

Comparative Analysis of Using Different Parts of Speech in the Ukrainian Texts Based on Stylistic Approach

Alina Dmytriv¹, Svitlana Holoshchuk¹, Lyubomyr Chyrun² and Roman Holoshchuk¹

¹ Lviv Polytechnic National University, S. Bandera Street, 12, Lviv, 79013, Ukraine

² Ivan Franko National University of Lviv, University Street, 1, Lviv, 79000, Ukraine

Abstract

The work aims to analyse words of different parts of speech in Ukrainian texts to identify the speaker's purpose of using certain parts of speech to express his opinion fully. Our analysis enables better recognition of written texts and the flow of the author's thoughts by considering words of different parts of speech. In work, such analogous systems as Intelligent Ukrainian Text Processing System, Large Electronic Dictionary of the Ukrainian Language (VESUM), lang-uk microservices, NER annotated corpus, and tonal dictionary of the Ukrainian language is considered. The system is designed by incorporating use-case, states and activities diagrams together with program implementation tools such as Python, MySQL and Tkinter. In addition, the software which analyses Ukrainian texts and calculates the frequency of words of different parts of speech is presented. It also demonstrates the results of frequency comparison of other parts of speech based on texts of different styles and then creates diagrams showing its statistics.

Keywords

Ukrainian language, morphological analysis, parts of speech, Ukrainian text

1. Introduction

Recognition of Ukrainian-language texts is only on the initial phase of its development [1-6]. Since the Ukrainian language belongs to a synthetic group of languages (i.e. syntactic relations within sentences are expressed by inflexion), it complicates the automatic detection and correction of errors, automated analysis and synthesis of oral speech, automatic translation and more [7-14]. The mentioned issues need ways of creating solutions that stimulate a wide range of possible research in this area [15-20]. Modern investigations focus on developing available tools for Ukrainian language recognition [21-29]. Most developers perform this work on a volunteering basis and provide online access to their libraries [15-20, 30-33]. They aim to enable every interested person to be involved in this project. The Ukrainian language takes 16th place among the most popular language on Wikipedia and trails behind with the 32nd on the Internet. Developing high-quality Ukrainian-language NLP programs is relevant and needed in Ukraine [34-42]. A large amount of untranslated professional literature, which makes Ukrainians read everything in original and creates difficulties for some of them, is another factor that fosters research in this area. In addition, such programs would provide an opportunity to analyse Ukrainian social networks and the media, facilitating faster and greater identification of helpful information and various problems [15-20]. Therefore, providing free access to the created libraries for NLP of the Ukrainian language is a significant process. It gives extra information for developing computational linguistics of the Ukrainian language.

Modern programmes developed for the analysis of the Ukrainian text do not cover the whole spectrum of the problems in the field [1-7]. An important question associated with better machine

COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022, Gliwice, Poland
EMAIL: alinadmtriv@gmail.com (A. Dmytriv); svitlana.l.holoshchuk@lpnu.ua (S. Holoshchuk); Lyubomyr.Chyrun@lnu.edu.ua (L. Chyrun); roman.o.holoshchuk@lpnu.ua (R. Holoshchuk)
ORCID: 0000-0003-0141-6617 (A. Dmytriv); 0000-0001-9621-9688 (S. Holoshchuk); 0000-0002-9448-1751 (L. Chyrun); 0000-0002-1811-3025 (R. Holoshchuk)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

recognition of the Ukrainian language is to develop new programmes and improve the existed ones based on the analysis and modification of relevant analogues for the work of English texts [43-62].

The purpose of the work is to analyse words of different parts of speech in Ukrainian-language texts to determine whether the author prefers to use certain parts of speech to express his opinion fully.

The research object is the construction analysis of the Ukrainian texts considering words of different parts of speech. The research subject is the analysis of the frequency and sequence of words used from different parts of speech in Ukrainian-language texts. The analysis can help better understand the writing of Ukrainian-language texts and the presented course of the author's thoughts by using words of different parts of speech. The stylistic analysis based on statistic analysis and machine learning technology can become a part of the statistical and semantic analysis of the Ukrainian texts, automatic translation, construction of N-grams, templates of the structure of the sentences, etc. [63-68].

2. Related Works

Analogues' characteristics, advantages, and disadvantages are searched and analysed to explore the mentioned question. It includes the following resources and systems, which are considered and described in detail [15-20]:

- Intelligent system of Ukrainian Text Processing;
- Large electronic dictionary of the Ukrainian language (VESUM);
- Microservices lang-uk;
- Corpus of NER-annotations;
- Tonal dictionary of the Ukrainian language.

2.1. Intelligent system of Ukrainian Text Processing

The intelligent system of Ukrainian text processing was developed in 2019 at the Taras Shevchenko National University of Kyiv [15]. It implements specific linguistic tasks related to the processing of the Ukrainian language, namely the preliminary processing of the text, morphological and lexical analysis of the text. To create such a system, the authors analyse the available means of text processing in natural language and consider using them in the Ukrainian language. In this system, for preliminary analysis of the Ukrainian text, NLTK library tools are selected. They are used for tokenisation, detection of sentence boundaries, deletion of non-text elements (tags, meta-information), e-mail highlighting, file name selection, compilation, words written with an interval between letters, removal of stop words, recognition of nominal entities. Regarding the morphological analysis and determination of words belonging to specific parts of speech, this program gives the following characteristics: lemma, stem, part of speech, cases, gender, number, being / non-being, time, person and species.

The advantage of the developed system is that it provides opportunities for the complex processing of Ukrainian texts at three levels of analysis. The disadvantages of the developed system are suboptimal algorithms for composing words written with spaces between letters (the algorithm uses a dictionary to identify words written with spaces between letters). A length limit is introduced to reduce the words search, and the words more extensive than the boundary conditions may contain errors. Also, it may not consider an incorrect analysis of non-vocabulary words, which is a disadvantage of pymorphy2, based on the morphological module of the system and simple lexical analysis.

2.2. Large electronic dictionary of the Ukrainian language

The large electronic dictionary of the Ukrainian Language (VESUM) is a dictionary of word changes in the Ukrainian language. Its main components are the register, the code of word change classes and rules for generating word forms based on these codes and using elements of program logic [16]. Consider an example of the representation of words lem in this dictionary. The following code represents the word *близнюк* [twins]: *близнюк /n20.a.p.ke.<*, where *n20* are nouns of the second declension of the masculine gender, the key *a* is the ending *-a* in the genitive singular, the key *p* is the plural form, the key *ke* is the ending *-e* in the accusative case.

VESUM performs the tasks of morphological analysis and synthesis. The first is to lemmatise (reduce a single word form to a lemma) and assign the appropriate grammatical tags. The second one involves generating all word forms from a particular lemma with the proper grammatical features-tags.

Distinctive characteristics of VESUM [16]:

1. Machine-readable format;
2. Open project;
3. Dynamic nature;
4. Large registers (more than 401 thousand lemmas, from which more than 6 million-word forms are generated);
5. Exhaustive coverage of proper names (almost 53 thousand);
6. Representation of twists (nearly 8 thousand), abbreviations, slangs, rarely used words;
7. Division of vocabulary into 13 word-changing classes, which partially coincide with the traditional parts of speech;
8. The dictionary does not contain accents;
9. Compact system of declension codes and tags for words;
10. Conjugation of complex names;
11. Replacement options for twists;
12. Information on case management;
13. Representation of homonyms.

The large electronic dictionary of the Ukrainian language is the basis for building a linguistic analysis of texts of the Ukrainian language. It is already used to check spelling, grammar and style, construct word vectors, implement a full-text search, and create a corpus. In addition, it can be used for compiling various types of dictionaries, linguistic research, development in the field of computational linguistics, reference functions, etc.

