

OLEZIA BARKOVSKA, ANTON HAVRASHENKO,
VLADYSLAV KHOLIEV, OLENA SEVOSTIANOVA
Kharkiv National University of Radio Electronics

AUTOMATIC TEXT TRANSLATION SYSTEM FOR ARTIFICIAL LLANGUAGES

The growing number and variety of artificial languages leads to the need and relevance of creating automatic dictionaries for their translation in order to facilitate human communication. Such languages include languages where vocabulary, phonetics, and grammar have been specifically designed to achieve specific goals and to communicate with a group of people by interests or place of residence. These languages can be distributed among people of certain professions or among neighboring nations. Examples are slang and surzhik. The common for them is that there is a basic language (literary), the intersection in spelling and meaning of words and phrases with which is quite large. The main goal of the project is to create a system of automatic translation of words and texts from / into arbitrary languages, including hybrid, artificial and slang ones. The proposed model shows the interaction and partial interdependence of the creation and adjustment modules and the translation module of the dictionary, which is explained by tacking the approach of reverse propagation of the translation error. To perform experiments and analyze the performance of the proposed approach to the organization of automatic translation of texts from and into arbitrary language, a software application was developed, which includes a subprogram of initial word processing for dictionary organization, one for creating a working dictionary and one for two-way improvement of created dictionary by the inclusion of new texts in order to improve the quality of translation, including the search for word phrases, idiom, and translation for them, the subprogram of dividing the dictionary into sub-dictionaries with a small percentage of text, the subprogram of the translator itself. To test and analyze the results of the proposed generalized model, three types of source texts were used: literary poetry translation, literary prose translation, literal prose translation. The results of the experiments showed that the proposed approach provides a high level of translation (up to 98,8%) in similar languages (between such languages as Ukrainian-Russian, or Ukrainian - Ukrainian-Russian surzhik with equal word order in the sentence), especially with a literally translated source text. It has become known that the use of artistic texts to generate dictionaries is possible, but not very effective.

Keywords: dictionary, processing, model, text, language, slang, computer system.

ОЛЕСЯ БАРКОВСЬКА, АНТОН ГАВРАШЕНКО,
ВЛАДИСЛАВ ХОЛІВ, ОЛЕНА СЕВОСТЬЯНОВА
Kharkiv national university of radio electronics, Kharkiv, Ukrain

СИСТЕМА АВТОМАТИЧНОГО ПЕРЕКЛАДУ ТЕКСТУ ДЛЯ ШТУЧНИХ МОВ

Зростання кількості та різноманіття штучних мов призводить до необхідності та актуальності створення автоматичних словників для їх перекладу із метою полегшення людського спілкування. До таких мов відносяться мови, де лексика, фонетика і граматики були спеціально розроблені для втілення певних цілей і спілкування групи людей за інтересами або місцем проживання. Прикладами є сленги та суржики. Спільним для них є те, що для них існує базова мова (літературна), перетин за написанням і значенням слів та словосполучень із якою є досить великий. Метою роботи є створення узагальненої моделі системи автоматичного перекладу текстів з довільної на довільну мови, включаючи штучні та гібридні мови. Запропонована модель показує взаємодію та часткову взаємозалежність модулів створення та коригування словнику та модулю перекладу, що пояснюється підходом зворотнього розповсюдження помилки перекладання. Для виконання експериментів було розроблено програмний застосунок, що включає в себе підпрограму первісної обробки тексту для організації словнику, підпрограму створення робочого словнику, підпрограму двостороннього покращення створеного словнику шляхом доповнення новими текстами для підвищення якості перекладу, включаючи пошук словосполучень, фразеологізмів, та перекладу для них, підпрограму розвитку словника на підсловники при малому відсотку перетину тексту, підпрограму перекладача. Для тестування та аналізу результатів роботи запропоновано узагальнену модель, використано три типи початкових текстів: художній переклад поезії, художній переклад прози, дослівний переклад прози. Результати показали, що запропонований підхід дає високий рівень перекладу (до 98,8%) для подібних мов (таких, як українська та російська, або українська та українсько-російський суржик із однаковим порядком слів у реченні), особливо при дослівно перекладеному початковому тексті. Стало відомо, що використання художніх текстів для генерації словників можливе, але не є ефективним.

Ключові слова: словник, обробка, модель, текст, мова, сленг, комп'ютерна система.

Introduction

Lexicography is a very old type of linguistic activity, as evidenced by the first dictionaries that appeared as early as the late twelfth century, to perform informative (allowing to access the accumulated knowledge in the shortest way – via notation) and normative functions (through fixating the meaning and usage of words, they contribute to improvement and unification of language as a means of communication), as well as to reflect the knowledge that society has in a given era. Further development of computer and computing technologies has led to the emergence of a new field in lexicography – computer (or electronic) lexicography, which provides the ability to use lexicographic resources through computers or other electronic devices. In a broad sense, the main object of computer lexicography research is the electronic dictionary [1,2]. Similar terms include "machine dictionaries", "automatic dictionaries", "automated dictionaries", "computer dictionaries" [3]. The basis for the creation of electronic dictionaries are science-intensive methods in the field of linguistic analysis, word processing, lexicographic databases, machine translation (statistical machine translation, neural machine translation, machine translation based on rules, hybrid machine

translation), etc.

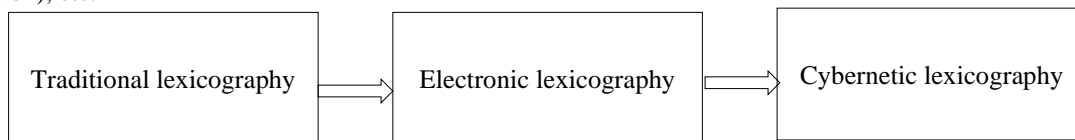


Fig.1. Evolution of human linguistic activity

The wide spread and mass access to the Internet marked the beginning of the development of such field of lexicography as cybernetic lexicography (cyberlexicography) (Figure 1), which involves the use of the Internet to compile a dictionary. Many sources on pedagogical or philological sciences provide detailed classifications of dictionaries. One such example is [4].

