Information System for Converting Audio Ukrainian-Language Text into Written one Based on NLP Methods and Machine Learning

Yuriy Tyshchuk¹, Victoria Vysotska¹², Mykhaïlo Luchkevych²

¹ Lviv Polytechnic National University, S. Bandera Street, 12, Lviv, 79013, Ukraine
² Osnabrück University, Friedrich-Janssen-Str. 1, Osnabrück, 49076, Germany

Abstract
Speech recognition provides various ways to analyse and process a user's recorded voice. It allows people to control different systems that are using one of the types of speech recognition. Speech-to-text conversion is also a type of speech recognition that uses unscripted conversational data for further processing. This system involves several steps to process an audio file using electro-acoustic tools, sound filtering algorithms to find only relevant sounds, electronic datasets for the chosen language, and mathematical models that find appropriate words for a list of phonemes. People whose professions are related to typing large amounts of text using the keyboard can significantly speed up, facilitate the work process and reduce stress using systems that convert Speech-To-Text. In addition, such systems help businesses because remote work is becoming increasingly popular, and companies need tools for translating recorded audio from meetings to text for further analysis and systematization. The work's study object is converting audio in the Ukrainian language into its textual representation using NLP methods and machine learning. The scope of the research is audio file processing algorithms for finding relevant sounds and recognizing phonemes, as well as mathematical models for identifying words using an array of found phonemes. The work aims to design and develop an information system for converting audio in the Ukrainian language into its textual representation. According to the developments and calculations presented in work, namely: analysis of algorithms, areas of application and review of analogue problems in the first stage, the system analysis of the information system in the second stage and analysis and selection of relevant technologies and software development tools in the third stage, the information system for converting audio in the Ukrainian language into its textual representation was implemented in the form of a web application called «Ukrainian Speech-to-text», which is a technology for accurate and easy analysis of Ukrainian-language audio files and their other transcription into text. The application supports uploading files from the file system, recording them using a microphone, and saving the analysed data. The system is ready for use.

Keywords
text-to-speech, speech recognition, Ukrainian-language, information system.

1. Introduction

With the development of technology, the improvement of algorithms, and the increase in computer work, society needs a fast and accessible voice control or translation of recorded voice into written text. Programs that implement such algorithms are beneficial for the elderly and people with disabilities and
for people who have to print a lot of text. The purpose of the work is to design and develop an information system to ensure the conversion of Ukrainian-language text into written. To do this, you must solve the following tasks:

- Investigate the sub-programs that implement such algorithms that benefit the elderly, people with disabilities, and people who have to print a lot of text. Analyze programs that implement such algorithms help the elderly, people with disabilities, and those who have to print a lot of text. Ject area, uses, algorithms and ready-made solutions, and then determine the approach to implementing the system.

- Make a systematic analysis of the goal, build a tree of goals, a matrix of pairwise comparisons, a context diagram and its decomposition, and a hierarchy of goals.

- Consider and select technologies and software to implement the information system.

- Implement an information system for converting audio Ukrainian-language text into written, using selected technologies.

- Justify the economic feasibility of developing an information system.

The object of research is converting a sound Ukrainian-language text into a written one. The research subject is the methods and algorithms used to transcribe audio into text. The practical significance of the obtained results lies in implementing the information system, which allows the ability to transcribe Ukrainian-language audio text into written. The developed method is tested and ready to use.

2. Related Works

2.1. Methodological principles of research

Language is the most powerful way of communication through which people express their thoughts and feelings. Features of speech differ in each language. However, even when communicating in the same language, the pace and dialect of each person are different. It makes it difficult for some people to understand the message being conveyed. Sometimes long speeches are also difficult to understand for reasons such as different pronunciations, tempo, etc. Language recognition is an interdisciplinary field of computational linguistics that assists in developing technologies that allow speech recognition and translation into text. This can be achieved by developing a speech recognition system: speech-to-text, which allows the computer to translate voice requests and dictation into text. AI speech-to-text is a field of computer science that specializes in giving the computer the ability to recognize and transcribe oral language into text. Speech-to-text differs from simple voice recognition because the software learns to understand and recognize spoken words.

In contrast, voice recognition software focuses on determining the voice patterns of individuals. A system that will translate speech into a text requires a combination of specially trained algorithms, computer processors and sound capture equipment (microphones). Trained algorithms parse a continuous complex acoustic signal into discrete language units called phonemes. A phoneme is the smallest single sound unit into which human speech can be broken [1]. These are the minimum sound units native speakers may perceive as different from creating significant differences between words. For example, English speakers understand that "though" and "go" are two other words because their first consonant sounds differ. However, their vowel sounds are the same. A language can have more or fewer phonemes than letters or graphemes. For example, there are only 26 letters in English, but some dialects contain 44 different phonemes. What further complicates the situation, the acoustic properties of this phoneme depend on the speaker and the context in which it sounds. For example, the sound "l" at the end of the word "ball" is acoustically closer to the vowel sound "o" than to the sound "l" at the beginning of the word "loud" in many English dialects. Algorithms that divide acoustic signals into phonemes must take into account the context. The workflow of the system for converting speech into text using AI consists of the following steps in [2]:

256

COLINS’2022, Volume II: Workshop. Gliwice, Poland, May 12-13, 2022, ISSN 2523-4013
http://colins.in.ua, online
1. A microphone captures sounds emanating from a person's mouth. Sounds are converted from analogue signals to digital files.
2. The software analyzes audio files bit by bit, down to hundredths of a second, looking for known phonemes.
3. Identified phonemes are run through a database of common words, phrases and sentences.
4. The software uses complex mathematical models to determine the most likely words/phrases of which the text is composed.

Anyone with a smartphone and an Internet connection can access speech-to-text software. The need for this feature has also grown sharply in the corporate and consumer markets. We can divide the primary demand for AI speech-to-text into the following sources [3]:

1. Customer service
   1. Many businesses rely on chatbots or AI assistants in customer service as a first tier to reduce costs and improve customer service. Because many users find the voice chat better and more comfortable, efficient and accurate text-to-text software can dramatically improve online customer service quality.
   2. AI chatbots with advanced speech recognition capabilities can reduce the burden on call centre workers. Acting as the first line of service, they can determine the intentions or needs of the interlocutor and redirect him to the appropriate service or resource.

2. Search for content
   1. Again, the growing use of mobile devices is increasing the demand for speech recognition algorithms. The number of potential users has increased dramatically due to the general availability of voice-to-text services provided free of charge on iOS and Android platforms.

3. Electronic documentation
   1. There are many services and areas where speech-to-text services will facilitate documentation. For example, doctors need it to more quickly and efficiently maintain medical records of patients and record diagnoses.
   2. Judicial systems and government agencies can use this technology to reduce costs and increase the efficiency of record keeping. Businesses can also use it during important meetings and conferences to keep minutes and other special needs.
   3. The COVID-19 pandemic of 2020 has revealed a new way to use speech-to-text technology. Due to many remote meetings and video conferencing, the audio-to-text feature allows companies to receive information, summarize discussions and obtain analytical data by recording conversations.

4. Content consumption
   1. Global content availability is a huge incentive for introducing speech-to-text technology. As online streaming displaces traditional forms of entertainment, the demand for digital subtitles is growing. Real-time subtitles have a vast market, as content is broadcast worldwide to viewers with different language levels.
   2. There is great potential for using text-to-speech and live entertainment technologies, such as sports broadcasts. Instant captions can be a revolutionary solution, increasing accessibility and user engagement.

