УДК 519.688 https://doi.org/10.31713/vt3202017

Бачишина Л. Д., к.е.н., доцент, Рощенюк А. М., к.пед.н., доцент
(Національний університет водного господарства та природокористування, м. Рівне)

ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ TYPESCRIPT ТА ANGULAR ПРИ РОЗРОБЦІ ВЕБСЕРВІСУ ПЕРЕТВОРЕННЯ ДРУКОВАНОГО ТЕКСТУ В АУДІО

Проведено дослідження технологій оптичного розпізнавання, їх структури, методів попередньої обробки та сегментації зображень текстових символів, що застосовуються для переведення друкованих матеріалів та сканованих зображень у цифровий формат. Проаналізовано технології синтезу мови та сфери їх застосування, визначені переваги та недоліки кожної з них. Досліджено використання систем оптичного розпізнавання символів для синтезу мови. Виконано дослідження штучних нейронних мереж, їх складових та алгоритмів навчання в контексті перетворення друкованих текстів у аудіоформат. Визначено переваги застосування вебсервісу Angular та технології TypeScript для трансформації друкованого тексту в аудіофайл. Розроблено вебсервіс, який розпізнає сканований образ, перетворює документ в цифровий формат та озвучує результати сканування. Сервіс призначений для широкої аудиторії користувачів та дозволяє вирішувати потреби, що пов'язані з навчанням, побутом та бізнесом. Перевагою додатка є зручний інтуїтивний інтерфейс та озвучування українською, англійською та російською мовами, що дозволить використовувати його у закладах освіти різного рівня, від дитячого садочка до університету. Використання аудіокниг та озвучування різноманітних матеріалів у інклюзивній та онлайн-освіті є актуальним в сучасному суспільстві. Програма допоможе вивести процес навчання на новий рівень, сприятиме полегшенню процесу створення онлайн-курсів та матеріалів для інклюзивної освіти. Вебзастосунок представлений у двох версіях: для персонального комп'ютера та як мобільний додаток.

***Ключові слова*: вебсервіс; аудіофайл; синтез мовлення; оптичне розпізнавання символів; нейронні мережі.**

Вступ. Переваги використання інформаційних технологій у наш час не піддаються жодному сумніву. За короткий час людство зробило величезний прорив у цьому напрямку. Інтернет-технології надійно увійшли в усі сфери діяльності людства. Можливості мобільних пристроїв удосконалюються щоденно. Перебуваючи далеко від стаціонарного комп'ютера або ноутбука людина може виконувати безліч операцій: вести бізнес, здійснювати закупівлі, читати, створювати комп'ютерні ігри тощо. Проте, окремі види діяльності потребують швидкої конвертації друкованих текстів у аудіофайли. Не зручно читати тексти з екрану, коли ви сидите за кермом автомобіля, займаєтеся спортом або просто перебуваєте на пішій прогулянці. Хоча цей час можна було б також використовувати для отримання інформації, що дуже важливо при сучасному темпі життя. Інклюзивна та онлайн-освіта теж потребують переведення матеріалів у аудіоформат. Такий матеріал можна завантажувати в програвач або мобільний телефон і слухати майже в будь-який час і в будь-якому місці.

Метою проєкту є реалізація ґрунтовного аналізу технологій оптичного розпізнавання символів, синтезу мовлення та розробка вебсервісу для перетворення друкованого тексту в аудіофайл, в якому користувач може перетворити текстовий документ у аудіоформат, з використанням методів, які базуються на нейронних мережах.

Засоби та технології розробки вебсервісу. При розробці програмного продукту використовувався TypeScript – мова програмування для розробки вебзастосунків. Цей засіб має кілька переваг над JavaScript і добре підходить для розробки великих додатків. Програми написані з використанням TypeScript легко відлагоджуються та структуруються. TypeScript представляє мову програмування на основі JavaScript. Проте TypeScript – це не просто новий JavaScript, це строго типізована і компільована мова; вона реалізує концепції, які властиві об'єктно-орієнтованим мовам; дозволяє швидше і простіше писати великі складні комплексні програми, відповідно їх легше підтримувати, розвивати, масштабувати і тестувати ніж на стандартному JavaScript; розвивається як opensource-проєкт і, як і багато проєктів, хоститься на GitHub; є кросплатформною, а це значить, що для розробки можемо використовувати як Windows, так і MasOS або Linux [3].

