

O'ZBEKISTON RESPUBLIKASI FANLAR AKADEMIYASI

O'ZBEKISTON RESPUBLIKASI AXBOROT TEXNOLOGIYALARI
VA KOMMUNIKATSIYALARINI RIVOJLANTIRISH VAZIRLIGI

**INFORMATIKA VA ENERGETIKA
MUAMMOLARI**
O'zbekiston jurnali

Узбекский журнал
**ПРОБЛЕМЫ
ИНФОРМАТИКИ И ЭНЕРГЕТИКИ**

Uzbek Journal
**OF THE PROBLEMS OF
INFORMATICS AND ENERGETICS**

2

2023

FAN VA TEXNOLOGIYA

O'ZBEKISTON RESPUBLIKASI FANLAR AKADEMIYASI
АКАДЕМИЯ НАУК РЕСПУБЛИКИ УЗБЕКИСТАН
O'ZBEKISTON RESPUBLIKASI RAQAMLI TEXNOLOGIYALAR
VAZIRLIGI
МИНИСТЕРСТВО ЦИФРОВЫХ ТЕХНОЛОГИЙ РЕСПУБЛИКИ
УЗБЕКИСТАН

ИНФОРМАТИКА
VA ENERGETIKA
MUAMMOLARI

2-2023

ПРОБЛЕМЫ
ИНФОРМАТИКИ
И ЭНЕРГЕТИКИ

*Журнал под таким названием издается с января 1992 г.
по 6 номеров в год*

Fan va texnologiyalar nashriyot-matbaa uyi

ТАШКЕНТ– 2023

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Ш.Х. ФАЗЫЛОВ (*главный редактор*)
Х.М. МУРАТОВ (*зам. главного редактора*)
Б.М. АЗИМОВ (*ответственный секретарь*)
О.З.ТОИРОВ (*ответственный секретарь*)
С.Ф.АМИРОВ
Р.А.ЗАХИДОВ
Х.З.ИГАМБЕРДИЕВ
М.А. ИСМАИЛОВ
О.Х.ИШНАЗАРОВ
М.М. КАМИЛОВ
М.М.КАРИМОВ
М.М.МУСАЕВ
Т.Х.НАСЫРОВ
Н. РАВШАНОВ
М.А.РАХМАТУЛЛАЕВ
Л.Ф.СУЛЮКОВА
Р.Х.ХАМДАМОВ

Адрес редакции:

100173, г.Ташкент, ул.Фозилтепа, 226
Телефоны +998 90 927 72 13, +998 90 965 43 28
e-mail: informatika-energetika@mail.ru

*Редактор: Ш. Кушербова
Технический редактор: А. Муридимова
Компьютерная верстка: Ш. Мирзиёева*

Журнал зарегистрирован Агентством по печати и информации Республики Узбекистан 22.12.2006 г.
Регистрационный №0046, Изд. лиц. АЛ №149, 14.08.09, Изд. №3-75. Сдано в набор 07.10.2023. Подписано в печать,
27.11.2023. Формат 60×84 1/8.
Гарнитура литературная. Печать офсетная. Бумага офсетная. Усл-печ. л. - 12,5. Уч. - изд. л. 8,28. Тираж 150. Заказ № 26
Цена договорная.

Издательство «Fan va texnologiyalar nashriyot-matbaa uyi»: 100173, г.Ташкент, ул.Фозилтепа, 226
Отпечатано в типографии Научно-исследовательского института развития цифровых технологий и
искусственного интеллекта при Министерстве цифровых технологий
Республики Узбекистан: 100125, Ташкент ш., Буз-2 мангеси, 17А уй.

«Fan va texnologiyalar nashriyot-matbaa uyi»

ТАШКЕНТ – 2023

СОДЕРЖАНИЕ

Информатика и управление

Т.Р.Нурмухаммедов, Ш.И.Юлдашев. Алгоритм прогнозирования подходящего управления для сложных систем на основе теории распознавания образов.....	3
Ж.Ш. Норкулов. Моделирование и оптимальное управление движением гидронасоса гидроприводов испытательного стенда рабочих органов машинных агрегатов.....	9
Б.Ф.Абдурахимов, Ж.Р.Абдураззоков. Использование криптоанализа на основе глубокого обучения (DEEP LEARNING) с целью получения информации о ключе упрощенного S-AES.....	17
Х.Х.Аминов. Метод решения задачи о выпуске газа с конца элементарного участка газопровода в окружающую среду.....	27
С.А.Саидов, Г.К.Аннакулова. Математическая модель управляемого сервомеханизма.....	37