2.3. Microservices lang-uk

Microservices lang-uk can efficiently run and use the essential tools developed by the lang-uk team [20]. Technically, this is implemented using Swagger and Docker technologies.

At the moment, there are the following services:

- Tokenisation;
- Ukrainian, Russian and English NER - allows NER-marking of token text using models trained with the MITIE library for Ukrainian, Russian and English. The microservice was developed by Mykhailo Chaly;
- Lemmatization – using the capabilities of the nlp-uk library, it is based on the NLP_UK library. It allows lemmatising the input text according to the dict_uk dictionary, including its tokenisation. The microservice was developed by Andriy Rysin;
- Language recognition using the capabilities of the WILD library - allows you to recognise the language of input text from a list of 156 languages used on the Internet, using the library wiki-lang-detect.

The lang-uk-ms project allows you to run all microservices simultaneously and access them through a web interface. The web service (Docker image) was created to carry out such operations of processing Ukrainian texts as an assessment of coherence of the text (the size of the text should make at least three sentences); selection of nominal groups; search for co-reference pairs.

2.4. Corpus of NER-annotations

The corpus of NER-annotations contains 229 texts from the Ukrainian Brownian corpus for 217,381 tokens out of 6,751 marked named entities [18]. NER is a name that indicates a unique entity. These include names of persons, places, organisations, works, websites, etc. It consists of one or more words. Any NER entity must contain at least one word with a capital letter or be written in another language (exceptions are the cases with an error, or the entire text is reduced to one register). But some entities may also contain lowercase words [19]. Total in the case: 229 texts; 217381 tokens; 6751 NER entities.

2.5. Tonal dictionary of the Ukrainian language

The tonal dictionary of the Ukrainian language contains 3442 words of the Ukrainian language, which have a non-neutral tone (-2, -1, 1, 2) [17]. Data are obtained from two sources:

- The file tone-dict-uk-manual.tsv is obtained by averaging the assessments of several experts;
- The tone-dict-uk-auto.tsv file is generated by automatically expanding the tone-dict-uk-manual dictionary using the ML model applying the word vectors word2vec and lex2vec and minor post-processing by humans.

The data format is tab-separated with the following columns - word and discrete key (from the range: -2, -1, 0, 1, 2). If possible, all words are reduced to the basic grammatical form in the dictionary, and common root adjectives replace adverbs.

3. Materials and Methods

3.1. UML-diagrams

UML diagrams are constructed for the selected topic, namely use-case, states and activities charts are created. Now let's take a closer look at each diagram separately. The diagram of variants of use is presented in Fig. 1. It contains the following actors:

- User – a person who uses the system to analyse the Ukrainian text;
- VESUM database – a large electronic dictionary of the Ukrainian language, which contains word changes of the Ukrainian language, the register, the code of word change classes and the rules for generating word forms based on these codes;
- System database – a created database for the developed system.

Then this diagram contains the following use-cases:

- Analyse text - performs the primary function of the system. The user enters the text and begins to analyse the entered text according to the following processes;
- Check the text language – checks the entered text (whether the text is written in Ukrainian);
- Tokenize – selects words from the entered text;
- Perform morphological analysis - a process that determines part of speech of all text words;
- Build sentence schemes – according to certain parts of speech it builds schemes of all sentences in the entered text;
- Calculate the frequency of parts of speech in sentences – the frequency of occurrence of grammatical conversion is calculated in each sentence separately, and then all the calculated frequencies are summed;
- Calculate the frequency of combinations of grammatical conversion – the frequency of combinations is calculated in each sentence separately, and then all the calculated frequencies are summed;
- Save everything in the system database – all the results of the calculated frequencies and constructed schemes are saved in the system database;
- Display the result of the analysis – displays the obtained result, namely graphs of frequencies of parts of speech and frequencies of word combinations of grammatical conversion, as well as the most common sentence schemes in the text;
- Build graphs – designs graphs of frequency occurrence of speech parts and frequencies of word combinations of different parts of speech based on obtained data from the system database;
- Get frequency results – send a query to the database to get the calculated frequencies of different parts of speech and frequencies of word combinations of other parts of speech;
- Display sentence schemes – identifies the most commonly used sentence schemes, which are built based on sentences from the entered text;
- Get constructed diagrams – send a query to the database to obtain created diagrams of the text sentences.

The state diagram is presented in Fig. 2. This diagram shows the transition of the system from one state to another. The activity diagram is illustrated in Fig. 3. This diagram shows the user's transition from one activity to another when using the system.