Для розробки front-end було застосовано Angular. Angular представляє фреймворк від компанії Google для створення клієнтських додатків. Насамперед він націлений на розробку SPA-рішень

(Single Page Application), тобто односторінкових додатків. В цьому плані Angular є спадкоємцем іншого фреймворка AngularJS. Водночас Angular – це не версія AngularJS, а принципово новий фреймворк. Angular надає таку функціональність, як двостороннє зв'язування, що дозволяє динамічно змінювати дані в одному місці інтерфейсу при зміні даних моделі в іншому (шаблони, маршрутизація та ін.). Однією з ключових особливостей Angular є те, що він використовує як мова програмування TypeScript [2], яка має ряд переваг перед традиційним JavaScript. Це, зокрема, швидкість розробки, надійність, швидкодія програми та застосування будь-якого сучасного браузера для роботи з програмою.

Процес перетворення зображення чи тексту в цифровий документ складається з кількох етапів, один з яких – оптичне розпізнавання символів (ORC), тобто електронне чи механічне переведення зображення в електронний документ. Отриманий в результаті оптичного розпізнавання файл можна використовувати для подальшого опрацювання текстовими або графічними редакторами.

ORC набуло широкого розповсюдження у вирішенні питань перетворення книг та документів у електронний вигляд, автоматизації систем обліку в бізнесі або для публікацій тексту на вебсторінках [6]. Оптичне розпізнавання символів дозволяє редагувати текст, зберігати його у більш компактній формі, друкувати або демонструвати матеріал без втрати якості, аналізувати інформацію, а також застосовувати до тексту електронний переклад, форматування чи здійснювати синтез мовлення цього тексту, робить можливим пошук конкретних слів або фраз. За час свого існування людство створило значну кількість рукописних та друкованих книг. Деякі з них збереглись у єдиному екземплярі, що виводить важливість перетворення їх у електронний вигляд на новий рівень. Електронні копії книжок можна читати на телефоні, планшеті чи іншому пристрої.

Системи оптичного розпізнавання символів складаються з кількох блоків, серед них: блок сегментації (локалізації та виділення) елементів тексту, блок попередньої обробки зображення, блок виділення ознак, блок розпізнавання символів, блок обробки результатів розпізнавання. Ці алгоритмічні блоки відповідають послідовним операціям обробки та аналізу зображень. Спочатку здійснюється виділення текстових областей, рядків та розбиття зв'язних текстових рядків на окремі знакомісця, кожне з яких відповідає одному текстовому символу. Після розбиття (іноді до або у процесі розбиття) символи,

представлені у вигляді двовимірних матриць пікселів, піддаються згладжуванню, фільтрації з метою усунення шумів, нормалізації розміру, а також іншим перетворенням з метою виділення утворюючих елементів або чисельних ознак, які у подальшому використовуються для розпізнавання.

Розпізнавання символів проходить у процесі порівняння виділених характерних ознак з еталонними наборами та структурами ознак, що формуються та запам'ятовуються у процесі навчання системи на еталонних та/або реальних прикладах текстових символів.

На фінальному етапі смислова або контекстна інформація може бути використана як для вирішення невизначеностей, що виникають при розпізнаванні окремих символів, які мають ідентичні розміри, так і для корегування помилково прочитаних слів та фраз. Серед систем ORC найбільш відомими є ABBYY FineReader, Brainware, CuneiForm.

Отриманий після розпізнавання текст можна озвучити. Для цього використовують технологію, що називається синтез мовлення. У широкому сенсі, синтез мовлення – це відновлення форми мовного сигналу за його параметрами [4]. Синтезом мовлення насамперед називають усе, що пов'язано зі штучним утворенням людської мови з тексту або зображення з використанням аналогової або цифрової техніки. Безпосереднє перетворення тексту у звук відбувається голосовим движком. З таким способом синтезування ми часто зустрічаємось, навіть не звертаючи на це уваги: у інформаційно-довідкових системах, для людей з особливими потребами (слабозорих або таких, що не можуть розмовляти); при оголошеннях у громадських місцях, у робототехніці, акустичному діалозі людини з комп'ютером. Синтез мови застосовується в багатьох сферах діяльності людства – це й голосові асистенти, розумні будинки тощо.

Для синтезу мовного сигналу широко застосовують штучний інтелект (нейронні мережі). Відомо, що цілий ряд компонентів синтезу мови можна замінити нейронними мережами, значно покращивши якість існуючих алгоритмів[7]. Існує два способи перетворення тексту у мову – конкатенативний та параметричний. Перший засновано на записі коротких фрагментів, що в подальшому об'єднуються, утворюючи речення. Запис виходить чіткий та зв'язний, але беземоційний. Конкатенаційні системи складаються з величезних баз даних, розробка їх довготривала. При параметричному синтезі з тексту спочатку виділяють лінгвістичні ознаки витягуються ознаки, потім – від-

повідний мовний сигнал: кепстральні коефіцієнти частоти та лінійну спектрограму. Параметричний синтез менш затратний ніж конкатенативні системи, проте на практиці виникає багато артефактів, що призводять до приглушеної мови з «металевим» звуком, який зовсім не схожий на природне звучання. Вирішити описані проблеми можна з допомогою нейронних мереж.

Нейронні мережі – це потужний інструмент, що може апроксимувати функцію довільної складності, тобто деякий набір вхідних даних X перетворити у набір вихідних даних Y. У контексті нашої задачі X – це текст, Y – звуковий сигнал. При використанні нейронних мереж для перетворення тексту у мову отримують систему, яка навчається на основі передбачень, що генеруються самою мережею. Штучна нейронна мережа моделює роботу біологічної нейронної мережі, та побудована із набору з'єднаних вузлів, які називають штучними нейронами (рис. 1).

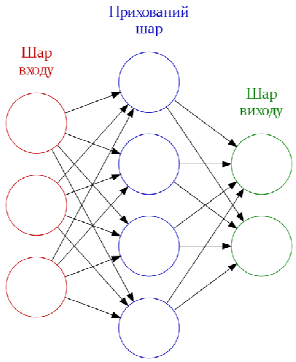


Рис. 1. Схема штучної нейронної мережі

Нейрони можуть передавати сигнали один одному, якщо між ними існує зв'язок. Той нейрон, що отримує сигнал здатний його обробити та передати іншим, що пов'язані з ним. Нейрон, що має мітку *j*, та отримує вхід $p_j(t)$ від нейронів-попередників, складається з таких частин:

- збудження $a_j(t)$, яке залежить від дискретного параметра часу *t*;
- порогу θ_j, що залишається без змін, якщо функція навчання не замінить його;
- функції збудження f, яка обраховує нове збудження у заданий час t+1 за попередніми значеннями $a_j(t)$, θ_j та мережевого входу $p_j(t)$, даючи у результаті відношення:

$$a_j(t+1) = f(a_j(t)), p_j(t), \theta_j, \quad (1)$$

- функції виходу f_{out}, що обраховує вихід з активації

$$o_j(t) = f_{out}(a_j(t)). \quad (2)$$

У нейрона входу немає попередників, він представляє собою інтерфейс входу для всієї мережі. Таким ж чином, у нейрона виходу немає наступників, тому він виконує роль інтерфейсу виходу для всієї мережі.

Мережа складається зі з'єднань, кожне з яких передає вихід нейрону *i* до входу нейрону *j*. Таким чином *i* є попередником *j*, а *j* є наступником *i*. Кожне з'єднання має вагу ω_{ij}.

Функція поширення обчислює вхід $p_j(t)$ до нейрону *j* з виходів $o_i(t)$ нейронів-попередників, і зазвичай має вигляд:

$$p_j(t) = \sum_i o_i(t)\omega_{ij}. \quad (3)$$

Правило навчання – це правило або алгоритм, який змінює параметри нейронної мережі, щоб заданий вхід до мережі видавав придатний вихід. Цей процес навчання зазвичай полягає в зміні ваги та порогів змінних мережі.

Такі системи мають здатність навчатись, розвиватись та збільшувати продуктивність за результатами своєї роботи та на основі прикладів. Навчання нейронної мережі в загальному означає вибір однієї моделі із запропонованої множини моделей. У нашому випадку під час навчання в якості вхідних послідовностей використовують записані звукові сигнали. Після тренування нейронної мережі можна генерувати фрази.

Найбільш поширеними методами синтезу людської мови є WaveNets, Tacotron 2, DCTTS, Deep Voices.