Энергетика

О.З.Тоиров, С.Э.Уроков, Ш.А.Ходжаева. Блок схема скалярного управления и алгоритм определения энергетических показателей частотно-регулируемого элек-тропривода пластичного питателя.....	43
Р.Х.Рахимов, К.А.Самиев, Б.А.Хикматов, М.С.Мирзаев. Экспериментальные исследования устройства солнечной сушилки с принудительной конвекцией и функциональной керамической композитной пленкой.....	54
М.М. Хамдамов. Математическое моделирование маломощного ветрогенератора с вертикальной осью вращения в рамках программного пакета Ansys Fluent.....	61
О.З. Тоиров, Ш.Ш. Атоев, Т.М. Саъдуллаев. Энергосберегающий режим работы и функциональные схемы хлопкоочистительной установки на базе частотно-регулируемых электроприводов.....	73

Информационные и телекоммуникационные технологии

З.З.Шамсиев. Алгоритмизация формирования программ подготовки авиадиспетчеров управления воздушным движением.....	82
Б.Елов, М.Абжалова, Р.Алаев. Учебный корпус узбекского языка и его возможности.....	89

МУНДАРИЖА
Информатика ва бошқарув

Т.Р.Нурмухаммедов, Ш.И.Юлдашев. Тимсолларни аниқлаш назариясига асосланган мураккаб тизимлар учун мослашувчан бошқаришни башоратлаш алгоритми.....	3
Ж.Ш. Норкулов. Машинатракторли агрегатларнинг ишчи органлари гидроритмаларини синаш стендидаги гидронасос ҳаракатини моделлаштириш ва оптимал бошқариш.....	9
B. F. Abduraximov, J. R. Abdurazzoqov. Soddalashtirilgan aes (S-AES) ning kaliti haqida ma'lumotga ega bo'lish maqsadida chuqur o'qitish (DEEP LEARNING)ga asoslangan kriptotahlildan foydalanish.....	17
X. X. Aminov. Gaz quvurining elementar qismidan atrof-muhitga gaz chiqishi muammosini hal qilish usuli.....	27
C. A. Saidov, G. K. Annakulova. Boshqariluvchi servomekhanizmining matematik modeli.....	37

Энергетика

O. Z. To'irov, S. E. Urokov, Sh. A. Xodjaeva. Частотавий бошқарилувчи электр юритмали пластинкали таъминлагичнинг скаляр бошқарув структура схемаси ва энергетик кўрсаткичларини аниқлаш алгоритми.....	43
R. X. Raximov, K. A. Samiyev, B. A. Hikmatov, M. S. Mirzayev. Funktsional keramika kompozit plyonkali majburiy konveksiyali quyosh quritgich qurilmasida tajriba tadqiqotlari.....	54
M. M. Hamdamov. ANSYS Fluent paket dasturi yordamida verikal o'qli kam kuchli shamol turbinasini matematik modellashirish.....	61
O. Z. To'irov, Sh. Sh. Atoev, T. M. Sa'dullayev. Частотавий бошқарилувчи электр юритмалар асосида пахта тозалаш қурилмасининг энергия тежамкор иш режими ва функционал.....	73

Ахборотли ва телекоммуникацияли технологиялар

Z. Z. Shamsiyev. Havodagi harakatni boshqaruvi aviatsiya dispetcherlarining o'qitish dasturlarini shakllantirish algoritimizatsiyasi.....	82
B. Elov, M. Abjalova, R. Alayev. O'zbek tili korpusi va uning imkoniyatlari.....	89

EDUCATIONAL CORPUS OF THE UZBEK LANGUAGE AND ITS OPPORTUNITIES

The educational corpus is a language corpus based on school textbooks and dictionaries, and is a structural type of the Uzbek National Corpus. According to the "Concept of the National Corpus of the Uzbek language", the educational corpus of the Uzbek language was created in the framework of the practical project AM-FZ-201908172 "Creation of the educational corpus of the Uzbek language", the first presentation was held on April 23, 2021. Since then its database has constantly been enriched and developed. It is well-known that language corpus is an important tool in language education the process that involves the creation of dictionaries, as well as various research, diachronic and synchronous study of language, the development of speech competence, vocabulary as well as speech patterns. The creation of the educational corpus is a unique technological tool in the study of the native language and its use as a foreign one. Therefore, the educational corpus of the Uzbek language is the first major stage of the "Concept of creating the National Corpus of the Uzbek language".

This article discusses the factors, principles, models, databases, architecture, information systems and opportunities for creating an educational corpus of the Uzbek language.

Keywords: educational corpus, lexicography, dictionary of word's frequency, terminological dictionary, database, model, information system. The article is written in the framework of the practical project AM-FZ-201908172 "Creation of the educational corpus of the Uzbek language".