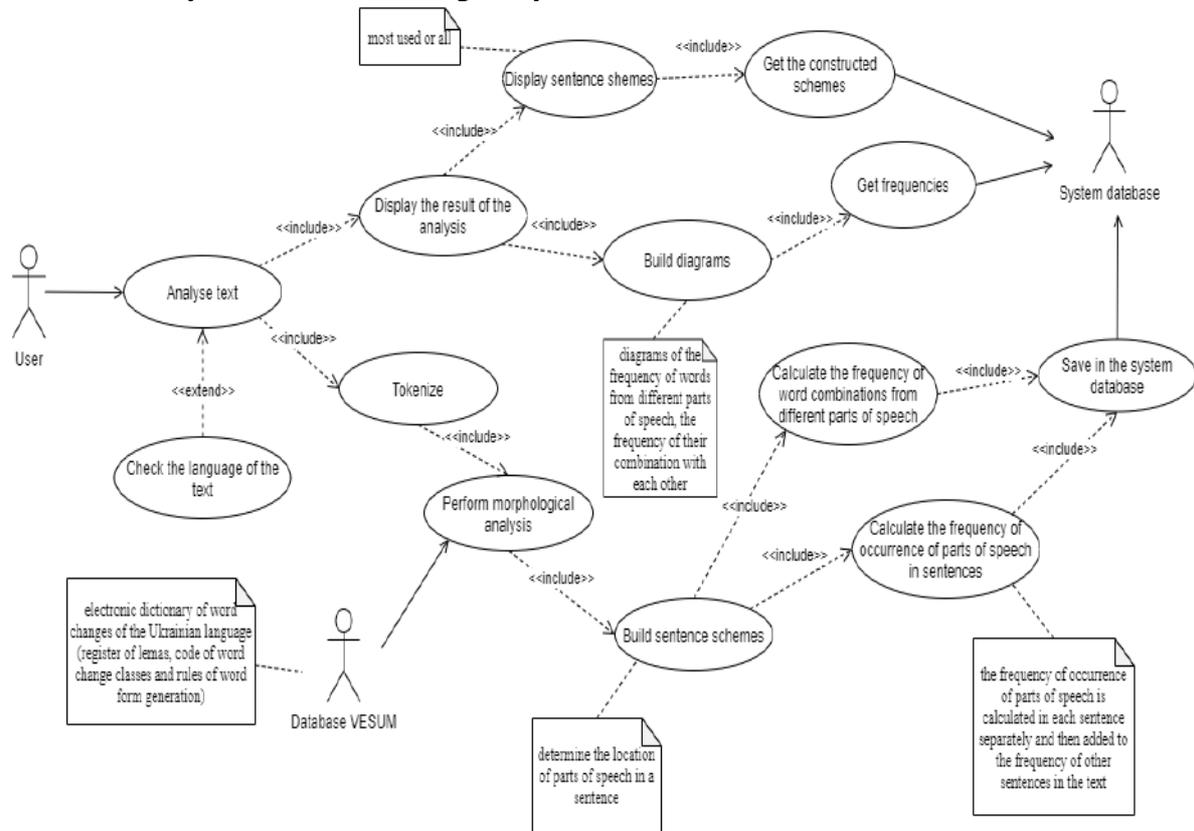


Figure 1: Use-case diagram

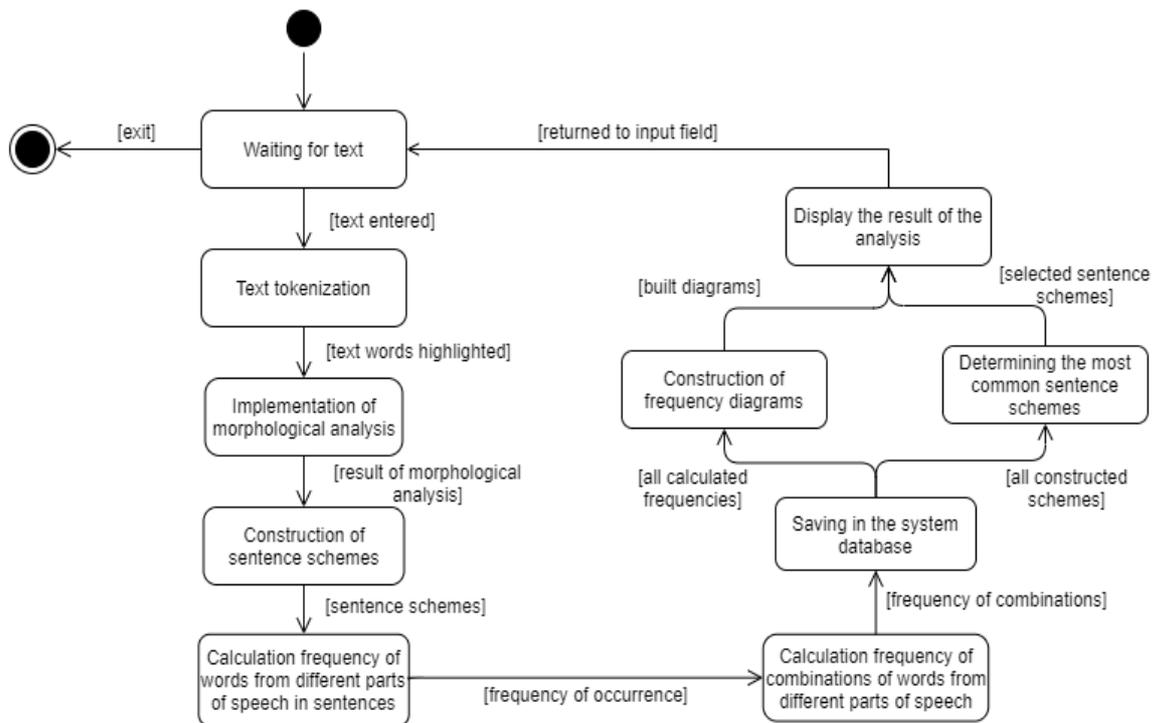


Figure 2: State diagram

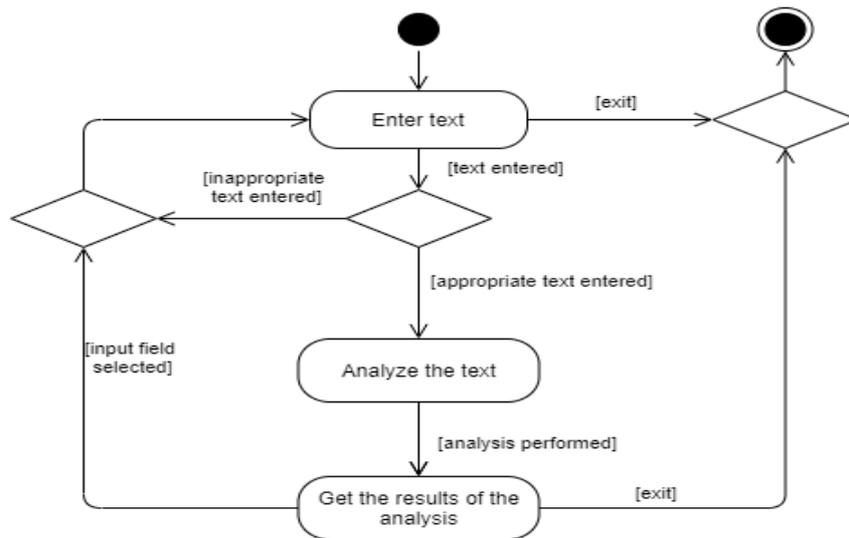


Figure 3: Activity diagram

3.2. Description of the basic functionality

First, the user sees the program window with empty fields. There is a field for entering text and a button to start the analysis of the text. If the text is not entered in Ukrainian, the program clears the field, notifies the user of the error, and asks him to enter the text again. Immediately after the analysis, the bottom panel displays two frequency graphs and diagrams of the most commonly used sentence schemes. Next, if the user continues the text analysis by entering the new text and pressing the button, the bottom panel is cleared and filled with new graphs and diagrams. And so on until the user finishes work and closes the program window. It should be noted that all the analysis results are stored in the system database.

3.3. Means of implementation

The system is developed using the following tools:

- UI elements (Tkinter);
- MySQL Database;
- Back-end (Python).

4. Experiments

The structure of the software contains the following files:

- *database.py* contains all the necessary arrays of letters for morphological analysis;
- *wordInfo.py* has functions for obtaining the basic morphological information (part of speech, word form and its case);
- *tableMorphAnalysis.py* includes operations for creating a graphical interface with morphological analysis of words from the text;
- *wordFrequency.py* contains functions for subtracting the frequencies of different parts of speech in the entered text and building graphs for them.

5. Results

A developed program analyses words of different parts of speech in Ukrainian texts. The program builds sentence schemes, which represent the sequence of used parts of speech and displays a list of

frequencies of occurrence of speech part in the text. Below you can find a Ukrainian text with an explanation of how the programme analyses it:

Темної ночі, коли за вікном проходять міські вечори (маєток злетів на гору й царить над містом), коли сині димки здіймаються над цегельнею й обивателі, як миші,— за підвороття, у канареечний замок, темної ночі в мойому надзвичайному кабінеті збираються мої товариші. Це новий синедріон, це чорний трибунал комуни.

[Temnoi nochi, koly za viknom prokhodiat miski vechory (maietok zletiv na horu y tsaryt nad mistom), koly syni dymky zdiimaiutsia nad tselhelneiu y obyvateli, yak myshi,— za pidvorittia, u kanareiechnyi zamok, temnoi nochi v moiomu nadzvychainomu kabinetі zbyraiutsia moi tovaryshi. Tse novyi synedriion, tse chornyi trybunal komuny].

First, you enter the text in the field and then wait for the programme to start. The execution time might be long, as the program has to process an extensive number of words and build appropriate graphics. The results of the text processed by the program are presented in Figs. 4-7.

```
Text:
Темної ночі, коли за вікном проходять міські вечори (маєток злетів на гору й царить над містом), коли сині димки здіймаються над цегельнею й обивателі, як миші,— за підвороття, у канареечний замок, темної ночі в мойому надзвичайному кабінеті збираються мої товариші. Це новий синедріон, це чорний трибунал комуни.

Sentences schemes in the text:

Sentence scheme 1
3 1 9 7 1 2 3 1 0 1 1 1 7 1 9 0 7 1 0 9 3 1 2 7 1 9 1 9 1 7 1 7 0 1
3 1 7 0 3 1 2 3 1

Sentence scheme 2 9 3 1 9 1 1 1

Frequency of appearance of parts of speech in the text:
{'fraction': 0.41, 'preposition': 0.17, 'noun': 0.95, 'verb': 0.07, 'adjective': 0.28, '-': 0.12}
```

Figure 4: Sentence schemes and the word frequency

Word	Part of speech	Word form	Genus	Syllables	Prefix	Root	Suffix	Ending	Letters	Vowel	Consonants	ь	й	Apostrophe	Hyphen
ночі	іменник	однина	жіночи	но-чі	-	ноч	-	і	4	2	2	0	0	0	0
коли	частка	-	-	ко-ли	-	кол	-	и	4	2	2	0	0	0	0
за	прийменник	-	-	за	за	за	за	за	2	1	1	0	0	0	0
вікном	іменник	однина	середні	ві-кном	-	вік	ном	-	6	4	2	0	0	0	0
проходять	дієслово	інфінітив	-	про-хо-дят	-	проха	проходять	ь	9	5	3	1	0	0	0
міські	прикметник	однина	чоловіч	місь-кі	-	міс	ьк	і	6	3	2	1	0	0	0
вечори	іменник	однина	чоловіч	ве-чо-ри	-	вечр	вечор	и	6	3	3	0	0	0	0
маєток	іменник	однина	чоловіч	ма-є-ток	-	має	ток	-	6	3	3	0	0	0	0
злетів	іменник	однина	чоловіч	зле-тів	-	злет	ів	-	6	4	2	0	0	0	0
на	прийменник	-	-	на	на	на	на	на	2	1	1	0	0	0	0

Figure 5: Morphological analysis

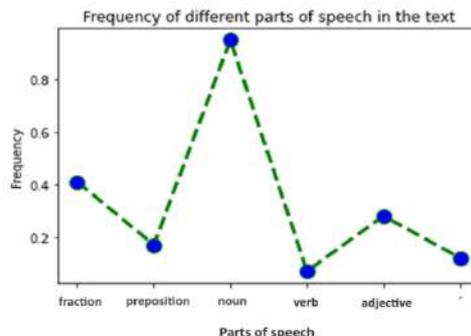


Figure 6: Frequency of different parts of speech used in the text

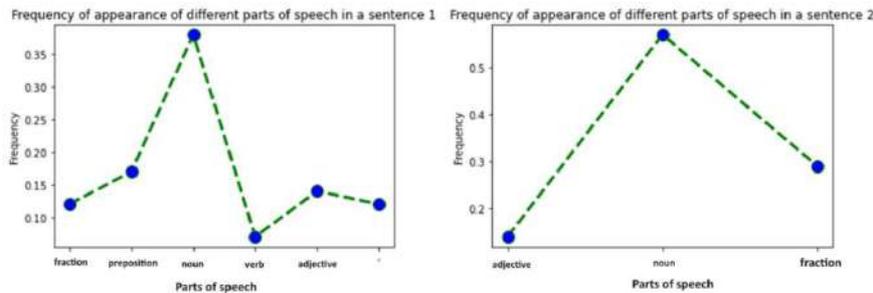


Figure 7: Frequency of different parts of speech used in sentences

Based on the conducted analysis, we may assume that a noun has the most significant frequency of use. It means that the nouns take the most remarkable occurrence in sentences compared with other words. The frequency is about 0.4 for the fraction and the adjective. The author gives additional semantic nuances to individual words or sentences using fractions, and also, he provides other features to certain words using adjectives. Thus, verbs and prepositions are present in the text, showing the lowest frequency. Prepositions indicate relationships between different words in a sentence and combine them. Therefore, it can be concluded that the author does not make many combinations of words using prepositions. The verbs take the least number of occurrences indicating that the author does not emphasise the performance of any action. We should also mention that some words are not recognised to belong to any part of speech.

6. Discussion

We have chosen five texts of different styles to test the program's analytics: publicistic, belles-lettres, scientific, official, and conversational styles. The program calculates the frequency of parts of speech throughout the text and separately for each sentence. In its final stage, the results of all texts are compared. Diagrams for a publicistic text are presented in Fig. 8-9:

В п'ятницю, третього жовтня, в селі Привільному Дубенського району Рівненської області пройшов кам'яний дощ, який завдав шкоди врожаю та трохи пошкодив будинки жителів цього села. На сьогодні роботи з відновлення цілісності будинків завершені. Місцеві мешканці повернулися до нормального життя.

[V piatnytsiu, tretoho zhovtnia, v seli Pryvilnomu Dubenskoho raionu Rivnenskoï oblasti proishov kamiani doshch, yakyi zavdav shkody vrozhaiu ta trokhy poshkodyyv budynky zhyteliv tsoho sela. Na sohodni roboty z vidnovlennia tsilisnosti budynkiv zavshereni. Mistsevi meshkantsi povernulysia do normalnoho zhyttia].

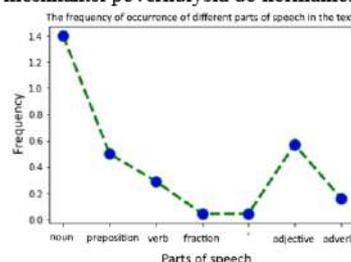


Figure 8: Frequency of different parts of speech in a publicistic text

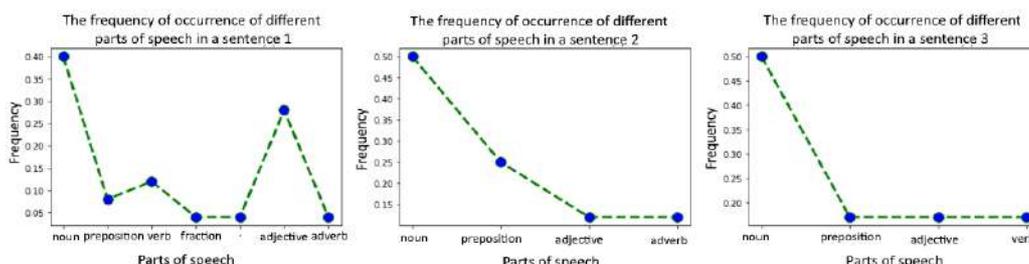


Figure 9: Frequency of parts of speech used in sentences: publicistic text

Below you can find diagrams of the text in a belles-lettres style (see Fig. 10-11):

Я дивлюсь на портрети: князь хмурить брови, княгиня — надменна зневага, княжата — в темряві столітніх дубів. І в цій надзвичайній суворості я відчуваю весь древній світ, всю безсилу грандіозність і красу третьої молодості минулих шляхетних літ.

[Ya dyvlius na portrety: kniaz khmuryt brovy, kniahynia — nadmenna znevaha, kniazhata — v temriavi stolitnikh dubiv. I v tsii nadzvychainii suvorosti ya vidchuvaiu ves drevnii svit, vsiu bezsilyu hrاندioznist i krasu tretoi molodosty mynulykh shliakhetnykh lit].

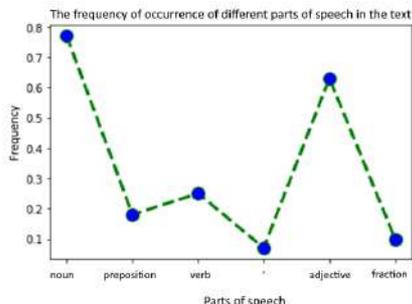


Figure 10: Frequency of different parts of speech in a belles-lettres text

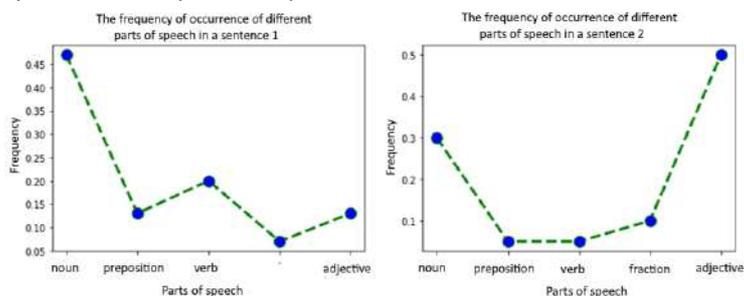


Figure 11: Frequency of different parts of speech used in sentences: belles-lettres style

Another style which is considered is a scientific one. The text in a scientific style is given below and the diagrams for it are presented in Fig. 12-13:

Податки – одна з найважливіших фінансових категорій. Історично це найдавніша форма фінансових відношень між державою та членами суспільства.

[Podatky – odna z naivazhlyvishykh finansovykh katehori. Istorychno tse naidavnisha forma finansovykh vidnoshen mizh derzhavoiu ta chlenamy suspilstva].

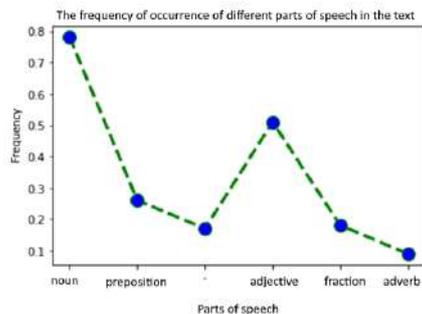


Figure 12: Diagram of the frequency of different parts of speech in a scientific text

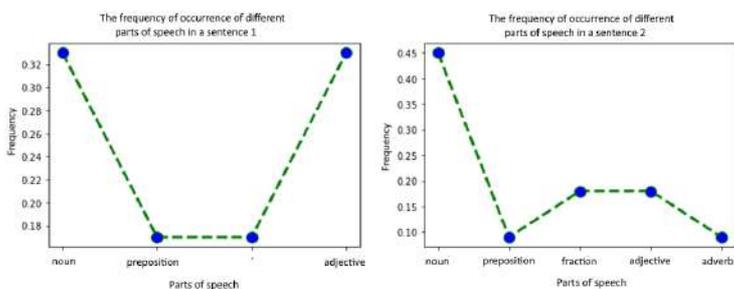


Figure 13: Diagrams of the frequency of parts of speech used in sentences of a scientific text

The next text represents an official style and the diagrams for the frequency of used parts of speech are presented in Fig. 14 - 15:

Держави визнають право дитини на відпочинок і дозвілля, право брати участь в іграх і розважальних заходах, що відповідають її віку, та вільно брати участь у культурному житті, займатися мистецтвом. Держави поважають і заохочують право дитини на всебічну участь у культурному і творчому житті.

[Derzhavy vyznaiut pravo dytyny na vidpochynok i dozvillia, pravo braty uchast v ihrakh i rozvazhalnykh zakhodakh, shcho vidpovidaiut yii viku, ta vilno braty uchast u kulturnomu zhytti, zaimatysia mystetstvom. Derzhavy povazhaiut i zaokhochuiut pravo dytyny na vsebichnu uchast u kulturnomu i tvorchomu zhytti].

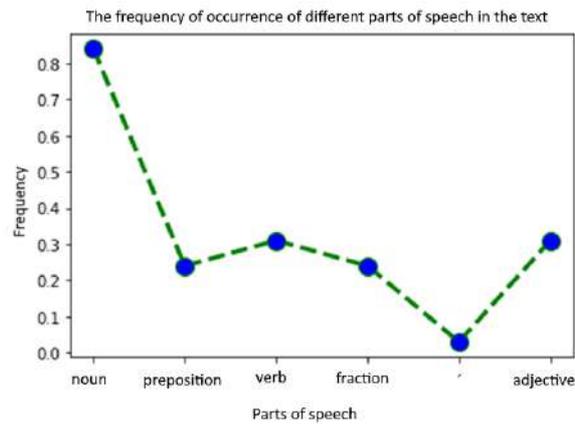


Figure 14: Diagram of the frequency of different parts of speech in an official text

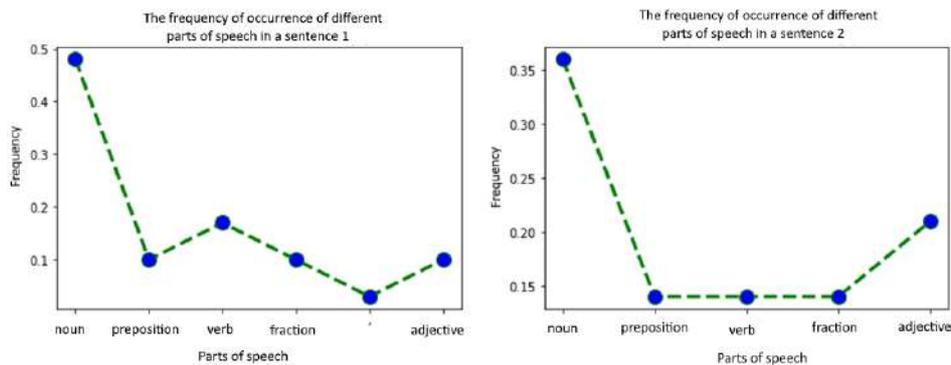


Figure 15: Diagrams of the frequency of parts of speech used in sentences of the official text

The conversational style is represented in a text below. The diagrams of it are shown in Fig. 16-17:

Грицю!- сказав батько. Га!- сказав Гриць. Видиш оту хату? Видзу. Пам'ятай собі, се школа. Ба,- сказав Гриць. Сюди будеш ходити вчитися. Ба,- сказав Гриць.

[Hrytsiu!- skazav batko. Ha!- skazav Hryts. Vydysht otu khatu? Vydzu. Pamiatai sobi, se shkola. Ba,- skazav Hryts. Siudy budesh khodyty vchytysia. Ba,- skazav Hryts].

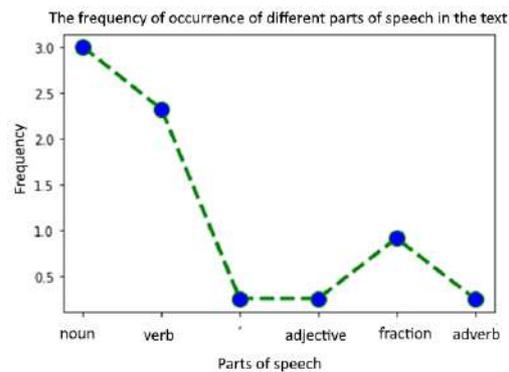


Figure 16: Diagrams of frequency of different parts of speech in a conversational text

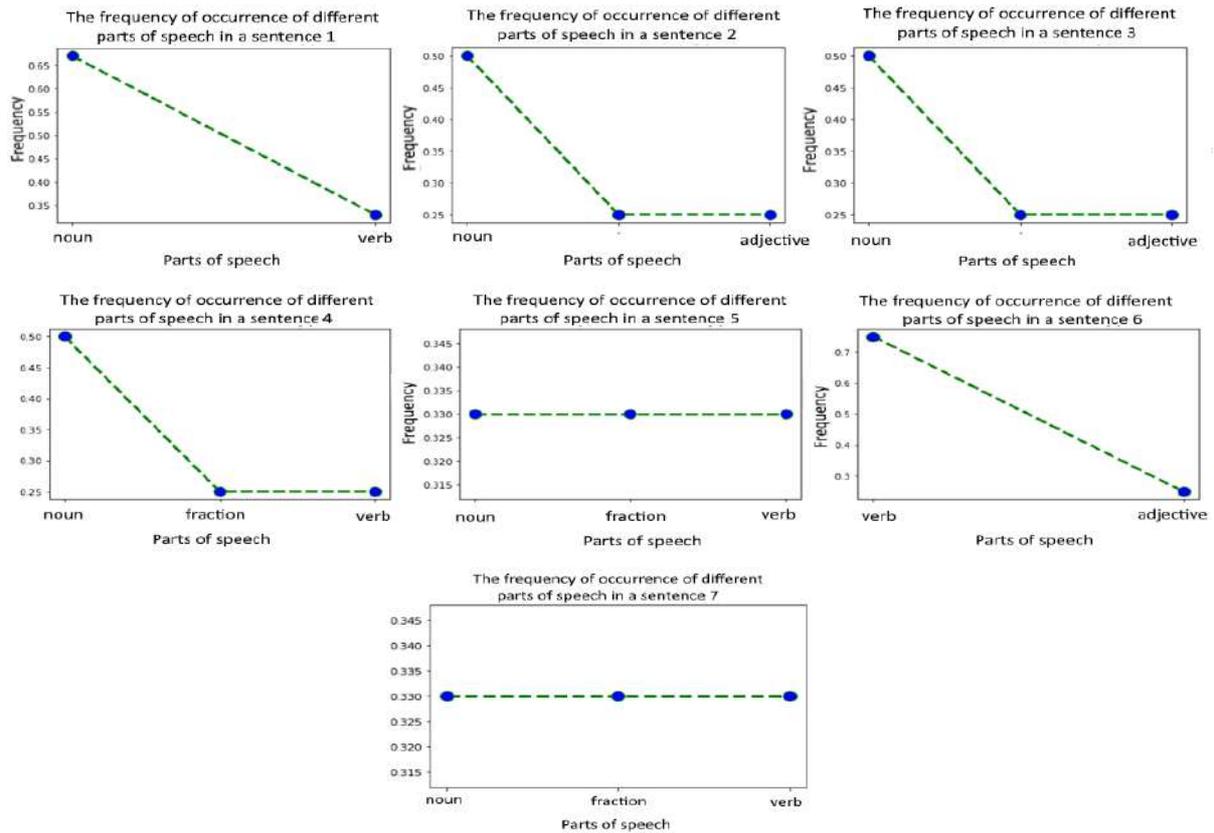


Figure 17: Diagrams of the frequency of parts of speech used in sentences: conversational style

Having analysed the texts with the application of the program functionality, we built a line diagram that shows the frequency of different parts of speech used in texts of different styles (Fig. 18). Its results demonstrate that the text of the conversational style has the highest number of the used parts of speech. In other texts, the fluctuations between the frequencies of occurrence are approximately the same. It is easy to recognise that verbs have the most significant frequency of use in the conversational text (a conversation between two people discussing their further actions). Texts of other styles have the lowest number of verb use compared with conversational text. Instead, other parts of speech such as nouns, adjectives and prepositions are often used. The frequency of verbs, numerals, and adverbs in these texts is about 0.

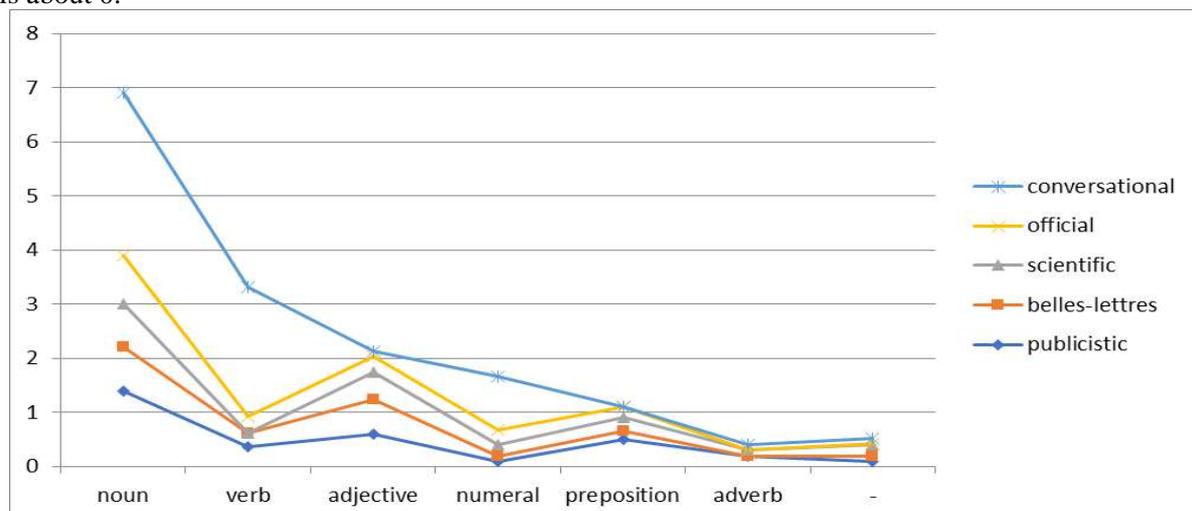


Figure 18: Frequency of parts of speech used in texts of different styles

7. Conclusions

The literature sources related to the morphological analysis of Ukrainian words and the frequency use of different parts of speech in Ukrainian texts are considered in our research. We have investigated the available analogue systems and applied the following: the intelligent system of Ukrainian text processing, large electronic dictionary of the Ukrainian language (VESUM), microservices lang-uk, NER-Annotations corpus and tonal dictionary of the Ukrainian language. The analysis of their features and characteristics and their advantages and disadvantages is conducted in detail. It proves the relevance and topicality of the project, which is to analyse the use of words of different parts of speech in Ukrainian texts. A system analysis is chosen to perform our research. We constructed UML diagrams to represent the general purpose of the developed system: use-case, states and activities. Each diagram shows the system from different aspects to better understand its processes. In addition, the basic functionality of the system is presented. As the implementation tools for creating a program, which will analyse Ukrainian texts, the Python programming language, MySQL database, and Tkinter graphical interface were chosen.

A program that analyses the use of words of different parts of speech in Ukrainian texts was developed. The program builds sentence schemes, which represent the sequence of used parts of speech and displays a list of frequencies of occurrence of parts of speech in the text. In addition, it displays frequency graphs for the text and separately for each sentence and presents a morphological analysis of each word in the entered text. Then, a control example of the developed program is shown; each element of the program is described, and the entered text analysis is presented. The program has a convenient and simple interface, is easy to use, and contains all design requirements. Accordingly, it displays all the information about the word and the calculated frequency that it has processed.

We analysed the obtained results, namely the frequencies of using words of different parts of speech in the texts of publicistic, belles-lettres, scientific, official, and conversational styles. It can be concluded that the developed algorithms for morphological analysis, based on the rules of the Ukrainian language, contain many inaccuracies as only a small number of exceptions for many Ukrainian words are taken into account. In addition, it should also be noted that an algorithm for calculating the word combinations frequencies of different parts of speech has not been developed. It means that the program needs further improvement, building better algorithms and methods, and designing a coherent structure. It also shows that the recognition of the Ukrainian language is a complex process and requires more research, which provides a good starting point for discussion and further investigation.

8. References

- [1] S. L. Holoshchuk, *Istorychni peredumovy rozvytku korpusnoyi linhvistyky* [Historical preconditions for the development of corpus linguistics], *International Academy Journal: Web of Scholar* 6 (15) (2017) 80-84.
- [2] V. Lytvyn, V. Vysotska, P. Pukach, I. Bobyk, D. Uhryn, *Development of a method for the recognition of author's style in the Ukrainian language texts based on linguometry, stylemetry and glottochronology*, *Eastern-European Journal of Enterprise Technologies* 4(2-88) (2017) 10-19. doi: 10.15587/1729-4061.2017.107512.
- [3] V. Vysotska, O. Kanishcheva, Y. Hlavcheva, *Authorship Identification of the Scientific Text in Ukrainian with Using the Lingvometry Methods*, in: *Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2018*, pp. 34-38. doi: 10.1109/STC-CSIT.2018.8526735.
- [4] V. Vysotska, V.B. Fernandes, V. Lytvyn, M. Emmerich, M. Hrendus, *Method for Determining Linguometric Coefficient Dynamics of Ukrainian Text Content Authorship*, *Advances in Intelligent Systems and Computing* 871 (2019) 132-151. doi: 10.1007/978-3-030-01069-0_10.
- [5] V. Vysotska, *Ukrainian Participles Formation by the Generative Grammars Use*, *CEUR workshop proceedings Vol-2604* (2020) 407-427.
- [6] V. Vysotska, S. Holoshchuk, R. Holoshchuk, *A comparative analysis for English and Ukrainian texts processing based on semantics and syntax approach*, *CEUR Workshop Proceedings Vol-2870* (2021) 311-356.

- [7] K. Tymoshenko, V. Vysotska, O. Kovtun, R. Holoshchuk, S. Holoshchuk, Real-time Ukrainian text recognition and voicing, CEUR Workshop Proceedings Vol-2870 (2021) 357-387.
- [8] A. Dmytriv, V. Vysotska, M. Bublyk, The Speech Parts Identification for Ukrainian Words Based on VESUM and Horokh Using, in: Proceedings of the IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT) :, 22-25 Sept., Lviv, Ukraine, 2021, Vol. 2, pp. 21–33. doi: 10.1109/CSIT52700.2021.9648813.
- [9] L. Savytska, N. Vnukova, I. Bezugla, V. Pyvovarov, M. Turgut Sübay, Using Word2vec Technique to Determine Semantic and Morphologic Similarity in Embedded Words of the Ukrainian Language, CEUR Workshop Proceedings Vol-2870 (2021) 235-248.
- [10] M. Sazhok, A. Poltieva, V. Robeiko, R. Seliukh, D. Fedoryn, Punctuation Restoration for Ukrainian Broadcast Speech Recognition System based on Bidirectional Recurrent Neural Network and Word Embeddings, CEUR Workshop Proceedings Vol-2870 (2021) 300-310.
- [11] O. Cherednichenko, O. Kanishcheva, Readability Evaluation for Ukrainian Medicine Corpus (UKRMED), CEUR Workshop Proceedings Vol-2870 (2021) 402-412.
- [12] A. Luchykh, O. Taran, O. Palchevska, N. Sharmanova, G. Demydenko, Corpus-Driven Approach to Ukrainian E-Anecdotes Study, CEUR Workshop Proceedings Vol-2870 (2021) 424-434.
- [13] V. Starko, Implementing Semantic Annotation in a Ukrainian Corpus, CEUR Workshop Proceedings Vol-2870 (2021) 435-447.
- [14] H. Sytar, O. Vietrov, V. Diachenko, Synonymizer of the Ukrainian Language: Stage of Creation, Features of Database Update and Software Implementation, CEUR Workshop Proceedings Vol-2870 (2021) 448-458.
- [15] N. Tmienova, B. Sus, System of Intellectual Ukrainian Language Processing, CEUR Workshop Proceedings Vol-2577 (2019) 199-209. URL: <http://ceur-ws.org/Vol-2577/paper16.pdf>.
- [16] Large Electronic Dictionary of the Ukrainian Language (VESUM) as a tool of NLP. URL: https://www.researchgate.net/publication/344842033_Velikij_elektronnij_slovník_ukrainskoi_movi_VESUM_ak_zasib_NLP_dla_ukrainskoi_movi_Galaktika_Slova_Galini_Makarivni_Gnatuk
- [17] Ukrainian tonal dictionary. URL: <https://github.com/lang-uk/tone-dict-uk>.
- [18] Brown Corps of the Ukrainian language. URL: <https://github.com/brown-uk/corpus>.
- [19] NER-text markup. URL: <https://github.com/lang-uk/ner-uk/blob/master/doc/README.md>.
- [20] Microservices lang-uk. URL: <https://lang.org.ua/uk/services/>.
- [21] V. Vysotska, Linguistic Analysis of Textual Commercial Content for Information Resources Processing, in: Proceedings of the Modern Problems of Radio Engineering, Telecommunications and Computer Science, TCSET, 2016, pp. 709-713. doi: 10.1109/TCSET.2016.7452160.
- [22] P. Zhezhnych, A. Shilinh, V. Melnyk, Linguistic analysis of user motivations of information content for university entrant’s web-forum, International Journal of Computing 18 (2019) 67-74.
- [23] Lytvyn Vasyl, Vysotska Victoria, Dosyn Dmytro, Holoschuk Roman, Rybchak Zoriana, Application of Sentence Parsing for Determining Keywords in Ukrainian Texts, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2017, pp. 326-331. doi: 10.1109/STC-CSIT.2017.8098797.
- [24] Y. Burov, V. Vysotska, P. Kravets, Ontological approach to plot analysis and modeling, CEUR Workshop Proceedings Vol-2362 (2019) 22-31.
- [25] V. Lytvyn, V. Vysotska, O. Veres, I. Rishnyak, H. Rishnyak, Content linguistic analysis methods for textual documents classification, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2016, pp. 190-192. doi: 10.1109/STC-CSIT.2016.7589903.
- [26] O. Bisikalo, V. Vysotska, Linguistic analysis method of Ukrainian commercial textual content for data mining, CEUR Workshop Proceedings Vol-2608 (2020) 224-244.
- [27] V. Vysotska, V. Lytvyn, V. Kovalchuk, S. Kubinska, M. Dilai, B. Rusyn, L. Pohreliuk, L. Chyrun, S. Chyrun, O. Brodyak, Method of Similar Textual Content Selection Based on Thematic Information Retrieval, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2019, pp. 1-6. doi: 10.1109/STC-CSIT.2019.8929752.
- [28] V. Lytvyn, V. Vysotska, I. Peleshchak, T. Basyuk, V. Kovalchuk, S. Kubinska, L. Chyrun, B. Rusyn, L. Pohreliuk, T. Salo, Identifying Textual Content Based on Thematic Analysis of Similar Texts in Big Data, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2019, pp. 84-91. doi: 10.1109/STC-CSIT.2019.8929808.

- [29] S. Kubinska, V. Vysotska, Y. Matseliukh, User Mood Recognition and Further Dialog Support, in: Proceedings of the IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, 2021, vol. 2, pp. 34–39. doi: 10.1109/CSIT52700.2021.9648610.
- [30] V. Lytvyn, N. Sharonova, T. Hamon, V. Vysotska, N. Grabar, A. Kowalska-Styczen, Computational linguistics and intelligent systems, CEUR Workshop Proceedings 2136 (2018).
- [31] V. Lytvyn, N. Sharonova, T. Hamon, O. Cherednichenko, N. Grabar, A. Kowalska-Styczen, V. Vysotska, Preface, CEUR Workshop Proceedings Vol-2362 (2019).
- [32] V. Lytvyn, V. Vysotska, T. Hamon, N. Grabar, N. Sharonova, O. Cherednichenko, O. Kanishcheva, Preface, CEUR Workshop Proceedings Vol-2604 (2020).
- [33] N. Sharonova, V. Lytvyn, O. Cherednichenko, Y. Kupriianov, O. Kanishcheva, T. Hamon, N. Grabar, V. Vysotska, A. Kowalska-Styczen, I. Jonek-Kowalska, Preface, CEUR Workshop Proceedings Vol-2870 (2021).
- [34] V. Husak, O. Lozynska, I. Karpov, I. Peleshchak, S. Chyrun, A. Vysotskyi, Information System for Recommendation List Formation of Clothes Style Image Selection According to User's Needs Based on NLP and Chatbots, CEUR workshop proceedings Vol-2604 (2020) 788-818.
- [35] O. Romanovskiy, N. Pidbutska, A. Knysh, Elomia Chatbot: The Effectiveness of Artificial Intelligence in the Fight for Mental Health, CEUR Workshop Proceedings 2870 (2021) 1215-1224.
- [36] A. Yarovy, D. Kudriavtsev, Method of Multi-Purpose Text Analysis Based on a Combination of Knowledge Bases for Intelligent Chatbot, CEUR Workshop Proceedings 2870 (2021) 1238-1248.
- [37] N. Shakhovska, O. Basystiuk, K. Shakhovska, Development of the Speech-to-Text Chatbot Interface Based on Google API, CEUR Workshop Proceedings Vol-2386 (2019) 212-221.
- [38] D. Aksonov, A. Gozhij, I. Kalinina, V. Vysotska, Question-Answering Systems Development Based on Big Data Analysis, in: Proceedings of the IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), 22-25 Sept., Lviv, Ukraine, 2021, Vol. 1. pp. 113–118. doi: 10.1109/CSIT52700.2021.9648631.
- [39] V. Lytvyn, V. Vysotska, A. Rzhetskyi, Technology for the Psychological Portraits Formation of Social Networks Users for the IT Specialists Recruitment Based on Big Five, NLP and Big Data Analysis, CEUR Workshop Proceedings Vol-2392 (2019) 147-171.
- [40] J. Deriviere, T. Hamon, A. Nazarenko, A scalable and distributed NLP architecture for web document annotation, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 4139 (2006) 56–67.
- [41] M. Boyè, T. M. Tran, N. Grabar, NLP-oriented contrastive study of linguistic productions of alzheimer's and control people, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8686 (2014) 412–424.
- [42] C. Shu, D. Dosyn, V. Lytvyn, V. Vysotska, A. Sachenko, S. Jun, Building of the Predicate Recognition System for the NLP Ontology Learning Module, in: Proceedings of the International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS, 2, 2019, pp. 802-808. doi: 10.1109/IDAACS.2019.8924410.
- [43] V.-A. Oliinyk, V. Vysotska, Y. Burov, K. Mykich, V. Basto-Fernandes, Propaganda Detection in Text Data Based on NLP and Machine Learning, CEUR workshop proceedings Vol-2631 (2020) 132-144.
- [44] I. Balush, V. Vysotska, S. Albota, Recommendation System Development Based on Intelligent Search, NLP and Machine Learning Methods, CEUR Workshop Proceedings Vol-2917 (2021) 584-617.
- [45] T. Batura, A. Bakiyeva, M. Charintseva, A method for automatic text summarization based on rhetorical analysis and topic modeling, International Journal of Computing 19(1) (2020) 118-127.
- [46] H. Schöpfer, W. Kersten, Using Natural Language Processing for Supply Chain Mapping: a Systematic Review of Current Approaches, CEUR Workshop Proceedings 2870 (2021) 71-86.
- [47] M. Zanchak, V. Vysotska, S. Albota, The Sarcasm Detection in News Headlines Based on Machine Learning Technology, in: Proceedings of the IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), 22-25 Sept., Lviv, Ukraine, 2021, Vol. 1. pp. 131–137. doi: 10.1109/CSIT52700.2021.9648710.
- [48] N. Kholodna, V. Vysotska, S. Albota, A Machine Learning Model for Automatic Emotion Detection from Speech, CEUR Workshop Proceedings Vol-2917 (2021) 699-713.

- [49] D. Nazarenko, I. Afanasieva, N. Golian, V. Golian, Investigation of the Deep Learning Approaches to Classify Emotions in Texts, *CEUR Workshop Proceedings Vol-2870* (2021) 206-224.
- [50] I. Bekhta, N. Hrytsiv, Computational Linguistics Tools in Mapping Emotional Dislocation of Translated Fiction, *CEUR Workshop Proceedings Vol-2870* (2021) 685-699.
- [51] I. Spivak, S. Krepych, O. Fedorov, S. Spivak, Approach to Recognizing of Visualized Human Emotions for Marketing Decision Making Systems, *CEUR Workshop Proceedings Vol-2870* (2021) 1292-1301.
- [52] P. C. Thoumelin, N. Grabar, Subjectivity in the medical discourse: On uncertainty and emotional markers [La subjectivité dans le discours médical: Sur les traces de l'incertitude et des émotions], *Revue des Nouvelles Technologies de l'Information E.26* (2014) 455-466.
- [53] N. Grabar, L.O. Dumonet, Automatic computing of global emotional polarity in French health forum messages, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9105 (2015) 243-248.
- [54] Z. Kochuieva, N. Borysova, K. Melnyk, D. Huliieva, Usage of Sentiment Analysis to Tracking Public Opinion, *CEUR Workshop Proceedings Vol-2870* (2021) 272-285.
- [55] N. Bondarchuk, I. Bekhta, Quantitative Characteristics of Lexical-Semantic Groups Representing Weather in Weather News Stories (Based on British Online Press), *CEUR Workshop Proceedings Vol-2870* (2021) 799-810.
- [56] O. Artemenko, V. Pasichnyk, N. Kunanets, K. Shunevych, Using sentiment text analysis of user reviews in social media for e-tourism mobile recommender systems, *CEUR workshop proceedings Vol-2604* (2020) 259-271.
- [57] V. Bobicev, O. Kanishcheva, O. Cherednichenko, Sentiment Analysis in the Ukrainian and Russian News, in: *First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 2017 pp. 1050-1055.
- [58] S. Bhatia, M. Sharma, K. K. Bhatia, P. Das. Opinion target extraction with sentiment analysis, *International Journal of Computing* 17(3) (2018) 136-142.
- [59] K. Shakhovska, N. Shakhovska, P. Vesely, The sentiment analysis model of services providers' feedback, *Electronics (Switzerland)* 9(11) (2020) 1-15.
- [60] V. Turchenko, L. Grandinetti, A. Sachenko, Parallel batch pattern training of neural networks on computational clusters, in: *Proceedings of the International Conference on High Performance Computing & Simulation (HPCS)*, 2012, pp. 202-208, doi: 10.1109/HPCSim.2012.6266912.
- [61] P. Kossakowski, P. Bilski, Analysis of the self-organizing map-based investment strategy, *International Journal of Computing* 16(1) (2017) 10-17.
- [62] M. Maree, M. Eleyat, Semantic graph based term expansion for sentence-level sentiment analysis, *International Journal of Computing* 19(4) (2020) 647-655.
- [63] S. Shrivastava, K. V. Lakshmy, C. Srinivasan, On the Statistical Analysis of ZUC, Espresso and Grain v1, *International Journal of Computing* 20(3) (2021) 384-390.
- [64] V. Turchenko, V. A. Golovko, A. Sachenko, Parallel Batch Pattern Training of Recirculation Neural Network, in: *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics, ICINCO*, 2012, 1, pp. 644-650.
- [65] I. Paliy, A. Sachenko, Y. Kurylyak, O. Boumbarov, S. Sokolov, Combined approach to face detection for biometric identification systems, in: *Proceedings of the IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, 2009, pp. 425-429. doi: 10.1109/IDAACS.2009.5342946.
- [66] Md Maksudur R. Mazumder, C. Phillips, Partitioning known environments for multi-robot task allocation using genetic algorithms, *International Journal of Computing* 19(3) (2020) 480-490.
- [67] M. Patil, T. Abukhalil, S. Patel, T. Sobh, UB SWARM: hardware implementation of heterogeneous swarm robot with fault detection and power management, *International Journal of Computing* 15(3) (2016) 162-176.
- [68] V. M. Hung, V. Mihai, C. Dragana, I. Ion, N. Paraschiv, Dynamic computation of haptic-robot devices for control of a surgical training system, *International Journal of Computing* 17(2) (2018) 81-93.