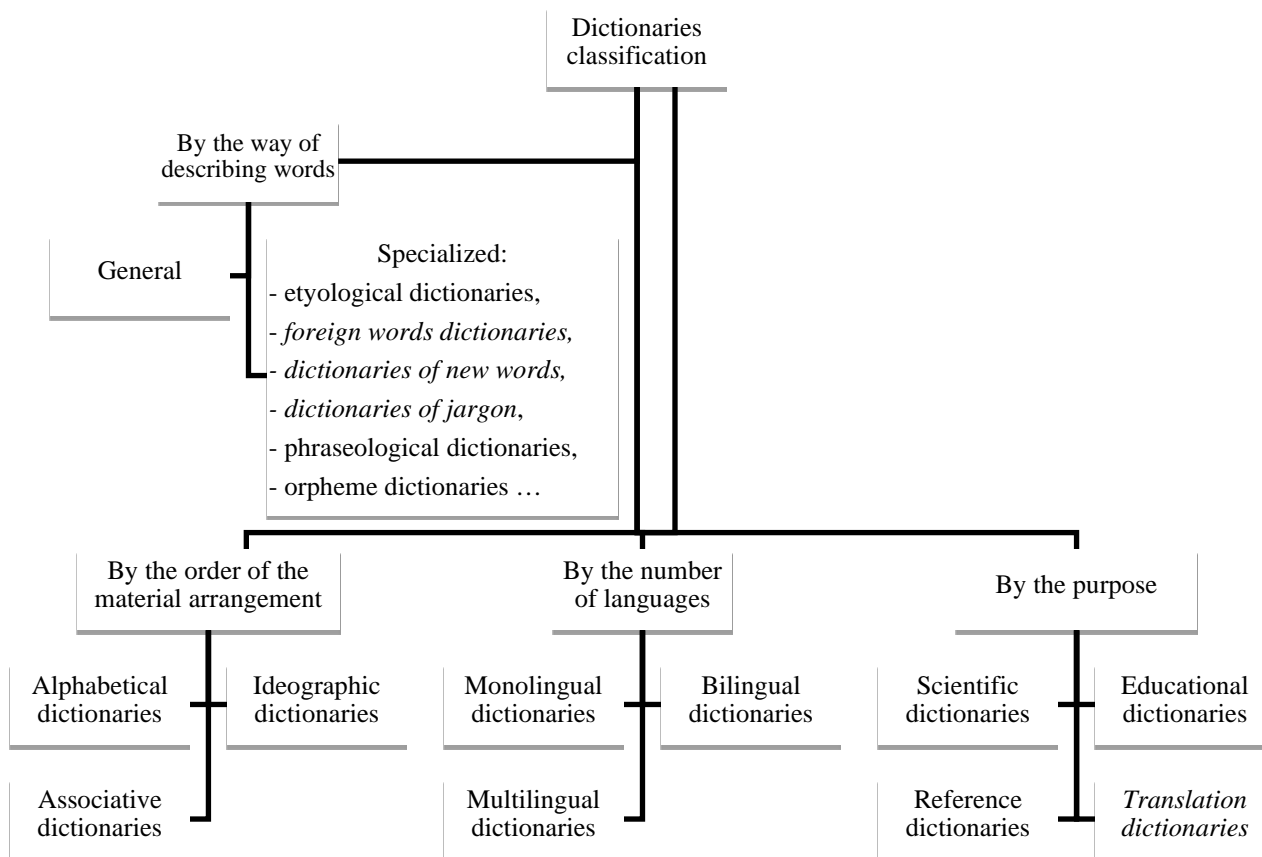


Fig.2. Systematization of dictionaries

The choice of using a particular dictionary depends on the task at hand, such as translating a text or interpreting a term. Thus, translation dictionaries are designed to compare and transition from one language system to another (Ukrainian-Belarusian, English-French, etc.), while special dictionaries, such as regional and private dialects, reveal lexical meanings within a single dialect or dialect. (dictionary of jargon and slang) (Figure 2) [5, 6].

Related works

Nowadays, it takes a lot of time to translate slang or surzhik, as it is difficult to find a specialized translator. All popular translators can only translate into a certain number of languages (Table 1), and it is quite difficult to scale such dictionaries by adding new words for translation.

Table 1

Analysis of the analogues

Comparison criterion	Google translate	Yandex. translator	www.dictionary.com	ABBYY Lingvo
Number of supported languages.	104	95	1	20 (220 vocabularies)
Access	Free.	Free.	Free.	Paid.
Supported platforms	Windows, Linux; separate apps for IOS and Android.	Windows, Linux; separate apps for IOS and Android.	Windows, Linux, Android, IOS	Windows, Android, IOS. Requires separate installation for each platform.
Offline availability	Online only.	Online only.	Online only.	Both online and offline

				modes is available.
--	--	--	--	---------------------

Using the table 1, one can understand that Google translate and Yandex translator have the most advantages, however they only provide the translation from and into official state languages, without supporting the possibility of expanding with new artificially created words and translating slang vocabulary and artificial languages. Given that the number of artificial languages today is very large and constantly growing, new slangs are being generated due the influence of languages on each other and depend on a certain area, the creation of separate dictionaries for each of them is not appropriate. Slang is similar to its language of origin, with the addition or change of a small number of words. However, it is not always convenient to understand and communicate freely with a group of people who have a certain slang. In artificial languages, vocabulary, phonetics and grammar are specially designed to embody certain goals, for example, to communicate with a group of people affected by a particular book, movie, game, and so on. It is purposefulness that distinguishes artificial languages from natural ones [7].

Considering all the above, the creation of an automatic translator of slang and artificial words into official languages is an urgent and demanded task in today's world.

The analysis of scientific works devoted to machine translation shows that to mobile applications for instant translation in order to expand the vocabulary [8], real-time natural language translators for a limited number of foreign languages [9] enough attention is paid. Translators are often a built-in module in NLP systems to determine the mood and emotional state of dialogue participants in different languages [10, 11]. In spite of practical application of translators, the basic algorithms in most proposed are such algorithms and numerical methods as Knut-Morris-Prata [12], TF-IDF, word bag, neural networks, etc. Most studies are conducted for popular languages, such as English, French, Spanish, but make mistakes in the translation of languages that are first provided to the system, or little known. Multilingual translators are promising, where the target language is not only English, but any other language.

The goal of the given work

The growing number and variety of artificial languages leads to the need and relevance of creating automatic dictionaries for their translation in order to facilitate human communication. Such languages include languages where vocabulary, phonetics, and grammar have been specifically designed to achieve specific goals and to communicate with a group of people by interests or place of residence. These languages can be distributed among people of certain professions or among neighboring nations. Examples are slang and surzhik. The common for them is that there is a basic language (literary), the intersection in spelling and meaning of words and phrases with which is quite large.

The object of this study is an automatic dictionary that can be used to translate and interpret slang or artificial vocabulary, as one that is created at the intersection of several languages and has many common (borrowed) spellings and meanings of words from different languages.

