UDC 004.912
https://doi.org/10.31891/csit-2022-4-2

OLESIA BARKOVSKA
Kharkiv National University of Radio Electronics

PERFORMANCE STUDY OF THE TEXT ANALYSIS MODULE IN THE PROPOSED MODEL OF AUTOMATIC SPEAKER'S SPEECH ANNOTATION

The global spread and use of remote and online learning systems at various educational levels puts forward a number of requirements for existing systems and needs for expansion of functionality. The current problem in Ukraine is the unstable operation of the energy infrastructure due to frequent hostile shelling, so it is problematic for residents of Ukraine to join online classes on time, to listen to lectures by lecturers and teachers completely, to take part in conferences and master classes in full. This determines the need to provide the opportunity of familiarization with educational materials at a convenient time in a form convenient for understanding and mastering. The lecture recording provides access to audio files that are intended for listening, but are not intended for printed reproduction. Therefore, the expansion of existing digital educational platforms with the possibility of forming an annotation (summary, abstract) of a lecture and presenting it in the form of text-and-graphic materials for further use by course students on paper media is an urgent task and can improve the quality assessment of a remote educational resource from the point of view of the content and methodological aspect. The aim of the study is to create a generalized hybrid model of automatic annotation of the speaker's speech, which provides for the possibility of recognizing the speech, transforming the available data into text and, at the last stage, summarizing the given text, keeping only the important meaningful part of a lecture. The desired aim was achieved due to the creation of a generalized hybrid model of automatic annotation of input audio data, taking into account the effectiveness and features of existing methods of automatic text annotation obtained after converting speech into text. The uniqueness of this study is the use of marker words at the stage of text summarization, as well as the comparison of the efficiency of data processing at different stages of operation of this model when using different hardware. The results of computational experiments on graphics processing units with the Turing architecture showed that when the scope of input data increases by almost 30 times, the time also increases proportionally, but the use of a more powerful graphics processing unit NVIDIA Tesla T4 gives an speedup of more than 2.5 times compared to the graphics processing unit NVIDIA GeForce GTX GPU 1650 Mobile for both English and Ukrainian languages. For texts in the Ukrainian language, the text compression obtained (the ratio of the word count of the input text array to the word count in the resulting annotation) is 89.7%, for English – 94.15%. The proposed use of marker words showed an increase in the logical connection of input information internally, but obliges speakers to use predefined marker words to preserve the structure of the annotation formed.

Keywords: annotation, text, input data, language, abstracting, calculation, graphics processing unit, summarization.

ОЛЕСЯ БАРКОВСЬКА
Харківський національний університет радіоелектроніки

ДОСЛІДЖЕННЯ РОБОТИ МОДУЛЮ АНАЛІЗУ ТЕКСТУ У ЗАПРОПОНОВАНІЙ МОДЕЛІ АВТОМАТИЧНОГО АНОТУВАННЯ ПРОМОВИ СПІКЕРА

Глобальне поширення та використання систем дистанційного та он-лайн навчання на різних освітніх рівнях висуває ряд вимог до існуючих систем та потребує розширення функціоналу. Проблемою сьогодення в Україні є нестабільна робота енергетичної інфраструктури через часті ворожі обстріли, тому, приєднуватися до онлайн занять вчасно, слухати повноцінні лекції лекторів та учителів, приймати участь у конференціях та майстер-класах у повному обсязі, жителям України є проблематичним. Це обумовлює необхідність забезпечити можливість ознайомлення із навчальними матеріалами у зручний час узручному для розуміння та засвоєння вигляді. Запис лекції забезпечує доступ до звукових файлів, які припускаються прослуховування, але не призначені для друкованого відтворення. Тому, розширення існуючих цифрових освітніх платформ можливістю формування анотації (резюме, реферату) лекції та подання її у вигляді текстографічних матеріалів для подальшого використання слухачами курсу на паперових носіях, є завданням актуальним та здатне підвищити оцінку якості дистанційного освітнього ресурсу з погляду змістовно-методологічного аспекту. Метою дослідження є створення узагальненої гібридної моделі автоматичного анотування промови спікера, яка надає можливість розпізнавання мовлення, перетворення наявних даних в текст і останнім етапом проведення сумаризації даного тексту, зберігаючи лише важливу змістовну частину лекції. Поставлену мету було досягнуто завдяки створенню узагальненої гібридної моделі автоматичного анотування вхідних аудіо даних, враховуючи ефективність та особливості існуючих методів автоматичного анотування тексту, отриманого після конвертації промови у текст. Новизною даного дослідження є використання слів маркерів на етапі сумаризації тексту, а також порівняння ефективності обробки даних на різних етапах роботи даної моделі при використанні різного апаратного забезпечення. Результати обчислювальних експериментів на графічних процесорах із архітектурою Turing показали, що при збільшенні обсягів вхідних даних майже у 30 разів, час також збільшується пропорційно, але використання більш потужного графічного процесора NVIDIA Tesla T4 дає прискорення більше ніж у 2.5 рази порівняно із графічним процесором NVIDIA GeForce GTX 1650 Mobile як для англійської, так і для української мови. Для текстів українською мовою отримане стиснення тексту (відношення кількості сліх вхідного текстового масиву до кількості слів в отриманій анотації) становить 89,7%, для англійської мови – 94,15%. Запропоноване використання слів-маркерів показало підвищення логічного зв'язку вхідної інформації між собою, але зобовязує спікерів використовувати попередньо визначені слова-маркери для збереженн структури сформованої анотації.