B.Elov, M.Abjalova, R.Alayev

O'zbek tili korpusi va uning imkoniyatlari

Korpus maktab darsliklari va lug'atlari asosida tuzilgan til korpusi bo'lib, O'zbek tili milliy korpusi korpusning turi hisoblanadi. "O'zbek tili milliy korpusi konsepsiyasi"ga muvofiq, o'zbek tili korpusi AM-FZ-201908172 "O'zbek tili korpusini yaratish" amaliy loyihasi doirasida yaratilgan bo'lib, ilk taqdimot 2021-yil 23-aprelda bo'lib o'tgan. Hozirda korpusning ma'lumotlar bazasi doimiy ravishda boyitib borilmoqda. Ma'lumki, til korpusi til ta'limining muhim vositasi bo'lib, lug'atlar yaratish, shuningdek, tilni turli xil tadqiqotlar, diaxronik va sinxron o'rganish, nutq kompetensiyasini, lug'at va nutq shakllarini rivojlantirishni o'z ichiga oladi. O'quv korpusining yaratilishi ona tilini o'rganish va undan chet tili sifatida foydalanishda o'ziga xos texnologik vosita hisoblanadi. Binobarin, o'zbek tilining o'quv korpusi "O'zbek tili milliy korpusini yaratish konsepsiyasi"ning birinchi yirik bosqichidir.

Kalit so'zlar: O'quv korpusi, leksikografiya, so'zlar chastotasi lug'ati, terminologik lug'at, ma'lumotlar bazasi, model, axborot tizimi. Maqola AM-FZ-201908172 "O'zbek tili milliy korpusini yaratish" amaliy loyihasi doirasida yozilgan.

Учебный корпус узбекского языка и его возможности

Показано, что учебный корпус представляет собой языковой корпус на основе школьных учебников и словарей и является структурным типом Узбекского национального корпуса. Согласно «Концепции Национального корпуса узбекского языка», учебный корпус узбекского языка создан в рамках практического проекта АМ-FZ-201908172 «Создание учебного корпуса узбекского языка». Первого презентация состоялась 23 апреля 2021 г. С тех пор ее база данных постоянно пополняется и развивается. Общеизвестно, что языковой корпус является важным инструментом языкового образования, процесса, который предполагает создание словарей, а также различные последования, дихроническое и синхронное изучение языка, развитие речевой компетентности, словарного запаса и речевых моделей. Созданный учебный корпус является уникальным технологическим инструментом в изучении родного языка и использовании его в качестве иностранного. Поэтому учебный корпус узбекского языка представляет собой первый крупный этап «Концепции создания Национального корпуса узбекского языка». В данной статье рассматриваются факторы, принципы, модели, базы данных, архитектура, информационные системы и возможности создания учебного корпуса узбекского языка.

Ключевые слова: учебный корпус, лексикография, словарь частотности слов, терминологический словарь, база данных, модель, информационная система. Статья написана в рамках практического проекта АМ-FZ-201908172 «Создание учебного корпуса узбекского языка».

Introduction. The profound development of the modern information technology has opened the door to endless possibilities for using the functional capabilities of language. Machine translation, automatic editing and analysis, speech synthesizers that transcribe written text, speech recognition programs that translate spoken speech into written text, electronic dictionaries, mobile linguistic applications, various types of thesauruses and language ontology are all can be a solid proof of our point. In particular, the creation of a plethora of modern electronic dictionaries and their use has proven to be effective in capturing language opportunities. The role of language corporations, which need to be created on a global scale, is invaluable in language education and the creation of electronic dictionaries [1–16].

The computer technology usage in corpus linguistics determines the period of the foundation and development of electronic linguistics which makes it possible for the scientists to follow the frequency of word use on the basis of language corpus, the development of terminology, the study and analysis of sentence structure, the creation of n-language bases for translation programs, and the study of methodology. The ability to create dictionaries has shifted from a card index to an automated process, creating unprecedented conveniences [1]. When it comes to compiling dictionaries, the usefulness of language corpora, especially the creation of frequency dictionaries, is important in the formation of terminology in a particular field. Also, translator dictionaries using the translation database of bilingual or multilingual parallel corpus serve as a major resource in the formation of linguistic databases of automatic translation programs.

In general, a multi-purpose and multifunctional language corpus is the National Corpus of a particular language. At the base of such a corpus the followings are estimated:

- 1) the presence of a large volume of texts belonging to all styles of language;
- 2) linguistic, morphological and syntactic tagging of texts;
- 3) availability of case metadata is a requirement.

These requirements ensure that the National Corpus becomes a necessary linguistic and linguodidactical technological tool. The educational corpus is a special type of the national corpus, mainly teaching language in the educational process; it is the main tool in matters such as conducting various practical exercises in the educational process, the use of text elements in the research process, the formation of terminology in the field of education and the formation of the educational thesaurus.

The corpus is a large source for compiling dictionaries. Over time, the corpus has become a powerful information resource. Corpus-based computer-generated dictionaries are being created and processed faster than ever before. For example, there are several dictionaries based on the Russian Language National Corpus. It is noteworthy that the glossary of electronic dictionaries (keywords) and dictionary articles based on language corporations with dynamic variability do not have the tendency to "wear out".

Charles Fillmore commented on the popularity of the text corpus:

"I can make two comments. First, I don't think there can be a corpus that contains information about all areas of vocabulary and grammar, no matter how big it can be. The second, every corpus I have read, no matter how small, has shown me facts that I could never have found otherwise". According to Fillmore, there is no limit to the size and scope of language corpora, and with such unlimited capabilities, other resources such as systems and softwares cannot compete with it. Therefore, the creation of corpus-based dictionaries reflects the viability of the language and the real vocabulary in the dictionaries.

In order to develop the field of **terminology**, the terms of a particular field is formed from the texts of millions of different fields in the database. For example, the English-Latin terms for engineering, banking, finance, construction, information technology, and international relations can be defined by a single corpus search. One of the results of the penetration of computer technology into all fields over the last decade is, firstly, the need to create different types of electronic lexicographic sources, and secondly, the digitalization of texts in different fields.

There are two reasons for creating corpus-based terminological dictionaries:

1) A large amount of written and oral texts in various fields are summarized in the corpus. This is especially true of the general corpus, the national corpus of natural language. Annotation, that is, the grammatical and semantic interpretation of each word, reveals the interpretation of a particular word in the field, and gives rise to a general or terminological feature.

2) Determining the frequency of words used in texts belonging to a particular field makes it possible to distinguish between active terms and inactive terms.

The set of terminological bases is the basis for any research and it can be considered as an important factor in the development of computer lexicography.

In the "Schemes of corpus objects" section of the article there are diagrams of words, phrases, and text objects on the basis of the educational corpus of the Uzbek language given.

Schemes of objects of the corpora. In the formation of the lexicographic database of the Uzbek language educational corpus, the database was modeled using the SQL-SERVER 2019 database management system on the basis of 1,2,3-normal forms of the Code. Figures 1, 2, and 3 below show the Words, Sentence, Text, and Source objects, and the relationship between the related objects.

Method analysis. It is known that in almost all public schools of the country, a student reads a sentence from primary school and identifies nouns, adjectives, numbers, verbs, adverbs, rhymes. In corpus linguistics, the word is a category tag. Word-tagging (in English it is called part-of-speech tagging (POS tagging or PoS tagging or POST), in Russian it is called *chasterechnaya razmetka* (частеречная разметка)) is a stage of automatic text processing, the function of which is the grouping of words used in the text and grammar in order to determine the properties. POS-tagging with this function is one of the first steps in automatic text analysis.

In corpus linguistics, tagging word groups, tapping grammatical categories, and eliminating ambiguities in word classification are used not only to refer to a word based on its lexical form, but also to its categorical tag according to its expression in the text

and it is important to consider the possibility of combining with other words in a sentence (paragraph and/or phrases). Identifying parts of speech is a difficult process. This is because it is not possible to tag all Uzbek words universally within 12 categories. The same word can be grammatically related to more than one grammatical function depending on the context of its realization in the sentence structure and the semantic valence of the N-gram words. For example: In the sentences “*Shifoxonaga bemorni keltirishdi*” (“The patient was brought to the hospital”) and “*Shifoxonaga bemor odamni keltirishdi*” (“The ill patient was brought to the hospital”), the word patient is a category of noun according to the sign of the category (the word “*bemor*” answers the question “Who?” in the first example) and whereas in the second example it has got the function of the Adjective (as it answers the question what kind of?). Out of 11,000 borrow words in the Uzbek dictionary, 66 similarly polyfunctional words were identified.

