

ЛЕГА МИХАЙЛО

Національний університет «Львівська політехніка»

<https://orcid.org/0009-0008-5925-9777>e-mail: mykhailo.o.leha@lpnu.ua

КОБИЛЮХ ЛЕСЛАВ

Національний університет «Львівська політехніка»

<https://orcid.org/0009-0006-5653-4943>e-mail: leslav.b.kobyliukh@lpnu.ua

МОДЕЛЬ СИСТЕМИ МОНІТОРИНГУ МОВЛЕННЯ З ВИКОРИСТАННЯМ СИСТЕМ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ РОЗПІЗНАВАННЯ УКРАЇНСЬКОЇ МОВИ

Знання й удосконалення мовних умінь для рідної мови є першочерговим з погляду самоідентифікації громадянина, і також важливим для збереження культурної спадщини, для можливості підвищити ефективність спілкування, сприйняття інформації тощо. Тому метою роботи стало розроблення моделі ІТ-супроводу удосконалення мовних навичок. Для цього було здійснено аналіз доступних інформаційних систем для удосконалення мовлення, а також досліджено можливості використання методів та засобів штучного інтелекту для розпізнавання мовлення українською і подальшого застосування у такій системі. Було проаналізовано наявні аналоги систем розпізнавання звуку, що використовують методи штучного інтелекту. На основі результатів дослідження, засобами UML було розроблено структурну модель інформаційної системи моніторингу та аналізу мовлення українською. Також запропоновано напрями для удосконалення такої системи моніторингу з використанням методів та засобів штучного інтелекту.

Ключові слова: розпізнавання звуку, українська мова, штучний інтелект, акустичні моделі, мовні моделі, транскрибування.

LEHA MYKHAILO, KOBLYUKH LESLAV

Lviv Polytechnic National University

MODEL OF SPEECH MONITORING SYSTEM WITH ARTIFICIAL INTELLIGENCE SYSTEM FOR OF UKRAINIAN LANGUAGE RECOGNITION

One of the means of improving speech is control over the use of specific vocabulary - the absence or presence of certain word usage. At one of the stages of functioning of the speech monitoring and analysis system, a necessary step is the user's speech recognition in real-time. During recognition, speech is transcribed, and the result is saved as a text file that can be edited, searched, etc. The Ukrainian language has several unique features that can affect the performance of speech recognition systems, for example, complex phonetics, morphology and syntax. Therefore, the task of developing a speech monitoring system that would transcribe speech in real-time using modern methods and tools is important, and it is advisable to start analyzing relevant methods and developing a model of such a system. A promising direction for solving speech recognition problems is the use of artificial intelligence systems, the use of such systems is becoming increasingly widespread, and their implementation is being improved. Speech recognition is based on acoustic and speech models. Acoustic modeling works out the relationship between linguistic units and audio signals, language modeling is used to distinguish words that sound the same or similar. The proposed speech monitoring and analysis system has among its main functions speech control, where the recognition results are compared with a dictionary, which can be configured to include or exclude specific words, and slang. If a match between the entered text and the dictionary is detected, the app will trigger an alarm that can be sent to the user's device, such as a phone or wristband. The vibration level can be adjusted and can also be ignored for certain words or phrases. External specialists can be involved in filling out such a dictionary, if it is, for example, about control of speech recovery in the event of an injury, or specialists in the education of children with special needs, if the system will be used to control the acquisition of necessary vocabulary. The use of methods and means of artificial intelligence is appropriate for the development of an information system that will be used to analyze speech in Ukrainian to control the quality of vocabulary, in this case, the system should respond to the use of profanity. The use of such an information system (in the form of a smartphone application) will help reduce the use of unwanted words and increase the culture of communication, improving the speech environment.

Keywords: sound recognition, Ukrainian language, artificial intelligence, acoustic models, language models.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Одним із засобів удосконалення мовлення є контроль за вживанням специфічної лексики – за відсутністю чи наявністю певних слововживань у мовленні. Система, що здійснюватиме моніторинг та аналіз мовлення може застосовуватись з різною метою, наприклад:

- Вивчення нових слів: система може допомогти користувачам вивчити нові слова, відстежуючи їх вживання та надаючи візуальні та аудіо підказки.
- Покращення словникового запасу: система може допомогти користувачам покращити свій словниковий запас, пропонуючи синоніми та альтернативи для слів, які використовуються занадто часто.
- Усунення мовленнєвих помилок: система може допомогти користувачам усунути мовленнєві помилки, такі як граматичні помилки або неправильна вимова.
- Контроль вживання ненормативної лексики: система може допомогти користувачам контролювати вживання ненормативної лексики, подаючи вібросигнал або інші повідомлення, коли користувач використовує небажані слова.

На одному з етапів функціонування системи моніторингу та аналізу мовлення необхідним кроком є власне розпізнавання мовлення користувача (speech-to-text) у режимі реального часу. Розпізнавання мовлення – це здатність машини або програми розпізнавати вимовлене та перетворювати його у машиночитний формат. В ході розпізнавання, мовлення транскрибується, а результат зберігається як текстовий файл, який можна

редагувати, здійснювати пошук тощо. В царині освіти актуальним прикладом застосування процесів транскрибування є створення відкритих освітніх ресурсів (Open Educational Resources). Національний університет «Львівська політехніка» з 2023 року бере участь в міжнародному освітньому проєкті «Відкриті освітні ресурси та Україна: Інформатика» (OER with Ukraine Informatics) за програмою проєктів DAAD «Україна цифрова: забезпечення академічної успішності у часи кризи». Разом з іншими учасниками проєкту, викладачами та науковцями вишів Харкова, Києва, Дніпра, політехніки готують освітні матеріали – понад п'ятдесят відеолекцій у 2023 році з царини комп'ютерних наук. Ці лекції партнер проєкту – Німецька національна науково-технічна бібліотека – розміщує у відкритому доступі на порталі <https://av.tib.eu/> (рис. 1) у вільному доступі за ліцензією CC BY (така ліцензія передбачає можливість копіювання та розповсюдження зі вказанням авторства).

Українська мова має кілька унікальних особливостей, які можуть впливати на роботу систем розпізнавання мовлення, наприклад, складні фонетика (окрім іншого, в українській мові спостерігається пом'якшення, яка впливає на вимову приголосних), морфологія (різні відмінки іменників, дієвідмінювання дієслів і моделі узгодження прикметників) та синтаксис (українська мова має вільний порядок слів, що ускладнює розпізнавання розмовної мови порівняно з мовами з фіксованим порядком слів). Тому важливим є завдання розроблення такої системи моніторингу мовлення, яка б здійснювала транскрибування мовлення в режимі реального часу з використанням сучасних методів та засобів, що доцільно почати аналізу відповідних методів та розроблення моделі такої системи.

The screenshot displays the Tib-av portal interface. At the top, there is a search bar and a search button. Below the search bar, the main title of the page is "Організація безпечного передавання у комп'ютерних мережах" (Organization of safe transmission in computer networks). The central part of the page features a video player with a green background. The video title is "Брандмауер та проху-сервер" (Firewall and Proxy Server). The video content shows a diagram illustrating the connection between a "Корпоративна мережа" (Corporate network) and the "Internet". The diagram includes a "Сервер-посередник" (Proxy server) and a "Брандмауер" (Firewall). The video player has a subtitle "to the Internet, to cache information resources" and a language selection menu with options: "off", "Ukrainian", "English", and "Ukrainian (auto-generated)". To the right of the video player, there are tabs for "SERIES", "TRANSCRIPT", and "ANNOTATIONS". Below the tabs, there is a list of related videos with titles like "Організація безпечного передавання у комп'ютерних мережах" and "Принципи організації передавання у безпроводних мережах". At the bottom of the page, there are logos for "Lviv Polytechnic National University" and "Projekt 'Open Education Resources with Ukraine'", along with a user profile for "Burov, Yevhen" and a "Recommendations" section.

Рис. 1. Сторінка порталу Tib-av з прикладом лекції, транскрибованої англійською

Аналіз досліджень та публікацій Формулювання цілей статті

У сучасному суспільстві технології голосових помічників стали невіддільною частиною повсякденного життя. Такі системи, як Siri від Apple, Alexa від Amazon та Google Assistant, стають все більш популярними завдяки своїй зручності та функціональності. Ці голосові помічники використовують автоматичне розпізнавання мови (automatic speech recognition, ASR) для розуміння і виконання команд користувача природною мовою. В галузі опрацювання природної мови набули поширення моделі BERT, BING, GEMINI, CHATGPT та інші. Дослідники вивчають можливість застосування цих моделей для розпізнавання українського мовлення [1], [2].

Перспективним напрямом вирішення завдань розпізнавання мовлення є використання систем штучного інтелекту, використання таких систем набуває значного поширення, а їх реалізація удосконалюється. Проаналізуємо доступні сервіси для транскрибування мовлення (табл. 1)

Розпізнавання мовлення ґрунтується на акустичних та мовних моделях. Акустичне моделювання опрацьовує зв'язок між лінгвістичними одиницями та аудіосигналами, мовне моделювання застосовується для розрізнення слів, що звучать однаково або подібно. Основні характеристики моделей розпізнавання звуку подано у таблиці 2.

Таблиця 1

Сервіси для транскрибування мовлення

Сервіс	Функція	Опис	Посилання
Descript Overdub	Озвучення тексту вашим голосом	Розбиває відео на кадри та текст, дає можливість редагувати звукову доріжку, в тому числі й озвучити текст вашим голосом.	[3]
Transkriptor	Транскрипція онлайн-зустрічей	Автоматично транскрибує зустрічі в Google Meets та зберігає їх у Google Docs. Розбиває запис на теми, щоб ви могли швидко зрозуміти, що коли обговорювалося.	[4]
Stenograf	Створення субтитрів	Сервіс, який дозволяє транскрибувати аудіо в текст, створювати субтитри, редагувати текст	[5]
Whisper	Транскрибування в реальному часі	Зручний і безкоштовний сервіс для транскрибування аудіо в текст, який може бути корисним для журналістів, творців подкастів і відео, а також для всіх, кому потрібно перетворювати записи голосу на тексти.	[6]

Таблиця 2

Моделі розпізнавання звуку

Тип моделі	Опис	Переваги	Недоліки	Посилання
<u>Акустичні моделі</u>				
HMM (Hidden Markov Models)	Статистичні моделі, які описують послідовність звуків	Прості	Неточні	[7]
GMM (Gaussian Mixture Models)	Статистичні моделі, які описують розподіл звуків.	Більш точні, ніж HMM	Складні	[8]
DNN (Deep Neural Networks)	Нейронні мережі, які описують зв'язок між звуками та текстом.	Найбільш точні	Найскладніші	[9]
<u>Мовні моделі</u>				
n-gram	Моделі, які описують ймовірність послідовності слів.	Прості	Неточні	[10]
RNN (Recurrent Neural Networks)	Нейронні мережі, які описують зв'язок між словами в реченні.	Більш точні, ніж n-gram	Складні	[11]

Методи та засоби штучного інтелекту застосовуються і для удосконалення процесу розпізнавання української мови [12], [13]. Тому власне використання систем штучного інтелекту для розпізнавання мовлення українською буде покладене в основу розроблюваної системи моніторингу мовлення.

Виклад основного матеріалу

Розроблення функціоналу системи моніторингу та аналізу мовлення розпочнемо із дослідження аналогів – наявних програм та додатків для IT-супроводу розвитку мовних умінь.

Duolingo – безплатний додаток на смартфон для вивчення мов, який має певний інноваційний підхід до вивчення, поєднуючи різноманітні інтерактивні вправи та ігри. Додаток містить понад 5500 різних вправ та тестів на різні тематики, включаючи граматику, аудіювання, письмо та читання. Duolingo також містить ігри, щоб зробити процес вивчення більш захопливим та ефективним. Додаток дозволяє користувачам відстежувати свій прогрес та отримувати рекомендації щодо покращення навичок мови. Duolingo використовує інтерактивний підхід до вивчення, використовуючи різні типи завдань, такі як вибір правильної відповіді, переклад з мови, аудіювання та тестування знань. Додаток містить різні рівні складності, що дозволяє користувачам починати з простих завдань та поступово переходити до складних, що допомагає краще засвоювати матеріал.

LanguageTool – безплатний додаток для перевірки граматики, орфографії та стилю написання текстів, який підтримує українську мову. Додаток допомагає письменникам та редакторам поліпшувати якість текстів дає змогу перевіряти текст на граматичні помилки, орфографію, стиль написання та пунктуацію. Додаток дає змогу користувачам додавати власні правила та словники. LanguageTool аналізує кожне слово в тексті та порівнює його зі словником та правилами, щоб визначити, чи містить текст помилки. Якщо помилки виявлені, додаток пропонує варіанти виправлення, які користувач може прийняти або відхилити.

P.I.D – додаток для Android, який допомагає користувачам вивчати українську мову. Додаток складається з великої кількості розмовних фраз та висловлювань для різних життєвих ситуацій. Крім того,

додаток містить вимову фраз та висловлювань, що дозволяє користувачам вивчати правильну вимову української мови. Додаток також містить тестові завдання для перевірки знань користувачів. Р.І.Д. дозволяє користувачам швидко знайти потрібні розмовні фрази та висловлювання за допомогою пошуку за ключовими словами або за темами.

«Мова – ДНК нації» – українська розробка, має ігрові елементи, інтерактивні вправи та завдання. На сайті є різноманітні вправи та тести на знання граматики, наголосів, фразеологізмів, орфографії тощо.

АнтиСуржик – додаток для Android, який допомагає користувачам коригувати українську мову та боротися з використанням суржику в мовленні. Додаток містить близько 1500 прикладів неправильного використання української мови з пропозиціями для їх правильного використання. Основний функціонал сформований в ігровій формі у вигляді кросвордів. Також в додатку є статистика, де користувач може відстежувати свої досягнення та прогрес.

Перевагами наведених прикладів є різноманіття методів та засобів для вивчення та удосконалення мови, зокрема, української, але жоден із них не функціонує в режимі реального часу і не має можливості задавати власну базу пошукуваної лексики. Спираючись на результати аналізу функцій, доступних розглянутим технологіям супроводу удосконалення мовлення, на рис. 2. подаємо структурну модель розробленої системи моніторингу мовлення.

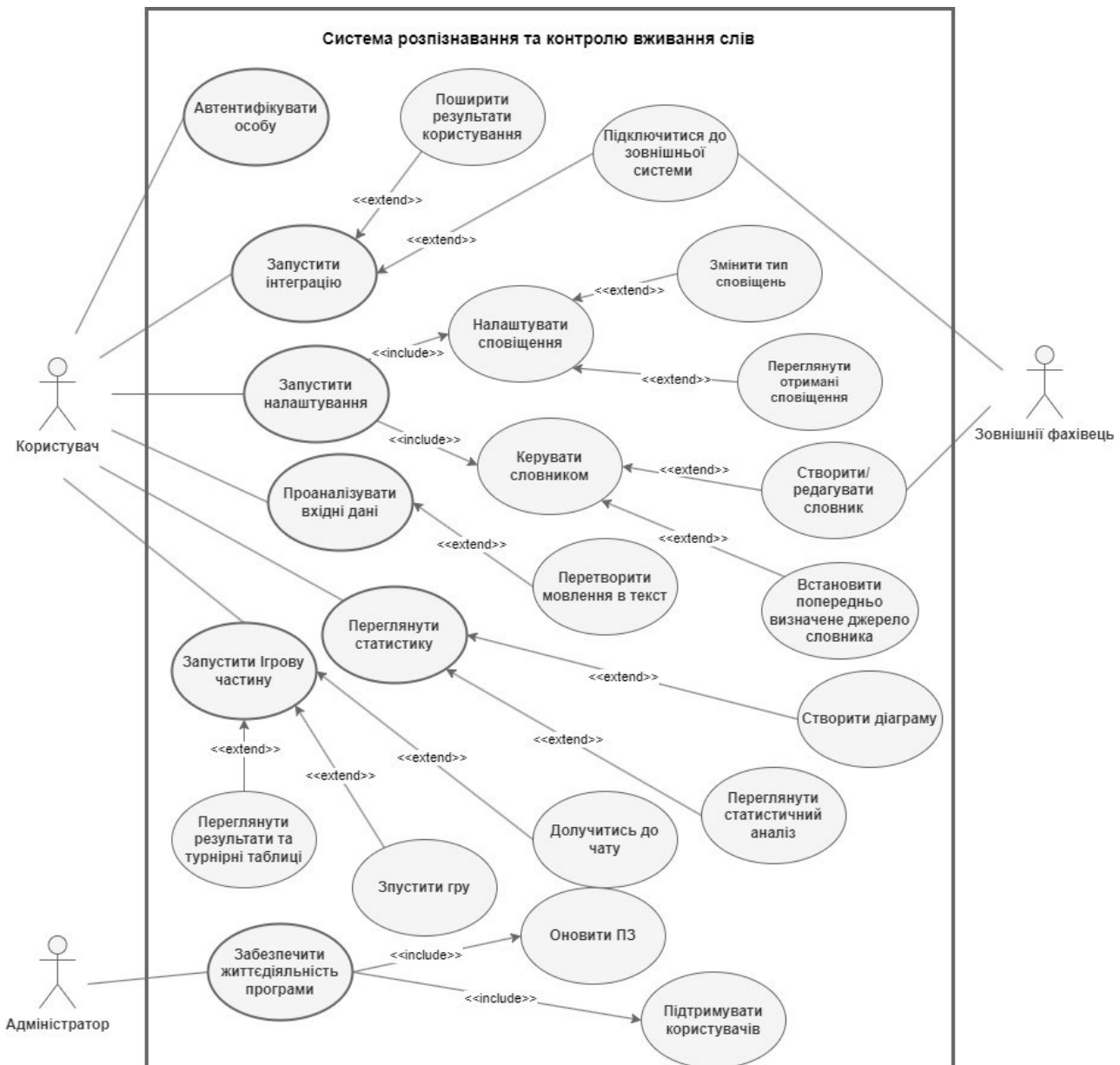


Рис. 2. Діаграма UML варіантів використання системи моніторингу та аналізу мовлення

Запропонована система моніторингу та аналізу мовлення серед основних функцій має контроль мовлення, коли результати розпізнавання порівнюються зі словником, що може бути налаштований для включення або виключення конкретних слів, категорій або сленгу. Якщо виявлено збіг між введеним текстом та словником, програма спрацьовує, активуючи сигналізацію, яка може бути відправлена на пристрій користувача, наприклад, телефон або браслет. Рівень вібрації може бути налаштований, а також може

ігноруватися для певних слів або виразів. До наповнення такого словника можуть бути залучені зовнішні фахівці, якщо йдеться, наприклад, про контроль відновлення мовлення в разі травми або фахівці освіти дітей з особливими потребами, якщо систему використовуватимуть для контролю засвоєння необхідної лексики.

Також розроблювана система накопичує статистику про вживання слів, включаючи їх частоту, категорії, час та контекст вживання – з подальшою візуалізацією. Ці дані корисні для відстеження прогресу чи самоаналізу. Програма відстежує кількість вживань слів користувачем протягом дня і збирає статистичні дані про їх вживання. Ці дані дозволяють користувачам переглядати свій прогрес у володінні мовою та аналізувати свої мовленнєві звички. Користувачі можуть переглядати цю статистику за різними періодами, що дозволяє їм зробити висновки про те, як вони прогресують у вивченні мови. Візуалізація результатів сприяє більш ефективному аналізу мовленнєвих звичок та плануванню подальших кроків у їх вдосконаленні.

Систему можна застосовувати також інтерактивно – для змагань з мовлення в ігровій формі. У програмі вбудована можливість ігрового спілкування, де користувачі можуть приєднатися до гри, спрямованої на уникнення вживання небажаних слів. У цій грі за кожне вживання небажаного слова з картки користувача знімаються певні бали або надається штраф. Такий підхід стимулює учасників гри активно працювати над покращенням своєї мовленнєвої поведінки, оскільки вони мають відчуття конкуренції та можливість виграти або втратити ресурси у грі. Це робить процес навчання цікавішим і вмотивованим.

Система може бути персоналізована для кожного користувача, налаштовуючи словник, рівень вібрації, типи графіків та інші параметри. Крім того, систему можна інтегрувати з іншими системами, такими як системи електронного навчання, мовні тренажери або чат-боти. Це дозволяє використовувати дані про вживання слів для персоналізації навчального контенту, надання мовних рекомендацій та покращення комунікативних навичок. Система контролю та аналізу мовлення може бути корисною для людей різного віку та професій, допомагаючи їм покращити мовленнєві навички, розширити словниковий запас, усунути мовленнєві помилки та контролювати вживання ненормативної лексики.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

Застосування методів та засобів штучного інтелекту є доцільним для розроблення інформаційної системи, що використовуватиметься для аналізу мовлення українською з метою контролю якості лексики, у цьому випадку – система повинна реагувати на вживання ненормативної лексики (лайка, сленг, суржик тощо). Використання такої інформаційної системи (у формі додатка для смартфона) сприятиме зменшенню вживання небажаних слів та підвищенню культури спілкування, оздоровленню мовленнєвого середовища.

Поза переліченими функціями, система моніторингу та аналізу мовлення може використовувати додаткові технології для покращення якості та ефективності розпізнавання та аналізу мовлення. Наприклад, акустична обробка мови може включати фільтрацію шуму та підсилення сигналу, щоб забезпечити краще розпізнавання мовлення в умовах шуму або перешкод. Фонетичний аналіз удосконалив процес розрізнення слів, які звучать схоже.

Крім розпізнавання мовлення та аналізу мовлення в режимі реального часу, є й інші сфери, де методи та засоби штучного інтелекту можуть бути імплементовані у роботу такої системи.

Аналіз настроїв – це процес визначення настроїв або емоцій, що стоять за фрагментом тексту. Методи аналізу настроїв на основі правил або лексикону, зазвичай, спираються на попередньо визначені списки позитивних, негативних і нейтральних слів, а також набір правил для комбінування оцінок настрою. Втім, таким методам іноді не вистачає глибини й контекстуального розуміння, яке забезпечує аналіз настроїв на основі машинного навчання. Використовуючи моделі машинного навчання система могла б автоматично аналізувати почуття у мовленнєвому контенті користувачів. Це розширить розуміння користувачів про їх стиль спілкування та допоможе покращити емоційний стан.

Автоматичне визначення теми передбачає використання методів вилучення ключових слів на основі частоти слів або інших евристичних заходів. Однак ці методи можуть бути не такими ефективними для охоплення базових тем у порівнянні з можливостями методів на основі машинного навчання. Алгоритми машинного навчання, такі як Latent Dirichlet Allocation (LDA), можуть використовуватися для автоматичного визначення основних тем, які обговорюються у дискусії користувача. Це може допомогти користувачам краще зрозуміти суть їхньої комунікації та підвищити її чіткість.

Покращення якості аудіо – з цією метою доцільно застосовувати алгоритми глибокого навчання для усунення фонового шуму, покращення чіткості мовлення та навіть адаптування до різних середовищ і умов запису.

Окрім ширшого застосування методів та засобів штучного інтелекту, функціонал системи моніторингу та аналізу мовлення можна модернізувати декількома способами, щоб покращити зручність у використанні та продуктивність. Деякі можливі покращення можуть охоплювати:

- Інтерфейс користувача: Можна розробити більш зручний і візуально привабливий інтерфейс за допомогою бібліотеки GUI, такої як Tkinter, PyQt або Kivy. Це дозволить користувачам взаємодіяти з програмою за допомогою кнопок, повзунків та інших графічних елементів замість текстового інтерфейсу.

- Фільтрування файлів: Покращення параметрів навігації та фільтрації файлів, дозволить користувачам фільтрувати аудіо файли на основі додаткових критеріїв, таких як тривалість файлу, розмір файлу або певні ключові слова, присутні в текстовому форматі файлу.

- Елементи керування: Реалізувавши розширені елементи керування для відтворення аудіо, як-от пауза, продовження відтворення, зупинка і регулювання гучності. Це забезпечить користувачам більше

контролю над відтворенням аудіо та покращить їхній загальний досвід користування.

- Асинхронне відтворення: Можна зробити відтворення аудіо асинхронним, щоб програма могла виконувати інші завдання на фоні під час відтворення аудіо, як-от відображення візуалізації аналізу даних або оновлення інших елементів інтерфейсу користувача.
- Аналіз аудіо: Інтегрування можливості аналізу аудіо, наприклад визначення настрою, темпу або гучності, та надання користувачеві аналізу даних параметрів розширить функціонал даної системи.
- Редагування аудіо: Дозволити користувачам виконувати основні завдання з редагування аудіо, наприклад обрізати, розділяти на менші та об'єднувати аудіофайли. Це дозволить користувачам налаштовувати аудіо-файли та керувати ними безпосередньо в програмі.

Література

1. Motyka V., Vysotska V., Chyrun L., Markiv O., Chyrun S., Kolyasa L. System Project for Ukrainian-language Feedback Tonality Analysis in the Health Care Field Based on BERT Model (2023) International Scientific and Technical Conference on Computer Sciences and Information Technologies. DOI: 10.1109/CSIT61576.2023.10323605
2. Synekop O., Lytovchenko I., Lavrysh Y., Lukianenko V. Use of Chat GPT in English for Engineering Classes: Are Students' and Teachers' Views on Its Opportunities and Challenges Similar? (2024) International Journal of Interactive Mobile Technologies, 18 (3), pp. 129 – 146 . DOI: 10.3991/ijim.v18i03.45025
3. Fix recorded speech as easy as typos with Overdub. URL: <https://www.descript.com/overdub>.
4. Transcribing Lecture Videos: Step-by-Step Guide. URL: <https://transkriptor.com/transcribe-lecture-videos/>.
5. Subtitles ready in minutes!. URL: <https://uk.stenograf.io> (date of access: 20.03.2024).
6. Radford A., Wook Kim J, Xu T., Brockman G., McLeavey C., and Sutskever I. Robust speech recognition via large-scale weak supervision. In Proceedings of the 40th International Conference on Machine Learning (ICML'23), Vol. 202. JMLR.org, 2023, Article 1182, 28492–28518.
7. Konduru A.K., Mazher Iqbal J.L. Emotion recognition from speech signals using digital features optimization by diversity measure fusion (2024) Journal of Intelligent and Fuzzy Systems, 46 (1), pp. 2547 - 2572. DOI: 10.3233/JIFS-231263
8. Rios-Urrego C.D., Rusz J., Orozco-Arroyave J.R. Automatic speech-based assessment to discriminate Parkinson's disease from essential tremor with a cross-language approach (2024) Digital Medicine, 7 (1), art. no. 37. DOI: 10.1038/s41746-024-01027-6
9. Sun C., Chen B., Chen F., Leng Y., Guo Q. Speech Keyword Spotting Method Based on Swin-Transformer Model (2024) International Journal of Computational Intelligence Systems, 17 (1), art. no. 61. DOI: 10.1007/s44196-024-00448-1
10. Rakib M., Hossain Md.I., Mohammed N., Rahman F. Bangla-Wave: Improving Bangla Automatic Speech Recognition Utilizing N-gram Language Models (2023) ACM International Conference Proceeding Series, pp. 297 - 301. DOI: 10.1145/3587828.3587872
11. Fedorov E., Nechyporenko O., Chychuzhko M., Chychuzhko V., Leshchenko R. NEURAL NETWORK-BASED METHODS FOR FINDING THE SHORTEST PATH AND ESTABLISHING ASSOCIATIVE CONNECTIONS BETWEEN OBJECTS (2023) Radioelectronic and Computer Systems, 2023 (2(106)), pp. 54 - 64. DOI: 10.32620/REKS.2023.2.05
12. Basystiuk O., Melnykova N., Rybchak Z. Multimodal Learning Analytics: An Overview of the Data Collection Methodology (2023) International Scientific and Technical Conference on Computer Sciences and Information Technologies. DOI: 10.1109/CSIT61576.2023.10324177
13. Pronina O., Piatyko O. The recognition of speech defects using convolutional neural network (2023) CEUR Workshop Proceedings, 3358, pp. 73 - 86.