Ключові слова: анотування, текст, вхідні данні, мова, реферування, обчислення, графічний процесор, сумаризація

Introduction

Information presented in text form is a valuable source of knowledge; however, it often needs to be effectively processed to get as much benefit as possible. Every year, the issue of creating an annotation (summary,

abstract) becomes more and more relevant [1, 2, 3]. For this purpose, it is necessary to compress text fragments to a shorter version, reduce the amount of the initial text while preserving key informational elements and content at the same time. Since it is a time-consuming and, as a rule, labor-intensive task to make annotation manually, the issue of automating this process is becoming increasingly popular in academic research.

An important component of the information space for remote education remains online lectures with experts, which can be held as a Q&A session and deal with questions from course students. Further access to online lecture materials should be convenient for understanding and mastering [4]. The lecture recording provides access to audio files that are intended for listening, but are not intended for printed reproduction.

Therefore, the expansion of existing digital educational platforms with the possibility of forming an annotation (summary, abstract) of a lecture and presenting it in the form of text-and-graphic materials for further use by course students on paper media is an urgent task, since it can improve the quality of presentation of educational information and the conditions of working therewith, as well as improve the quality assessment of a remote educational resource from the point of view of the content and methodological aspect [5].

That is why one of the areas of research is speech processing and conversion of audio files into text material, while keeping only important and relevant information. The key challenges include topic determination, interpretation, abstract generation, and its quality assessment. The most important tasks include identifying key phrases and using them to select sentences that will be included in the annotated text.

Text abstracting is the task of compressing a text fragment into a shorter version, reducing the amount of the original text while preserving key informational elements and content at the same time. Since manual text summarization is a time-consuming and, as a rule, labor-intensive task, the issue of automating the task is becoming increasingly popular and therefore is a strong motivation for academic research [6, 7].

There are important text summarization tasks related to NLP, such as classification of texts, answering to questions, summarization of legal texts, summarization of news and generation of headings. In addition, summarization can be integrated into these systems as an intermediate stage that contributes to reducing the length of a document [8].

In the age of big data, there has been an explosion in the amount of textual data from various sources. This text length is an invaluable source of information and knowledge that should be effectively summarized to be useful. The growing availability of documents requires comprehensive research in the domain of natural language processing for automatic text summarization. Figure 1 shows a diagram of a typical text summarization workflow.