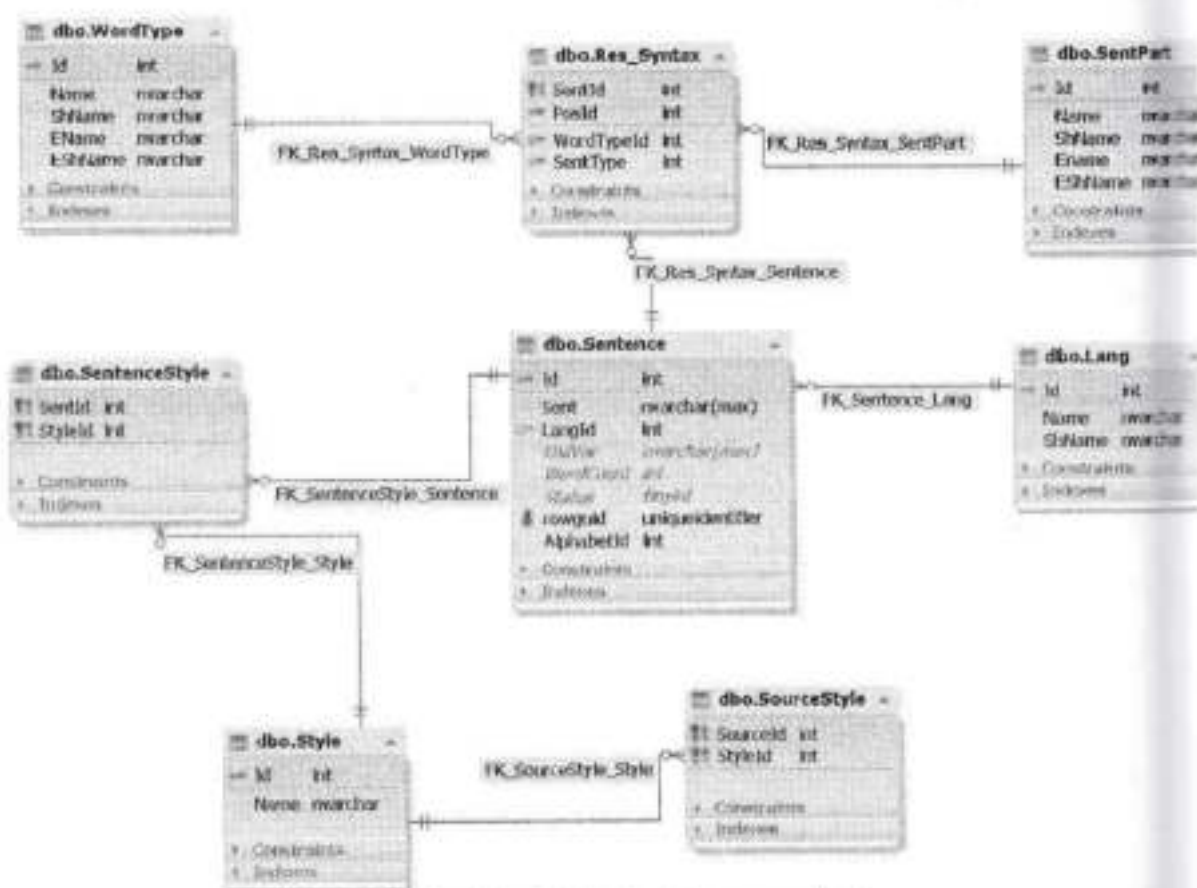


Fig. 1. Relational model of the Sentence object

It is not enough to include a list of words and their categories in the linguistic database for word grouping (STT). Linguists have to ponder about and explore the issues of the loss of consistency, as in the case of the definition of the above word group, or finding a set of polyfunctional, homonymous [Abjalova, Yuldashev, 2021] or polysemous words in a sentence. Also, many words in the Uzbek language do not belong to a certain category. Given the problems that exist in every natural language, STT relies on several methods.

Methods of tagging word groups. In most cases, the following methods or algorithms are based on the tagging of word groups:

- rule-based approach
- stochastic (or statistical) method.

According to the rule-based method, word grouping is done in two stages:

In the first stage, the PoS-tagger (annotated, spelling, morphological or orthographic) relies on the dictionary. It uses a dictionary to determine the grammatical category of each word.

In the second stage, a series of homonymous words is written by hand, and a long list of rules for determining the function of such words in a sentence is developed.

A clear example of the method of automatic generation of rules is the method of American linguist Eric Brill.

Features of rule-based ST tags

Rule-based PoS tags have the following features:

- These tags are based on knowledge.
- Rules are created manually.
- Information is encoded in the form of rules.
- The rules will be limited.

For digital technology, infinity is abstract, and compounds and symbols such as "...", ":", "etc." indicate ambiguity, not the length of the list. Therefore, in computer linguistics, such ambiguities are not allowed, but the nature and characteristics of each linguistic unit must be clearly stated in the software database.

- Language modeling is based on the rules of tagging.