The main goal of the project is to create a model of the system of automatic translation of words and texts from / into arbitrary languages, including hybrid, artificial and slang ones.

To achieve this goal, the following tasks must be solved:

- implementation of the possibility to use the prepared dictionary for translation;
- expansion of the vocabulary;
- creating new dictionaries based on existing ones;
- development of approach and creation of inverted dictionaries.

Proposed model

The paper proposes a generalized model of the system of automatic translation of texts from and into arbitrary language (Figure 3). This system primarily provides the ability to translate even from / into artificial languages, such as Esperanto, Talos, Interlingua, Lojban and others.

The given model shows the interaction and partial interdependence of the creation and adjustment modules and the translation module of the dictionary, which is explained by tacking the approach of reverse propagation of the translation error.

The texts submitted to the input of the proposed model must be presented in two arbitrary languages and have the same number of sentences. Sentences are counted by the number of dots ("."), Exclamation mark ("!"), Question mark ("?") And three dots ("..."). In addition, each sentence must have at least one word. For example, if there are only spaces, commas, quotation marks, and other punctuation marks between the exclamation mark and the question mark, this will not be considered a sentence.

In addition, sentences with the same sequence numbers will be considered translations of each other, ie changing the order of sentences will lead to incorrect operation of the program.

Let's explain the work of the module for creating working dictionaries, which will be the basis for the translation of source texts.

Dictionaries use an algorithm that calculates the relative probability of translation based on the location of words in a sentence, the percentage of corresponding sentences in which both words occurred, and their similarity using the trigram method and the Knut-Maurice-Pratt algorithm [13-15].

However, since it is not always possible to find the same text in both selected languages, the paper proposes to build a dictionary through existing ones.

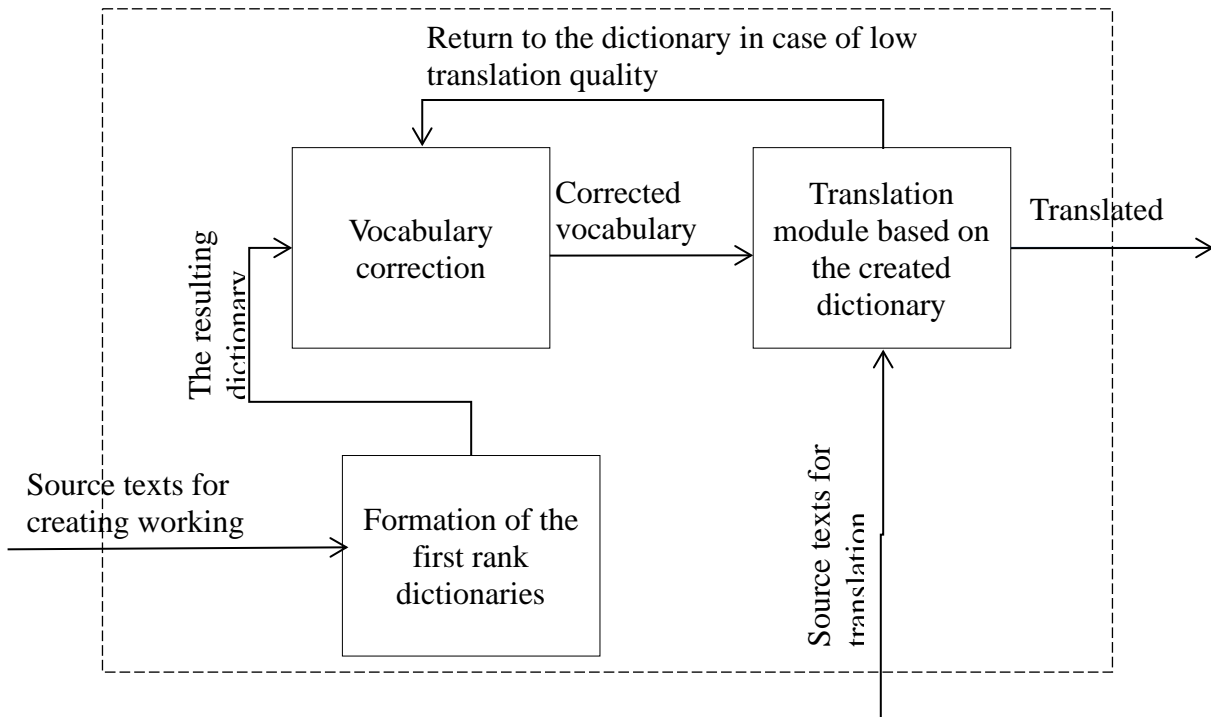


Fig.3. Generalized model of the system of automatic translation of texts from and into arbitrary language

Let languages 1 and 2 be known languages, languages 3 and 4 be slang. In addition, there is a two-way dictionary between languages 1-2, and there is also a two-way dictionary between languages and their slang. Let us mark this on the graph, where each language is a vertex and the dictionary is an edge (Figure 4).

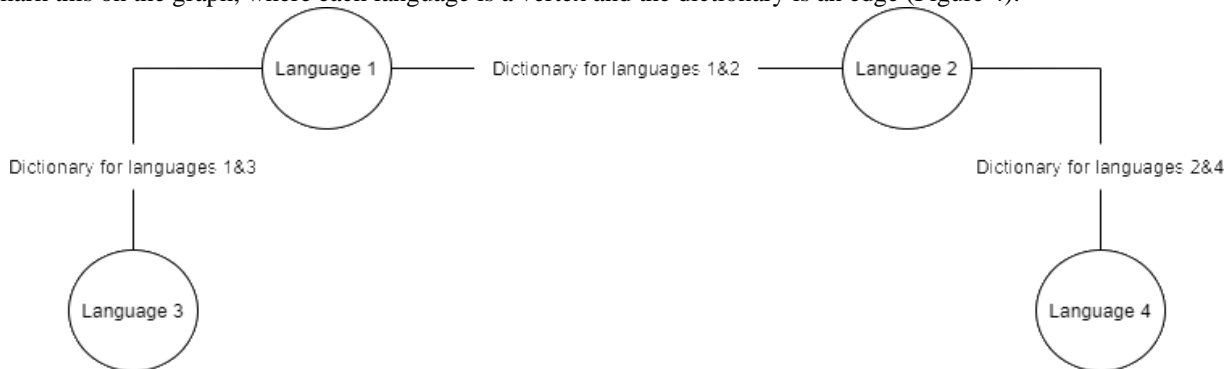


Fig.4. The graph of languages and dictionaries between them in the initial position

Since there are edges between languages 3 and 1 and between 1 and 2, for each word from the 3rd language we will try to find a translation into language 2 using a dictionary from the first to the second language. Since this is not an exact translation, the probabilities of translations will be multiplied by each other, and multiplied by a correction factor k , $0.8 \leq k \leq 1$ ($k = 1$ if all probabilities are less than one). Let's mark the new dictionary between languages 2 and 3 with a dotted line on the scheme (fig. 5).

