Information System of Psycholinguistic Text Analysis

Viktoriia Vasyliuk, Yuliia Shyika [0000-0003-2474-0479], Tetiana Shestakevych [0000-0002-4898-6927]

Lviv Polytechnic National University, Lviv, Ukraine
tetiana.v.shestakevych@lpnu.ua

Abstract. The progressive global computerization demands the creation of new information systems and improvement of the existing ones that will provide assistance in scientific researches of the 21st century in the field of psychology, in particular, and will help them to take a new form and not to lose their scientific value during active usage. In recent years, there has also been increasing interest in psycholinguistic research. Creation of an information system for the analysis of Ukrainian-language texts for psychological purposes can increase demand for Ukrainian product. The article deals with the peculiarities of the psycholinguistic analysis of a text and designing of the appropriate information system of the Thematic Apperception Test support. The information system will be available for psychologists and applied linguists. The creation of new categories of lemmas for more accurate results of the Thematic Apperception Test, based on recent research, will be possible.

Keywords: psycholinguistics, text, psycholinguistic text analysis, Thematic Apperception Test, information system.

1 Introduction

Linguistics and linguistic studies hold a special place among sciences, because language is a dynamic phenomenon and every person possess speech. Among the linguistic sciences, psycholinguistics is particularly relevant as far as it is a science that combines the basic concepts of linguistics and psychology; science which helps to understand how the human brain works. According to S. Kuranova, psycholinguistics is a science that studies the processes of acquisition, interpretation and formation of speech in their interaction with the language system, and develops models of speech activity and psychophysiological speech organization of a person and tests them by means of psychological experiments [1]. Ch. Osgood, who is one of the first scholars in the field of psycholinguistics, argued that psycholinguistics deals with the encoding and decoding of language signals, correlating them with the state of the participants of the communication [2]. Psycholinguistics also examines specific circumstances: how internal and external factors influence the development of language skills, for example, the impact of hearing and vision impairment on language learning, or how brain damage can affect various aspects of speech. Within psycholinguistics, the new scientific field of text linguistics has been developed. This field investigates the principles and rules of

text formation as well as types and varieties of speech. Text is the main subject of the research here.

Since it has been scientifically proved that a mental state of a person directly influences the product of a speech act, scientists have started to research specific materials [3]. The speech of a fiction character has been investigated by Y. Kozachenko [4]. The average length of sentences is measured to obtain information about the emotional level and verbal intelligence of the speaker. These measures show the emotional state of the speaker at the moment when the text was created. In this aspect A. Peleshchyshyn also investigated manipulation examples in Wikipedia Talk Pages [5], discovering the specific speech features of different categories of speakers; S. Fedushko also developed a computer linguistic method of offensive language filtering [6]. V. Vysotska has been carrying scientific researches in the field of content analysis [7-10], paying special attention to the automatic processing of texts, developing the new methods of linguistic analysis of Ukrainian-language content; O. Levchenko has worked on sentiment analysis of Tweets [11]. These studies in the field of content analysis and information system are of great importance in the national studies of applied linguistics.

2   Projective Psychodiagnostic Techniques

Projective technique involves respond of a person to the stimuli designed in such a way, that responding to situations, a person subconsciously projects these situations on him / herself. Projective techniques give an opportunity to examine human subconscious and are a powerful tool in psychological diagnosing. The most popular projective tests are the *Rorschach test* (based on perceptions of inkblots), *Select face* (developed to determine the level of anxiety of preschool children), *16 associations* by Carl Jung and TAT, the *Thematic Apperception Test* (based on the narratives about pictures). A simplified version of the Thematic Apperception Test is available online and employs the Linguistic Inquiry and Word Count (hereinafter LIWC), a software for psycholinguistic text analysis designed by James W. Pennebaker. However, the online version of Thematic Apperception Test (https://www.utpsyc.org/TATintro/) is relevant for native speakers, as well as for foreigners who can express their thoughts and feelings fluently in English.

The analysis of psychodiagnostic projective techniques revealed the need for an information system for Thematic Apperception Test for the needs of Ukrainian-speaking consumers [12]. That is why it is advisable to create such an informative system of Thematic Apperception Test, which will apply similar to LIWC principles of text study, will provide previous and further work with a psychologist and will be adapted to the Ukrainian language. This means that for its implementation it is necessary to identify such psycholinguistic characteristics of the text, which will indicate a certain psycho-emotional state, and to give the information system opportunities to apply these characteristic to the text.

3 Psycholinguistic Analysis of the Text

The core of the information system is the psycholinguistic analysis of the text, so it is necessary to describe and systematize all methods and ways of text analysis, lemmatization and analysis of lemmas.

Lemmatization is a process that involves determining the part of speech of a word, and further stemming that involves finding the root of the word by removing endings and suffixes. A breakthrough in this field in the Ukrainian language was made by the creation of the Great Electronic Dictionary of the Ukrainian Language (hereinafter GEDUL), which contains almost 400,000 lemmas and is constantly updated. Besides the lemmatization, the resource also replaces words that are not spelled according to the current orthography, provides information about rare words, abbreviations and slang. The advantage of GEDUL is its free access on the Internet.

Besides the lemmatization, the information system will also make statistical analysis of the text. This will be done using the coefficients created by the psycholinguists and psychotherapists around the world. These coefficients include [12]:

- the coefficient of emotionality of the text (the ratio of the number of adjectives to the number of the words multiplied by two);
- the Trager coefficient (the ratio of the number of verbs to the number of adjectives - indicates emotional stability / instability);
- the coefficient of certainty of the action (the ratio the number of verbs to the number of nouns – indicates the socialization of the speaker);
- the coefficient of aggressiveness (the ratio of the number of verbs and verb forms to the total number of words);
- the coefficient of directivity (indicates the speaker's determination, readiness for action);
- the coefficient of vocabulary diversity (the ratio of the number of different words to total number of words in the text);
- logical coherence (ratio of the number of function words to number of sentences);
- embolism (indicates a speech disfluency – the ratio the number of emboli to the total number of words in the text).

4 The Model of Information System of Thematic Apperception Test Usage

The model of functioning of the information system for the needs of the client of psycholinguistic analysis services (and taking into account demands of persons with special needs [13, 14]) can be shown by activity diagram (Fig. 1). To start a person needs to log in and enter his/her personal and customer information (e-mail) to get the test results. Then the person has to take the test, writing the text description to the picture. After that, the program performs two parallel independent processes: statistical analysis of the text and lemmatization as well as further statistical analysis of the lemmas. Then,

the program uses statistical data obtained from previous parallel processes (statistical text analysis, lemmatization process and coefficients for calculating text characteristics) to calculate various measures. The result of all the above-described processes is the conclusion about the psycho-emotional state of the subject. The results of the test are sent to the e-mails of the user and the customer, which were provided during the authorization. All written texts are stored in a database for further research by psychologists and applied linguists.

The users of such information system are: a person under investigation, psychologist, IT specialist and a customer of psycholinguistic text analysis services. The usage of the information system begins with the customer of the service, which motivates a person to take the test. The psychologist assesses the psycho-emotional state of the person on the basis of the conversation, using his/her knowledge and experience. Then, the person takes a Thematic Apperception Test: writes the text, describing a picture. The entered text is analysed according to indicators and coefficients that help to investigate emotions and intentions; the psychologist is involved in the analysis of the text. The program shows the results of psycholinguistic analysis of the text which are sent to both, the person who took the test and the customer of the service. The IT specialist supports the functioning of the entire information system of Thematic Apperception Test. This information system, which shows results, will be adapted for native speakers of the Ukrainian language and will be opened for further improvement by scientists (psychologists and applied linguists). All these and other characteristics and functions of the information system will be convenient for all participants of the process: potential customers, users, psychologists and linguists.

The algorithm of the information system usage for the needs of the psychologist is the following:

Step 1. Authorization of the person and psychologist in the system.

Step 2. The person enters the text.

Step 3. The psychologist evaluates the psycho-emotional state of the person based on his judgment and experience.

Step 4. Carrying out of the psycholinguistic analysis of the text.

Step 4.1. Lemmatization and stemming.

Step 4.2. Statistical analysis of lemmas.

Step 4.3. Statistical analysis of the text.

Step 3.4. Calculation of coefficients.

Step 5.Updating of the database of psycholinguistic analysis of texts.

Step 6. Output of the results of psycho-emotional diagnosis in retrospect.
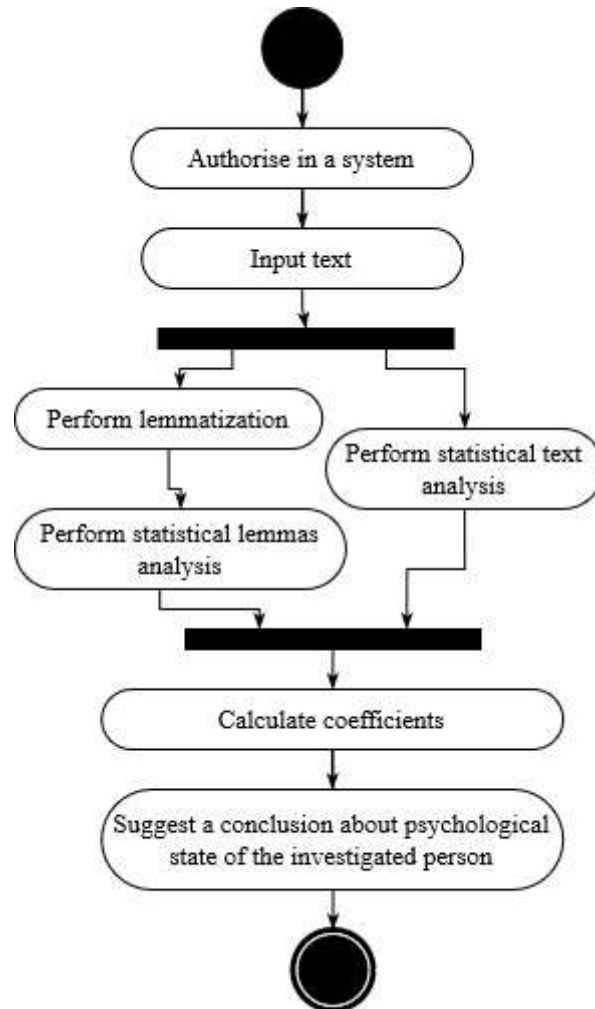
Step 7. Shutting down.

Fig.1. Activity chart for the TAT IS model

5 Entity-Relation Model of Database of Thematic Apperception Test Results

To develop a project of an information system of Thematic Apperception Test, a database that will store all changes, data and results should be developed. In order to visualize the results, an entity-relation model database should be created.

The entity "Customer" in the described model has four attributes: full name, gender, date of birth, contacts. Providing this data allows to identify the person and see changes and progress in retrospect of a specific individual. Gender and age are important for understanding the psychological background of the person, and contacts are essential

for sending results. The entity "Psychologist" has three attributes: the code of the specialist, the full name and the position. A specialist code allows to identify a psychologist, as well as to see time and person under the investigation. Gender and the age of the psychologist are irrelevant in this case, unlike the position, since the specialists in the field of psychology have specific range of positions and responsibilities. This information is presented in the entity-relation model of the project of Thematic Apperception Test information system in Fig. 2.
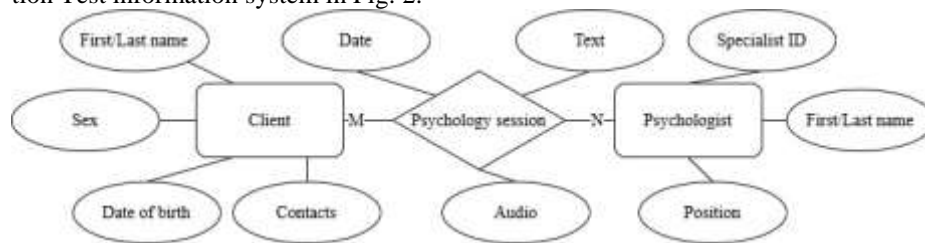

Fig. 2. Entity-relation model of the database of Thematic Apperception Test

6   Structure of the Database of Thematic Apperception Test Results

A creation of a database that will store all the information is needed before the development of the website. To create a database, Microsoft Access 2016 is used. After the analysis of different aspects, nine related tables were created: client, embols, embols_dict, evaluation, picture, psychologist, session, state, and statistics.

Statistics of the text analysis will be held in the table "Statistics" (Fig. 3.6.). This table represents all characteristics of the text most fully and is the most informative. It contains the following fields:

- - id_stat and id_embol are identifiers with text datatype;
- - words - contains the information about the total number of words in the text;
- - sentences - contains the information about the total number of sentences in the text;
- - verbs - contains the information about the total number of verbs and verb forms in the text;
- - nouns - contains the information about the total number of nouns in the text;
- - adjectives - contains the information about the total number of adjectives in the text;
- - condition_words - contains the information about the total number of conditional words;
- - different words - contains the information about the number of different words in the text for further calculation of the vocabulary variety factor;
- - serv_words - contains the information about the number of functional words.

The table "Session" (Fig.3.12) holds the information about the identifiers of other tables: evaluation, statistics, client, psychologist, and picture. The field "Text_" holds

the text description of the picture; the field "Date_" holds the information about the date of the test; the field "Audio_" allows attaching the audio recording of the conversation with the psychologist, namely the description of the picture. Audio material is an important element for future applied studies of non-verbal characteristics such as speech rate, pauses, voice changes, etc.

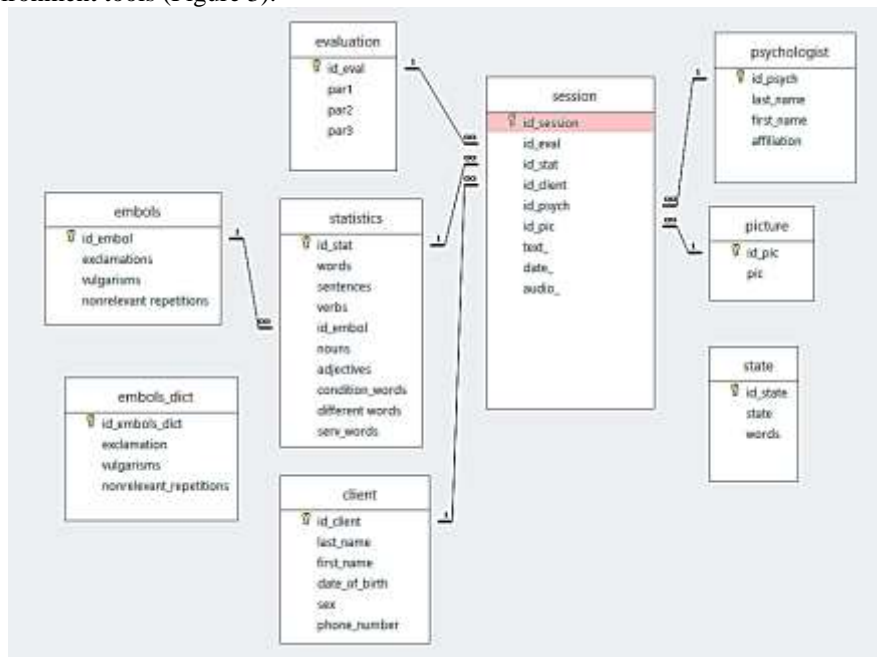All relations between tables can be shown by using the Microsoft Access 2016 environment tools (Figure 3).

Fig. 3. Table relations

7 Frontend Project for Thematic Apperception Test

To pass the test, a comfortable environment should be provided. To develop the page, it was decided to use the appropriate Internet resources. The most useful one is Wix, a platform with a huge variety of structural elements which help to create animated websites with a simple, intuitive interface. The platform was used to develop an information system with the webpage from Psychohelp as a frontend of the appropriate information technology.

The site consists of 5 main pages: registration, introduction, test, results (as part of the test page), and feedback. Animated video in calming colours (light blue and violet) was chosen as the background for all pages. Also, in the lower right corner of each page, there are buttons which redirect to social media (Facebook, Twitter, and Instagram). These are the accounts of the platform on which the site was developed and which can help to improve it.

The registration page contains various fields for entering information, including name, surname, date of birth, gender, email, and email of psychologist for sending results. To choose a date of birth, you should click on the appropriate field; to choose a gender you should put a tick. The button "Log in" saves data and redirects the user to the input page.

The introductory page contains the information about the test and user instruction. On the left side of the page there is a picture from Thematic Apperception Test. The Thematic Apperception Test contains 30 different pictures, from which the psychologist should choose 20 according to the gender and age of the person. In this case, since this is only a part of the online testing, a picture suitable for psychodiagnostic analysis regardless of the gender and age of the subject was chosen. It is necessary to take a good look at the picture and try to imagine what preceded the moment that is shown on the picture and what will happen next. It is also important to think about what depicted people feel, what emotions and thoughts they experience. It is important to take at least 10 minutes to complete the test. The button "Start test" redirects the user to the new page (Fig. 4).
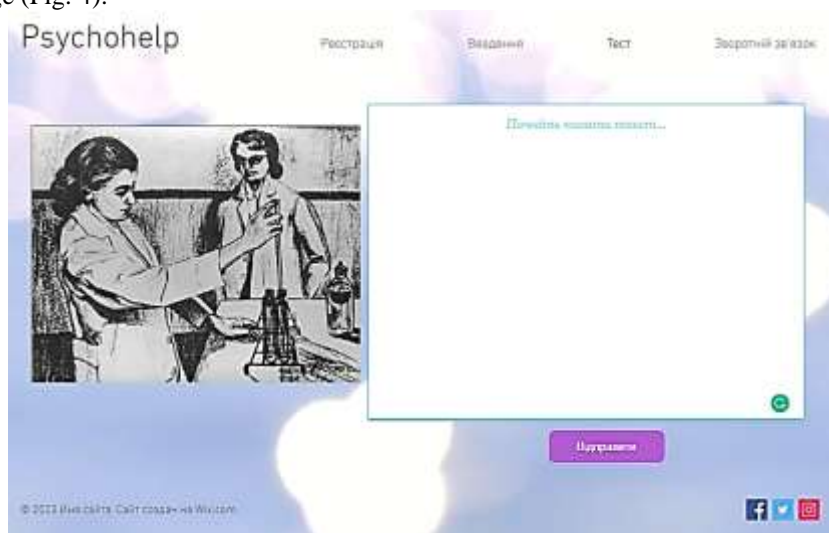

Fig. 4. Webpage of the test

The test page contains a picture and a field for a person to enter the text in Ukrainian. The submitted text is processed and the results are shown (Fig. 5). The person can send them to personal and psychologist's mail by pressing the button "Send". The results are structured into a table that contains the following columns: coefficient, norm, value, description. The Coefficient column lists the indicators that software needs to calculate. In this case, the coefficient of emotionality, Trager index, certainty of action, directivity, vocabulary diversity, aggressiveness, logical connectedness and embolism are shown.

Fig. 5. Test results

These indicators require a certain norm - a neutral index which is indicated for each coefficient in the Norm column. The next Value column shows the calculated coefficient of the text that is entered by the person. The description of the value of this indicator is given according to the result (value) and abnormality. For example, if the norm of text emotionality is equal to 1 and the index of the analyzed text is 3.55, it means that the emotionality is increased. The norm of the vocabulary diversity is 0.7 and is equal to 0.75 in the text, which is within the norm. It indicates well developed imagination and a high level of verbal intelligence. Normal coefficient of aggression is 0.6, and the person shows 0.15, which means the absence of aggression and ambitions. The Trager index is twice the norm, which indicates emotional instability and dynamism (sudden mood swings, etc.). The coefficient of directivity correlates with the measure of aggression, which means that they are psychologically interdependent. They are calculated independently, but they validate each other and the result itself - the object is not prone to activity. The coefficient of embolism indicates the overall speech culture. In this case it is quite low, and the selected words contain only exclamations, but not profanity or irrelevant repetitions. That is why we can speak about the fluency of speech, the high level of verbal intelligence of the person, who is the author of the text.

All pages (registration, introduction, test, results and feedback) as well as the content are open to editing and improvement, since the project of informative system of Thematic Apperception Test in the Ukrainian language is a wide field of research of different specialists, and therefore new discoveries will be made in the fields of psychology, psychodiagnostics, psycholinguistics, and text linguistics. That is why both the project and the site are open for improvements according to the innovative discoveries in these and related scientific fields.

8   Conclusions

The theoretical foundations of psycholinguistics, discourse analysis, text linguistics and psycholinguistic analysis of linguistic material have been considered. The investigations of national and foreign scientists in these fields as well as examples of practical use of psycholinguistic text analysis have been described. During the analysis of theoretical aspects, a list of coefficients necessary for the evaluation of the text for psychodiagnostic purposes has been formed. The analysis of projective techniques revealed the need for the information system for a thematic apperception test for the needs of Ukrainian-speaking consumers. That is why it is advisable to create an information system of thematic apperception test that will use the similar principles of text analysis (for example LIWC) and coefficients, will involve previous and further work with a psychologist and will be adapted to the Ukrainian language.

Information systems provide the development of various models, so UML tools have been used to create a model of functioning of the Thematic Apperception Test information system. The model of an information system usage has been created for the needs of the customer.

Since the core of the whole information system is psycholinguistic analysis of the text, a detailed analysis of the coefficients and all components of the calculations has been carried out. The presented coefficients have been calculated on the basis of the text written during Thematic Apperception Test, and conclusions, based on the results, have been made.

Thus, it has been proved that the calculation of psycholinguistic indicators plays an important role and can provide assistance in reaching correct conclusions about psychoemotional state of a person.

Based on psycholinguistic research, a project of an informative system of Thematic Apperception Test has been created. The project involved the creating of the test environment as well as of the database to store data. The database has been created using Microsoft Access 2016, *Psychohelp* and *Wix*, a website builder.

The website and database show the results of information system design and are the basis for the implementation of Thematic Apperception Test in the Ukrainian language. Both website and database are available for further improvements and updating, because new investigations in the fields of psycholinguistics and psychology will provide more opportunities for text analysis, and therefore the results of these investigations can be immediately implicated in the project.

The designed information system is to be a helpful tool for scientific researchers conducted in linguistics, psychology, and psycholinguistics, as well. This tool is easy to improve according to the new scientific findings. The information system has a practical value, too, it is a mean to demonstrate the possibilities and peculiarities of text analysis for students of linguistics at *Mathematical linguistics* course, among others.

References

1. Kuranova, S. I.: Fundamentals of psycholinguistics. Academia, Kyiv (2012).

2. Osgood, Ch.: Psycholinguistics, Cross-Cultural Universals, and Prospects for Mankind. Praeger Publishers (1988).
3. Nytspol, V.: Conceptual metaphor in the serial killer character's discourse in the 20th-century American fiction. In: European Applied Sciences, vol. 4, pp. 52-54 (2017).
4. Kozachenko, Yu.: Psycholinguistic analysis of the internal monologue of a hero of a drama (based on Bertolt Brecht's drama "The Life of Galileo"). In: The trajectory of science, vol. 2, No. 12. pp. 4.1-4.6 (2016).
5. Yakovyna, V., Peleshchyshyn, A., Albota, S.: Discussions of Wikipedia Talk Pages: Manipulations Detected by Lingual-Psychological Analysis. In: Proceedings of the 1st International Workshop on Control, Optimisation and Analytical Processing of Social Networks (COAPSN-2019), Lviv, Ukraine, May 16-17, 2019, CEUR Workshop Proceedings, vol. 2392, pp. 309-320 (2019).
6. Korobiichuk, I., Syerov, Y., Fedushko, S.: The Method of Semantic Structuring of Virtual Community Content. In: Szewczyk, R., Krejsa, J., Nowicki, M., Ostaszewska-Liżewska, A. (eds) Mechatronics 2019: Recent Advances Towards Industry 4.0. MECHATRONICS 2019. Advances in Intelligent Systems and Computing, vol 1044. Springer, Cham (2020).
7. Lytvyn, V., Pukach, P., Bobyk, I., Vysotska, V.: The method of formation of the status of personality understanding based on the content analysis. In: Eastern-European Journal of Enterprise Technologies, vol. 5/2(83), pp. 4-12 (2016).
8. Chyrun, L., Vysotska, V., Kis, I., Chyrun, L.: Content Analysis Method for Cut Formation of Human Psychological State, Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP 2018, pp. 139-144 (2018).
9. Lytvyn, V., Vysotska, V., Rzheuskyi, A.: Technology for the Psychological Portraits Formation of Social Networks Users for the IT Specialists Recruitment Based on Big Five, NLP and Big Data Analysis. In: CEUR Workshop Proceedings, vol. 2392, pp. 147-171 (2019).
10. Chyrun, L., Kis, I., Vysotska, V., Chyrun, L.: Content monitoring method for cut formation of person psychological state in social scoring. In: Proceedings of the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies CSIT 2018, pp. 106-112 (2018).
11. Dilai, M., Levchenko, O.: Discourses surrounding feminism in Ukraine: a sentiment analysis of Twitter data. In: Proceedings of the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies CSIT 2018, pp. 47-50 (2018).
12. Vasyliuk, V.: Psycholinguistic text analysis for evaluation of person's emotional state. In: Abstracts of the 2nd International scientific and practical conference. SPC "Sci-conf.com.ua", pp. 427-430, Lviv (2019).
13. Shestakevych, T., Pasichnyk, V., Kunanets, N., Medykovskyy, M., Antonyuk, N.: The Content Web-Accessibility of Information and Technology Support in a Complex System of Educational and Social Inclusion. In: Proceedings of the 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies CSIT 2018, pp. 27-31. Lviv (2018).
14. Shestakevych, T., Pasichnyk, V., Nazaruk, M., Medykovskyy, M., Antonyuk, N.: Web-Products, Actual for Inclusive School Graduates: Evaluating the Accessibility. In: Shakhovska N., Medykovskyy M. (eds) Advances in Intelligent Systems and Computing III. CSIT 2018. Advances in Intelligent Systems and Computing, vol 871. Springer, Cham. (2019).