UDC: 519.8 https://doi.org/10.15407/jai2026.02.220

J. Pisarenko[1], **K. Karmazin**[2]
[1,2]V. M. Glushkov Institute of Cybernetics
of the National Academy of Sciences of Ukraine, Ukraine
40, Academician Glushkov ave., Kyiv, 03187
[1]newjulia1979@gmail.com
[2]kirillkarmazin2301@gmail.com
[1]https://orcid.org/0000-0001-8357-8614
[2]https://orcid.org 0009-0003-8852-8079

APPLICATION FOR AUTOMATIC SEMANTIC EXTRACTION FROM AUDIO AND TEXT INFORMATION

Abstract. The article considers the actual scientific and technical problem of developing an intelligent system for automatic summarization and analysis of text and audiovisual content. In the context of exponential growth of digital information volumes, there is an urgent need for tools that allow fast extraction of the content core from unstructured data. The authors propose a web application architecture based on the ASP.NET Core platform using modern natural language processing (NLP) models. Mechanisms for integrating cloud computing and intelligent APIs to ensure high accuracy of information compression are studied. The software solution is based on the client-server architecture with the implementation of real-time protocols for two-way data exchange. The paper details approaches to speech-to-text conversion with subsequent semantic analysis. The research results can be implemented in educational processes, business analytics, and government institutions to increase the efficiency of working with information flows.

Keywords: intelligent systems, natural language processing, automatic summarization, cloud services, neural networks, language models, ASP.NET Core, SignalR.

Introduction

The current stage of development of the global information society is characterized by an unprecedented intensification of the digital data accumulation. Enormous amounts of text, audio-, and video- information are generated every day, which creates the problem of "information overload" for specialists in various industries. This problem is particularly acute in the context of academic activities, journalism, management, and law, where the speed of processing large amounts of documents directly correlates with the efficiency of decision-making.

Traditional methods of manual analysis and note-taking of information no longer meet the requirements of the time due to their significant labor intensity and low speed. Therefore, the development of means of automating intellectual activity, in particular automatic summarization systems, is becoming one of the priority areas in the field of computer science and artificial intelligence. Modern advances in the field of deep learning and natural language processing (NLP) open up new opportunities for creating application systems that are capable of not only formally reducing text, but also interpreting its content, preserving the semantic integrity of the source material.

Problem statement

The main scientific problem is the need to create a holistic technological cycle that combines the stages of obtaining data from various sources (audio recordings, text documents, media streams), their initial processing (speech transcription), and further intellectual analysis to generate concise conclusions. Most existing solutions are fragmented: they either specialize exclusively in working with text or require significant computing resources to deploy local speech models. Thus, there is a need for a software package that would meet the following requirements:

1) ensuring multimodality, i.e. the ability to receive and process input data in various formats (text, audio-);

2) high quality semantic compression while preserving contextual connections;

3) the ability to work in real time via a web interface;

4) effective use of cloud architectures to minimize the load on user end devices.

Development of the concept of intelligent services of situational centers based on semantic analysis of information

The algorithm and software tool presented in this work for automatically extracting essence from audio and text data should be considered as an important stage in the development of the intellectual system "CONTROL_TEE" (Control of Technical-Ecological Events), the architecture of which is substantiated in detail in the works [1, 2, 3, 4]. The basic element of this system is a network of regional situational centers (RSCs), which act as coordination hubs in the "smart city" structure.

Real-time multi-agent system

Situation centers, drones, edge nodes, and streaming data client devices can be represented as agents [5]:

$$A = \{a_1, a_2, \dots, a_n\},$$

which interact through the state of the environment:

$$s_{t+1} = f(s_t, a_t^1, a_t^2, \dots, a_t^n).$$

The cooperative goal of the system is defined as ensuring consistency of actions (e.g., optimal allocation of data transmission frequencies, routing of data flows, management of traffic priorities), which is associated with rewards of various types.

A multi-agent architecture is proposed that promotes hierarchical management (central and regional centers) with decentralized decision-making at the levels of edge nodes, UAVs, and client devices (RL agents). This allows for scalable processing of large amounts of data in real time and adaptive response to changes in network and situational conditions. The multi-agent architecture scheme is shown in Figure 1.

According to the RSC concept, the main tasks are to ensure two-way communication between municipal services and performers (in particular, UAVs) to monitor the technogenic and ecological conditions of territories.