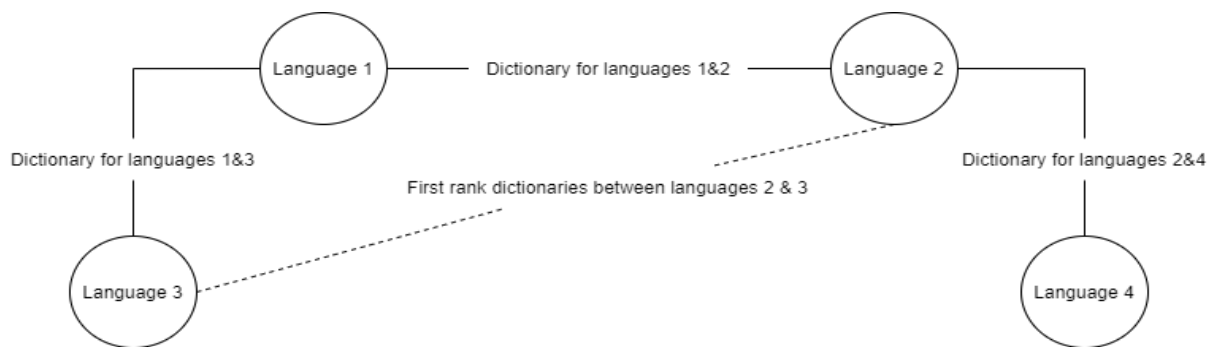


Fig.5. The graph of languages and dictionaries between them after constructing a dictionary from 3 to 2

Using the same idea, we will build a dictionary through languages 3–2–4. Then we get a dictionary with a correction factor $k = 0.64$.

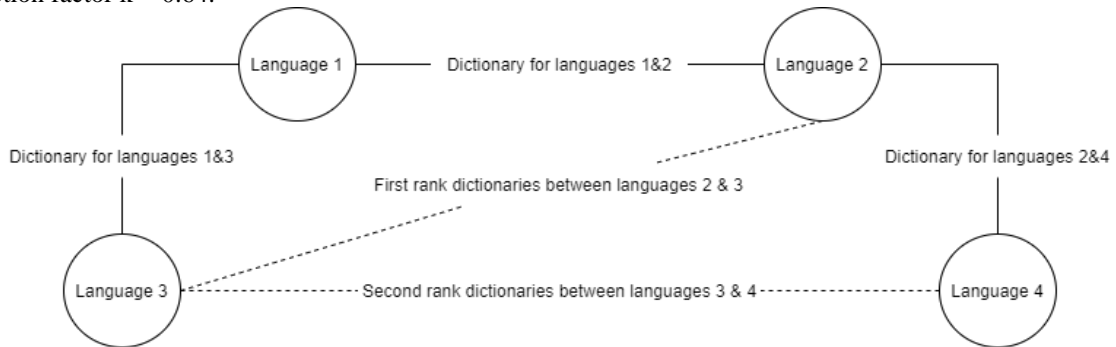


Fig.6. The graph of languages and dictionaries between them after constructing a dictionary from 3 to 4

This dictionary can also be further used to build new ones, but it is not recommended, as dictionaries with rank 3 or higher contain very low probabilities of relative translations, and their use is not appropriate (dictionary rank is defined as the sum of ranks of dictionaries that built it, increased by one) (Figure 6).

To improve the quality of newly constructed dictionaries, it is mandatory and necessary to supplement the finished dictionary with new texts. To do this, in previous iterations it is necessary to save not only the word, possible translations and the probability of these translations, but also the number of processed words and the total size of sentences on which the dictionary was built.

After building the dictionary, we can use it to translate texts. When generating a dictionary, it is possible to build not only a dictionary of word-for-word translations, but also a word-for-phrase and vice versa. In order to choose the best opportunity to translate phrases, the following algorithm was used (Figure 7).

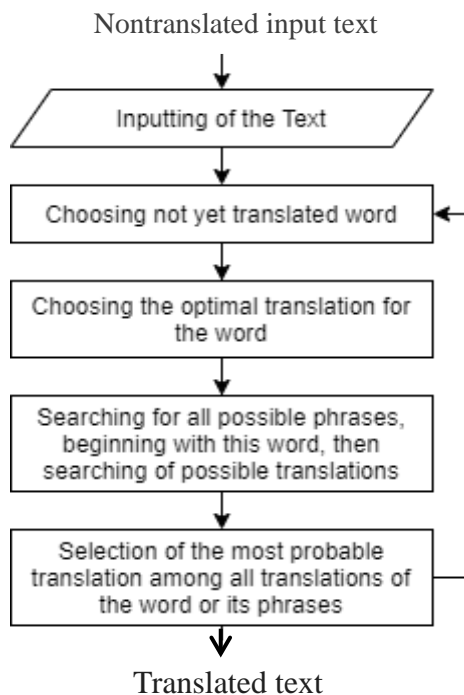


Fig.7. Algorithm for translating text with a phrases dictionary

As a result of the algorithm, the input text will be translated using dictionaries that were submitted to the translator input. The algorithm tries to maximize the total probability of correct translation by maximizing the sum of the probabilities of the relative translation of each word or phrase that have been recorded in dictionaries.

The results of the experiments

To perform experiments and analyze the performance of the proposed approach to the organization of automatic translation of texts from and into arbitrary language, a software application was developed, which includes a subprogram of initial word processing for dictionary organization, one for creating a working dictionary and one for two-way improvement of created dictionary by the inclusion of new texts in order to improve the quality of translation,

including the search for word phrases, idiom, and translation for them, the subprogram of dividing the dictionary into sub-dictionaries with a small percentage of text, the subprogram of the translator itself.

To test and analyze the results of the proposed generalized model, three types of source texts were used:

- literary poetry translation;
- literary prose translation;
- literal prose translation.

In addition, in Esperanto we will replace the ĉ, ĝ, ĥ, ĵ, ŝ, ŭ на c, g, h, j, s, u characters to avoid encoding problems.