Even the best automatic language recognition algorithms cannot achieve 100% accuracy. This figure is 95%, which Google Cloud Speech first completed in 2017. Numerous factors create this 5% error rate in the world's best text-to-speech services.

- Accents and dialects - even ordinary people often cannot understand what someone is saying in their native language because of local dialects and accents. Programming AI to detect all these nuances is a very difficult task.
- Context - homophones are words that have the same or similar sounds but different meanings. A simple example is "hospitable" and "hospitable". AI often has difficulty identifying these homophones in a sentence without learning these words in appropriate contexts.
- Input quality - background noise can greatly affect the AI's ability to convert speech to text accurately. If the user suffers from a disease such as a cold or sore throat, the changes they make to the language can often confuse the software.
- Visual signals - to convey a message, humans rely not only on the voice. Our words are often supplemented by facial expressions and gestures that reinforce and sometimes radically change the meaning of what is said. AI cannot decrypt these signals unless it is an advanced image and sound processing algorithm capable of analysing both data sets in video files.
- Low-resource languages - development of AI speech-to-text for languages that have less recorded data
- Mixing languages - in multilingual communities, people use words from several languages in one conversation. It creates difficulties for AI, as it must be able to handle lexical and grammatical models when switching between languages and use a wider overall set of possible models.

Mainly three algorithms are used for speech recognition [5]: The hidden Markov model (HMM), Dynamic time distortion and Artificial neural networks.

2.2. Hidden Markov Model (HMM)

The latent (hidden) Markov model (HMM) is a statistical model in which a Markov process with latent states predicts the modelled system. HMM can be represented as the simplest dynamic Bayesian network. The hidden Markov model is a set of states associated with transitions. It begins in a specific initial state. At each step of the discrete time, there is a transition to a new state, and then one source symbol is generated in this state. Transition selection and source symbol are random, controlled probability distributions. Fuel can be thought of as a "black box", where the sequence of source symbols generated over time is open, but the sequence of states visited over time is hidden from view. That is why it is called the hidden Markov model. When HMM is used for speech recognition, the state is interpreted as acoustic models that indicate which sounds are most likely to be heard during the respective speech segments. At the same time, transitions provide time constraints, meaning how states can follow one another in sequence. Because language always moves forward in time, transitions in a language application always move forward [5]. HMM-algorithms:
1. The forward algorithm helps recognize isolated words.
2. The Viterbi algorithm helps recognize continuous speech.
3. The direct-reverse algorithm helps learn HMM.

Restrictions on HMM:
1. Constant monitoring of personnel
2. Markov's assumption
3. Lack of formal methods for choosing the topology of the model
4. A large amount of training data is required
5. Assumptions about conditional independence

2.3. Algorithm for dynamic time scale transformation

The easiest way to recognize an isolated word pattern is to compare it with several stored word patterns and determine which one works best. Several factors complicate this task. First, different examples of this word will have slightly different durations. This problem can be solved simply by
normalizing the patterns and unknown speech so that they all last the same duration. However, another problem is that the pace of speech may be erratic. In other words, the optimal alignment between the template and the sample language may be nonlinear. The algorithm of dynamic transformation of the time scale is a well-known technique for finding the optimal alignment between two given (time-dependent) sequences under certain restrictions. Under certain limitations, it is intuitively clear that sequences deform non-linearly to match each other. Initially, this algorithm was used to compare different language models. The algorithm has been successfully used to automatically overcome temporal strains and different speeds associated with time-dependent data in areas such as data retrieval and information retrieval.

2.4. Artificial neural networks

An artificial neural network can be defined as a model of reasoning based on the human brain. The brain consists of an interconnected set of nerve cells called neurons, or basic information processing units. The human brain includes nearly 10 billion neurons and 60 trillion connections and synapses. The brain can perform its functions much faster than the fastest computers available today by using multiple neurons at once. Each neuron has a very simple structure, but a large number of such elements is huge computing power. A neuron consists of a cell body, a soma, several dendrites, and one axon. An artificial neural network consists of some very simple processors called neurons. They are analogous to biological neurons in the brain. Neurons are connected by suspended connections that transmit signals from one neuron to another. The output signal is transmitted through the output connection of the neuron. This connection is divided into several branches that transmit the same signal. Output branches end at the input connections of other neurons in the network.

The perceptron is the simplest form of neural network. It consists of a single neuron with adjustable synaptic weights and a rigid limiter. There are two types of artificial neural networks:

1. A direct propagation neural network is a network in which information moves in only one direction - forward from input nodes, hidden nodes (if any) and output nodes. There are no loops or loops in the network [6].
2. A recurrent neural network is a class of artificial neural networks in which connections between nodes form a time-oriented graph. It creates an internal state of the network, which allows it to exhibit dynamic behaviour over time.[7].

Advantages of artificial neural networks:

1. Artificial neural networks are nonlinear models that are easy to use and understand compared to statistical methods.
2. They are non-parametric models, while most statistical methods are parametric ones requiring a higher level of knowledge in the field of statistics.
3. Neural networks have several advantages, including the need for less formal statistical learning, the ability to detect complex nonlinear relationships between dependent and independent variables implicitly, the ability to detect all possible interactions between predictor variables, and the availability of multiple learning algorithms

3. Analysis of known methods and solutions

Nowadays, there are many APIs on the Internet for converting language into text. Some of them are available for free. Many others are available in SDKs and APIs designed for corporate clients. **Google Speech-to-text** (Fig. 1) [8], which supports more than 120 languages, is today the undisputed leader in language recognition. Voice search, audio-to-text transcription and other advanced services are available in many online Google services, such as Google Docs, Google Translate, etc.
IBM Watson (Fig. 2) [9] is another major player in the development of AI for language recognition. The AI of the Watson supercomputer is quite well known. It is an enterprise-oriented service with a wide range of applications, one of which is the conversion of speech into text. It currently supports 7 major languages.

**Figure 1:** Demo window view for Google Cloud Speech-To-Text

**Figure 2:** Demo window view for IBM Watson

Dragon Naturally Speaking (Fig. 3) has released one of the first products to convert speech into text in the consumer market. The company remains a reliable speech recognition software provider, especially among healthcare professionals. Modern versions of the software use advanced features of
artificial intelligence. In most cases, Dragon Anywhere, a cloud version of Dragon Naturally Speaking, allows you to convert speech to text on desktop and mobile devices. It is the best method for mobile devices. For users who need more advanced vocabulary, Dragon also offers Dragon Legal (recommended for law students) and Dragon Medical (recommended for medical students).

![Figure 3: Dragon NaturallySpeaking window view](image)

**Microsoft Dictate** (Fig. 4) [10] proves that even the best software for converting speech to text can be free and no worse than premium programs. Created in Microsoft Garage (a company division where employees work on their ideas in the form of projects), this multifunctional application uses the same advanced language recognition technology used in the virtual assistant Microsoft Cortana.

Dictate is a Microsoft Office add-on and works well with Word, PowerPoint and Outlook. Once installed, you can access it through the Dictation tab, which appears in the upper right corner of the toolbar. The application supports voice commands for most standard operations, such as entering or editing text, moving the cursor to a new line, and adding punctuation marks manually or automatically.

![Figure 4: Appearance panel for audio recording](image)

**Odrey** (Fig. 5) [11] can work both as a voice decoder and as a dictaphone. It means you can record a voice message using the app, play files or just download files to the app from your device. The application can transcribe the recorded language from both audio and video files.
Figure 5: Windows Odrey: a -- the view of the audio recording window; b -- view of the window with the list of broadcast texts

Table 1
Table of comparisons of the considered decisions

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Google Speech-to-text</th>
<th>IBM Watson</th>
<th>Dragon Naturally Speaking</th>
<th>Microsoft Dictate</th>
<th>Odrey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free version</td>
<td>Part</td>
<td>Part (500 minutes)</td>
<td>Part</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>File download support</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>File recording support</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Type of program</td>
<td>SDK</td>
<td>SDK</td>
<td>Desktop/mobile</td>
<td>Integrated with Microsoft Office</td>
<td>Mobile</td>
</tr>
<tr>
<td>Support for the Ukrainian language</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

A comparison table was created to understand better and analyze the considered solutions. It covered each product according to the following criteria: free version, file download support, file write support, program type and Ukrainian language support. According to this analysis, most products have Ukrainian language support. Still, they are all implemented as desktop, mobile, or SDK applications. For a better user experience, it was decided to create a web application because it does not require downloading and installing additional programs. It was also agreed that the developed information system would support two types of file processing: recording files using a microphone and downloading an audio file from the file system.

4. System analysis of the object of study

Systems analysis is a scientific method of cognition, a sequence of actions to establish structural relationships between variables or elements of the studied system. It is based on general scientific, experimental, natural, statistical, and mathematical methods. System analysis has become widespread in information technology, where the system is designed must be characterized in terms of its content. This method remains relevant at all stages of development, from the first ideas to the finished system.
The project life cycle begins with the transformation of an idea into specific technical requirements. The project involves a systems analyst who collects requirements from stakeholders.

Basic principles on which systems analysis is based [12]:

1. The focus of optimality - when designing a system, the main thing is to choose the most rational version of the system. From the view of systems analysis, the task is not to find a better solution than the current one but to find the best option possible.
2. The systematicity principle indicates the need to study objects and systems as a whole and consist of interconnected elements in the supersystems composition.
3. The principle of hierarchy - hierarchical construction is inherent in all systems. At the lower levels of the hierarchy, specific functions are depicted with a detailed description of the functionality. At the highest levels of the hierarchy, generalized information is used to implement the system's functions as a whole. It should also be noted that in almost any system. The hierarchical structure is not strict but is determined depending on the goals.
4. The principle of integration is a principle that indicates the need to combine individual elements and functions into one. This principle aims to study the integration patterns of systems that emerge by combining their parts.
5. The principle of the ultimate goal indicates the decisive role of the ultimate global goal in creating and studying systems.

The goal tree [13] is a visual graphic representation of the subordination and relationship of goals, which shows the division of the overall goal into sub-goals, objectives and individual actions. The main purpose of the goal tree is to achieve the goal. The main idea for building a "goal tree" is decomposition. Decomposition is a method of revealing the structure of the system, in which it is divided into separate components on one basis. The target tree for the system under development is shown in Fig. 6. The main goal is to create an information system for converting audio Ukrainian-language text into written. To accomplish this, you need to achieve the following levels of aim: "Theme Analysis", "System Design", and "Create a System". We need to follow these steps to achieve the goal of "Theme Analysis":

- Analyze specific solutions in the market. It will help identify any shortcomings to avoid them during development.
- Analysis of algorithms for creating a system will help determine the most optimal way to solve the problem.
- Analysis of available learning datasets will help speed up the system's learning and increase the accuracy with which the system translates speech into text.

The main quality criteria for this part was chosen "Relevance" - because nowadays, algorithms and approaches to solving problems of this kind are changing rapidly, and it is necessary to monitor research constantly. There are three sub-objectives for System Design:

- Construction of all necessary diagrams. It is needed to create class diagrams of type IDEF0, which will consider all the processes and hierarchy of the developed system. This chart is primarily needed to understand the primary goal and the procedures to be performed.
- Construction of interior architecture. It represents the architecture of modules, classes and infrastructure.
- Platform selection. You need to choose which platform is best for your system.

The criterion for the first two sub-goals is "Quality". Based on this, all diagrams should clearly describe the system being created and its processes, which will help develop a better product. "Practicality" means that the chosen platform should support the developed method to be practical for use. The last goal is "Creating a system". The developer will have a clear purpose, diagrams, and designed architecture at this stage. To achieve this goal, you need to complete the following sub-goals:

- Implementation of intermodular interaction. Since the created system will consist of different modules, the interaction between them must be flexible and fast. And also meet all the parameters of a good program.
• Software development. It is one of the most important stages because it is at this stage that the developer will implement the system using all the previous information.
• System testing. An equally important step during the software creation and after its release is testing all system functions.

The criteria for the last part are "Performance" and "Reliability". These criteria are very important because the product must work quickly and be reliable, i.e. tested.

**Figure 6:** The goals tree of the system of sound Ukrainian-language text conversion into a written one

After building the goal tree, the next step is to determine the type of information system. For this purpose, four main types of information systems were chosen: information-analytical, information-management, decision support system and information-advisory (Fig. 7). It is necessary to begin with a matrix of pairwise comparisons (table 2). Namely, we will define estimations of advantages of one criterion of the system over others.
Figure 7: Hierarchical structure of the problem of determining the type of IS

Table 2
Matrix of pairwise comparisons of criteria

<table>
<thead>
<tr>
<th>№</th>
<th>Criterion</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Rating</th>
<th>Vector priorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Topicality</td>
<td>1.0</td>
<td>2.0</td>
<td>1.0</td>
<td>2.0</td>
<td>1.32</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Practicality</td>
<td>1.0</td>
<td>2.0</td>
<td>0.5</td>
<td>1.0</td>
<td>1.0</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Quality</td>
<td>0.5</td>
<td>1.0</td>
<td>0.33</td>
<td>1.0</td>
<td>0.61</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Speed</td>
<td>1.0</td>
<td>3.0</td>
<td>1.0</td>
<td>3.0</td>
<td>1.78</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Reliability</td>
<td>0.5</td>
<td>1.0</td>
<td>1.0</td>
<td>0.33</td>
<td>0.7</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

Afterwards, pairwise comparison matrices were created for each of the selected types of information systems according to the quality criterion. These matrices are given in Tables 3-7.

Table 3
Matrix of pairwise comparisons for the Relevance criterion

<table>
<thead>
<tr>
<th>№</th>
<th>Type of information system</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Rating</th>
<th>Vector priorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Information and analytical</td>
<td>1.0</td>
<td>4.0</td>
<td>2.0</td>
<td>1.0</td>
<td>1.682</td>
<td>0.38</td>
</tr>
<tr>
<td>2</td>
<td>Information and management</td>
<td>0.25</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.11</td>
</tr>
<tr>
<td>3</td>
<td>Decision support system</td>
<td>0.5</td>
<td>2.0</td>
<td>1.0</td>
<td>0.5</td>
<td>0.84</td>
<td>0.19</td>
</tr>
<tr>
<td>4</td>
<td>Information and advisory system</td>
<td>1.0</td>
<td>2.0</td>
<td>2.0</td>
<td>1.0</td>
<td>1.41</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Table 4
Matrix of pairwise comparisons for the Practicality criterion

<table>
<thead>
<tr>
<th>№</th>
<th>Type of information system</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Rating</th>
<th>Vector priorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Information and analytical</td>
<td>3.0</td>
<td>3.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.45</td>
<td>0.49</td>
</tr>
<tr>
<td>2</td>
<td>Information and management</td>
<td>0.33</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.54</td>
<td>0.11</td>
</tr>
<tr>
<td>3</td>
<td>Decision support system</td>
<td>0.5</td>
<td>2.0</td>
<td>1.0</td>
<td>2.0</td>
<td>1.19</td>
<td>0.24</td>
</tr>
<tr>
<td>4</td>
<td>Information and advisory system</td>
<td>0.5</td>
<td>2.0</td>
<td>0.5</td>
<td>1.0</td>
<td>0.84</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 5

COLINS’2022, Volume II: Workshop. Gliwice, Poland, May 12-13, 2022, ISSN 2523-4013
http://colins.in.ua, online
### Matrix of pairwise comparisons for the Quality criterion

<table>
<thead>
<tr>
<th>№</th>
<th>Type of information system</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Rating</th>
<th>Vector priorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Information and analytical</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1,68</td>
<td>0,4</td>
</tr>
<tr>
<td>2</td>
<td>Information and management</td>
<td>0,5</td>
<td>1</td>
<td>1</td>
<td>0,5</td>
<td>0,71</td>
<td>0,16</td>
</tr>
<tr>
<td>3</td>
<td>Decision support system</td>
<td>0,5</td>
<td>1</td>
<td>1</td>
<td>0,5</td>
<td>0,71</td>
<td>0,16</td>
</tr>
<tr>
<td>4</td>
<td>Information and advisory system</td>
<td>0,5</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1,19</td>
<td>0,28</td>
</tr>
</tbody>
</table>

**Table 6**

### Matrix of pairwise comparisons for the Performance criterion

<table>
<thead>
<tr>
<th>№</th>
<th>Type of information system</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Rating</th>
<th>Vector priorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Information and analytical</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2,28</td>
<td>0,5</td>
</tr>
<tr>
<td>2</td>
<td>Information and management</td>
<td>0,33</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0,88</td>
<td>0,19</td>
</tr>
<tr>
<td>3</td>
<td>Decision support system</td>
<td>0,33</td>
<td>1</td>
<td>1</td>
<td>0,5</td>
<td>0,62</td>
<td>0,14</td>
</tr>
<tr>
<td>4</td>
<td>Information and advisory system</td>
<td>0,33</td>
<td>0,5</td>
<td>2</td>
<td>1</td>
<td>0,76</td>
<td>0,18</td>
</tr>
</tbody>
</table>

**Table 7**

### Matrix of pairwise comparisons for the Reliability criterion

<table>
<thead>
<tr>
<th>№</th>
<th>Type of information system</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Rating</th>
<th>Vector priorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Information and analytical</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1,19</td>
<td>0,29</td>
</tr>
<tr>
<td>2</td>
<td>Information and management</td>
<td>0,5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0,84</td>
<td>0,21</td>
</tr>
<tr>
<td>3</td>
<td>Decision support system</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0,25</td>
<td>0,25</td>
</tr>
<tr>
<td>4</td>
<td>Information and advisory system</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0,25</td>
<td>0,25</td>
</tr>
</tbody>
</table>

After calculating the values of priority vectors for the criteria of the developed system (Table 2), as well as priority vectors for each of the selected types of information systems (Table 3 – Table 7), the resulting priorities were calculated for each of the criteria (Table 8)

### Table 8

#### The resulting priorities for each of the criteria

<table>
<thead>
<tr>
<th>Criterion</th>
<th>System type</th>
<th>Topicality</th>
<th>Practicality</th>
<th>Quality</th>
<th>Speed</th>
<th>Reliability</th>
<th>Resulting priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information and advisory</td>
<td>0,31</td>
<td>0,17</td>
<td>0,28</td>
<td>0,18</td>
<td>0,25</td>
<td>0,23</td>
<td></td>
</tr>
<tr>
<td>Information and management</td>
<td>0,11</td>
<td>0,11</td>
<td>0,16</td>
<td>0,19</td>
<td>0,21</td>
<td>0,15</td>
<td></td>
</tr>
<tr>
<td>Information and analytical</td>
<td>0,38</td>
<td>0,49</td>
<td>0,4</td>
<td>0,5</td>
<td>0,29</td>
<td><strong>0,43</strong></td>
<td></td>
</tr>
<tr>
<td>Decision support</td>
<td>0,19</td>
<td>0,24</td>
<td>0,16</td>
<td>0,14</td>
<td>0,25</td>
<td>0,18</td>
<td></td>
</tr>
</tbody>
</table>

As a result, the highest result is 0.43, which corresponds to the information-analytical system. Therefore, choosing this type of system will help make the program more reliable, high quality and fast.

### 5. Concretization of system functioning

IDEF0 is a methodology for a graphical description of the organization's systems and processes as a set of interdependent functions. It allows you to explore the organisation's tasks without linking them to the objects that ensure their implementation [14]. The IDEF0 methodology is quite similar to the DFD business process description scheme. The difference is that IDEF0 shows not just inputs and outputs but uses three types of inputs. The first type is input, and the other two inputs are controls and mechanisms. The main elements in notation are blocks representing processes and arrows representing information and material resources. The block is a rectangle inside, which is the name of the process and its number. The title must be an active verb.

The arrows in the IDEF0 standard do not show data movement or sequence of events, as in DFD or WFD charts. Here they are designed to indicate the data and objects required for implementing this
process and what results from its implementation. Arrows can be straight or broken. In the diagram, they should be located vertically or horizontally, but diagonals are not allowed. Within the framework of this standard, there are four types of arrows: input, output, control arrows, and mechanism arrows.

The login arrow is data that is processed by the process to get the result. The input arrow is on the left side of the process. The management arrow is the rules, strategies, and standards defining the conditions and functions needed to make the right exit. The control arrow is at the top of the unit. It is often difficult to determine whether data is input or controlled. When data is changed or processed, it is input, if not - control. The arrow of the mechanism is the resources that support the execution of the function. The arrow of the mechanism enters the lower part of the process. The mechanism arrow may not be displayed on the model. The exit arrow is the result of the function. The exit arrow comes from the right side of the process. Each function must have at least one exit arrow. A context diagram is a type of IDEF0 diagram. This diagram is the most general description of the system and its interaction with the external environment. The context diagram consists of a single block describing the top-level function, its inputs, outputs, controls, and mechanisms, along with the model goal formulations. AllFusion Process Modeler was used to build a context diagram of the developed information system (Fig. 8). It depicts the primary goal of the system – "Turn Ukrainian-language audio into text."

"User's voice" is the data that will process the process and through which you can begin to convert audio into written text. The control arrows for this system are "Development Tool Standards" and "Speech-to-Text Algorithms". The resources required to perform the function are "Dataset", "Microphone", "Developer", "QA", "Database", "User" and "Analog-to-Digital Converter". After processing the input data, and mechanisms and subject to compliance with the control elements, the output will be "Ukrainian text". Each next chart consists of child blocks and arrows that provide the required detail of the parent block at this level. Thus, the child diagram describes the same subject area as its parent block. To create a child chart, you must decompose the primary process into subprocesses. The created top-level diagram will have three blocks: "Edit audio file", "Find phonemes", and "Convert a set of phonemes into the text" (Fig. 9). The first unit will be responsible for recording the audio file and processing it to obtain an "Audio file with corresponding sounds". Find Phonemes will be responsible for splitting the audio file into very small parts and then searching for phonemes for each

Figure 8: Context diagram of the process "Convert Ukrainian audio to text"
part of the audio file. "Convert a set of phonemes into the text" is necessary to create a full text from the array of found phonemes and display it to the user.

Figure 9: High-level diagram of the process "Convert Ukrainian-language audio to text"

Their decomposition was performed for a more detailed description of the upper-level processes of the developed system. Fig. 10 shows a diagram for the first subprocess: "Process audio file".

Figure 10: Detailed diagram of the "Process audio file" subprocess

The diagram has three blocks: "Start audio recording" will be responsible for the beginning of the user's voice recording, and "Save audio file" will be responsible for the end of recording and saving the
audio file. And "Filter sound" is needed to use an analogue-to-digital converter to analyze saved audio-relevant sounds. The control arrow for this chart will be the "Development Tool Standards", which are necessary for the system to work properly and user data to be safe. Required resources are "Developer", "User", "QA", "Microphone" and "Analog-to-Digital Converter". After the audio file has been recorded and saved and its contents have been filtered out, the system will output "Audio file with corresponding sounds". The decomposition of the process "Find phonemes" is shown in Fig. 11. The chart consists of three processes: "Break the audio file into small pieces", "Find the appropriate phonemes for each element", and "Merge phonemes". Find Phonemes will be responsible for breaking up the audio file. The following process, "Find matching phonemes for each element", will be responsible for finding the appropriate phoneme for each part of the audio file. The third process, Combine Phonemes, will be responsible for saving all found phonemes in the appropriate order. The arrows of control here are "Standards of development tools" and "Speech-to-text algorithms". They will be responsible for the flexibility of the system and for the correctness of the phoneme search algorithm, respectively. Resources are "Developer", which will be responsible for the development of modules, "QA", "Dataset", which will contain phonemes, and "Database", which will store the dataset. As a result of these three processes, the output system will receive a "Phoneme Array".

Figure 11: Detailed diagram of the "Find phonemes" subprocess

The decomposition of the process "Convert a set of phonemes into the text" is shown in Fig. 12. This diagram consists of two processes, namely: "Run through a mathematical model" and "Display text". The first process will be responsible for converting the phoneme array into text. To do this, he takes "Phone Array" as an input arrow, as well as resources "Dataset", which will be a set of well-known sentences, phrases or words and "Developer", which will be responsible for writing a mathematical model. "Display text" accepts "Converted phonemes into the text" and is responsible for displaying the result to the user. The resources for this process will be "Developer" and "QA". The management arrows
are again "Development Tool Standards" and "Speech-to-Text Algorithms". As a result of these two processes, the user will receive "Ukrainian text".

Figure 12: Detailed diagram of the subprocess "Convert a set of phonemes into the text"

The first step to building a process hierarchy diagram is to create a context diagram and a sufficient level of its decomposition and detail. The process hierarchy will show the relationship between all subprocesses of the system, which will help to plan the development process, dividing it into smaller, clearer tasks. AllFusion supports the creation of this chart by using the "Diagram" button and selecting "Add Node Tree" in the context window. Next, select the chart name and decomposition level in another context window. The created diagram will be similar to a tree, the top of which is the primary goal, and the branches coming from the top are processes. The hierarchy of processes for the information system for converting Ukrainian-language audio into text is shown in Fig. 13. The main goal is to "Convert Ukrainian-language audio into text." This process includes child processes that will have their subprocesses:

1) "Process audio file":
   • "Start audio recording".
   • "Save audio file".
   • "Filter sounds".
2) "Find phonemes":
   • "Break the audio file into small pieces".
   • "Find the appropriate phonemes for each element".
   • Combine Phonemes.
3) "Convert a set of phonemes into the text":
   • "Run through the mathematical model."
   • "Display text".
6. Selection and justification of the means of solving the problem

Based on the analysis of analogues conducted in the first section, it was decided that the system will be developed as a web application. This solution will help the user not to be tied to a specific platform, and the speed of the system will depend only on the characteristics of the server on which the product will be located. According to StatCounter, the most popular browsers since the beginning of 2021 were Google Chrome – 64.36%, Safari – 19.13%, and Edge – 4.07% [15]. The first version of Google Chrome was created in 2008 based on the open-source browser Chromium. It was created from scratch, recognizing that most websites are now not just pages but entire web applications with their infrastructure, development team, and so on. The advantages of this approach are speed, security, and stability. No less important feature is the simple design and clear interface. For security reasons, Chrome periodically downloads updates from two blacklists and alerts the user when they try to visit a malicious site from that list. Another important security parameter is that a separate process is created for each tab, which avoids situations where one tab can affect another. In addition, creating a separate process for each tab positively affects the browser's stability because even a critical failure in the process of one tab can not interfere with other processes. The disadvantages of this browser include the high load on RAM. It can be a critical problem for users with weak computers when working with many tabs.

Apple developed Safari, part of the macOS and iOS operating systems. This browser was first released in 2003, only for macOS, and only four years later, the browser was ported to Windows. The advantages of Safari include a simple design and compact interface. For developers, this browser has the function of viewing the DOM web page. Also, it supports most modern CSS and HTML functionality, which facilitates the process of developing the interface of web applications. Disadvantages include the low startup speed of the browser itself, as well as outdated security issues.

The third most popular Edge browser is Microsoft, which was released in 2015, making it the youngest of the other browsers on this list. Powered by Chromium gives users access to an extensive library of extensions. Edge was initially planned for Windows 10 only but was later ported to other Windows, macOS and Linux platforms. The advantages include speed because the browser uses the same technology as Google Chrome. Unfortunately, memory problems have not escaped this browser.

In addition to the considered browsers, there are many less popular analogues, which also have advantages and disadvantages. However, the developer's main criterion is the functionality supported by one or another version of the browser. This difference is because browsers use different technologies, which process and display HTML tags and CSS styles differently. The system will be developed as a
cross-browser web application to reach more users. Cross-browser is a web application property that displays and functions equally in all frequently used browsers.

Next, you should analyze which programming language is best for the application's server part. As of April 2022, the most used options are Java, Python, and Node.JS [16].

Java is an object-oriented programming language developed in 1996. This language is ideal for writing both large and small web applications. It was made possible by the developed infrastructure of this language. It also has server-side page rendering technology, which helps avoid using client frameworks such as Angular, React, and Vue. This technology is suitable for applications where functionality is important, and interface and design take a back seat.

Python is also an object-oriented programming language, but unlike Java, it has dynamic typing. The advantages of this programming language include advanced infrastructure, where you can find libraries for any purpose. The pros also include simple syntax. The disadvantages of Python are dynamic typing, which complicates the development of a complex multimodular project. Also, this programming language is interpreted, slowing down the project. Node.JS [22] is a programming language based on JavaScript, making it a common language. The advantage of Node.JS is the asynchronous query execution model and non-blocking I/O. The big drawbacks are the immaturity of this programming language, which means that the infrastructure is less developed than in Java or Python and the inability to handle difficult tasks. It is important to analyze the means which allow you to create web applications. For Java, in most cases, it is Spring. For Python, it is Django. And in the case of Node.JS, it is Express.

Spring [25] is a framework that includes different tools that make it easier to work with authentication, authorization, request processing, databases, microservices and more. In addition to a large selection of tools, Spring automates many processes and settings, which speeds up the writing of business code and project deployment. The advantage of this framework over other analogues includes a large audience, which makes it easier to get to know and work with Spring. Django [21] is an open-source framework for developing Python web applications. It contains basic tools, such as ORM, authentication, administrator page, forms, etc. The main benefits include default security, and the framework takes on attack prevention mechanisms such as SQL injection and CSRF. Unlike the previous framework, Django's ORM framework is outdated, making it difficult to handle complex transactions. Express [20] is an open web framework for Node.JS. It is implemented as free and open source MIT software. The advantages of this framework are flexibility, simplicity, scalability and productivity. Express.js and Node.js can be used to create APIs for single-page, multi-page, hybrid mobile and web applications. Therefore, according to the analysis of three different web frameworks, namely the following aspects: diversity of functionality, infrastructure, documentation, community, performance and quality, Java and Spring should be preferred as the most mature and advanced technologies.

Table 9
Advantages and disadvantages of the Java programming language

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>OOP is a concept that has been used for many years and facilitates application development</td>
<td>Like all high-level programming languages, Java has performance issues compared to other low-level programming languages.</td>
</tr>
<tr>
<td>Java is a high-level programming language with a clear syntax that will help you write and test code.</td>
<td>Although it is possible to develop a GUI in Java (an example is an IntelliJ IDEA), most of it requires a lot of time and a very experienced team</td>
</tr>
<tr>
<td>Java supports many libraries, many of which are already mature technologies, as this programming language is the standard in enterprise development.</td>
<td></td>
</tr>
<tr>
<td>Java is a cross-platform programming language that does not depend on the operating system developed.</td>
<td></td>
</tr>
</tbody>
</table>

272

COLINS'2022, Volume II: Workshop. Gliwice, Poland, May 12-13, 2022, ISSN 2523-4013
http://colins.in.ua, online
Table 10
Advantages and disadvantages of the Spring framework language

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>The framework offers autoconfiguration, which reduces the number of configuration files and the complexity of setting up the project</td>
<td>Since Spring configures many dependencies for the developer, keeping track of all unused dependencies is difficult.</td>
</tr>
<tr>
<td>Spring has a built-in Tomcat server, which reduces application restart time</td>
<td>To efficiently and quickly develop applications using Spring, you need to know the basics of many different framework parts.</td>
</tr>
<tr>
<td>The large Spring ecosystem allows you to easily and quickly create web applications, interact with databases and write your microservices.</td>
<td></td>
</tr>
</tbody>
</table>

In addition to the Spring framework, we must choose technologies for communication and interaction with the database, cloud storage, and testing to develop a full-fledged server part.

Java does not offer many database communication frameworks, but most are mature technologies. JDBC, Hibernate, and Spring Data [26] are the leading technologies that will help build a domain model and services for storage operations. If you analyze each of the proposed frameworks, it becomes obvious that Hibernate is a JDBC wrapper, and Spring Data, in turn, is a wrapper over Hibernate. Therefore, the only feature that plays a role in the choice is the efficiency and ease of writing code. And this benefits the latest technology - Spring Data because it greatly simplifies the process of writing code and has the function of auto-generating methods, using the correct names of the latter in the interface. In addition, configuring the framework becomes much easier because you need to describe the five properties in a special file, “application.properties”.

Table 11
Advantages and disadvantages of the Spring Data framework

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring Data, as part of the Spring framework, allows you to integrate into already written code using this framework easily</td>
<td>Because Spring Data takes over code generation, you need to know the basics well and keep track of generated queries.</td>
</tr>
<tr>
<td>Spring Data is much easier to use than Hibernate or JDBC because it performs many functions for the developer.</td>
<td>Entity classes are quite difficult to set up because you need to know exactly how to create relationships between classes and be able to optimize them</td>
</tr>
<tr>
<td>Spring Data supports auto-generation methods only with correctly named methods in the interface.</td>
<td>Very often, memory leaks due to an excessive number of generated requests or due to sub-optimal transactions</td>
</tr>
<tr>
<td>It has pretty good documentation.</td>
<td></td>
</tr>
</tbody>
</table>

Google Cloud Storage [17] was chosen to work with cloud storage - a technology for storing and accessing files in the Google Cloud Platform infrastructure. It combines performance, security, relative ease of use, and scalability, and you can store up to 15 gigabytes of files for free. To integrate with Java, the service offers a library with the same name: Google Cloud Storage. All interaction occurs through a special interface called "Storage", which combines many methods for flexible operation.

The main tools for test coverage are Junit, AssertJ and Mockito. They make it easier to write both modular and integration tests. Modular tests are tests of the smallest part of the code, isolated from other parts. Integration tests are tests in which not an isolated part of the functionality is tested, but the interaction between several program modules. Junit, in turn, contains a variety of functionality, but mostly it checks and confirms that the result meets certain expectations. Although Junit offers its class to test the results, it is not flexible enough and not comfortable to use, so the developers add AssertJ to the project. AssertJ is a library that greatly expands the capabilities of test validation and allows multi-
step validation. Each class contains certain dependencies that need to be created and initialized to test. This approach is not convenient for modular testing because we have to test an isolated piece of code. To replace objects, use the Mockito library, which allows you to configure and rotate fictitious data without creating or configuring the object itself. A relational database was chosen for the designed information system, as all models have a certain structure. The best option for use is PostgreSQL. It is an open-source database with advanced functionality that supports many new features, such as interacting with the JSON data type. Also in favour of this choice is that most web applications use this product as the main database, and it also has a driver for the Java programming language.

Having decided on the server-side programming language, you should also choose one of the frameworks for the client part. Among the main options: are Angular, React, and Vue.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Angular</th>
<th>React</th>
<th>Vue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stability</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Developers</td>
<td>Google</td>
<td>Facebook</td>
<td>Alibaba</td>
</tr>
<tr>
<td>Community</td>
<td>Great</td>
<td>Great</td>
<td>Not great</td>
</tr>
<tr>
<td>Programming language</td>
<td>TypeScript</td>
<td>JavaScript</td>
<td>JavaScript</td>
</tr>
<tr>
<td>Development speed</td>
<td>Normal</td>
<td>Normal</td>
<td>Speed</td>
</tr>
<tr>
<td>Component framework</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Good documentation</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Code reuse</td>
<td>Yes</td>
<td>No</td>
<td>Parts</td>
</tr>
<tr>
<td>Autonomy</td>
<td>It does not require the connection of third-party technologies</td>
<td>Requires connection of third-party technologies</td>
<td>Requires connection of third-party technologies</td>
</tr>
</tbody>
</table>

So, after analyzing the various parameters, the client framework was chosen Angular because it does not require the connection of third-party technologies, and good documentation, along with the reuse of code and TypeScript, will help write code quickly and make it easy to maintain.

Since the developed information system is a web application, you need to choose the technology for further project deployment on a remote server. Docker and Kubernetes are ideal for this.

Docker is a software platform designed to automate deployment and project management. It allows you to create a container for the application with all its dependencies and then run it on any Linux system. The advantages of Docker include abstraction of the application from the host, ease of scaling, isolation of the environment and ease of use of layers. However, this software platform has enough disadvantages. For example, it does not support inverse sadness, which means that new versions of the program may not support the configuration that was written for the old version. Also, it is important for applications with a heavy load to write a very clear and high-quality configuration, which is almost impossible for beginners.

Kubernetes is used to automate the deployment and management of container projects. It also supports Docker as a container launch environment. Project management includes automated planning, self-recovery capability, automatic deployment, load balancing, and automatic environment scaling. Its advantages include increased code performance, the deployed application's reliability, and the relative cheapness of using Kubernetes. The big disadvantage of this platform can be its complexity because Kubernetes has many different features.

Modern projects result from the work of many developers who can be in any part of the world. Therefore, in addition to local changes, the developer must be able to add these changes to the project code base. Thus, special software was created [18] - version control systems that help developers...
manage change, view the authors of changes and interact effectively with changes. In VSC, each developer can create their branch, where they can add, edit or delete code without changing the main project code or branches of other developers. Such systems make the development process clearer and more transparent. Version control systems are divided into two types:

1. Centralized version control systems are applications with a client-server architecture when the repository itself is only in one instance and on a remote server. The most obvious disadvantage of this version control system is the central repository. If it fails, all changes and work will be lost. It is why the following types of systems began to appear.

2. Distributed version control systems allow a copy of the repository to be kept by each developer working on the project. At the same time, the central repository in which changes from local repositories are sent is also allocated. Working with this version control system requires constant synchronization of the local repository with the central one. This type of system solves the problem with one central repository, which contains the entire history of changes, as well as increases the autonomy of the developer.

The central version control systems that are often used on projects are Git, Mercurial and SVN. Git is a free, distributed version control system originally designed to work on the Linux kernel. It is compatible with various protocols (HTTP, SSH, FTP), has cryptographic authentication, and has high speed. Immediately after installation, Git offers two utilities to work with it: Git Bash and Git UI. After the analysis, it becomes clear that the high speed, flexibility and integration with many GIT software is the best solution. The last important point in designing an information system is the design template choice because it will make the program flexible, reliable and easy to test.

![Scheme of operation of the distributed version control system](image)

Figure 14: Scheme of operation of the distributed version control system

The main template used in web development today is MVC. The MVC (Model-View-Controller) template [19] is a web application design pattern that includes several templates. When using MVC, the program data model is divided into three separate components, the user interface and the logic of user interaction with the system, so the modification of one of these components has minimal impact on others or does not do it at all. The MVC concept divides data, representation and processing of user actions into components/ A model is an object model of a specific subject area that is developed, which contains methods of working with this data in addition to the data itself. The peculiarity of the model is that it does not contain information about the visualisation methods, i.e., the presentation of this data. It does not interact directly with the user. View - is responsible for presenting and displaying data. It can be an API-level representation, i.e., exporting data in JSON, XML, or other in a supported format, or using various page rendering technologies. Controller - interacts directly with the user, transmitting data to the system. Uses the model to implement the expected functionality and passes it to the view. The received data are filtered and aggregated at this level, and user rights are checked.

275
Figure 15: MVC pattern scheme

7. Technical characteristics of selected software development tools

To choose a development environment, you must first analyze the selected technologies for the server part - Java, Spring, and Spring Data. For the client part, it is Angular, HTML, and CSS. One of the best development environments for such a list of technologies is IntelliJ IDEA Ultimate [23]. You need a license to use this product, but students can get one for free on the company’s website.

This integrated development environment has many advantages, including intuitive design, functionality, integration with Spring and a large library of plug-ins (Figure 16). It also supports the ability to run projects written in different programming languages: Java, Angular, Kotlin, Node.js, React, and more. Also, for easy operation, IntelliJ IDEA offers a flexible way to search for files by name and write code or method names. In addition, the environment has a built-in auto-code generation functionality (Figure 17), which facilitates the development process for Java developers.

Figure 16: Plugin window view
Intellij IDEA Ultimate was also chosen to work with the database. It has built-in functionality and support for connecting to various databases, sending requests and viewing schemas or data.

Also, a big advantage of this integrated development environment is the support of GIT (Figure 18), namely the UI with which you can easily perform most work with the version control system.

8. Description of the created software

8.1.1. Description of the database structure

During the design of the information system, a PostgreSQL database was selected, which will store information about the registered user and downloaded audio files. The scheme of the database is shown.
in Figure 19 [24]. The audio table will contain basic information about the recorded audio file: Sequence number, File path in Google Cloud Storage, Creation date, and User ID. The transcription table includes information on all possible transcriptions of the audio file: Sequence number, Transcription, and Audio file ID. The user table contains information about the user: User ID, login, Name, and Last name.

8.1.2. Functional purpose

The developed information system is focused on any user. Based on this, the functionality of converting Ukrainian-language audio recording into text is available to everyone. The user can also register in the developed information system and further login. He will then have access to a page with a list of all his audio files that have been translated into written text, where he will also be able to view information on each of these files.

8.1.3. Description of logical structure

This information system has a client-server architecture. The implementation was divided into several different modules:
• Module for working with audio files;
• User registration and authorization module;
• Module for working with the database;
• Module for working with cloud storage.

The module that works with audio files is designed to process an audio file and its subsequent transformation into written text. After processing the audio file, an array with different text options can be returned. A separate module was also developed for the user registration and authorization system. The user registration process requires a password, login, and first name. This entry will then be saved in the database. Before this happens, the password will be encrypted using the bcrypt algorithm. After registration, the user can log in using the authorization system. The login function checks whether there is a user with such login and password. Also, to check the password, you can not just get it from the database because it was previously hashed. Therefore, the entered password is first hashed and compared with what is contained in the database. If the user does not exist, the system will notify you by displaying an error on the screen. Otherwise, a jwt authorization token will be created and placed in the request header. In the future, this token will be placed in each request, and the system will be able to check its validity to give the user access to some pages.

The database interaction module is responsible for all operations performed with the database, namely: connecting, saving, updating, and reading data about recorded audio files or the user. For this purpose, the Spring Data framework was used, which generates SQL-queries for further sending to the database. The last module is the module for interaction with the cloud environment. It is used to efficiently store and read saved audio files efficiently. It uses the imported Google Cloud Bucket library.

8.1.4. Used technical means

The developed system is a web application, so you need to have a tool with Internet access and a browser. For example, for a laptop or computer, the requirements will look like this:
• Operating system: Windows, MacOS, Linux;
• Processor: Intel Pentium 4 / Athlon 64;
• Memory: 512 MB.
8.1.5. Input and output data

The input for the system is an audio file that can be downloaded or recorded. After processing it as a result of the web application, the user will receive the following initial data: information about the audio file and options for converting it into written text.

9. User manual

9.1.1. Introduction

The information system being created will be a good tool for people who have difficulty typing large amounts of text on the keyboard or listening to the sent audio file. The program is a web application that provides mobility and ease of use.

9.1.2. General information about the system

The program is called "Ukrainian Speech-to-text". The client was developed in the TypeScript programming language using the Angular framework and the IntelliJ Idea Ultimate development environment. The server part uses the Java programming language and the Spring framework. The IntelliJ Idea Ultimate development environment was also used to write the server part. The program exists as a web application, so it is not tied to the operating system or browser. Users will be able to record or download audio and translate it into text, and authorized users will be able to view all recorded audio files.

9.1.3. Classes of tasks to be solved

The main task solved by the developed information system is the transformation of Ukrainian-language audio into written text. The user can download or record an audio file using a microphone, which the system transforms into written Ukrainian text and displays on the appropriate page. An additional task is to save and view the list of downloaded audio files for users who have registered in the system. A list of possible translations into written text will also be available.

9.1.4. Description of the main characteristics and features of the program

The peculiarity of the program is that it is designed as a web application that provides easy access to it from any device with the Internet and a browser. Also, unlike many solutions, this application contains two options for creating an audio file for further processing, which facilitates the use process. The program is available around the clock because it is located on a cloud server, which also increases its reliability. The program's interface is intuitive, so it does not require special knowledge. The functionality of the list of downloaded audio files includes viewing and downloading files.

9.1.5. Information on functional limitations for use

You can use the program when connected to the World Wide Web and if you have a browser of new versions. The user must provide permission to use the microphone for audio recording functionality. It is also recommended to have a good microphone and to pronounce words. The same applies to the file that the user will download. It must be high quality and with clear pronunciation.
The application has been tested in new web browsers: Google Chrome 101.0.4951.54 and Microsoft Edge 101.0.1210.39. Other web browsers may have different site interfaces.

10. Analysis of the control example

To confirm that the system is fully operational, the results of the implemented web application were presented. After the user opens the main page in the web browser, he will be offered two options for downloading the audio file. For the first option (Fig. 20-21), there will be one field for downloading the file and the button "Upload" after clicking, which the file will be sent to the server and analyzed. For the second option, you need to click on the "Record" button, which will open the second element of the table in which there will be a button, the beginning of the audio recording (Fig. 22).

Figure 20: File download page

Figure 21: Page view after selecting a file

Figure 22: File recording page

To start recording audio, press the "Start Audio Recording" button, and to end it, the "Stop Audio Recording" button (Fig. 23). After pressing the button, the system will finish recording the audio file.
and create two fields (Figure 24). Use the first field to listen to the recording or download it. In the second field, the user must enter the name of the recorded file. If you leave this field blank, the system will generate the name based on the hash code of the audio file.

![Stop Audio Recording](image)

**Figure 23:** Page view after recording an audio file

![Page view after recording an audio file](image)

**Figure 24:** Appearance of the page after recording the audio file

In both cases, the "Upload audio" button will be available, after which the system will send the data to the server, where it will be processed. Then the user will be directed to a new page (Figure 25), which will specify the file name as well as a table of possible textual content of the audio file.

![Audio file information page](image)

**Figure 25:** Audio file information page

In addition, the user has access to the registration page (Fig. 26). This page contains two fields where you need to enter the login and password, which will then be used to log in.
The authorization page contains two fields: login, password, and the "Sign in" button, after which the user will be directed to the main page. The system generates a jwt token, which provides access to the page with a list of downloaded audio files. The program will add a new button to the header, "Your list" (Fig. 27), to open the page with the audio list. After clicking, the user will be directed to a new page.

The page with the list of all files downloaded by the user is conditionally divided into two parts. The first part on the right contains the button "Your audio files" (Fig. 28). After clicking it, the program will display a list with the names of all files (Fig. 29).
To view the information, you need to click on any name from the list, after which the program will show the data in the middle of the page (Figure 30): the name of the audio file and possible transcriptions. To exit the web application, you need to click the "Logout" button, which is also in the program's header, and then the user will be directed to a new page (Fig. 31). The text box with information will be located. Also, the header itself will be changed, and it will contain new buttons (Fig. 32), namely: "Login", "Register", and "Upload audio".

**Figure 29:** Page layout after clicking "Your audio files"

**Figure 30:** View the page after clicking on one of the titles

**Figure 31:** Page after exit

283
In order to better analyze and find the dependencies, the dependence of processing accuracy on the number of words and the dependence of the processing speed of the file on the number of words in it was analyzed. The dictation text on the Ukrainian language "Birth of Tradition" was used during the analysis. It contains 185 and many complex words with different accents, which will complicate the test and improve it. For each of the results, an audio file with a different number of words was tested in stages: 13, 36, 72, 96, 108, 140, and 185 words. As a result of the first analysis (Fig. 33), the average accuracy percentage is 90 per cent. It is also possible to notice that the word number does not affect the algorithm's accuracy. And the decrease in the percentage is not large and was due to the word complexity and low quality of the microphone, and hence the recorded file. For the second analysis of the dependence of the processing speed of the audio file on the number of words in it (Fig. 34), we can conclude that although the program returns quite good results, its speed can be an obstacle to a good user experience. It is especially true for large files because their processing takes more than 20 seconds, so in the future, improving the results will need to increase the performance of the cloud server and the processing algorithm.
11. Conclusion

In recent years, text-to-speech software has become increasingly used by individuals and organizations as it becomes more accessible. It has become more convenient and less expensive, and if you gain enough experience, it can significantly increase the speed of typing and make the user more productive. It is a necessary tool for people with accessibility problems or those trying to prevent recurrent stress disorders due to too frequent typing. Modern AI in automatic language recognition can give accurate results in 95% of cases. An error of 5% may seem trivial, but in some programs, such as defence or the automotive sector, such errors can have serious consequences.

High-quality training kits are part of the solution. With more sophisticated algorithms, decrypted, annotated data sets with broad acoustic and linguistic coverage are needed to improve AI accuracy in this area. The task of creating an information system for converting Ukrainian audio text into written text is the solution. As a result, the information-analytical type of system was chosen for creation.

The process of converting speech into text, areas of use, and algorithms are analyzed. In addition, off-the-shelf solutions were considered: Google Speech-to-text, IBM Watson, Dragon NaturallySpeaking, Microsoft Dictate and Odrey. Based on this, a comparison table was created, and the requirements for the developed information system were selected.

First, a goals tree was built, and on its basis, the pairwise comparisons matrices were calculated, and the type of information system was chosen, namely information-analytical. After that, a context
diagram, a child diagram, which was created by the decomposition and subprocess diagrams for each process in the child diagram were constructed.

In addition, a table of comparisons has been created for the following parameters: free version, file download support, file recording support, program type and Ukrainian language support. As a result, it was decided that the program would be developed as a web application with support for downloading and recording audio files. A goal tree was created, which helped to define the primary goal of the system, as well as to formulate the second-level goals and criteria. Using the formed criteria, the analysis method of hierarchies was applied, by means of which the type of the developed information system was determined: information-analytical. In addition, construct the IDEF0 context chart, the child chart that was created as a result of the context chart decomposition, and the subprocess chart for each process in the child chart. This detailed description of the processes helped pinpoint the processes' inputs, resources, management, and interoperability.

The last point is a diagram of the hierarchy of processes. Although the nodes in this tree are the same processes as the IDEF0 charts, this way of depicting the approaches helps to more accurately assess the order of work and the specific steps that need to be taken at each stage to complete the project. During the writing, the leading technologies used in web development were researched, and the ones best suited for the created information system were selected: Java, Spring Data, Junit, Mockito, PostgreSQL, Angular, HTML, CSS, Docker, Kubernetes. All this should provide quality, flexibility, reliability and clear design. In addition, architectural programming patterns were analyzed, and Model-View-Controller was selected as the most popular and reliable template in web development. The Intellij IDEA Ultimate development environment was chosen for the development of the information system, as it supports the operation and launch of all selected technologies. It also makes working with the version control system and database easier. It makes it easier to write code, as well as future debugging.

The most popular browsers, programming languages, frameworks, databases, version control systems, deployment systems and design templates are analyzed. Then select the most appropriate for the development of information systems. It was decided that the system will be developed as a web application using Java, Spring, Spring Data, Google Cloud Storage, PostgreSQL, Junit, Mockito, AssertJ, Angular, GIT, Docker, Kubernetes and MVC design template.

The functional purpose, logical structure, technical means used and input initial data are described, which as a result, is a characteristic of the developed software. The user manual provides general information about the program, classes of tasks and its limitations.

The information system is implemented, as well as the description of the database structure, functional purpose, logical structure, technical means, and input and output data. Program features and functional limitations are also considered at the end for the dependence analysis of accuracy and processing speed on the word number. The research of the example demonstrates the created information system, as well as all its functionality. In addition, the accuracy dependence on the word number and the processing speed dependence on the number of words were analyzed. The result of the work is a tested and ready-to-use information system for converting Ukrainian-language audio text into written, which provides the following main functions:

• Download an audio file from the system.
• Record an audio file using the microphone.
• Processing and saving the audio file.
• View transcription results.

12. References