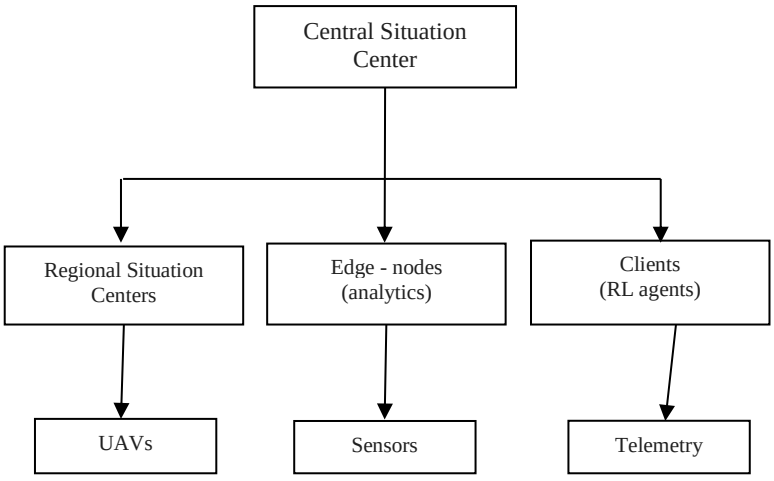

Fig. 1. Multi-agent architecture diagram

However, a critical challenge for such centers is processing exponential volumes of unstructured information. In this context, the proposed application acts as an intelligent filter and semantic core of the RSC:

1) *integration into the "smart city" ecosystem*: the proposed solution allows transforming RSCs from simple data collection centers into strategic decision-making centers. Semantic extracting from the operational reports, audio- messages from emergency services, or citizen testimonies allows the system operator to instantly receive a summary of the situation without wasting resources on studying complete data sets. This correlates with the information storage model described in [1], where data should not only be stored but also structured for quick access;

2) *educational and training aspect*: special attention in the materials [1] is paid to 3D training tools and personnel training for work in the "Control_TEE" system. The described application complements this learning paradigm. Introducing automatic

summarization tools into training courses allows future situation center operators to more quickly master technical documentation, operating instructions for complex systems, and crisis response protocols. The ability to generate short summaries from long lecture or instructional materials significantly increases the intensity and quality of the educational process within the functioning of the RSC;

3) *creating convenient services for the population*: within the paradigm of a convenient city, situational centers should ensure transparency and accessibility of services. A text analysis application can be used as a service component for citizens, enabling them to obtain concise conclusions about complex bureaucratic documents, environmental reports, or urban development plans stored in the RSC. Thus, the technology of automatic content extraction becomes a connecting link between the technical data of the monitoring system and the end user. So, if the works [1, 2, 3, 4, 5] define the organizational and architectural principles for creating situational centers, then this study offers a specific semantic analysis tool that fills these centers with real intellectual functionality.

Analysis of recent research and publications

In today's conditions of rapid growth in the volume of unstructured data, the issue of automated processing of multimedia and text information is becoming critically important for the scientific and business communities. Analysis of existing solutions in the field of summarization and transcription allows us to identify key vectors for the development of intellectual analysis technologies.

Among the most significant tools on the global market, the Sembly AI platform, developed by Aitera, stands out. This service is based on advanced speech recognition algorithms and automatic generation of accurate transcripts highlighting key moments of meetings. However, despite the high accuracy and speed of processing large amounts of data, the use of such intelligent systems is accompanied by significant costs, which limits their accessibility for small businesses.

An alternative approach is to use specialized machine learning platforms for text analytics, such as MonkeyLearn. This tool focuses on text classification and sentiment analysis through an intuitive interface. Similar functionality, but with a developer focus through advanced APIs, is provided by Aylien and TextRazor. The latter is noted for its high precision in semantic analysis and speech recognition. IBM also made a significant contribution to the development of the industry with its Watson Natural Language Understanding (NLU) service, which provides deep categorization and extraction of key phrases, integrating into large-scale corporate ecosystems.

For tasks of highly specialized transcription of audio information, the Otter.ai service has become widely used. Its advantage is the real-time ability to convert audio events into structured text documents with speaker identification.

Thus, despite the technological perfection of existing solutions, they have a number of significant disadvantages:

1) language barrier: most market leaders are focused on the English-speaking segment, leaving Ukrainian language support at the level of beta versions or ignoring its specifics, which leads to low quality of semantic extraction;

2) functional limitation: some services specialize exclusively in text or only in audio, which does not allow for a comprehensive approach to processing diverse sources of information within a single interface;

3) economic factors: The high cost of plans and the complexity of integration make these products inaccessible to a wide range of Ukrainian users.