For a literal translation of prose, we use the texts shown in Figures 8a and 8b.

Автор эсперанто, Лазарь Лудовик Заменгоф, в 1878 будучи в последнем классе гимназии закончил первый проект язык. И учась в университете, Заменгоф dum протяжении полных семи лет до 1885 года дорабатывал свой язык. Наконец, в 1887 в Варшаве появился первый учебник под заголовком "Международный язык доктора эсперанто". Автор издал учебник не под своим именем, а под псевдонимом "Д-р Эсперанто", т.е. "Тот кто надеется". Теперь и сам язык назван этим псевдонимом.

В 1913 году под руководством известного французского писателя Тристана Бернард эсперанто проэкзаменовали в специальной комиссии. Определённый литературный текст был переведён на многие языки, между прочим также и на эсперанто, а затем тот же текст снова был переведён другим переводчиком, не знающим оригинальный текст, на язык оригинала. Результат показал, что самый совершенный перевод был сделан с помощью эсперанто, потому что текст, переведённый обратно с эсперанто был наиболее близок к оригиналу. Первый всеобщий конгресс эсперанто был организован в Болонье (Франция) в 1905 году.

a)

La autoro de esperanto lazary ludovik Zamenhof en 1878 esan nel laste klaso de gimnazio endi et une projekto del lingvo. Trastudan nel universiteto, Zamenhof dum plene sep yaros ad la yaro 1885 prosesari e sue lingvo. Enden, nel yaro 1887 en Varshava apero la une manuelo sub la titulo "Internacia Lingvo de Doktoro Esperanto". La autoro no eldoni e manuelo sub sue namo, bat sub la pseudonimo "D-ro Esperanto", s_e "Kvele espera". Prezentan la lingvo same es namate segun eze pseudonimo.

Nel yaro 1913 sub la gvido de populare franse skriero Tristan Bernard os exzameni et esperanto en spesiale comiseo. Definite literature texto esi tradukite al multe lingos, inter otre gam al esperanto, ve posten la same texto esi retradukite per otre tradukistos, inkonante et originale texto, al lingvo del originalo. La rezulto montri, ke la plu perfekte tradukos esari farite per helpo de esperanto, kar la retradukite texto from esperanto esi la plu proxime al originalo. La une universale kongreso de esperanto esari organizate en Bolongne-sur-Mer (Fransia) nel yaro 1905.

b)

Fig.8. Corresponding texts in different languages: a) in Russian, б) in Esperanto

After using the program we get the following two dictionaries.

```
1878 : 1878 0.8381 будучи 0.74405 последнем 0.74405 классе 0.68809 в 0.1625 |
1885 : 1885 0.95 до 0.75 года 0.75 лет 0.7 дорабатывал 0.7 |
1887 : 1887 0.88651 варшаве 0.73651 появился 0.67301 наконец 0.60952 в 0.17143 |
1905 : 1905 0.73572 году 0.16667 франция 0.11667 болонье 0.1 в 0.083333 |
1913 : 1913 0.89106 руководством 0.6821 году 0.2 под 0.18667 в 0.17333 |
ad : семи 0.8 полных 0.75 лет 0.75 протяжении 0.7 до 0.7 |
al : эсперанто 0.12957 текст 0.11442 и 0.11333 также 0.11167 прочим 0.11 |
aperi : появился 0.73651 варшаве 0.67301 учебник 0.56032 первый 0.2 под 0.17143 |
autor : автор 0.74564 издал 0.2 лазарь 0.18155 эсперанто 0.1375 учебник 0.09375 |
bat : др 0.74722 псевдонимом 0.6 а 0.175 эсперанто 0.175 именем 0.1625 |
bernard : проэкзаменовали 0.8 бернад 0.6821 специальной 0.6821 эсперанто 0.18667 тристана 0.
bolongnesurmer : франция 0.72857 1905 0.72857 болонье 0.65715 году 0.16667 в 0.15 |
comiseo : комисси 0.14667 специальной 0.13333 проэкзаменовали 0.10667 эсперанто 0.093332 бер
de : эсперанто 0.57168 в 0.32715 первый 0.24187 язык 0.059225 был 0.05797 |
de_esperanto : эсперанто 0.18448 был 0.11594 в 0.10972 лазарь 0.066668 первый 0.065278 |
definite : определённый 0.8 литературный 0.7729 текст 0.18667 был 0.18 переведён 0.17333 |
del : язык 0.48552 оригинала 0.2 оригинальный 0.086667 незнакомим 0.083333 переводчиком 0.08
doktoro : эсперанто 0.15714 доктора 0.14286 язык 0.12857 международный 0.11428 заголовком 0.1
dro : т_е 0.74722 кто 0.74722 надеется 0.69444 тот 0.6 эсперанто 0.175 |
dum : университете 0.75 протяжении 0.7 замэнгоф 0.6 на 0.1875 в 0.175 |
e : под 0.37778 дорабатывал 0.2 года 0.1875 свой 0.1875 не 0.18472 |
eldoni : не 0.8 издан 0.69444 своим 0.69444 учебник 0.1875 под 0.1875 |
en : в 0.46519 эсперанто 0.34362 первый 0.27375 язык 0.057 |
enden : наконец 0.8 1887 0.67301 в 0.18571 варшаве 0.14286 появился 0.12857 |
endi : проект 0.74405 язык 0.2 первый 0.175 закончил 0.1625 гимназии 0.15 |
es : назван 0.8 этим 0.67301 сам 0.54604 язык 0.17143 псевдонимом 0.14286 |
esan : последнем 0.8 классе 0.74405 будучи 0.68809 гимназии 0.68809 1878 0.1625 |
esari : с 0.26154 сделан 0.893 был 0.17464 эсперанто 0.16631 в 0.1 |
esi : текст 0.14193 эсперанто 0.13768 был 0.12609 переведён 0.1 на 0.1 |
espera : надеется 0.69444 кто 0.1625 тот 0.15 т_е 0.1375 эсперанто 0.125 |
esperanto : эсперанто 0.61158 был 0.098575 текст 0.057645 в 0.04792 |
et : язык 0.38643 эсперанто 0.095278 комисси 5e-06 незнакомим 5e-06 |
exzameni : специальной 0.8 комисси 0.74106 проэкзаменовали 0.6821 эсперанто 0.16 бернад 0.1
eze : этим 0.54604 псевдонимом 0.17143 назван 0.11428 язык 0.085715 сам 0.057142 |
farite : помощью 0.8 сделан 0.72642 потому 0.72642 с 0.72107 эсперанто 0.19131 |
franse : тристана 0.8 писателя 0.74106 бернад 0.74106 французского 0.6821 эсперанто 0.17333
fransia : 1905 0.8 франция 0.65715 году 0.18333 болонье 0.15 в 0.13333 |
from : наиболее 0.76321 близок 0.72642 к 0.68963 обратно 0.17391 переведённый 0.16522 |
gam : также 0.8 прочим 0.7729 между 0.74581 и 0.19333 языки 0.18 |
gimnazio : проект 0.8 закончил 0.68809 первый 0.1875 язык 0.1875 гимназии 0.1625 |
guido : известного 0.8 руководством 0.74106 французского 0.74106 писателя 0.6821 под 0.17333
haino : потому 0.8 помощью 0.72642 с 0.6829 эсперанто 0.19131 текст 0.18261 |
k : esi 0.76923 proxime 0.73445 al 0.70769 from 0.70167 texto 0.67692 |
klance : nel 0.8 esan 0.75595 laste 0.75595 1878 0.71191 klaso 0.71191 |
komisii : esperanto 0.75789 exzameni 0.75228 en 0.71579 os 0.70456 et 0.53334 |
kongress : universale 0.8 une 0.74286 kongreso 0.73809 la 0.68571 de 0.68571 |
kto : esperanto 0.8 dro 0.75278 s_e 0.75278 pseudonimo 0.71191 kvele 0.70555 |
lazary : de 0.8 esperanto 0.7619 la 0.72381 lazary 0.71191 autoro 0.56845 |
let : la 0.8 ad 0.75 yaro 0.75 yaros 0.7 1885 0.7 |
literaturnyi : literature 0.8 texto 0.77419 definite 0.77376 esi 0.74839 tradukite 0.72258
lyudovik : esperanto 0.8 de 0.7619 lazary 0.75595 autoro 0.72381 ludovik 0.71191 |
moyu : inter 0.8 lingos 0.77376 otre 0.77376 multe 0.74753 sam 0.74753 |
```