Програмна реалізація розробки вебсервісу. У якості середовища розробки вебсервісу було обрано Visual Studio Code. Дане середовище є кросплатформним, добре працює на таких операційних системах, як Windows, Linux, macOS, що, безперечно, є значною перевагою. Visual Studio Code містить має налагоджувач, інструменти для роботи з Git, підсвітку синтаксису та засоби рефакторингу. А також надзвичайно широкий спектр можливостей кастомізації інтерфейсу та функціоналу, що дозволяє кожному користувачу налаштувати дане середовище під свій смак та потреби.

У процесі роботи над розробкою вебзастосунку було створено наступні компоненти: header, main, footer. Компонент header відпові-

дає за верхню частину сайту, на якому розміщені логотипи, кнопки навігації та контакти, що дозволяють зв'язатись з автором. Компонент є адаптивним та правильно відображається на екранах з будь-якою шириною. Компонент main відповідає за основний контент сторінки. На сторінці, яку відображає браузер, він розташований між компонентами header та footer. Компонент footer відповідає за нижню частину сторінки. В ньому продубльовані кнопки навігації сайту та посилання на пошту, соціальні мережі. Вищезгадані компоненти розроблені з допомогою HTML та CSS. За допомогою мови стилізації CSS було видозмінено зовнішній вигляд кожного з елементів у компонентах: колір, розмір, шрифт та товщину шрифту тексту, колір фону, відступи між елементами, їх розмір. Реалізовано ефекти при наведенні курсору миші на окремі елементи. Створено медіа-запит для максимальної ширини 768 пікселів та іншими стилями для окремих елементів.

Вебзастосунок складається з трьох сторінок: головної (Home) та двох інформаційних (About та How it works). У верхній частині сторінки Home (рис. 2) розміщено логотипи, навігаційні кнопки, що забезпечують перехід між сторінками сервісу, та посилання на пошту і соціальні мережі. В центральній частині сторінки розміщено основний контент, що складається з кнопки вибору файлу та мови розпізнавання, короткої інструкції для користувача, кнопок початку розпізнавання та синтезу мовлення, зони для виводу результату сканування. У нижній частині сторінки продубльовано навігаційні кнопки.

Рис. 2. Головна сторінка

Після вибору необхідного файлу зображення за допомогою кнопки «Choose your file», вибору мови розпізнавання тексту (укр., рос., англ.) та натискання кнопки «Enter» відбувається розпізнавання тексту на цьому зображенні. Коли процес розпізнавання закінчується, результат виводиться у відведене поле, під написом «Result». Під полем для виводу текстового результату розміщено кнопку «Voice Synthesis», натискання на яку генерує виконання перетворення тексту в аудіоформат.

Сторінка About має інформаційний характер і містить короткий опис технології оптичного розпізнавання символів. Наступна сторінка, яка називається How it works, також має інформаційний характер і містить інформацію про етапи, які необхідно виконати для перетворення друкованого тексту в аудіофайл (рис. 3).

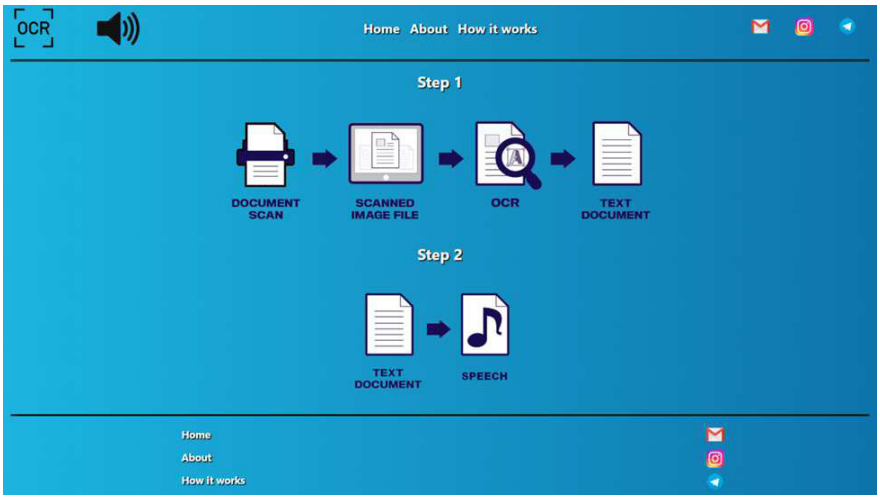


Рис. 3. Інформаційна сторінками з описом порядку виконання роботи

Аналіз отриманих результатів та висновки. Було проведено дослідження технологій оптичного розпізнавання символів, їх структури, методів попередньої обробки та сегментації зображень текстових символів, співставлення зображень та шаблонів, статистичні характеристики та постобробку результатів розпізнавання. Проаналізовано технології синтезу мови та сфери їх застосування, визначено переваги та недоліки кожної з них.

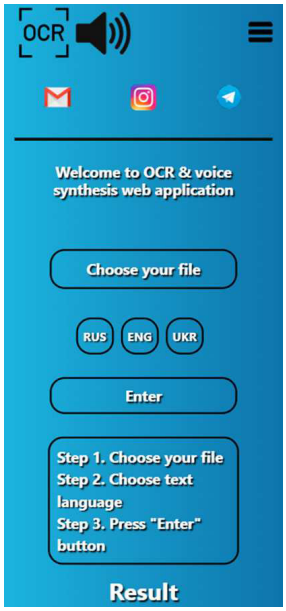
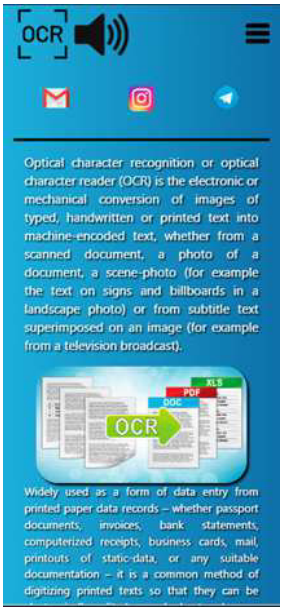


Рис. 4. Мобільна версія головної сторінки



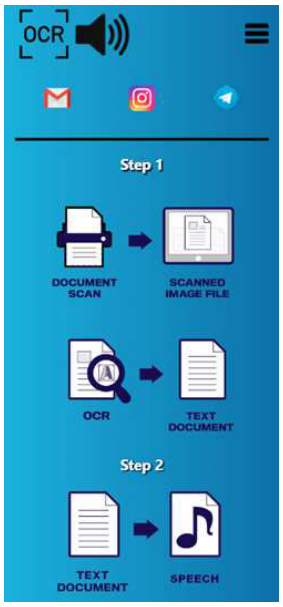


Рис. 5. Мобільна версія інформаційної сторінки з описом технологій та порядку роботи

Виконано дослідження штучних нейронних мереж, їх складових та алгоритмів навчання в контексті перетворення друкованих текстів у аудіоформат. Виконано програмну реалізацію вебсервісу для перетворення друкованого тексту в аудіофайл за допомогою нейронних мереж з використанням мови програмування TypeScript та технології створення вебресурсів Angular.

1. Людовик Т. Н. Использование речевых баз данных большого объема при синтезе речи в системах искусственного интеллекта. URL: http://masters.donntu.org/2015/fknt/bakalenko/library/ ludovick.pdf (дата звернення: 20.07.2020). **2.** Сайт про програмування. Що таке Angular. URL: https://metanit.com/web/angular2/1.1.php (дата звернення: 20.07.2020). **3.** Сайт про програмування. Що таке TypeScript. URL: https://metanit.com/web/typescript/1.1.php (дата звернення: 20.07.2020). **4.** Синтез мовлення. URL: https://uk.wikipedia.org/wiki. (дата звернення: 20.07.2020). **5.** Фримен А. Angular для профессионалов. СПб.: Питер, 2018. 800 с. **6.** Что такое OCR и для чего оно используется? URL: http://artismedia.by/blog/chto-takoe-ocr-i-dlya-chego-ono-ispolzuetsya/ (дата звернення: 20.07.2020). **7.** Моделі нейронних мереж. URL:

https://studme.com.ua/1246122010028/neural/models.htm (дата звернення: 20.07.2020).

REFERENCES:

1. Lyudovik T. N. Ispolzovanie rechevyih baz dannyih bolshogo obyema pri sinteze rechi v sistemah iskusstvennogo intellekta. URL: http://masters.donntu.org/2015/fknt/bakalenko/library/ ludovick.pdf (data zvernennia: 20.07.2020). **2.** Sait pro prohramuvannia. Shcho take Angular. URL: https://metanit.com/web/angular2/1.1.php (data zvernennia: 20.07.2020). **3.** Sait pro prohramuvannia. Shcho take TypeScript. URL: https://metanit.com/web/typescript/1.1.php (data zvernennia: 20.07.2020). **4.** Syntez movlennia. URL: https://uk.wikipedia.org/wiki. (data zvernennia: 20.07.2020). **5.** Frymen A. Angular dlia professyonalov. SPb.: Pyter, 2018. 800 s. **6.** Chto takoe OCR y dlia cheho ono yspolzuetsia? URL: http://artismedia.by/blog/chto-takoe-ocr-i-dlya-chego-ono-ispolzuetsya/ (data zvernennia: 20.07.2020). **7.** Modeli neironnykh merezh. URL: https://studme.com.ua/1246122010028/neural/models.htm (data zvernennia: 20.07.2020).

Bachyshyna L. D., Candidate of Economics (Ph.D.), Associate Professor, Roshcheniuk A. M., Candidate of Pedagogic Sciences (Ph.D), Associate Professor (National University of Water and Environmental Engineering, Rivne)

USE OF TYPESCRIPT AND ANGULAR TECHNOLOGIES IN DEVELOPMENT WEB SERVICE CONVERTING PRINTED TEXT INTO AUDIO

The research of optical pattern recognition technologies, their structure, methods of pre-processing and segmentation of text symbols' images used for transformation of printed materials and scanned images into digital format has done. The technologies of speech synthesis and their area of use have been analysed. The advantages and disadvantages of each have been identified. The use of optical pattern recognition systems for speech synthesis has been studied. Research of artificial neural networks, their components and learning algorithms at the context of converting printed texts into audio format has done. The benefits of using the Angular web service and TypeScript technology to transform printed text into the audio file

have been identified. The web service for recognizing the scanned image has been created. It converts the document into digital format and sounds the scan results. This service is intended for a wide users' audience and helps to solve the needs related to education, everyday life and business. The advantage of this application is a convenient intuitive interface and sounding in Ukrainian, English and Russian. This allows using it in educational institutions of various levels, from kindergarten to university. The use of audiobooks and the translation of different materials in inclusive and online education is very important in the modern society. The application will help bring the learning process to a new level and will make the process of creating online courses and materials for inclusive education simple and fast. The web service is presented in two versions: for the personal computer and the mobile application.
***Keywords*: web service; audio file; speech synthesis; optical pattern recognition; neural networks.**

Бачишина Л. Д., к.э.н., доцент, Рощенюк А. М., к.пед.н., доцент
(Национальный университет водного хозяйства и природопользования, г. Ровно)

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ TYPESCRIPT И ANGULAR ДЛЯ СОЗДАНИЯ ВЕБ-СЕРВИСА ПРЕОБРАЗОВАНИЯ ПЕЧАТНОГО ТЕКСТА В АУДИО

Проведено исследование технологий оптического распознавания образов, их структуры, методов предварительной обработки и сегментации изображений текстовых символов, используемых для перевода печатных материалов и сканированных изображений в цифровой формат. Проанализированы технологии синтеза речи и сферы их применения, определены преимущества и недостатки каждой из них. Исследовано использование систем оптического распознавания символов для синтеза речи. Выполнены исследования искусственных нейронных сетей, их составляющих и алгоритмов обучения в контексте преобразования печатных текстов в аудиоформат. Определены преимущества применения веб-сервиса Angular и технологии TypeScript для трансформации печатного текста в звуковой файл. Создан веб-сервис, который распознает отс-

канированный образ, превращает документ в цифровой формат и озвучивает результаты сканирования. Сервис предназначен для широкой аудитории пользователей и позволяет решать потребности, связанные с обучением, бытом и бизнесом. Преимуществом приложении есть удобный интуитивный интерфейс и озвучивание на украинском, английском и русском языках, что позволит использовать его в учебных заведениях различного уровня, от детского сада до университета. Использование аудиокниг и озвучивания различных материалов в инклюзивном и онлайн-образовании является актуальным в современном обществе. Программа поможет вывести процесс обучения на новый уровень, будет способствовать облегчению процесса создания онлайн-курсов и материалов для инклюзивного образования. Веб-приложение представлено в двух версиях: для персонального компьютера и мобильное приложение.
***Ключевые слова*: веб-сервис; аудиофайл; синтез речи; оптическое распознавание символов; нейронные сети.**