Therefore, the current task is to develop a Ukrainian-oriented software product that will combine the capabilities of multi-format data processing with high-quality semantic analysis specifically for the Ukrainian language.

The purpose of the study

The purpose of this work is to design and practical implementation of an automated system based on modern cloud technologies and intelligent linguistic models for rapid extraction of essence from diverse information sources. Achieving the goal involves solving

problems related to the development of client-server architecture, configuring data transfer protocols, and validating the quality of the obtained information compression results.

Description of the software package and the results obtained

The developed system is built on a modular principle on the .NET platform. The main server component uses ASP.NET Core, which allows achieving high throughput and cross-platform compatibility. For real-time interaction with the user, the SignalR library has been implemented, which provides instant transmission of processing statuses and analysis results without the need for a full page refresh.

The logic of the application's operation is divided into several sequential stages:

1) initialization and loading: the user loads a file or enters text information through the client interface. The system validates the format and transfers the object to the server controller;

2) preprocessing: in the case of audio files, a speech recognition module is used. The program uses integration with cloud services (for example, OpenAI Whisper or similar), which ensure high transcription accuracy even in the presence of background noise;

3) semantic analysis and summarization: the received text array is passed to the analytical core. The API of modern large language models (LLM) is used. The system generates a specific query (prompt) that directs the model to isolate theses, key points and final conclusions;

4) post-processing and visualization: the generated result is structured in a readable form and transmitted to the client via a secure communication channel.

Special attention in the architecture is paid to security and scalability. The use of cloud computing (for example, AWS or Azure) allows the system to dynamically distribute the load. The use of Entity Framework Core provides reliable work with the database to save the history of user requests, which is critical for long-term information monitoring projects.

Software implementation of the system and algorithmic support

The software implementation of the developed VoxSumm application is based on the creation of a flexible web environment for automating the processes of obtaining informative conclusions from audio and text sources. The system architecture provides for integration with cloud artificial intelligence services, in particular OpenAI for semantic analysis and Deepgram for high-speed audio-to-text conversion.

The central element of the software implementation is the development of the atomic Summz object, which acts as a container for all data related to the process of extracting the essence: the initial source, intermediate transcripts, and final analysis results.

The algorithm for functioning and calculating resources is shown in Figure 2.

To ensure economic stability and scalability of the system, a mechanism for tokenizing transactions has been implemented. The process of processing each Summz object goes through a developed cost estimation algorithm, which is shown in Figure 2, the module's use case diagram is shown in Figure 3.

The general algorithm includes the following steps: 1) source type identification: the system determines whether the input data is text (e.g., copied text, .docx, .txt files) or audio information (e.g., files or recordings from a microphone); 2) audio preprocessing: if an audio file is available, its duration in minutes is calculated; 3) text array analysis: the system counts the number of characters and words in the text.

The module architecture is shown in Figure 4.

Mathematical model for cost calculating

Based on the defined parameters and established rates (for example, 500 tokens for 1 minute of audio and 1 token for 1 word in text), a formula was derived for calculating the approximate cost of processing the request in monetary equivalent.

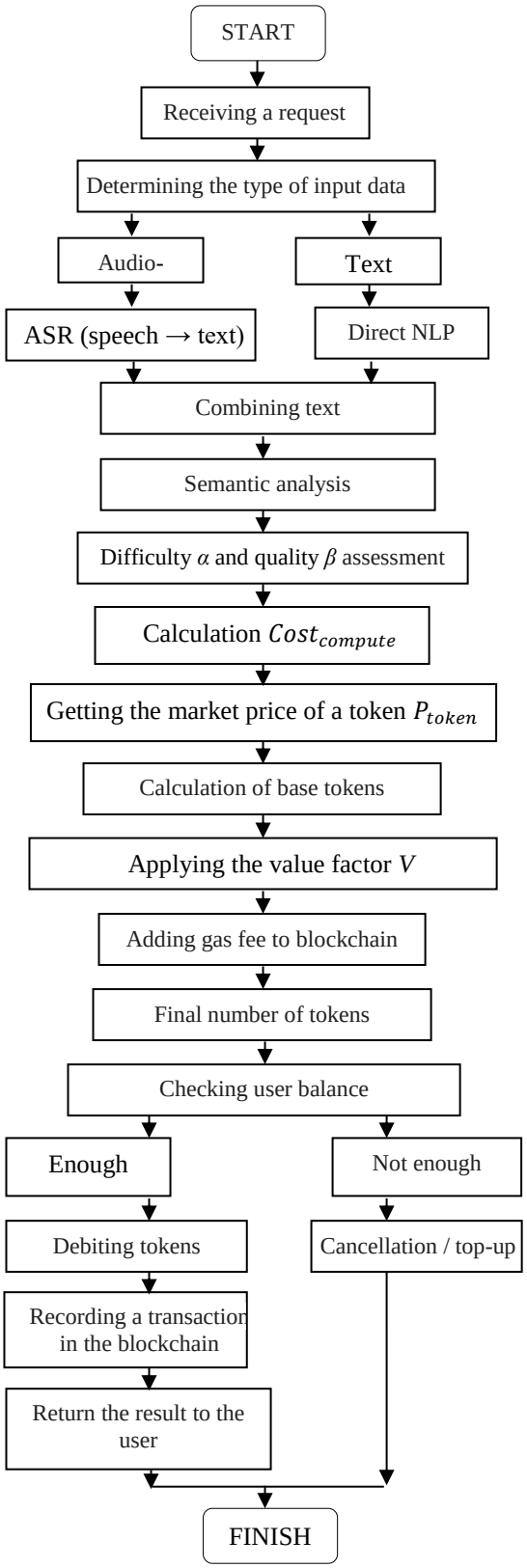

Fig. 2. Cost calculation algorithm

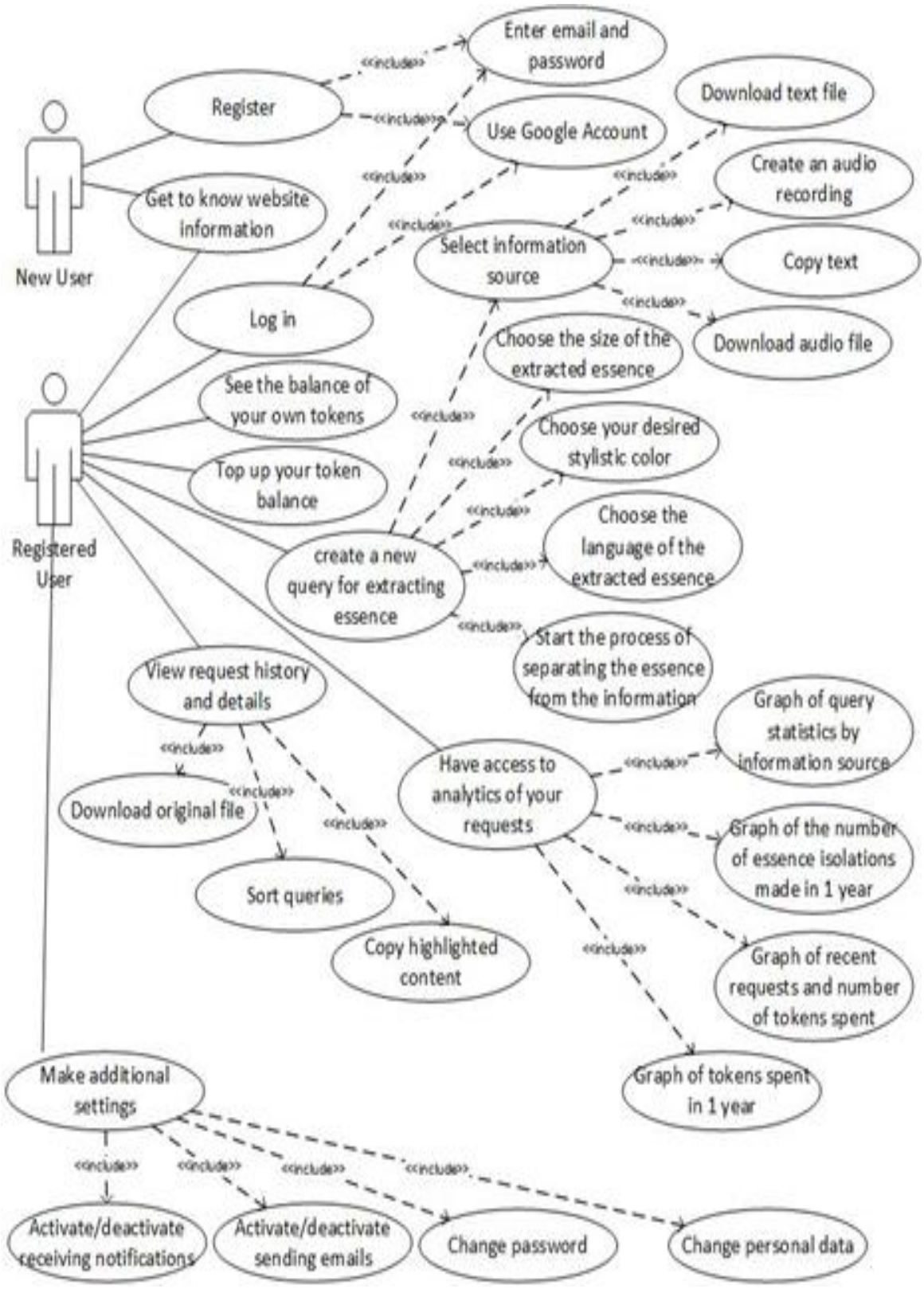

Fig. 3. Use case diagram

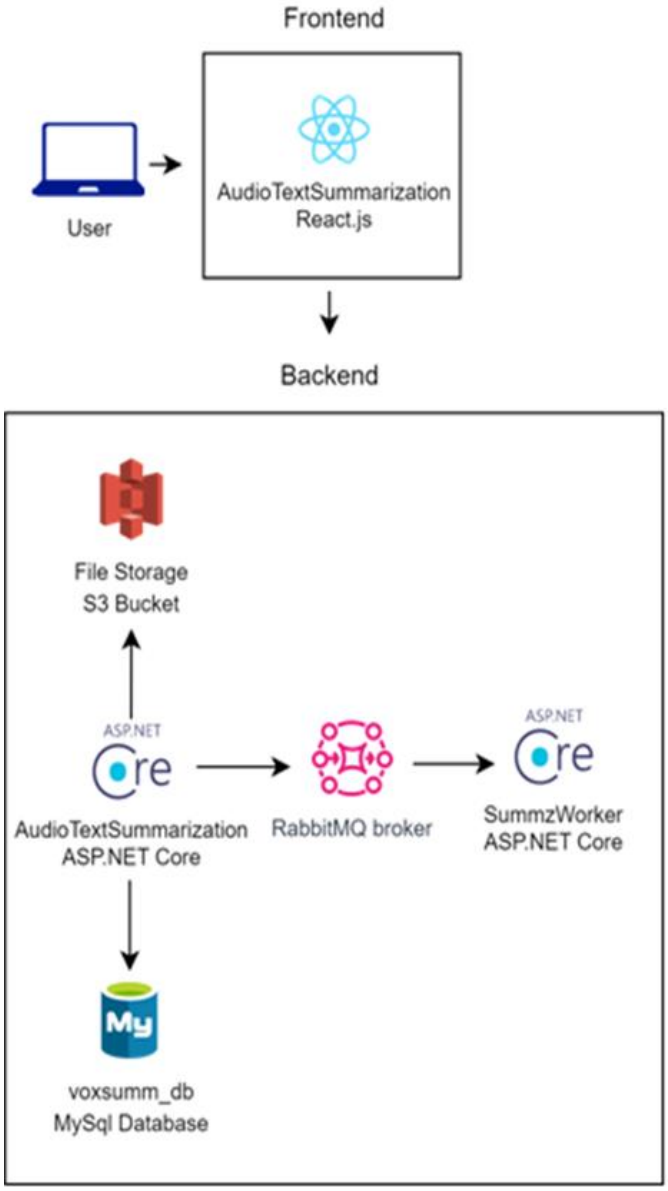

Fig. 4. Architecture of the module for extracting essence from audio and text sources

1. We denote the main mathematical variables: A – audio duration (second), T – text volume (tokents or characters), C_{asr} – cost of speech recognition (per second), C_{nlp} – cost of text processing (per token), C_{sem} – cost of semantic analysis (per token), α – complexity coefficient (language, noise, domain), β – quality coefficient (accurace, depth of analysis).

2. Computation cost (off-chain) audio → text:

$$Cost_{audio} = A \cdot C_{asr} \cdot \alpha .$$

Cost of text processing computation:

$$Cost_{text} = T \cdot (C_{nlp} + C_{sem}) \cdot \beta.$$

Total computation cost:

$$Cost_{compute} = Cost_{audio} + Cost_{text}$$

3. Cost in tokens.

Let: P_{token} – market price of the token, k – service margin (profit + infrastructure).

Cost in tokens:

$$Tokens_{required} = \frac{Cost_{compute} \cdot (1+k)}{P_{token}} .$$

4. Dynamic token pricing model (tokenomics).

To stabilize the system, incorporate supply and demand: $D(t)$ – demand (number of requests), $S(t)$ – token supply, γ - market sensitivity.

Price of the token:

$$P_{token}(t) = P_0 \cdot \left(1 + \gamma \cdot \frac{D(t)}{S(t)}\right),$$

where P_0 – base (initial) token.

5. Value-based Adjustment (value-based pricing). Semantic services vary in value:

simple summary ≠ legal analysis. Let: V – value coefficient (0,5 ÷ 5). Cost in tokens:

$$Tokens_{final} = Tokens_{required} \cdot V .$$

6. Blockchain costs.

Add: G – gas fee, L – network load. Вартість блокчейну:

$$Cost_{blockchain} = G \cdot L.$$

Thus, we finalize cost in tokens:

$$Tokens_{total} = Tokens_{final} + \frac{Cost_{blockchain}}{P_{token}}.$$

Let's write it down:

$$Tokens_{total} = \frac{[A \cdot C_{asr} \cdot \alpha + T \cdot (C_{nlp} + C_{sem}) \cdot \beta] \cdot (1+k) \cdot V}{P_{token}} + \frac{G \cdot L}{P_{token}} .$$

The described formula is integrated into the software module "Summz Value Calculator", which allows the user to estimate the required token balance in real time before starting complex computational operations.

The software interface is implemented according to the SPA (Single Page Application) principle using modern methods of DOM manipulation and interaction via REST API.

The system also supports CI/CD mechanisms to ensure continuous functionality updates and stability under high load conditions. For the convenience of users who do not have deep technical training, a system for saving query history and visual analytics in the form of resource usage graphs for time periods (month, last 20 queries, etc.) has been implemented.

Conclusions

To develop the concept of the intelligent system "Control_TEE" and deploy the RSC, a module was implemented for automatically extracting the essence from audio and text information.

A software solution for automating the processes of intelligent information processing has been designed and implemented.

Unlike existing tools, the proposed complex combines the capabilities of audio transcription and deep semantic text analysis in a single interface.

The main scientific and practical results of the work are:

1) justification for choosing an architecture based on ASP.NET Core and SignalR to create responsive intelligent adaptive analytics systems;

2) development of a methodology for combining various NLP services to achieve a cumulative effect in the quality of summarization;

3) proving the effectiveness of the cloud deployment model for systems that require significant computing power to operate neural networks.

Further research will be aimed at expanding support for specialized terminology dictionaries (medicine, law, technology) to increase the accuracy of analysis of narrow-profile texts, as well as optimizing the cost of using the API by implementing caching methods and choosing less energy-intensive models [6].

References

1. Pysarenko, Y. V., Melkumyan, K. O., Varava, I. V., Koval, O. S., & Chumakova, N. F. (2022). On the organization of regional situational centers of the intellectual system "CONTROL_TEA" using UAVs. Artificial Intelligence, (1), PP. 275-281. https://doi.org/10.15407/jai2022.01.275

2. Pisarenko, J., & Melkumyan, E. (2019). The structure of the information storage “CONTROL_TEA” for UAV applications. IEEE 5th International Conference on Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), 274–277. https://doi.org/10.1109/APUAVD47061.2019.8943938.

3. V.G. Pisarenko, N.V. Nogin, A.S. Kryachok, J.V. Pisarenko, I.A. Varava, A.S. Koval. (2022) About complex intelligent technologies for techno-ecological events control in the water area. Prombles in programming, 3-4, PP. 437-445. https://doi.org/10.15407/pp2022.03-04.437

4. V.G. Pisarenko, J.V. Pisarenko, O.E. Gulchak, T.I. Chobotok, A.G Boyko. (2021) Practical experience in the technical systems creating with the artificial intelligence elements. Artificial Intelligence, (1), PP. 95-101. https://doi.org/10.15407/jai2021.01.095

5. Писаренко Ю.В., Кармазін К.В. Використання ШІ для обробки та аналізу великих обсягів даних у реальному часі з інтеграцією мультиагентних систем // Artificial Intelligence. – 2025. – № 4. – PP.134-145 https://doi.org/10.15407/jai2025.04.134

6. Adam Freeman. Pro ASP.NET Core 6: Develop Cloud-Ready Web-Applications Usng MVC, Blazor, and Razor Pages. – Apress. – 2022. – 1253 p. ISBN 9781484279564

The article has been sent to the editors 20.05.26.
After processing 30.05.26.
Submitted for printing 30.06.26