Fig.9. The result of the application. A dictionary for 202 words of the working Russian-Esperanto dictionary has been created

As a result of work among 202 words in both dictionaries 5 corresponding translations for each word with a probability of more than 0.3 were found, and they were arranged in descending order of probability. Translations with a lower probability are considered as noise. Translations with the highest probability are considered as correct translation (Figure 9).

The analysis of the obtained translation result showed that 50% of words in the dictionary have translations with the highest probability and correspond to the actual translation of the word. 90% of words also have their actual translation in the dictionary among the five options offered, although not with the highest probability.

To test the literary translation, we will use the beginning of Bulgakov's novel "The Master and Margarita". We use an excerpt from the text shown in Figures 10a and 10b.

Однажды весной, в час небывало жаркого заката, в Москве, на Патриарших прудах, появились два гражданина. Первый из них, одетый в летнюю серенькую пару, был маленького роста, упитан, лыс, свое приличную шляпу пирожком нес в руке, а на хорошо вышитом лице его помещались сверхъестественных размеров очки в черной роговой оправе. Второй – плечистый, рывчатый, выхристый молодой человек в заломленной на затылок клетчатой кепке – был в ковбойке, жеваных белых брехах и в черных тапочках. Первый был не кто иной, как Михаил Александрович Берлиоз, председатель правления одной из крупнейших московских литературных ассоциаций, сокращенно именуемой МАССОЛИТ, и редактор толстого художественного журнала, а молодой спутник его – поэт Иван Николаевич Понирев, пишущий под псевдонимом Бездонный. Попав в тень чуть зеленых лип, писатели перемогли долго бросились к пестро раскрашенной будочке с надписью "Пиво и вода". Да, следует отметить первую странность этого страшного майского вечера. Не только у будочки, но и во всей аллее, параллельной Малой Бронной улице, не оказалось ни одного человека. В тот час, когда уж, кажется, и сил не было дышать, когда солнце, раскалив Москву, в сухом тумане валилось куда-то за Садовое кольцо, – никто не пришел под липы, никто не сел на скамейку, пуста была аллея.

a)

Foje en Moskvo, dum malnormale varmega printempa sunsubiro, du civitanoj venis la gardenplacon de la Patriarha lageto. La unua, vestita per griza somera kompleto, estis malalta, diketa, kalva, sian malfrivolan capelon li portis en la mano, kaj sur lia bone razita vizaĝo vastis kolosaj okulvitroj en nigra korna muntumo. La dua, largasultra juna viro kun senorda rufeta hararo kaj kvadrata kaskedo sur la nuko, surhavis buntan cemizon, cifitan blankan pantalonon kaj nigrajn sportosuojn. La unua estis neniu alia ol Mihaelo Aleksandrovic Berlioz, redaktoro de dika beletra revuo kaj la prezidanto de Massolit, unu el la plej grandaj Moskvaj literatur'asocioj, lia akompananto estis la juna poeto Ivano Nikolaic Ponir'ov, verkanta sub la pseudonimo Senhejmlulo. Atinginte la ombren de apenau verdigintaj tilioj, la literaturistoj tuj jetis sin al bunta budo kun surskribo "Biero kaj Trinkakoj". Nu, oni notu la unuan strangacon de tiu terura maja vespero. Ne nur antau la budo, sed en la tuta aleo paralela al la Malgranda Bronnaja strato, ne estis ec unu homo. Je horo, kiam oni apenau povis spiri la varmegan aeron, kiam la suno, ardiginte Moskvon, en seka nebulo estis falanta ien trans Sadovaja-strato, neniu venis sub la tiliojn, neniu sidigis sur benkon, malplena estis la aleo.

b)

Fig.10. Corresponding texts in different languages: a) excerpt from the novel in the original language; b) an excerpt from the novel translated into Esperanto

As a result, we obtain the following dictionaries (Figure 11).



Fig.11. The result of the application

As a result of work among 304 words in both dictionaries 5 corresponding translations for each word with a probability of more than 0.3 were found, and they were arranged in descending order of probability. Translations with a lower probability are considered as noise. Five such words were found. Translations with the highest probability are considered as a correct translation.

The analysis of the obtained translation result showed that 15% of words have translations in the dictionary with the highest probability and correspond to the actual translation of the word. 40% of words also have their actual translation in the dictionary among the five options offered, although not with the highest probability (Figure 11).