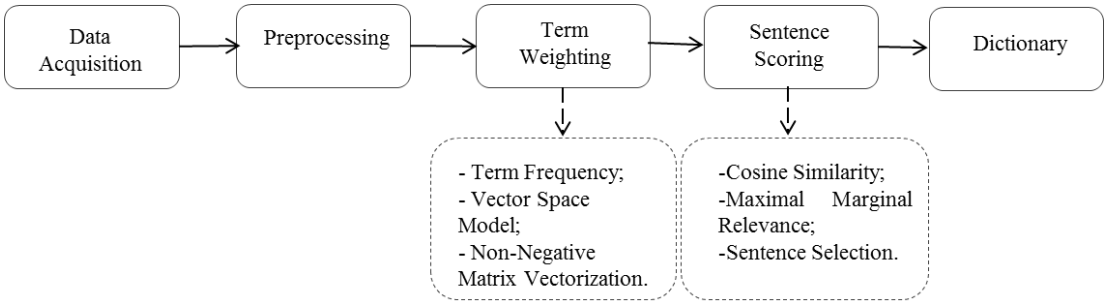

Fig. 1. Text summarization workflow

Most of existing approaches to text summarization model the problem as a classification problem that decides whether to include a sentence in the summary or not. Other approaches have used information on the topic, latent semantic analysis (LSA), sequence-to-sequence models, reinforcement learning and adversarial processes.

Related works. Research task rationale

The relevance of the work described above consists in increasing the efficiency and relevance of forming an annotation of a speaker's (lecturer's, expert's, teacher's, etc.) speech. The analysis of existing methods of annotating text data also proves the existing interest in NLP methods and in particular in methods of text summarization when performing academic research. The analysis of the problem area has shown that there are two general approaches to automatic abstracting (Table 1):

extraction (extractive approach) [9]. When extracting, the content is extracted from the input data, but the extracted content is not changed in any way. The methods of this approach characterize the existence of a function of evaluation of the importance of information block. As a rule, the importance of a sentence is determined by the importance of the words therein;

abstraction (abstractive approach) [10.]. Abstractive methods build an internal semantic presentation of the original text, and then use this representation to create an abstract. It involves the generation of new words and phrases that do not appear in the input text to report the most useful information from the original text.

Table 1

Generalization of analysis of the features of extractive and abstractive approaches to abstracting

	Extractive method of abstracting	**Abstractive method of abstracting**
Advantages	More simple than the abstractive approach, since it is based on copying pieces of the input text based on the determination of key phrases; it is easier to ensure basic levels of grammar and accuracy.	provides for application of additional knowledge due to the use of deep learning; the resulting abstract is closer to an abstract that can be generated by a human, since it uses a semantic analysis of the entire input text.
Disadvantages	does not provide for paraphrasing, inclusion of additional knowledge for high-quality summarization.	requires deep knowledge of the developer in the domain of artificial intelligence and computer linguistics.

At present, there are some high-tech solutions from different companies, but each of them has its advantages and disadvantages as discussed below (Table 2).

Dragon Anywhere is voice recognition software. This solution allows the user for dictating large documents without limitation on the time of dictation or numbers of pages. If a mistake is made during dictation, there is an option to correct it or edit the previous sentence using simple voice commands, such as “correct”. The correction menu that appears will provide a contextual list of alternative phrases to choose from.

Table 2

Overview of existing solutions in the domain of STT and annotation of text arrays

	Dragon Anywhere	**Amazon Transcribe**	**QuillBot**
Advantages	high accuracy of voice recognition (~ 99%); no word count limit; several ways to exchange documents.	high accuracy of voice recognition; possibility of interaction with other solutions of the Amazone ecosystem.	there is an option to add a browser extension; ease of use.
Disadvantages	lack of text summarization option; possibly cutthroat prices; it may take time to learn the built-in commands.	high cost of use; lack of the text summarizing option (there is an option of using separate modules, which will lead to an increase in the cost of use); an understanding of the AWS ecosystem is required.	works only with English language; lack of ability to dictate the text; has limitations when using the free version.

Amazon Transcribe is an automatic speech recognition service that makes it easy to add speech-to-text options to any application. Transcribe functions allow for obtaining audio, creating and reviewing easy-to-read transcripts, improving accuracy with customization and filtering content to ensure customer privacy.

QuillBot is a paraphrasing and summarizing tool that helps millions of students and professionals to reduce their time of writing by more than half by using the most advanced AI to rewrite any sentence, paragraph or article. It has both free and premium version. There is also access to use the API.

Aims and tasks of the work

The aim of the study is to create a generalized hybrid model of automatic annotation of the speaker’s speech, which provides for the possibility of recognizing the speech, transforming the available data into text and, at the last stage, summarizing the given text, keeping only the important meaningful part of a lecture.

Since the reliability of information contained in the educational resources of remote courses is one of the key requirements for digital educational platforms, cutting down the emergence of false or distorted data during the conversion of audio sequence into text data for further semantic analysis is the primary aim of the work.

The uniqueness of this study is the use of marker words at the stage of text summarization, as well as the comparison of the efficiency of data processing at different stages of operation of this model when using different hardware [11, 12, 13].

To achieve the desired aim, the following tasks should be solved:

- ✓ creation of a generalized hybrid model of automatic annotation of input audio data;
- ✓ review and analysis of existing methods of automatic annotation;
- ✓ adaptation of input text annotation methods for different computing architectures;
- ✓ evaluation of the timing of operation of the text analysis module;
- ✓ analysis of the results obtained.

Results and Discussion

The paper proposes a generalized hybrid model of automatic annotation of input audio data (Figure 2).

The automatic speech recognition (ASR) module accepts input of sound recording in WAV format, cleans the audio sequence using a deep neural network, and converts the cleaned audio sequence into text [14, 15].

The text analysis module accepts input of the deliverables from the speech recognition (ASR) module in the form of a JSON object. Text filtering takes place at the stage of transition of the JSON object from the ASR module

to the text summarization module. Next, the selection of key characteristics of the text and the extraction of the most significant fragments of the text using the mT5 model (pre-trained multilingual transformer for 101 languages), which is an extension of the Text-to-Text Transfer Transformer (T5) model.

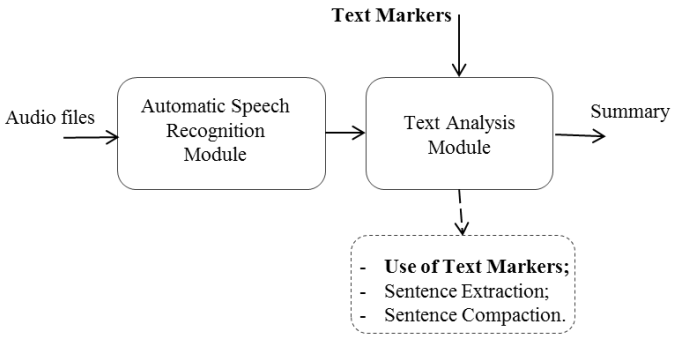

Fig.2 . Generalized hybrid model of automatic speaker's speech annotation

This solution was trained in 101 languages on a corpus of Common Crawl web pages, and supplemented with the XL-SUM dataset (covering 45 languages, highly abstract, concise and high-end as evidenced by human and internal evaluation). The data in different languages was sampled so that the balance between rare and popular web page languages could be adjusted.

When presenting the experiments, the results of summarizing texts for the Ukrainian and English languages were studied. Please find the results of benchmarking the evaluation of XL-SUM test sets according to the ROUGE metric in Table 3.

Table 3

Benchmarking of XL-SUM test sets

Language	ROUGE-1	ROUGE-2	ROUGE-3
English	37.601	15.1536	29.8817
Ukrainian	23.9908	10.1431	20.9199

The paper proposes an idea for creating a service for filtering the input text using marker words as described below.

The primary idea of the Text markers sub-module is to break the input text into fragments. That is, a separate json file is created with the "chopped" text between two word markers. This functionality creates an option of abstracting the separate fragments of the text without mixing unrelated information, which can lead to the loss of the sense of information. After the abstracting stage, the data received is "glued" into one document of the following type: "Word marker: content". According to the study conducted, this improvement is aimed at increasing the logical connection of input information internally.

Please find the scheme of performance of this solution in Figure 3. The sequence of stages of the text analysis module with modification by adding text markers is as follows: text and text markers are input, the text is parsed and cut into "pieces", then for each "piece" separately the summarization process is launched, and the process finishes with the stage of gluing the data into a single document.

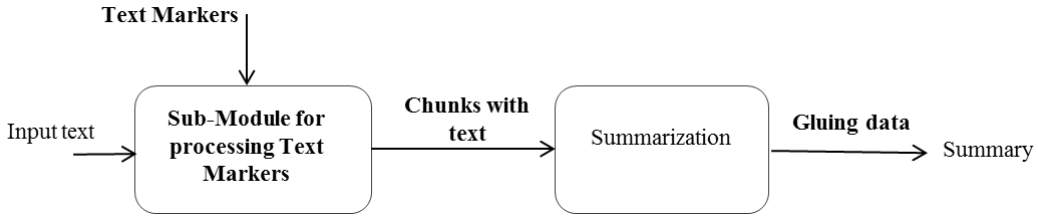

Fig. 3. Scheme of modification of the text analysis module due to the use of text markers

The performance evaluation of the text analysis module demonstrated quite high results for the task of abstracting texts in the Ukrainian and English languages. This approach to summarization uses the abstractive method using parallel computing.

Please find the results of comparison of the input and output text in Table 4. The text obtained from the ASR module is used for comparison.

Table 4

Comparison of text compression for summarization completed

Language	Initial word count	Final characters count	Text compression, %
Ukrainian	136	14	89.7
English	188	11	94.15

Please find below the examples taken for comparison of the performance of the summarization module in the table above.

In general, it is necessary to point out the quite high quality of text abstracting even in Ukrainian. It should be noted that study in the NLP domain for text summarization has been conducted for a relatively long time for many languages, but the leader in terms of high rates is English.

Computational experiments have been conducted with the use of computers with different performance. The following hardware has been used as an available estimator on a personal computer – central processor Intel Core i7-9750H (2.6-4.5 GHz), graphics processing unit NVIDIA GeForce GTX 1650 Mobile. The characteristics of the hardware on the remote cloud solution are as follows – central processor Intel Xeon 2.30GHz, graphics processing unit NVIDIA Tesla T4.

Please find the time spent by the text analysis module for processing the input data in the Ukrainian language in Table 5.

Table 5

Comparison of the time spent on summarization of the text in Ukrainian

Word count	Time spent on a personal computer (Ukrainian), sec	Time spent on a cloud solution (Ukrainian), sec
25	0.4	0.37
136	3.2	0.92
725	13.6	5.13

According to the results shown in the diagram (Figure 4), there is a trend to increase in the data processing time with the growth of the text dictionary, which is the expected result. It is possible to see a time reduction in data processing for a more powerful graphics card, namely NVIDIA Tesla T4, as compared to NVIDIA GeForce GTX 1650 Mobile PC graphics card.

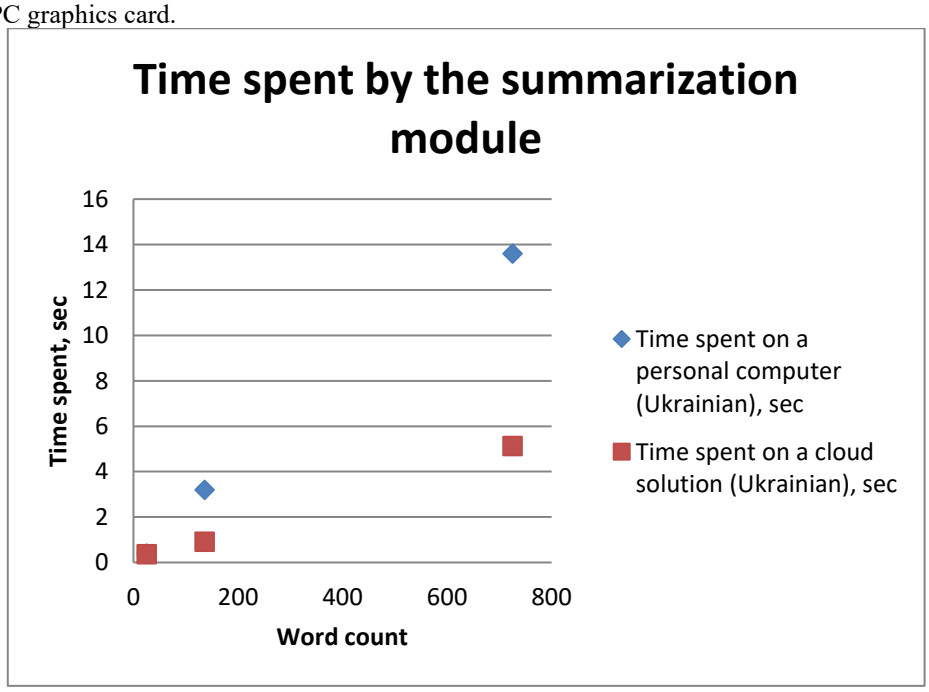

Fig. 4. Diagram of time spent for processing text in Ukrainian

Please find the time spent by this module for processing input data in English in Table 6.

Table 6

Comparison of the time spent on summarizing the text in English

Word count	Time spent on a personal computer (English), sec	Time spent on a cloud solution (English), sec
32	0.38	0.34
187	3.3	0.94
713	12.7	4.98

According to the results obtained, a diagram was built demonstrating the time reduction in data processing for the Ukrainian language. In the same way as in the case above, using a more powerful graphics card can show a significant increase in data processing for larger text content. Please find the diagram in Figure 5.

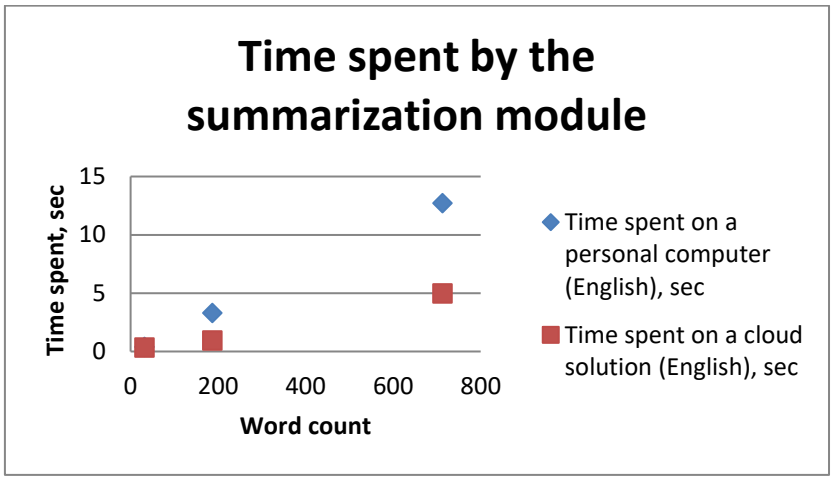

Fig. 5. Time consumption diagram for processing text in English

Therefore, we can make a conclusion on the effectiveness of speeding up data processing in this module with a more powerful video card. For the Ukrainian language, there is a significant speedup with a larger scope of input data, namely if we take into account the test results of 136 and 725 input words, the average speedup will be about 34%. Based on the results for 725 input words, the more there are input words, the higher is the speedup from a more powerful graphics card.

As to the processing of text in English, the result is slightly faster and the speedup is observed as well for a more powerful video card. The average speedup result for the input of 187 and 713 is 34% as well.

The studies conducted to improve the performance of the text analysis module due to the use of text markers proves that the solution developed compensates for the problem of the loss of context of a document and additionally with the use of parallel computing, does not critically load the system due to the distribution of independent annotation for selected text "pieces" that have been separated by the user with pre-determined text markers. However, the proposed approach creates a limitation for the speaker, namely it compels the speaker to use marker words.

Conclusions

The expansion of existing digital educational platforms with the possibility of forming an annotation (summary, abstract) of a lecture and presenting it in the form of text-and-graphic materials for further use by course students on paper media is an urgent task and can improve the quality assessment of a remote educational resource from the point of view of the content and methodological aspect. The aim of the study is to create a generalized hybrid model of automatic annotation of the speaker's speech, which provides for the possibility of recognizing the speech, transforming the available data into text and, at the last stage, summarizing the given text, keeping only the important meaningful part of a lecture. The desired aim was achieved due to the creation of a generalized hybrid model of automatic annotation of input audio data, taking into account the effectiveness and features of existing methods of automatic text annotation obtained after converting speech into text. The uniqueness of this study is the use of marker words at the stage of text summarization, as well as the comparison of the efficiency of data processing at different stages of operation of this model when using different hardware. The results of computational experiments on graphics processing units with the Turing architecture showed that when the scope of input data increases by almost 30 times, the time also increases proportionally, but the use of a more powerful graphics processing unit NVIDIA Tesla T4 gives an speedup of more than 2.5 times compared to the graphics processing unit NVIDIA GeForce GTX GPU 1650 Mobile for both English and Ukrainian languages. For texts in the Ukrainian language, the text compression obtained (the ratio of the word count of the input text array to the word count in the resulting annotation) is 89.7%, for English – 94.15%. The proposed use of marker words showed an increase in the logical connection of input information internally, but obliges speakers to use predefined marker words to preserve the structure of the annotation formed.

Further research and improvement of the proposed generalized model of automatic annotation of the speaker's speech is the possibility of deploying this model in cloud solutions such as Amazon Web Services or Google Cloud Platform to prevent data loss in the event of war or natural disasters. Cloud solutions ensure the reliability of data storage and processing due to the creation of data snapshots and replication thereof between servers that have different geographical locations. And additionally in the event of unavailability of the necessary hardware – the use of dedicated capacities of cloud solutions.

References

1. N. Bharti, S. N. Hashmi and V. M. Manikandan, "An Approach for Audio/Text Summary Generation from Webinars/Online Meetings," *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)*, 2021, pp. 6-10, doi: 10.1109/CICN51697.2021.9574684.
2. M. Kirmani and A. K. Shukla, "Systematic review of methods used in text summarization," *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2022, pp. 1048-1052, doi: 10.1109/ICACITE53722.2022.9823831.
3. Zhou, M., Duan, N., Liu, S., & Shum, H. Y., "Progress in neural NLP: modeling, learning, and reasoning*", Engineering*, 6(3), 275-290, https://doi.org/10.1016/j.eng.2019.12.014
4. Barkovska, Olesia, Viktor Khomych, and Oleksandr Nastenko. "Research of the text processing methods in organization of electronic storages of information objects," *Innovative Technologies and Scientific Solutions for Industries*, 1(19), (2022), https://doi.org/10.30837/ITSSI.2022.19.005
5. Barkovska, O., Pyvovarova, D., Kholiev, V., Ivashchenko, H., & Rosinskiy, D., "Information Object Storage Model with Accelerated Text Processing Methods", *In COLINS,* 2021 (April), pp. 286-299.
6. Ramezani, Majid, and Mohammad-Reza Feizi-Derakhshi. "Automated text summarization: An overview." *Applied Artificial Intelligence,*2014, 28.2, pp. 178-215.
7. Rahul, S. Adhikari and Monika, "NLP based Machine Learning Approaches for Text Summarization," *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC),* 2020, pp. 535-538, doi: 10.1109/ICCMC48092.2020.ICCMC-00099.
8. El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). "Automatic text summarization: A comprehensive survey", *Expert Systems with Applications*, 165p.
9. R. Mishra, V. K. Panchal and P. Kumar, "Extractive Text Summarization - An effective approach to extract information from Text," *2019 International Conference on contemporary Computing and Informatics (IC3I)*, 2019, pp. 252-255, doi: 10.1109/IC3I46837.2019.9055636.
10. D. Singhal, K. Khatter, T. A and J. R, "Abstractive Summarization of Meeting Conversations," *2020 IEEE International Conference for Innovation in Technology (INOCON)*, 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298305
11. 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*, 8(3), March 2020, pp.690-694, DOI:10.30534/ijeter/2020/13832020
12. Barkovska O., Pyvovarova D. and Serdechnyi V., "Pryskorenyj alghorytm poshuku sliv-obraziv u teksti z adaptyvnoju dekompozycijeju vykhidnykh danykh". [Accelerated word-image search algorithm in text with adaptive decomposition of input data], *Systemy upravlinnja, navighaciji ta zv'jazku,* 4 (56), 2019, 28-34. (in Ukrainian)
13. O. Barkovska, P. Rusnak, V. Tkachov and T. Muzyka, "Impact of Stemming on Efficiency of Messages Likelihood Definition in Telegram Newsfeeds," *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek),* 2022, pp. 1-5, doi: 10.1109/KhPIWeek57572.2022.9916415.
14. Alshemali, B., & Kalita, J., "Improving the reliability of deep neural networks in NLP: A review", *Knowledge-Based Systems*, 191, 105210.
15. I. Mykhailichenko, H. Ivashchenko, O. Barkovska and O. Liashenko, "Application of Deep Neural Network for Real-Time Voice Command Recognition," *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek),* 2022, pp. 1-4, doi: 10.1109/KhPIWeek57572.2022.9916473.

Olesia Barkovska **Олеся Барковська**	Ph.D., Associate Professor of Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine https://orcid.org/0000-0001-7496-4353 e-mail: olesia.barkovska@nure.ua	Доцент к.т.н., доцент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна