntiCam												1	
Aqua-Lung													
Authentication												1	
Aviation													
Barometer	1												
BASIC												1	
Beurs													
Bunsen-Rosco Law							1				1		
Capitalism													
Catalysis				1							1		
Catharsis													
Computer												1	
Deontology													1
Differential evolution												1	
Diffusion			1								1		
Digital electronic computer												1	
Ellipse													
Exponent													
Golden ratio													
Imperium													
Integral													
Isomerism		1									1		

Figure 7. Distribution of the number of terms across all languages by field of activity and date

Online contacts with citizens are established through the “Public tribune” section. The development of the NTIS and its web portal creates broad opportunities for the creation of a specialized knowledge base focused on the analysis of terms and increasing the efficiency of term creation in Azerbaijan.

Possibilities and prospects for the implementation of the NTIS. The results obtained from the efforts made towards the synthesis and intellectualization of the national terminological information system determined the possibilities and prospects for the implementation of the system:

- It will provide e-terminology services to users;
- The free participation of people in the term creation process, their comments and suggestions, and their participation in online forums dedicated to the discussion of the prepared terminological dictionaries will give impetus to the formation of citizen terminology;
- As a result of collecting terminological dictionaries prepared in various fields in a single information system, the efficiency of

terminological creation will increase (harmonization of terms, etc.);

- As a result of transparency and easy accessibility, wide opportunities will be created for the protection of the copyrights of scientists and specialists who have prepared terminological dictionaries;
- By processing feedback information on terminological dictionaries and individual terms, it will be possible to conduct various content analyses and assessments, and organize ratings;
- The integration of the national terminological information system with similar systems of international and foreign countries will be accelerated;
- It will create a great opportunity for the formation and development of the terminological base of the Azerbaijani language;
- It will play the role of an information base for research conducted in the field of terminology;
- It will support the implementation of terminology activities in our country based on international and national standards, etc.

RESULTS

The main scientific results presented for defense based on studies carried out on the topic of dissertation are formulated as the following provisions:

1. The scientific and technological problems of terminological informatics were analyzed, the main development directions were determined, and the importance of conducting scientific research in these directions was substantiated [1, 2, 4, 5, 6, 12, 16,18];
2. The conceptual model and architectural and technological principles of the NTIS were developed, its functional subsystems were analyzed, and its capabilities and prospects were determined [3,7,10];
3. A system supporting decision-making based on OLAP technologies and a polycubic OLAP-model were developed for the assessment of the national terminological information system [13,14];
4. A method based on statistical analysis was developed for the automation of term creation processes within the NTIS [8, 9, 11, 15].

5. A method based on the population growth model was developed for modeling dynamic processes within the NTIS [17];
6. A method based on a fuzzy clustering algorithm and a Naive Bayes classification model for the intellectual analysis of terminological dynamic processes has been developed [20];
7. A method based on a statistical frequency model for assessing the popularity level of individuals' speeches based on NTIS has been developed [19].

The following scientific works on Dissertation materials have been published:

1. Əliquliyev R.M., Qurbanova Ə.M. Virtual məkanda Azərbaycan dilinin problemləri və həlli yolları // Terminologiya məsələləri, - Bakı: - 2009, - s.11-18
2. Əliquliyev R.M., Qurbanova Ə.M. Terminoloji informasiya sistemi və inkişaf perspektivləri // Terminologiya məsələləri, - Bakı, - 2010, - s. 27-33.
3. Gurbanova A. Conseptual essences of creation of terminological information system in Azerbaijan // The Third International Conference Problems of Cybernetics and Informatics (PCI), vol.1, - 2010, - Baku, - pp. 157-159.
4. Qurbanova Ə.M. Terminoloji informatika sahəsində aparılan elmi-texnoloji araşdırmalar və onların müqayisəli təhlili // İnformasiya cəmiyyəti problemləri, - Bakı, - 2013, №2(8), - s. 3-12.
5. Qurbanova Ə.M. Qloballaşma şəraitində Azərbaycan dilinə terminoloji təhlükələr haqqında // İnformasiya cəmiyyəti problemləri, - Bakı, - 2015, №2, - s. 87-95.
6. Qurbanova Ə.M. Terminoloji informatika: formalaşma mərhələləri və inkişaf tendensiyaları // İnformasiya cəmiyyəti problemləri, - Bakı, - 2016, №1, - s. 71-77
7. Qurbanova Ə.M. Verilmiş predmet sahəsində terminoloji informasiya sisteminin işlənilməsi // İnformasiya cəmiyyəti problemləri, Bakı, - 2016, №2, - s.74–79.
8. Qurbanova Ə.M. Predmet sahəsinin terminlərinin semantik şəbəkəsinin avtomatik qurulması texnologiyaları // İnformasiya texnologiyaları problemləri, - Bakı, - 2018, №1, - s. 84–91.

9. Qurbanova Ə.M. Termin yaradıcılığı işlərinin avtomatlaşdırılması metodlarının təhlili // İnformasiya texnologiyaları problemləri, - Bakı, - 2018, №2, - s. 92-100.
10. Alguliyev R., Gurbanova A. The Conceptual Foundations of National Terminological Information System // International Journal of Education and Management Engineering, MECS Press, - 2018, 4, - p. 19-30.
11. Гурбанова А.М. Преимущества автоматизации терминографической работы // «TurkLang» VI International Conference on Computer Processing of Turkic Languages, - Tashkent (Uzbekistan), - 2018, - с. 265-272.
12. Alguliyev R., Yusifov F., Gurbanova A. Protection of Azerbaijani language in e-government and development prospects // «TurkLang» VI International Conference on Computer Processing of Turkic Languages, - Tashkent (Uzbekistan), - 18-20 October 2018, - p. 282-288.
13. Набибекова Г.Ч., Гурбанова А.М. Система поддержки принятия решений в среде национальной терминологической информации // Международная конференция «Развитие информатизации и государственной системы научно-технической информации», - Минск, - 2019, - с. 41-45
14. Alguliyev R., Nabibayova G., Gurbanova A. Development of a Decision Support System with the use of OLAP-Technologies in the National Terminological Information Environment // Internationa Journal of Modern Education and Computer Science, Vol.11 (6), - 2019, - p. 43-52. (Scopus)
15. Qurbanova Ə.M. Mətnlərdən terminlərin avtomatik çıxarılması metodlarının araşdırılması və müqayisəli analizi // İnformasiya texnologiyaları problemləri, - Bakı, - 2021, №1, - s. 55-69.
16. Alguliyev R., Yusifov F., Gurbanova A. Protection of Azerbaijani Language in e-government platform // e-Journal of Linguistics, Vol. 15, No.2, - 2021, - p. 155-161.
17. Алигулиев Р.М., Гурбанова А.М. Математическое моделирование динамических процессов в рамках терминологической информационной системы // Информационные технологии, 2022, № 5, с. 226-233.
18. Gurbanova A.M. Problems and Prospects for Minority Languages in the Age of Industry 4.0. // The 4th International Symposium on Computer Science, Digital Economy and Intelligent Systems (CSDEIS2022), The Springer serie Lecture Notes on Data Engineering and Communications

Technologies, - 2023, - Wuhan (China), volume 158, pp. 722-734. (Scopus)

19. Alguliyev R., Aliguliyev R., Gurbanova A. A method of evaluating the popularity of individuals' speeches based on the National Terminological İnformation System // IEEE 17th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan, - 2023, - pp. 29-33, (Scopus)
20. Gurbanova A., Alakbarova I., Mammadzade S. Development of a method for intellectual analysis of terminological dynamic processes // Proceedings of the 2nd Information Technologies and Its Applications Conference, Part 3, – Baku, Azerbaijan: – 23-25 april 2024, – pp. 1–15.

The personal role of applicant in works published with coauthors:

[1] Analysis of research conducted on the status, level of use and problems of the Azerbaijani language in the virtual space and proposing a new approach to solving the problem.
[2] Analyzing the current situation and problems in the field of terminology in Azerbaijan, identifying the possibilities of applying ICT in this direction, and proposing a new approach to terminology activities.
[10] Development of a conceptual model of the national terminological information system, identification of the opportunities and prospects created by the system.
[12, 16] Identification of opportunities for applying linguistic technologies to protect and develop the Azerbaijani language in the e-Azerbaijani segment of the global information space, development of a conceptual model of the Azerbaijani language ecosystem in the e-government environment, identification of functional subsystems of the ecosystem.
[13, 14] The research question is to develop a system that supports decision-making based on OLAP in NTIS and analyze the system's data.
[17] Analysis of the work carried out on mathematical modeling of dynamic processes within the framework of NTIS and proposal of a new approach.
[19] Research on the work carried out on automating the assessment of speech popularity level based on NTIS and proposing a new approach.
[20] Analysis of the work carried out on the automation of the analysis and evaluation of dynamic processes within the framework of NTIS and proposal of a new approach.

The defense of the dissertation will be held on 7 February 2024 at 16^{00} at the meeting of the ED 1.35 Dissertation Council operating under the Institute of Information Technology of the Ministry of Science and Education of the Republic of Azerbaijan.

Address: Az 1141, Baku city, B.Vahabzadeh street, 9a.

The dissertation can be viewed in the library of the Institute of Information Technology of the Ministry of Science and Education of the Republic of Azerbaijan.

Electronic versions of the dissertation and abstract are posted on the official website of the Institute of Information Technology of the Ministry of Science and Education of the Republic of Azerbaijan.

The abstract was sent to the necessary addresses on 28 December 2024.

Printed: 26.12.2024

Paper format: $60 \times 84^{1/16}$

Volume: 37196 characters

Printing: 100 copies