Such a low result was obtained because words that occur only once are taken as the most probable translations of those that stand in their place in another language. However, among non-noise words that occur two or more times, more than 40% have the most probable correct translation, and more than 80% contain the correct translation among those suggested. In conclusion, if a literary translation of the text is to be used, a large amount of original text is required.

To test the translation of poetry, we use the text shown in Figures 12a and 12b.

As a result, we obtain the following dictionaries (Figure 13).

As a result of work among 181 words in both dictionaries 5 corresponding translations for each word with a probability of more than 0.3 were found, and they were arranged in descending order of probability. Translations with a lower probability are considered as noise. Only 1 noise word appeared in the given text. Translations with the highest probability are considered as a correct translation.

The analysis of the obtained translation result showed that 25% of words in the dictionary have translations with the highest probability and correspond to the actual translation of the word. 60% of words also have their actual translation in the dictionary among the five options offered, although not with the highest probability. (Figure 13).

Among all three tested variants, the best among the words that occur only once is a literal translation of the

text. Words that are used more than once are more likely to be translated. In addition, all the words that were considered noise were in Esperanto. This may be due to the specifics of the language, as this language contains non-translatable words, such as articles.

Над землёю новым чувством веет,
И призыв разносится по свету;
Он на крыльях, словно лёгкий ветер,
Облетает быстро всю планету.
Не мечом – орудьем разрушенья –
Он сплотить желает все народы:
Людям, утомившимся в сраженьях,
Обещает мир он и свободу.
Все, кто верит в счастье дней грядущих,
Собрались под мирные знамена,
Под звездой идеи всемогущей
Стали мы сильнее и сплочённой.
Стены недоверья вековые
Разобшили страны и народы;
Но сметет препятствия любые
Светлая любовь – сестра свободы.
Овладевши языком общенья,
Хорошо друг друга понимая,
Будут жить народы, без сомненья,
Как семья единая земная.
Полные энергии и страсти,
Не устанем яростно трудиться,
Чтоб мечта, которой нет прекрасней,
Для людей могла осуществиться.

a)

En la mondon venis nova sento,
tra la mondo iras forta voko;
per flugiloj de facila vento
nun de loko flugu gi al loko.
Ne al glavo sangon soifanta
gi la homan tiras familion:
al la mond' eterne militanta
gi promesas sanktan harmonion.
Sub la sankta signo de l' espero
kolektigas pacaj batalantoj,
kaj rapide kreskas la afero
per laboro de la esperantoj.
Forte staras muroj de miljaroj
inter la popoloj dividitaj;
sed dissaltos la obstinaj baroj,
per la sankta amo disbatitaj.
Sur neutrala lingva fundamento,
komprenante unu la alian,
la popoloj faros en konsento
unu grandan rondon familian.
Nia diligenta kolegaro
en laboro paca ne lacigos,
gis la bela songo de l'homaro
por eterna ben' efektivigos.

b)

Fig.12. Corresponding texts in different languages: a) excerpt in Russian; b) an excerpt translated into Esperanto

Texts in the following pairs of languages were used to test the translation in Cyrillic languages: Russian and Ukrainian; Ukrainian and Old Ukrainian; Ukrainian and Russian-Ukrainian surzhik; Ukrainian and Russian-game surzhik.



Fig.13. The result of the application

The efficiency and example of the application is demonstrated for the Ukrainian-Old Ukrainian language pair, for which Taras Shevchenko's work "On Easter Day Upon the Straw" was used in the original and in an adapted form to generate a dictionary. The text is shown in Figures 14a and 14b.

На Великдень, на соломі Проти сонця, діти Грались собі крашанками, Та й стали хвалитись Обновами. Тому к святкам З лиштвою пошили Сорочечку, а тій стьожку, Тій стрічку купили; Кому шапочку смушеву, Чобітки шкапові; Кому свитку. Одна тільки Сидить без обнови Сиріточка, рученята Сховавши в рукава. - Мені мати купувала. - Мені батько справив. - А мені хрещена мати Лиштву вишивала. "А я в попа обідала!" Сирітка сказала.	На Великдень, на соломі Проти сонця, діти Грались собі крашанками, Та й стали хвалитись Обновами. Тому до свят З вишиванкою пошили Сорочку, а тій смужку, Тій стрічку купили; Кому шапочку хутряну, Чобітки з шкіри; Кому сорочку. Одна тільки Сидить без обнови Сиріточка, рученята Сховавши в рукава. - Мені мати купували. - Мені батько зробив. - А мені хрещена мати Сорочку вишивала. "А я в попа обідала!" Сирітка сказала.
a)	b)

Fig.14. Texts for dictionary generation: a) original poem by Taras Shevchenko, б) adapted for dictionary creation

After the program, we will delete the same words and get the result shown in Figure 15.

```

к : до 0.8 тому 0.7619 свят 0.7619 з 0.72381 вишиванкою 0.69571 |
лиштвою : вишиванкою 0.82 пошили 0.79524 з 0.7619 свят 0.74881 сорочку 0.7381 |
лиштву : сорочку 0.81429 мати 0.71071 вишивала 0.71071 хрещена 0.57143 |
свитку : шкіри 0.7819 чобітки 0.71429 хутряну 0.6619 шапочку 0.6381 кому 0.59643 |
святкам : свят 0.92857 до 0.7619 з 0.7619 тому 0.74881 вишиванкою 0.73381 |
смушеву : хутряну 0.81429 шапочку 0.77619 кому 0.77381 чобітки 0.7619 шкіри 0.70571 |
сорочечку : сорочку 0.91587 пошили 0.77857 а 0.7619 вишиванкою 0.73381 тій 0.72381 |
справив : зробив 0.83333 батько 0.61667 мені 0.4 |
стьожку : смужку 0.86429 тій 0.79524 стрічку 0.75238 а 0.72381 купили 0.71905 |
тільки : тільки 0.93333 одна 0.745 сидить 0.73667 без 0.64 обнови 0.57667 |
шкапові : шкіри 0.8019 чобітки 0.77619 шапочку 0.74286 хутряну 0.72381 кому 0.67262 |
    
```

Fig.15. The result of the application after deleting the same words

The result is eleven words, each of which is correctly translated and ten of which have a probability of translation greater than or equal to 0.8.

Conclusion

A model of the system of automatic translation of words and texts from / into arbitrary languages, including hybrid, artificial and slang ones was proposed. For analysis of the performance of the proposed approach to the organization of automatic translation system using the developed software application, the results were summarized, as shown in table 2.