STOCHASTIC TAGGING METHOD. This method is based on frequency or probability. Therefore, in some sources it is explained as a statistical or probabilistic method. As can be seen, the following methods are used for ST tags in simple stochastic tagging:

1. Frequency approach. In this approach, stochastic tags eliminate grammatical ambiguities based on the probability that a word will meet a particular tag in the text. It can also be said that a tag that is often used with a particular word in the collection (part of the text) under study is a tag that helps to inform the ambiguity of that word. For example, according to the level of usage, the linguistic unit that comes after the word in the consonant belongs to the group of verbs: "*kitobni o'qimoq, ismni yozmoq, mamlakatni aylanib chiqmoq* => {Wni +V}" (to read a book, to write a name, to travel around the country). The main problem with this approach to tagging is that it can result in a sequence of tags that are not cohesive in natural language. For example, in the example "*Do'stinga kitobni sovg'a qilish uchun sotib oldim*" ("I bought a book as a gift to a friend"), the word "gift" is combined with the compound "to do" on the right, but cannot be grammatically or semantically combined with the word "book" on the left. In such cases, incorrect information is generated during morphological analysis as a result of the disproportion of words that do not have cohesiveness in the matching.

2. Probability of a sequence of tags or the n-gram method. This approach to the stochastic method calculates the probability that a given sequence of tags will be used. This approach is also called the N-gram method because it is based on a sequence dimension, n (bigram - a sequence of two elements, trigram - three consecutive tags, 4 grams - a sequence of four tags). N-gram is a mathematical calculation tool widely used in automatic word processing.

The hidden Markov model is actively used in a stochastic way. In the 1960s, Baum L.E. and this method, developed by his colleagues, helps to take into account the probability of all the options that arise in the statistical process. For example, if in a given text the words belonging to the category of noun are more frequent than the conjunctions, then the homonym present in the same context is most likely not a conjunction, but a word belonging to the category of noun, the next is considered as a binder in probability. The N-gram is used to describe the context. An N-gram is a sequence of N-identifying elements, such as words or tags.

Features of the method of stochastic tagging of word groups

Stochastic PoS-taggers have the following features:

- This tagging is based on the probability of the tags being used sequentially.
- Educational corpus required.
- There is no possibility for words that do not exist in the corpus.
- In addition to the educational corpus, other types of language corpus can be used.

-The simplest ST tagging method, because this method selects the tags that are actively used in the language corpus.

Educational corpus architecture. The educational corpus is designed as an integrated web application that allows users to access and use the university's distributed information system to carry out their activities effectively. Educational corpus users can use a web browser to access the Internet or a local area network. In this case, to ensure security, it is necessary to ensure the confidentiality of traffic from the web browser to the web server, to authenticate users, to ensure the authentication of the web server. SSL certificates are used to provide traffic protection and authenticate the web server. SSL certificates allow encrypted data exchange between the user's browser and the server. SSL certificates can be purchased from Registration Centers (VeriSign, Inc.; The Go Daddy Group, Inc.; Thawte et al.).

IDS / IPS systems use a firewall to filter incoming and outgoing traffic to detect and eliminate attacks from the Internet or other networks. The architecture of authentication, identification and authorization of educational corpus users is shown in Fig. 4:

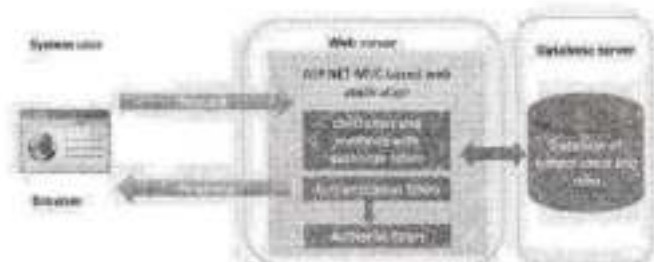


Fig. 4. Authentication, identification and authorization of educational corpus users

The security mechanisms shown in Fig. 5 below are used to provide security in the educational corpus web application layer.

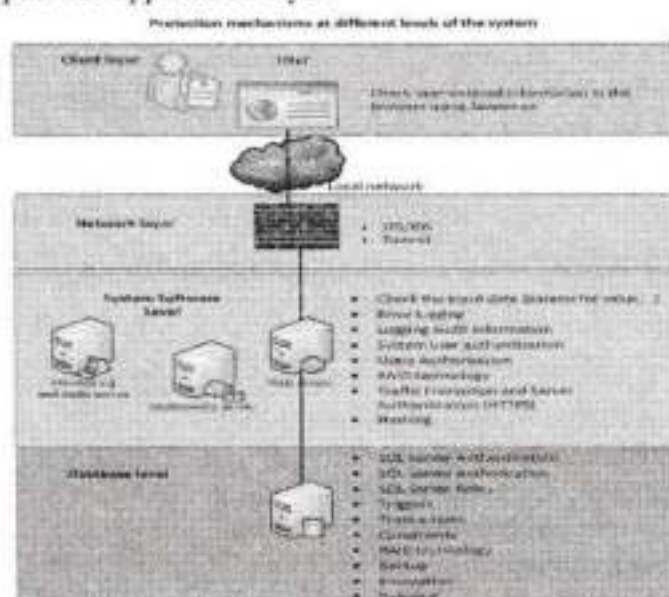


Fig. 5. Information security mechanisms at different levels of the information system

Structure and opportunities of the Uzbek language educational building.

The corpus consists of a corpus interface (1), a search engine (2) and a dictionary column (3). In the "Dictionaries" section there is an annotated dictionary, dictionaries of homonyms, synonyms, paronyms and antonyms as lexicographic resources. The first page of the website of the Uzbek language educational corpus contains basic information about the corpus and its creators, you can go to any page in the menu on the right (fig. 6,7).

Corpus search allows the user to identify the followings:

- 1) identify all forms of a particular word with an array of examples;

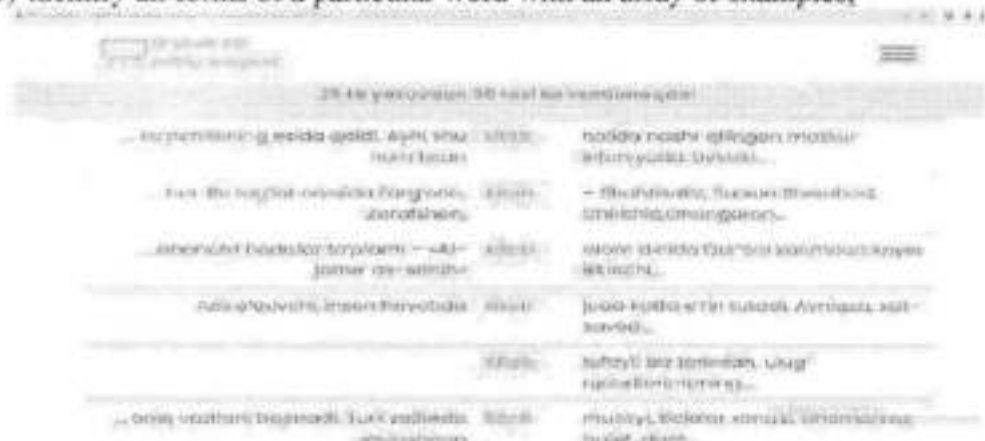


Fig. 6. An array of examples involving the word "book"

- 2) identify the source of the examples;
- 3) obtain information about the source;
- 4) get acquainted with the meaning of the word;
- 5) morphological analysis of the word, study of its division into syllables and listening to its pronunciation;
- 6) obtain a list of words that can be combined with the search word on the right and left (N-gram technology);
- 7) know the frequency or statistics of word usage;
- 8) determine the relationship of form and meaning of the word (homonyms, paronyms, antonyms, synonyms)



Fig. 7. The results of the search for "in our lives" in the "Dictionaries" section of the Uzbek language educational corpus

These possibilities are reflected in the corners of the corpus. As an example, we see the result by typing the word "in our lives" in the search bar in the "Dictionaries" section of the corps:

- 1) "word structure" - automatic morphological analysis reveals that the basis of the word "in our life" is "life";
- 2) "Word syllables" are highlighted;
- 3) an explanation of the word life is provided;
- 4) there are no contradictory forms;
- 5) the meanings of the lemma of life are presented: *living, life, life*;
- 6) it is found that there is no pronunciation (paronymy).

As can be seen from this example, the educational corpus has a synonymizer program, which provides a meaning from a database of synonyms, independent of other resources, based on the lemma of the word.

The size of the educational corpus. Currently, the volume of linguistic knowledge of the Uzbek language educational corps has the following indicators table.

About sources in database of Uzbek national corpus

No	Name of the source	Amount
1	Books	184
2	Internet texts	47 8908
3	Words that are tagged with word groups	36 897
4	Sentences	5 600 015
5	Context	5 562 202
6	The number of words explained	4575
7	Synonic line	993
8	Antonymic pairs	870
9	Paronymic pairs	558
10	Frazems	1384
11	Homonyms	1636
12	Gradual line	413



Fig. 8. Search results for "Dictionaries" in the Uzbek language educational corpus

Practical results. As a result of several years of research and practical efforts in 2020-2021, the project AM-FZ-201908172 "Uzbek language EDUCATIONAL CORPORA OF THE UZBEK LANGUAGE was created in the framework of the

practical project "Creation of an educational corpus" in the collaboration between Tashkent State University of Uzbek Language and Literature and Department of Computer Linguistics and the Department of Digital Technologies and Applied Linguistics and Linguodidactics. Today in this corpus there are such functions as:

- 1) automatic morphological analysis creation;
- 2) division of a word into syllables;
- 4) an explanation generator;
- 5) indication of antonyms;
- 6) synonymizer (a program to provide its meanings to the search word);
- 7) paronyms of the word with their explanation;
- 8) information from the "Mother tongue encyclopedic dictionary" related to the word searched;
- 9) indication of the phrase that contains the searched word;
- 10) demonstration of a series of hierarchies according to its various features.

Conclusion: 1. The language corpus is a system with electronic search capabilities to identify the features of national language units, a set of digitized written and spoken texts of a natural language.

2. It will be widely used by linguists, lexicographers, computer linguists, programmers, editors, translators, journalists, publishers, scientists, teachers, learners and any other professional.

3. It is advisable to rely on the corpus, where the database is regularly updated, so that language learning is not interrupted.

REFERENCES

1. The National Corpus of the Uzbek Language is an important cultural reality (conversation) // "Ma'rifat", 11.08.2021. № 32 (9357).
2. Abjalova M.A. Linguistic modules of editing and analysis programs. [text] : Monograph. Tashkent: Nodirabegim, 2020. –176 p. ISBN 978-9943-6939-0-6.
3. Abjalova M.A., Yuldashev A. Methods for Determining Homonyms In Homonymy And Linguistic Systems. ACADEMICIA: An International Multidisciplinary Research Journal. Vol. 11, Issue 2, February. Impact Factor: SJIF 2021 = 7.492 (<https://saarj.com>). ISSN: 2249-7137.
4. Asiryun A.K. Сравнение инструментов морфологической разметки// Научный взгляд в будущее, 10.30888/2415-7538.2017-07-01-027.
5. Boliang Zhang, Ajay Nagesh, Kevin Knight. Parallel Corpus Filtering via Pre-trained Language Models. 10.18653/v1/2020.acl-main.756
6. Brill E. Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging // Computational Linguistics. Vol. 21. 1995. P. 543–565. <http://acl.ldc.upenn.edu/J/195/195-4004.pdf>.
7. Elaheh Rafatbakhsh, Alireza Ahmadi. A thematic corpus-based study of idioms in the Corpus of Contemporary American English. Asian-Pacific Journal of Second and Foreign Language Education. 2019. 10.1186/s40862-019-0076-4.
8. Elaheh Rafatbakhsh, Alireza Ahmadi. The most frequent idioms used in contemporary American English: A corpus-based study. Applied Research on English Language, 10.22108/are.2019.114449.1389.
9. Erwin Tschirner, Jupp Möhring. A Frequency Dictionary of German. A Frequency Dictionary of German. 2019.10.4324/9781315620008.
10. Mansur Alp Toçoglu, Adil Alpkoçak. Lexicon-based emotion analysis in Turkish. Turkish Journal of Electrical Engineering and Computer Sciences. 2019.10.3906/elk-1807-41.

11. Nina Vyatkina. Corpora as open educational resources for language teaching. *Foreign Language Annals*, 10.1111/flan.12464.
12. Salima Mdhaffar, Yannick Estève. A multimodal educational corpus of oral courses: Annotation, analysis and case study. LREC 2020 – 12th. International Conference on Language Resources and Evaluation, Conference Proceedings.
13. Valeriy M. Mokienko, Elena I. Zykova. The school dictionary of Russian winged words as an object of lexicographical description. *Voprosy Leksikografii*. 2019.10.17223/22274200/16/9.
14. Zhao Shengxue. English corpus translation system based on FPGA and machine learning. *Microprocessors and Microsystems*, <https://doi.org/10.1016/j.micpro.2020.103464>.
15. Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М. 2009. Азбуковник.
16. Полицын С.А., Полицына Е.В. Применение корпуса текстов для автоматической классификации в комплексе инструментов автоматизированного анализа текстов. *Вестник ВГУ. Серия: Системный анализ и информационные технологии*. 2018. 10.17308/sait.2018.2/1224. <http://uzschoolcorpara.uz/> – Educational corpus of Uzbek language (15.05.2022).

Tashkent State University of
Uzbek Language and Literature
named after Alisher Navoi
National University of
Uzbekistan named
after Mirzo Ulugbek

Receipt date
25.08.2023