10.20535/2411-2976.22024.20-24

UDC 004.7

AUTOMATED LECTURE NOTE-TAKING SYSTEM

Svitlana V. Sulima, Artem G. Genash

Educational and Research Institute of Telecommunication Systems
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

Background. In the modern world, information technologies evolve rapidly, constantly altering our approaches to learning, work, and daily life. One significant aspect of this evolution is the automation of various processes, including education. Students and teachers are faced with large volumes of information that need to be processed, stored, and used in the educational process. Taking lecture notes is still an important task that requires a lot of time and effort, thus automating this process is both relevant and necessary. Automated lecture note-taking based on video and audio materials greatly facilitates the lives of students and teachers by providing quick access to structured information. The use of speech recognition and artificial intelligence technologies to create notes from lecture materials opens up new opportunities for effective learning. These systems can significantly save time, improve the quality and accuracy of notes, and ensure their accessibility to all participants in the educational process. These systems can not only create notes but also structure them by highlighting key points and providing easy access to information. This allows students to focus on understanding and comprehending the material rather than writing it down, thereby improving the quality of learning and knowledge acquisition.

Objective. The purpose of the paper is to simplify the note-taking process and improve its quality by developing a system for automated lecture note-taking based on video and audio materials, ensuring the efficient and rapid creation of structured notes from lecture materials.

Methods. Analysis of Literature and Contemporary Studies: Studying scientific articles, monographs, and dissertations related to the topic of automated lecture note-taking, speech recognition, and artificial intelligence. System Analysis: Defining system requirements, analysing possible approaches and tools for implementation. Experimental Method: Developing, implementing, and testing the system. Comparative Analysis: Evaluating the effectiveness of different speech recognition tools and AI models for creating notes. Modelling and Prototyping: Creating a system prototype, testing it, and improving it based on the obtained results.

Results. During the implementation of the automated lecture note-taking system, an effective Telegram bot was created, which uses "whisper-1" and "gpt-4" models to provide high-quality speech recognition and the generation of structured notes from video and audio materials.

Conclusions. The developed system of automated note-taking of lectures based on video and audio materials significantly simplifies the preparation of materials for students and teachers. Integration with Telegram and implementation of the system through a Telegram bot ensure cross-platform, accessibility and ease of use and at the same time provide an opportunity to avoid creating additional web or mobile applications for a wide range of users. The use of OpenAI's "whisper-1" model demonstrates high accuracy of speech recognition, which allowed improving the quality of transcriptions compared to other tools such as Vosk or FasterWhisper.

Keywords: *automated note-taking; speech recognition; Telegram bot; GPT models; audio and video processing; Whisper-1.*

Introduction

In today's world, information technologies are developing at an extraordinary pace, constantly changing our approaches to education, work, and everyday life. One important aspect of this development is the automation of various processes, including education. Today, students and lecturers face a large volume of information that needs to be processed, stored, and utilized in the educational process. Taking lecture notes remains a significant task that requires a lot of time and effort, making the automation of this process both relevant and necessary. Automated lecture note-taking based on video and audio materials significantly eases the lives of students and lecturers by providing quick access to structured information. The use of speech recognition and artificial intelligence technologies to create lecture notes opens new possibilities for effective learning. Such systems can save time, improve the quality and accuracy of notes, and ensure their availability to all participants in the educational process.

Automated lecture note-taking is crucial in today's world, where a significant amount of information is transmitted orally, particularly in the form of lectures, webinars, and meetings. Modern approaches to automated

note-taking rely on the use of speech recognition (Speech-to-Text or STT) and natural language processing (NLP) technologies.

Statistical Methods-Based Models: Early speech recognition systems, such as Hidden Markov Models (HMM), relied on probabilistic models to transcribe audio signals into text [1].

Deep Learning Methods: Nowadays, neural networks, particularly recurrent neural networks (RNN) [2], long short-term memory (LSTM) [3], and transformers [4], are increasingly used. These models demonstrate a high level of accuracy in real-time speech recognition.

Tools and Libraries for Speech Recognition:

- Google Speech-to-Text: One of the most popular commercial services with high accuracy and wide language support [5].
- Microsoft Azure Speech Service: An alternative solution from Microsoft offering high quality and flexibility of settings [6].
- Amazon Transcribe: A tool from Amazon that is convenient for large data volumes and integration with other AWS services [7].
- Vosk: An open-source offline tool that can use language models without an internet connection [8].
- Whisper: A model from OpenAI that shows high accuracy [9].

Natural Language Processing (NLP):

To automatically create meaningful and structured notes, it is important to use NLP technologies:

- Tokenization: Breaking text into individual words or phrases.
- Parsing: Analysing the syntax of the text to get structured information about the sentences.
- Semantic Analysis: Determining the meaning of words and sentences in context.
- Text Summarization: Used to automatically create shorter versions of texts without losing the main information [10].

Tools and Libraries for NLP:

- NLTK (Natural Language Toolkit): A library for text processing with a wide range of functions for various NLP tasks [11].
- spaCy: A fast tool known for its performance in text processing [12].
- Transformers by Hugging Face: Contains a wide range of models based on transformer architecture, including GPT, BERT, T5, and others [13].
- OpenAI GPT: A model that shows impressive ability in text generation and summarization [14].

Challenges of Automated Note-Taking:

- Recognition Accuracy: Different accents, speech speeds, and background noise can negatively affect speech recognition accuracy.
- Integration with Various Formats: Supporting different audio and video file formats, their conversion, and processing may require additional resources.
- Processing Large Volumes of Data: Models like GPT-4 have text size limitations, which can create difficulties when working with long lectures or seminars.

AI-Based Automated Note-Takers:

- Scribie [15]
- Otter.ai [16]

Some tools specialize in creating text summaries that have a clear structure and highlight key points. These platforms typically use algorithms to analyze text and extract key ideas. Platforms for Creating Structured Text Notes:

- Notion [17]
- Evernote [18]

Integrated platforms for automated note-taking combine several functions for automated note-taking, including speech recognition, structured note creation and storage. These platforms typically use cloud-based technology to provide access to notes from any device.

Examples of integrated platforms: Description [19], Rev [20].

Main part

Let's consider the main tools that support the recognition of the Ukrainian language, which is one of the most important work criteria.

Vosk is an open-source speech recognition tool that works offline.

Vosk has not achieved perfection in speech recognition accuracy, so results may be unpredictable. However, it can be customized for specific systems to improve accuracy.

Whisper is a speech recognition tool developed by OpenAI.

Whisper can work in both online and offline modes, making it convenient for a variety of applications. However, high requirements for computing resources can be an obstacle to using Whisper on less powerful devices.

FasterWhisper is an optimized version of Whisper aimed at increasing the speed of speech recognition without significant loss in accuracy [21].

However, as in the case of Whisper, FasterWhisper requires significant computing resources to ensure high

performance. Its effectiveness may vary depending on the complexity of the audio data and recording conditions.

InsanelyFastWhisper is another optimized version of Whisper, aimed at reducing the processing time of audio files as much as possible [22].

However, over-optimization can lead to a decrease in recognition accuracy, which should be considered when using this tool.

WhisperX is a version of Whisper optimized for large audio files and streaming data.

Whisper-1 is the latest version of OpenAI's Whisper model, which offers improved accuracy and performance compared to previous versions [23].

Whisper-1 requires significant computing resources, but provides high accuracy and reliability, making it one of the best speech recognition tools available today.

A comparative analysis of the speed and effectiveness of speech recognition tools was performed on a device with the characteristics shown in Fig. 1.

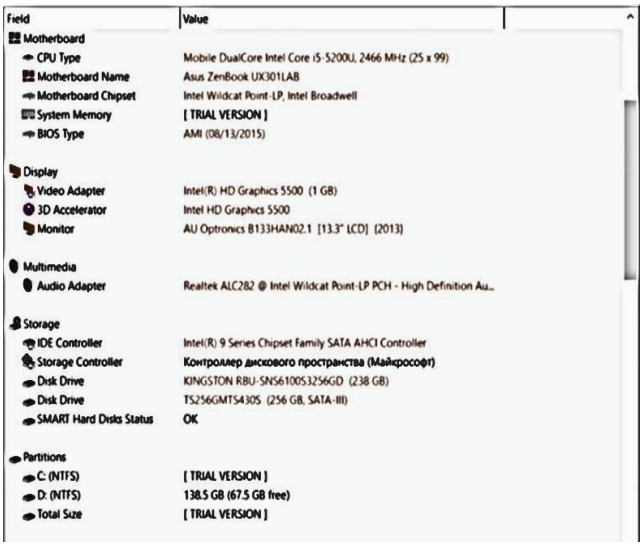

Fig. 1. Technical characteristics of the device used to test the speed and effectiveness of speech recognition tools

Table 1. Comparison of the speed of speech recognition tools: Vosk, Whisper, FasterWhisper, InsanelyFastWhisper WhisperX, whisper-1

Tool	Speed (words/sec)	Free/Paid
Vosk	10-20	Free
Whisper	30-40	Free
FasterWhisper	50-60	Free
InsanelyFastWhisper	80-100	Free
WhisperX	120-150	Free
Whisper-1	200-250	Paid

Table 2. Comparison of speech recognition accuracy for speech recognition tools: Vosk, Whisper, FasterWhisper, InsanelyFastWhisper WhisperX, whisper-1

Tool	Accuracy (WER)
Vosk	90-95%
Whisper	92-96%
FasterWhisper	94-98%
InsanelyFastWhisper	96-99%
WhisperX	98-99.5%
Whisper-1	99.5-99.9%

Analysis:

- Whisper-1 demonstrates the highest recognition speed and accuracy, making it the optimal choice for implementing an automated lecture note-taking system.
- WhisperX [24] offers high speed and accuracy, slightly inferior to whisper-1.
- InsanelyFastWhisper, FasterWhisper and Whisper have moderate speed and accuracy, which may be sufficient for some tasks.
- Vosk is the slowest and least accurate of the tools presented.

Additional factors:

- Cost: Some tools are free, while others require a paid subscription.
- Availability: Some tools may not be available for use in certain regions or platforms.
- Functionality: Some tools offer additional features in addition to speech recognition, such as automatic translation and others.

To select the optimal tool, several different models were tested, in particular: Vosk, Whisper, FasterWhisper, InsanelyFastWhisper, WhisperX.

After the testing was completed, the results were analyzed. Tools Vosk, Whisper, FasterWhisper, InsanelyFastWhisper, WhisperX were not effective enough in terms of speed and accuracy of speech recognition for the needs of this work.

Based on the results of testing and analysis, a decision was made to use the "whisper-1" model from OpenAI. This model is characterized by high speed and accuracy of speech recognition, which makes it the most optimal option for use in a note-taking system.

It is also important to consider that the "whisper-1" model is paid, but it is very affordable in terms of cost, which makes it an acceptable solution considering the value/quality ratio.

Therefore, the choice of the "whisper-1" model from OpenAI is the optimal solution for the implementation of

a system of automated note-taking of lectures based on video and audio materials. This tool meets all the project's requirements for speed, accuracy, cost and availability.

Telegram bot is used as the main graphical interface for the automated lecture note-taking system.

OpenAI's "whisper-1" model was chosen for speech recognition in the project "Automated lecture note-taking based on video and audio materials" due to its high accuracy and optimal speed. The integration of this model allowed for the efficient conversion of audio and video data into text, which is a critical step for the further creation of structured lecture notes.

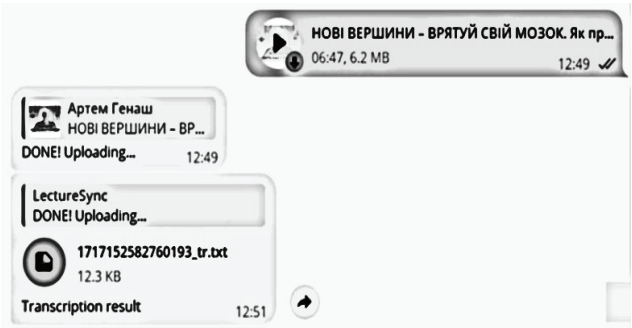

Fig. 2. Test of the "whisper-1" model for speech recognition.

```
Time taken for data\1717152582760193.mp3 : 32.62023711204529
```

Fig. 3. Performance speed result of "whisper-1" model for speech recognition.

The test results showed that the "whisper-1" model from OpenAI provides high accuracy and recognition speed, in accordance with the requirements of the project. An important advantage is the ability of the model to process a wide range of language options and accents, which makes the system universal for different users.

Thus, the application of the "whisper-1" model significantly increased the efficiency of processing audio and video lectures, which will allow providing users with high-quality and accurate text summaries.

Conclusions

The developed automated lecture note-taking system based on video and audio materials significantly simplifies the preparation of materials for students and teachers. Integration with Telegram and the implementation of the system through a Telegram bot ensure cross-platform availability, ease of use, and avoid the need to create additional web or mobile applications for a wide range of users. The use of the "whisper-1" model from OpenAI demonstrates high speech recognition accuracy, which has improved the quality of transcriptions compared to other tools such as Vosk or FasterWhisper. Adding support for other popular platforms such as Zoom or Microsoft Teams will expand the system's capabilities and attract new users. Further improvement of the system focused on eliminating existing shortcomings and expanding functionality will make it even more useful and accessible to a wide range of users.

References

1. Hidden Markov Models (HMM) [Online] – Retrieved from https://mi.eng.cam.ac.uk/~mjfg/mjfg_NOW.pdf
2. RNN [Online] – Retrieved from https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn
3. LSTM [Online] – Retrieved from https://www.simplilearn.com/tutorials/artificial-intelligencetutorial/lstm?tag=%22LSTM%22
4. Transformers [Online] – Retrieved from https://www.datacamp.com/tutorial/how-transformers-work
5. Google Speech-to-Text [Online] – Retrieved from https://cloud.google.com/speech-to-text
6. Microsoft Azure Speech Service [Online] – Retrieved from https://azure.microsoft.com/en-us/products/ai-services/ai-speech
7. Amazon Transcribe [Online] – Retrieved from https://aws.amazon.com/transcribe/
8. Vosk [Online] – Retrieved from https://alphacephei.com/vosk/
9. Whisper OpenAI [Online] – Retrieved from https://github.com/openai/whisper
10. Natural Language Processing NLP [Online] – Retrieved from https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP
11. NLTK Python [Online] – Retrieved from https://www.datacamp.com/tutorial/text-analytics-beginners-nltk
12. spaCy Python [Online] – Retrieved from https://realpython.com/natural-language-processing-spacy-python/
13. Transformers HuggingFace Python [Online] – Retrieved from https://www.datacamp.com/tutorial/an-introduction-to-using-transformers-and-hugging-face
14. OpenAI GPT Python [Online] – Retrieved from https://www.datacamp.com/tutorial/using-gpt-models-via-the-openai-api-in-python

15. Scribe [Online] – Retrieved from https://scribie.com/
16. Otter.ai [Online] – Retrieved from https://otter.ai/
17. Notion [Online] – Retrieved from https://www.notion.so/
18. Evernote [Online] – Retrieved from https://evernote.com/
19. Descript [Online] – Retrieved from https://www.descript.com/
20. Rev [Online] – Retrieved from https://www.rev.com/
21. Faster Whisper [Online] – Retrieved from https://github.com/SYSTRAN/faster-whisper
22. Insanely Fast Whisper [Online] – Retrieved from https://github.com/Vaibhavs10/insanely-fast-whisper
23. Whisper-1 OpenAI [Online] – Retrieved from https://platform.openai.com/docs/guides/speech-to-text/quickstart
24. WhisperX [Online] – Retrieved from https://github.com/m-bain/whisperX

Суліма С.В., Генаш А.Г.

Автоматизована система конспетування лекцій

Проблематика. У сучасному світі інформаційні технології стрімко розвиваються, постійно змінюючи наші підходи до навчання, роботи та повсякденного життя. Одним із важливих аспектів цієї еволюції є автоматизація різних процесів, у тому числі освітніх. Студенти та викладачі стикаються з великими обсягами інформації, які необхідно обробляти, зберігати та використовувати в навчальному процесі. Конспектування лекцій залишається актуальним завданням, яке потребує багато часу та зусиль, тому автоматизація цього процесу актуальна та необхідна. Автоматизоване конспектування лекцій на основі відео- та аудіо матеріалів значно полегшує життя студентів і викладачів, надаючи швидкий доступ до структурованої інформації. Використання технологій розпізнавання мовлення та штучного інтелекту для створення конспектів із лекційних матеріалів відкриває нові можливості для ефективного навчання. Ці системи дозволяють значно заощадити час, підвищити якість і точність конспектів, забезпечити їх доступність для всіх учасників навчального процесу. Ці системи можуть не тільки створювати нотатки, але й структурувати їх, виділяючи ключові моменти та забезпечуючи легкий доступ до інформації. Це дозволяє учням зосередитися на розумінні та розумінні матеріалу, а не записувати його, тим самим покращуючи якість навчання та засвоєння знань.

Мета дослідження. Метою даної роботи є спрощення процесу конспектування та підвищення його якості шляхом розробки системи автоматизованого конспектування лекцій на основі відео та аудіо матеріалів, що забезпечує ефективне та швидке створення структурованих конспектів із лекційних матеріалів.

Методика реалізації. Аналіз літератури та сучасних досліджень: Вивчення наукових статей, монографій та дисертацій, пов’язаних з темою автоматизованого конспектування лекцій, розпізнавання мовлення та штучного інтелекту. Системний аналіз: Визначення системних вимог, аналіз можливих підходів та інструментів для впровадження. Експериментальний метод: розробка, впровадження та тестування системи. Порівняльний аналіз: оцінка ефективності різних інструментів розпізнавання мовлення та моделей ШІ для створення нотаток. Моделювання та прототипування: створення прототипу системи, його тестування та вдосконалення на основі отриманих результатів.

Результати дослідження. Під час впровадження автоматизованої системи конспектування лекцій створено ефективний Telegram-бот, який за допомогою моделей «whisper-1» та «gpt-4» забезпечує якісне розпізнавання мовлення та генерацію структурованих нотаток із відео та аудіо. матеріалів.

Висновки. Розроблена система автоматизованого конспектування лекцій на основі відео та аудіо матеріалів значно спрощує підготовку матеріалів для студентів і викладачів. Інтеграція з Telegram та реалізація системи через Telegram-бот забезпечують кросплатформність, доступність та зручність використання та водночас дають можливість уникнути створення додаткових веб- чи мобільних додатків для широкого кола користувачів. Використання моделі OpenAI «whisper-1» демонструє високу точність розпізнавання мови, що дозволило підвищити якість транскрипції порівняно з іншими інструментами, такими як Vosk або FasterWhisper.

***Ключові слова**: автоматичне ведення нотаток; розпізнавання мовлення; Telegram-бот; моделі GPT; обробка аудіо та відео; Whisper-1.*