Table 2

Summary for the level of translation

Type of tested text	The percentage of words that have the most probable translation as their actual one	Percentage of words that have a valid translation among the five proposed.
A literal translation of prose between Russian and Esperanto	40,6%	90%
Literary translation of prose between Russian and Esperanto	15%	39,7%
Literary translation of poetry between Russian and Esperanto	25,2%	74,8%
Literal translation of prose between Russian and Ukrainian	98,8%	98,75%
A literal translation between Ukrainian and Old Ukrainian	98,65%	98,7%
Translation between Ukrainian and Ukrainian-Russian surzhik	98,72%	98,84%

Translation between Ukrainian and Russian-game surzhik	50,2%	70%
--	-------	-----

Testing has shown that the proposed approach provides a high level of translation (up to 98,8%) in similar languages, especially with a literally translated source text. At least three different sentences with separate words are needed for correct translation. It has become known that the use of artistic texts to generate dictionaries is possible, but not very effective.

The results of the experiments also showed that the size of the source texts has a significant impact on the quality of the constructed dictionaries – the larger the volume of the source pairs of texts, the better the dictionary is generated. This applies not only to fiction, but to all texts in which the order of words in a sentence has been changed. Some language pairs may not be compatible if their rules have a different word order, ie the efficiency of dictionaries between languages such as English-Ukrainian or Esperanto-Ukrainian will be lower than between Ukrainian and Russian.

The developed test application demonstrated a fairly good translation result (up to 98,8%) between similar languages, such as Ukrainian-Russian, or Ukrainian - Ukrainian-Russian surzhik.

References

1. V. Ryzhkova, "Possibilities of Computer Lexicography in Compiling Highly Specialized Terminological Printed and Electronic Dictionaries (Field of Aviation Engineering)," *2020 Ivannikov Memorial Workshop (IVMEM)*, 2020, pp. 40-42, doi: 10.1109/IVMEM51402.2020.00013.
2. B. Ranaivo-Malançon, S. Saeed and J. F. Wilfred Busu, "Discovering linguistic knowledge by converting printed dictionaries of minority languages into machine readable dictionaries," *2014 International Conference on Asian Language Processing (IALP)*, 2014, pp. 140-143, doi: 10.1109/IALP.2014.6973522.
3. G.R.Chumarina. Classification of electronic dictionaries in modern lexicography and lexicologists and features of their use [Electronic resource] *Baltic Humanitarian Journal*. 2013. №4. C.123-126.
4. L.V.Popova. Typologies and classifications of dictionaries. *Bulletin of the Chelyabinsk State University*. 2012. №20 (274). *Philology. Art history*. № 67. C.106-113
5. K. Manuel, K. V. Indukuri and P. R. Krishna, "Analyzing Internet Slang for Sentiment Mining," *2010 Second Vaagdevi International Conference on Information Technology for Real World Problems*, 2010, pp. 9-11, doi: 10.1109/VCON.2010.9.
6. F. Ren and K. Matsumoto, "Semi-Automatic Creation of Youth Slang Corpus and Its Application to Affective Computing," in *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 176-189, 1 April-June 2016, doi: 10.1109/TAFFC.2015.2457915.
7. Kazakov D. Artificial naturalness // *Science and life*. – 2017. – № 10. – С. 100–107.
8. C. Kongcharoen, S. Prasunpaengsri and T. Wongmeekaew, "Investigating applications of mobile translator for assisting english vocabulary learning," *2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media)*, 2017, pp. 1-5, doi: 10.1109 / UMEDIA.2017.8074120.
9. A. Santosa et al., "The Architecture of Speech-to-Speech Translator for Mobile Conversation," *2019 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA)*, 2019, pp. 1-6, doi: 10.1109 / O-COCOSDA46868.2019.9041196.
10. D. Mahajan and D. Kumar Chaudhary, "Sentiment Analysis Using Rnn and Google Translator," *2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2018, pp. 798-802, doi: 10.1109 / CONFLUENCE.2018.8442924.
11. A. Amin, I. Hossain, A. Akther and K. M. Alam, "Bengali VADER: A Sentiment Analysis Approach Using Modified VADER," *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019, pp. 1-6, doi: 10.1109 / ECACE.2019.8679144.
12. D. Angreani, D. P. I. Putri, A. N. Handayani and H. Azis, "Knuth Morris Pratt Algorithm in Enrekang-Indonesian Language Translator," *2020 4th International Conference on Vocational Education and Training (ICOVET)*, 2020, pp. 144-148, doi: 10.1109 / ICOVET50258.2020.9230139.
13. S Zaiceva and O Barkovska, Analysis of Accelerated Problem Solutions of Word Search in Texts, *The Fourth International Scientific and Technical Conference «COMPUTER AND INFORMATION SYSTEMS AND TECHNOLOGIES»*. Kharkiv: NURE. 2020. – p.66 DOI: <https://doi.org/10.30837/IVcsitic2020201445>
14. Olesia Barkovska, Oleg Mikhal , Daria Pyvovarova , Oleksii Liashenko , Vladyslav Diachenko and Maxim Volk, Local Concurrency in Text Block Search Tasks, *International Journal of Emerging Trends in Engineering Research*. - Volume 8. No. 3, March 2020. – P.690-694.
15. Barkovska O., Pyvovarova D. and Serdechnyi V., Pryslorenyj alghorytm poshuku sliv-obraziv u teksti z adaptivnoju dekompozycijeju vykhidnykh danykh. [Accelerated word-image search algorithm in text with adaptive decomposition of input data]. *Systemy upravlinnja, navigacijy ta zv'jazku* 4 (56), 28-34. (in Ukrainian)

Olesia Barkovska Олеся Барковська	Ph.D., Associate Professor of Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine e-mail: olesia.barkovska@nure.ua ; ORCID ID: 0000-0001-7496-4353, Scopus Author ID: 24482907700, https://scholar.google.ru/citations?hl=ru&user=Uj96xp4AAAAJ	доцент к.т.н., доцент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна
Anton Havrashenko Антон Гаврашенко	Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine e-mail: anton.havrashenko@nure.ua	магістрант, Харківський національний університет радіоелектроніки, Харків, Україна
Vladyslav Kholiev Владислав Холєв	Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine; e-mail: vladyslav.kholiev@nure.ua ; ORCID ID: 0000-0002-9148-1561	аспірант, Харківський національний університет радіоелектроніки, Харків, Україна
Olena Sevostianova Олена Севостьянова	Assistant Lecturer of Department of Electronic Computers, Kharkiv national university of radio electronics, Kharkiv, Ukraine e-mail: olena.sevostianova@nure.ua .	асистент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна