

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ комп'ютерних наук _____
(повна назва)

Кафедра _____ програмної інженерії _____
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти _____ другий (магістерський) _____

Дослідження методів аналізу неструктурованих текстів для заповнення форм

Виконав:
студент (ка) 2 курсу, групи ІПЗм-22-4

_____ Онацький Р. Б. _____

Спеціальність 121 – Інженерія програмного
забезпечення _____
(код і повна назва спеціальності)

Тип програми освітньо-наукова _____

Керівник доц. каф. ІІІ Турута О. П. _____
(посада, прізвище, ініціали)

Допускається до захисту
Зав. кафедри

_____ (підпис)

_____ З.В.Дудар _____
(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет комп'ютерних наук
Кафедра програмної інженерії
Рівень вищої освіти другий (магістерський)
Спеціальність 121 – Інженерія програмного забезпечення
Тип програми освітньо-наукова
Освітня програма Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)
«___» _____ 2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Онацькому Роману Богдановичу

1. Тема роботи «Дослідження методів аналізу неструктурованих текстів для заповнення форм»
Затверджена наказом по університету від 20.03.2024 р. № 250Ст
2. Термін подання студентом роботи до екзаменаційної комісії 19.06.2024
3. Вихідні дані до роботи описані методи аналізу неструктурованих текстів, порівняння й аналіз типів форм і методів заповнення їх текстом, аналіз інструментів для перетворення мовлення в текст в режимі реального часу, пояснювальна записка.
4. Перелік питань, що потрібно опрацювати в роботі мета роботи, аналіз предметної галузі і постановка задачі, огляд та аналіз літературних джерел з дослідження, дослідження теоретичне, дослідження практичне.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз предметної галузі	01.05.2024 – 03.05.2024	<i>Виконано</i>
2	Розробка постановки задачі	04.05.2024 – 06.05.2024	<i>Виконано</i>
3	Розробка ПЗ	07.05.2024 – 10.05.2024	<i>Виконано</i>
4	Проведення експерименту	13.05.2024 – 14.05.2024	<i>Виконано</i>
5	Оформлення пояснювальної записки	15.05.2024 – 27.05.2024	<i>Виконано</i>
6	Підготовка презентації та доповіді	01.06.2024 – 02.06.2024	<i>Виконано</i>
7	Попередній захист	05.06.2024	<i>Виконано</i>
8	Нормоконтроль, рецензування	14.06.2024	<i>Виконано</i>
9	Здача роботи у електронний архів	16.06.2024	<i>Виконано</i>
10	Допуск до захисту у зав. кафедри	19.06.2024	<i>Виконано</i>

Дата видачі завдання 27.04.2024 р.

Студент

(підпис)

Онацький Р. Б.

Керівник роботи

(підпис)

доц. каф. ПІ, Турута О. П.
(посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Пояснювальна записка містить 57 стор., 15 табл., 22 джерела, 24 рисунки.

АНАЛІЗ НЕСТРУКТУРОВАНИХ ТЕКСТІВ, ВЕБ ФОРМА, МОВЛЕННЯ, МОДЕЛЬ МОВЛЕННЯ-ДО-ТЕКСТУ, НЕЙРОМЕРЕЖА, ОБРОБКА ПРИРОДНОЇ МОВИ, САЙТ, ТЕКСТ

Об'єктом дослідження є методи аналізу неструктурованих текстів.

Метою кваліфікаційної роботи є дослідження та визначення ефективних методів аналізу неструктурованих текстів в контексті заповнення форм на вебсайтах.

У результаті роботи було здійснено аналіз предметної галузі, поставлена задача для дослідження, проведено дослідження методів аналізу неструктурованих текстів.

ANALYSIS OF UNSTRUCTURED TEXTS, NATURAL LANGUAGE PROCESSING, NEURAL NETWORK, SPEECH, SPEECH-TO-TEXT MODEL, WEB FORMS, WEBSITE, TEXT.

The object of the research is the methods of unstructured text analysis.

The aim of the master's research is to investigate and determine effective methods for the analysis of unstructured texts in the context of filling out forms on websites.

As a result of the work, an analysis of the subject area was carried out, a research task was formulated, and an investigation of methods for the analysis of unstructured texts was conducted.

Я, Онацький Роман Богданович, студент гр. ПЗм-22-4, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Дослідження методів аналізу неструктурованих текстів для заповнення форм», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не

містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Усі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

ЗМІСТ

Вступ.....	7
1 Аналіз предметної області та постановка задачі	8
1.1 Аналіз предметної області	8
1.2 Постановка завдання дослідження.....	9
2 Дослідження методів аналізу неструктурованих текстів.....	11
3 Експериментальне дослідження	19
3.1 Планування експерименту	19
3.2 Speech Recognition Anywhere.....	20
3.3 Mozilla Speech Secognition API.....	24
3.4 OpenAI Whisper API.....	28
4 Аналіз результатів дослідження	31
Висновки	38
Перелік джерел посилання.....	40
Додаток А. Перелік джерел посилання за науковими напрямками керівника та науковців кафедри Програмної Інженерії..	Error! Bookmark not defined.
Додаток Б. Результати перевірки на антиплагіат	Error! Bookmark not defined.
Додаток В. Слайди презентації.....	Error! Bookmark not defined.
Додаток Г. Тези доповіді для конференції	Error! Bookmark not defined.
Додаток Д. Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015	Error! Bookmark not defined.

ВСТУП

Серед багатьох актуальних проблем сучасної науки одне з найважливіших місць займає проблема автоматизованої обробки мовної інформації [1]. В наш час неструктуровані тексти є неабияким джерелом інформації і відіграють ключову роль в багатьох сферах. Люди і підприємства взаємодіють з величезними обсягами даних, зокрема у формі тексту на вебсайтах, соціальних мережах, електронних листах і багатьох інших ресурсах. Однак, величезний об'єм неструктурованої інформації створює виклики для ефективного аналізу та використання цих даних.

Методи аналізу неструктурованих текстів надають можливість конвертувати сирий текст у структуровані дані, зрозумілі для комп'ютера. Застосування інструментів природної обробки мови (NLP) дозволяє автоматизувати цей процес, роблячи його більш ефективним і точним. Також це дозволяє виявляти та аналізувати настрої, відгуки у соціальних мережах та на вебсайтах. Дослідження в цій сфері допомагають визначати популярність продуктів, надійність брендів, збільшувати здатність підприємств, розуміти та забезпечувати потреби клієнтів [2].

Однією з важливих областей дослідження може бути заповнення вебформ голосом, саме на це й спрямована ця магістерська робота. Це зручно для всіх користувачів, адже дозволяє економити багато часу і робить заповнення форм більш зручним. Також це відкриває нові перспективи для людей з обмеженими можливостями, наприклад, з обмеженим зором або руховими функціями. Такі технології мають стати кроком у напрямку створення більш інклюзивного інтернет-простору.

Велика перевага голосового заповнення форм розкривається і для користувачів мобільних пристроїв, особливо тих, хто перебуває в дорозі або не має можливості користуватися екраном. Це робить сайт більш доступним та зручним для ширшої аудиторії.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Аналіз предметної області

Предметною галуззю дослідження є неструктуровані тексти та їх аналіз. Цей термін означає дані, що не мають жорсткої організації або визначеної структури. Це можуть бути дані соціальних мереж, електронна пошта, відгуки користувачів, статті новин тощо. При аналізі таких текстів важливо розуміти природу мови, контекст і семантику. Основною задачею аналізу неструктурованих текстів є виділення ключової інформації для вирішення поставлених завдань. Для цього використовуються методи обробки природної мови (Natural Language Processing), таким чином, вони грають важливу роль у видобуванні та структуризації інформації з таких текстів.

Неструктуровані тексти є невіддільною частиною інформаційного простору, але їх обробка та аналіз є непростим завданням. Їх наявність спровокувала розвиток технологій аналізу і виникнення розгалуженої галузі, яка має на меті забезпечити можливість отримувати значущі дані з безлічі текстових джерел [3]. Незважаючи на досягнення в цій галузі, все ще існують певні складнощі, наприклад, розуміння контексту, іронії та амбігвітету в мові, а також питання етики та конфіденційності даних. Тим не менш, розвиток штучного інтелекту та глибокого навчання постійно збільшує можливості аналізу неструктурованих текстів. Поліпшується точність алгоритмів, розширюється обсяг аналізу мови та її контексту, розвиваються нові методи обробки.

З аналізом неструктурованих текстів пов'язані безліч застосувань у сучасному світі. Розглянемо кілька задач, до яких активно залучається такий тип технологій:

- системи моніторингу соціальних мереж, використовують аналіз текстової інформації для визначення громадської думки, виявлення трендів та взаємодії між користувачами [4];
- аналіз відгуків користувачів, компанії активно використовують його для вдосконалення своїх продуктів та послуг. Виявлення негативних або

позитивних відгуків, аналіз ключових слів дозволяє підприємствам адаптувати стратегію розвитку;

- автоматична обробка новинних статей, дозволяє автоматично категоризувати, резюмувати та аналізувати новини. Це допомагає користувачам завжди отримувати актуальну інформацію з різних джерел.

Також для даного дослідження важлива досить інноваційна галузь в області веброзробки – голосове заповнення форм. Ця технологія відкриває нові можливості для користувачів та власників вебсайтів, дозволяючи взаємодіяти з вебформами, використовуючи свій голос. Замість традиційного введення тексту, користувач може диктувати вміст полів форми, і система автоматично розпізнає та обробить його. Це надає користувачам можливість взаємодії без необхідності використання клавіатури. Основною перевагою такого підходу є його зручність для користувачів. Також голосове заповнення форм сприяє підвищенню швидкості та ефективності введення інформації. Замість того, щоб витратити час на введення тексту, користувач може швидко та легко надати необхідну інформацію.

Автоматичне заповнення форм голосом поділяється на два завдання – розпізнавання мови, а це і є задача аналізу неструктурованого тексту, яку було розглянуто раніше в цьому підрозділі та визначення поля, для якого було надиктовано текст і автоматичне його заповнення.

1.2 Постановка завдання дослідження

Задачею цього дослідження є вивчення, тестування та порівняння методів аналізу неструктурованих текстів, які можна використати для автоматичного заповнення форм на вебсайтах. Таким чином, в рамках дослідження необхідно виконати наступні етапи:

- провести аналіз методів аналізу неструктурованих текстів;
- визначити, які з методів підходять для заповнення вебформ голосом, здійснити їх порівняння;

- обрати методи, які будуть тестуватись в застосунку для автоматичного заповнення вебформ;
- розробити такий застосунок та провести експеримент, метою якого є тестування методів та способу їх впровадження;
- проаналізувати результати дослідження та надати рекомендації з використання методів аналізу неструктурованих текстів для голосового заповнення вебформ.

Результатом дослідження мають бути висновки щодо переваг застосування різних методів аналізу неструктурованих текстів для заповнення вебформ в режимі реального часу.

2 ДОСЛІДЖЕННЯ МЕТОДІВ АНАЛІЗУ НЕСТРУКТУРОВАНИХ ТЕКСТІВ

Першим етапом дослідження є відбір інструментів, за допомогою яких можливо надиктувати і розпізнати неструктурований текст. Знайдені інструменти наведено в таблиці 1.

Таблиця 1 – Інструменти методів аналізу неструктурованих текстів

Інструмент	Стислий опис
Speechnotes	Це вебінструмент для перетворення мовлення в текст, який дає змогу транскрибувати аудіо- та відеозаписи, надиктовувати нотатки, заповнювати форми в браузері Chrome за допомогою розширення. Завдяки таким функціям, як голосові команди для пунктуації та форматування, автоматичне використання великих літер і прості параметри імпорту/експорту, Speechnotes забезпечує ефективне та зручне диктування та транскрипцію [5].
Speech Recognition Anywhere	Розширення браузера Chrome, що надає повний голосовий контроль над браузером, зокрема перехід по сторінкам, фокусування на полях та надиктовку в них тексту [6].
Alan AI	Інструмент, що дозволяє вбудувати розпізнавання голосу в форми вашого сайту. Інтегрується в будь-який фронтенд фреймворк та надає UX статистику [7].
Mozilla Speech Recognition API	Надає API для розпізнавання тексту в реальному часі в браузерах Mozilla [8].
Whisper	Система автоматичного розпізнавання мовлення, заявляє покращену стійкість до акцентів, фонового шуму та технічної мови [9].
Dragon Anywhere	Мобільний диктофон професійного рівня Dragon Anywhere дозволяє легко створювати документи будь-якої довжини, редагувати, формувати та ділитися ними безпосередньо з вашого мобільного пристрою під час відвідування клієнтів, місця роботи чи місцевої кав'ярні [10].
Google Voice Typing	Ви можете вводити та редагувати голосом у Google Документах або Google Презентаціях [11].
Zapier	Це інструмент, який допомагає підключати додатки та автоматизувати робочі процеси без будь-якого складного коду. Для розпізнавання мовлення можна використати інтеграцію з Temi [12].

Кінець таблиці 1

Siri	Голосовий помічник компанії Apple, зокрема дозволяє надиктовувати текст [13].
Google Assistant	Голосовий помічник компанії Google, зокрема дозволяє надиктовувати текст [14].
Alexa	Голосовий помічник компанії Amazon, зокрема дозволяє надиктовувати текст [15].
Cortana	Голосовий помічник компанії Microsoft, зокрема дозволяє надиктовувати текст [16].

З отриманого списку інструментів одразу виключимо наступні, адже вони дозволяють надиктовувати текст, але не забезпечують автоматичне заповнення вебформи на основі цього тексту:

- Dragon Anywhere;
- Google Voice Typing;
- Zapier;
- Siri;
- Google Assistant;
- Alexa;
- Cortana.

Тепер проведемо аналіз та визначено переваги та недоліки інструментів, що залишилися.

Speechnotes та Speech Recognition Anywhere зорієнтовані на транскрипцію та голосове керування браузером.

Alan AI є інструментом для вбудовування розпізнавання голосу в вебсайти.

Mozilla Speech Recognition API надає API для розпізнавання тексту в браузерах Mozilla.

Whisper (Hugging Face): Моделі глибокого навчання для розпізнавання мовлення.

Виділено наступні критерії оцінки:

- а) вартість:

- 1) вартість інструменту: Оцінка фінансових витрат на придбання та підтримку;
 - 2) додаткові витрати: Врахування можливих додаткових витрат, таких як навчання персоналу чи інтеграція з іншими системами;
- б) продуктивність:
- 1) швидкодія: Оцінка швидкодії інструменту при обробці та розпізнаванні мовлення;
 - 2) масштабованість: Аналіз можливостей розширення та масштабування при збільшенні обсягу роботи;
- в) сумісність та Інтеграція:
- 1) інтеграція з іншими системами: Оцінка можливостей інтеграції з іншими інструментами чи платформами;
 - 2) сумісність з пристроями: Визначення того, як добре інструмент працює з різними пристроями та операційними системами;
- г) технічна складність:
- 1) простота використання: Оцінка зручності інтерфейсу та процесу використання;
 - 2) вимоги до обладнання: Аналіз того, яке обладнання потрібно для нормального функціонування;
- д) системні обмеження:
- 1) географічні обмеження: Визначення можливих обмежень за географічним принципом;
 - 2) мовні обмеження: Оцінка, наскільки добре інструмент пристосований до різних мов та акцентів.

На основі виділених критеріїв було сформовано шкали оцінювання та проведено оцінку (див. табл. 2 та 3).

Таблиця 2 – Аналіз шкал

Критерій	Шкала	Приклад	MIN або MAX
Вартість	Абсолютна шкала	Платний; Безкоштовно	MIN
Продуктивність	Порядкова шкала	Миттєва; Затримки на повільних з'єднаннях	MAX
Сумісність та Інтеграція	Порядкова шкала	Широкі можливості інтеграції; Обмежена	MAX
Технічна складність	Порядкова шкала	Інтуїтивний інтерфейс; Середній рівень складності	MIN
Системні обмеження	Порядкова шкала	Без обмежень Географічні обмеження	MIN

Таблиця 3 – Оцінка критеріїв інструментів

Критерії / Інструмент	Speechnotes	Speech Recognition Anywhere	Alan AI	Mozilla Speech Recognition API	Whisper
Вартість	Безкоштовно	Платний	Платний	Безкоштовно	Залежить від обсягу тексту
Продуктивність	Миттєва	Миттєва	Миттєва	Затримки на повільних з'єднаннях	Можливі лаги на великих обсягах
Сумісність та Інтеграція	Обмежена	Широкі можливості інтеграції	Широкі можливості інтеграції	Обмежена	Широкі можливості інтеграції
Технічна складність	Інтуїтивний інтерфейс	Інтуїтивний інтерфейс	Середній рівень складності	Інтуїтивний інтерфейс	Вимагає навичок в роботі з моделями
Системні обмеження	Без обмежень	Без обмежень	Географічні та мовні обмеження	Географічні обмеження	Географічні обмеження

Застосувавши експертну оцінку [17], перетворимо наші шкали у числові показники.

Для цього використаємо лінійну адитивну згортку з ваговими коефіцієнтами [18]. Помножимо кожен критерій на його ваговий коефіцієнт, а потім усі зважені

критерії підсумуємо і утворимо зважену цільову функцію, значення якої інтерпретується як «коефіцієнт якості» отриманого рішення.

Проаналізувавши особливості інструментів, виходячи з їх опису і практик застосування, отримуємо таблицю 4.

Таблиця 4 – Призначення коефіцієнтів за лінійною аддитивною згорткою

Критерії / Інструмент	Speechnotes	Speech Recognition Anywhere	Alan AI	Mozilla Speech Recognition API	Whisper
Вартість	0	1	1	0	0,5
Продуктивність	1	1	1	0,5	0,25
Сумісність та Інтеграція	1	0,5	1	1	0,5
Технічна складність	1	1	0,25	1	0,35
Системні обмеження	1	1	0,25	0,50	0,50

Вагові коефіцієнти для кожного критерію визначаються на основі їхньої важливості у виборі між альтернативами. Ось пояснення для обраної конфігурації вагових коефіцієнтів.

Вартість (30%): Вартість завжди є важливим фактором при виборі інструментів. Чим менше вартість, тим ефективніше можна розпоряджатися ресурсами.

Продуктивність (20%): Продуктивність визначає, наскільки швидко та ефективно інструмент виконує свої завдання. Важливо мати ефективні інструменти для забезпечення продуктивності процесів.

Сумісність та Інтеграція (10%): Якщо інструмент добре інтегрується з іншими системами та має високий рівень сумісності, це спрощує процес роботи та забезпечує зручність використання.

Технічна складність (20%): Чим менша технічна складність реалізації, тим легше використовувати і впроваджувати інструмент. Простота використання

сприяє точному плануванню роботи, зменшенню ризиків на всіх етапах використання.

Системні обмеження (20%): Важливо враховувати системні обмеження, оскільки вони можуть впливати на ефективність та можливості інструментів в конкретному середовищі.

Зазначимо векторний опис задачі (формули 1 та 2):

$$w(x) = \sum_{i=1}^n c_i f_i \quad (1)$$

де $w(x)$ – векторна функція, яка представляє результат згортки,

c_i – вагові коефіцієнти, що в сумі дають 1,

f_i – значення критеріїв з таблиці 4.

$$w(x) = 0.3f_1 + 0.2f_2 + 0.1f_3 + 0.2f_4 + 0.2f_5 \quad (2)$$

Розраховані показники наведено в таблиці 5.

Таблиця 5 – Визначення оцінок по теорії користності

Критерії / Інструмент	Speechnotes	Speech Recognition Anywhere	Alan AI	Mozilla Speech Recognition API	Whisper	Ваговий коефіцієнт
Вартість	0	0,3	0,3	0	0,15	0.3
Продуктивність	0,2	0,2	0,2	0,1	0,05	0.2
Сумісність та Інтеграція	0,1	0,05	0,1	0,1	0,05	0.1
Технічна складність	0,2	0,2	0,05	0,2	0,07	0.2
Системні обмеження	0,2	0,2	0,05	0,1	0,1	0.2
Результат	0,7	0,95	0,7	0,5	0,42	

Найкращі показники було отримано у Speech Recognition Anywhere, що дорівнює 0,95.

Далі проведемо аналіз отриманої таблиці за Парето. Для цього порівняємо кожен інструмент між собою. За Парето, альтернатива X домінує альтернативу Y , якщо для всіх критеріїв F_i , значення критерію для X не менше, ніж для Y . Умова нерівності говорить про те, що хоча б один критерій повинен бути суворо кращий або однаковий для X , ніж для Y . Тобто, альтернатива X домінує альтернативу Y , якщо вона принаймні така ж за кожним критерієм і краща за одним критерієм [19]. Результати аналізу зведено до таблиці 6.

Таблиця 6 – Результат аналізу Парето

Критерії / Інструмент	Speechnotes	Speech Recognition Anywhere	Alan AI	Mozilla Speech Recognition API	Whisper	Ваговий коефіцієнт
Вартість	0	0,3	0,3	0	0,15	0.3
Продуктивність	0,2	0,2	0,2	0,1	0,05	0.2
Сумісність та Інтеграція	0,1	0,05	0,1	0,1	0,05	0.1
Технічна складність	0,2	0,2	0,05	0,2	0,07	0.2
Системні обмеження	0,2	0,2	0,05	0,1	0,1	0.2
Результат	0,7	0,95	0,7	0,5	0,42	

З проведеного аналізу бачимо, що Speech Recognition Anywhere та Speechnotes перемагають Mozilla Speech Recognition API. При цьому, Speech Recognition Anywhere краще ніж Whisper.

Таким чином, для подальшого тестування впровадження та інтеграції в застосунок, що буде розроблено далі, було обрано інструменти Speech Recognition Anywhere, бо він показав найкращий результат в проведеному аналізі. Також, з

метою додаткової верифікації аналізу, інструменти Mozilla Speech Recognition API та Whisper теж обрано для подальшого тестування.

3 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ

3.1 Планування експерименту

Проведення експериментальної частини дослідження передбачено наступним чином. Обрано інструменти розпізнавання мовлення:

- Speech Recognition Anywhere;
- Mozilla Speech Recognition API
- OpenAI Whisper.

Для кожного планується побудувати вебзастосунок, в якому користувач зможе заповнити голосом певну форму. Провести тести, в ході яких визначити наступні критерії кожного інструменту:

- підтримувані мови;
- підтримувані браузері;
- точність розпізнавання мовлення;
- найкращі сценарії застосування;
- найкращі способи реалізації.

Також необхідно провести порівняння швидкості наступних режимах заповнення форми та оцінити приріст швидкості:

- ручний;
- напівавтоматичний;
- автоматичний.

Під напівавтоматичним режимом мається на увазі, що на формі вже є розпізнавання мовлення, але користувач має сам обирати поле (наприклад, за допомогою мишки або клавіатури), і потім надиктовувати текст. При автоматичному режимі користувачу достатньо промовити, в яке поле мають заноситись дані. Для досягнення такого функціоналу передбачається додатково використати ChatGPT API.

В результаті експерименту планується отримати таблиці з порівняльними характеристиками обраних інструментів за наведеними вище критеріями.

3.2 Speech Recognition Anywhere

Speech Recognition Anywhere є розширенням до браузерів Chrome та Edge, яке дозволяє повністю перевести керування браузером на голосове [3]. В ході експерименту розширення було встановлено в Chrome. Воно працює наступним чином: користувач запускає розширення, відкривається вкладка браузеру, яка слухає мовлення і на якій можна обрати мову, визначити персональні команди, ознайомитися з інструкцією й т.і. (див. рис. 1), і далі можна надиктовувати текст і керувати браузером.

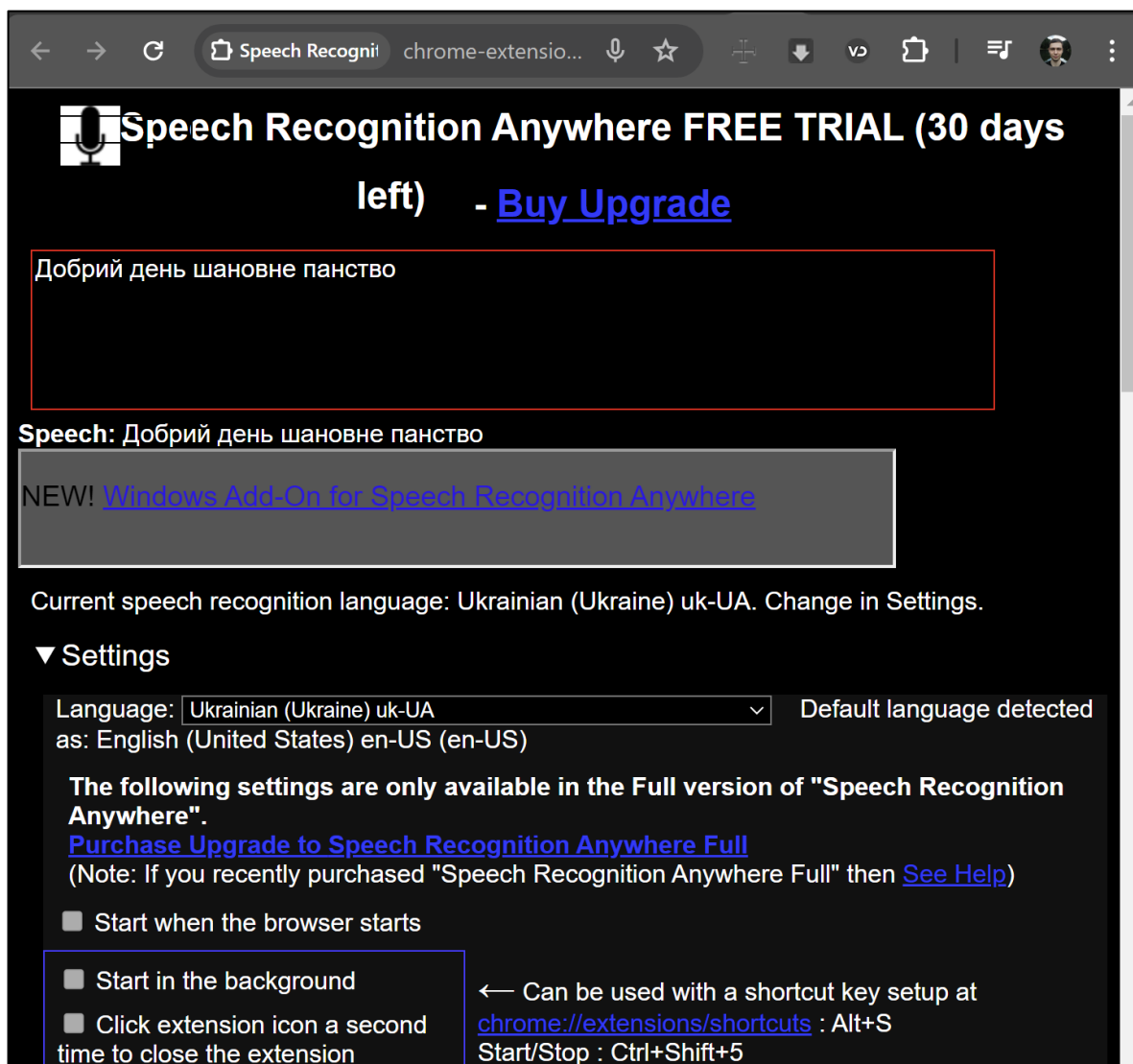


Рисунок 1 – Вкладка розпізнавання голосових команд

Доступна велика кількість мов, зокрема англійська та українська. Розширення розуміє наступні команди:

- відкриття/закриття вкладок;
- фокусування на полі за його назвою;
- очищення обраного поля:
- натискання на текст (наприклад, посилання у видачі Google);
- підтвердження (надсилання форми);
- переходи вперед та назад між сторінками.

В інструкції зазначено велику кількість команд, але всі вони англійські (див. рис. 2).

Edit
Select Highlight (to) (the) (last previous next all text field box none nothing start end) (n) (character letter word sentence paragraph line document paper)(s)
Deselect Unselect Unhighlight (all)
Copy Cut Paste (to from) (clipboard)
Clear (the) (field text textarea input box all)
Find (word or phrase)
Capitalize All (Word Or Phrase). Ex: Capitalize All Hello World
Capitalize (Word or phrase). Ex: Capitalize Hello world
Uppercase (word or phrase). Ex: Uppercase HELLO WORLD
Lowercase (word or phrase). Ex: Lowercase hello world
Scrolling
Scroll (page body document menu sidebar box) (all the way) (to the) up down left right write start top end bottom
Scroll (page body document menu box) up down left right (all the way) (to) (the) start top end bottom
Forms, Buttons and Links
(Press) (the) tab (key) (n time(s))
(Press) (the) shift tab (key) (n time(s))
Tab (previous left up app backward next right down forward) (n time(s))

Рисунок 2 – Перелік команд розширення

Спочатку було проведено тестування англійською. Протягом тестування розширення працювало задовільно, команди відпрацьовували коректно, але було виявлено певні труднощі з розпізнаванням мовлення. Це може бути пов'язано з

тим, що тестування відбувалось не носієм англійської, але, як побачимо пізніше, це не становило проблеми для Mozilla Speech Recognition API та Whisper. Як приклад, лише сконцентрувавшись і промовляючи слова максимально чітко й повільно, вдалося надиктувати фразу “I can fill out any form now” (див. табл. 7).

Під час тестування українською мовою, було виявлено проблеми. Команди для керування англійські, і навіть якщо користувач знає англійську, інструмент розпізнає лише деякі з них в україномовному режимі. Так, було втрачено функціонал фокусування на полі форми голосом, доводилося робити це вручну. Однак, розпізнавання мовлення працювало добре, навіть з оновленим правописом (слова етер, катедра і т.д.).

Таблиця 7 – Приклади розпізнавання фраз Speech Recognition Anywhere

Мова	Очікуваний результат	Отриманий результат
Англійська	I can fill out any form now	I can feel out any form now
		I can throughout any form now
		I can fill out any farmer
		I can fill out any form now
	Baker Street, 226	Baker Street, 226
	663897	663897
	Clear box (команда очистити поле)	Player box
		Clear unbox
		The air box
		Clear box
Українська	Clear (команда очистити поле)	Клея
		Не розпізнано
	Enter (команда підтвердження)	Розпізнано, натиснуто Enter
	Space (команда введення пробілу)	Розпізнано, додано пробіл

Кінець таблиці 7

	Submit (команда підтвердження)	Не розпізнано
	Я маю чудовий вебдодаток	Я маю чудовий веб-додаток
	Вулиця Плеханівська 63/8	Вулиця Плеханівська 63/8
	Етер	Етер
	Катедра	Катедра

Далі було проведено порівняння швидкості заповнення форми. Спочатку заповнення відбувало вручну – набором тексту на клавіатурі. Пройшовши відповідний тест [20], було визначено, що моя швидкість набоу в середньому дорівнює 30 слів на хвилину, що є трохи вищим за середні показники, згідно дослідженню “Computerised assessment of handwriting and typing speed” [21].

В наступну чергу форма заповнювалась голосом, але з диктуванням кожного поля окремо, фокусуючись на полях за допомогою мишки.

Далі було проведено заповнення голосом з фокусуванням на полях командою “Next”, тобто “наступне поле”. Результати наведено в таблиці 8.

Таблиця 8 – Середній час заповнення тестової форми з Speech Recognition Anywhere

Час заповнення вручну, сек.	Час заповнення голосом, сек.	
	Ручне фокусування на полях	Автоматичне фокусування на полях
43,516	24,233	26,572

Бачимо, що голосове заповнення вебформи має велику перевагу з заощадження часу. При цьому, використання автоматичного фокусування на полях зайняло більше часу, адже після кожного поля необхідно промовити команду і дочекатись, поки модель розпізнає та виконає її. Хоча такий спосіб може бути зручнішим, фокусування мишкою виявилось швидшим.

3.3 Mozilla Speech Recognition API

Mozilla Speech Recognition API – це частина проекту Mozilla DeepSpeech з відкритим кодом. Проект надає набір АПІ для розпізнавання мовлення, призначений для браузерів Mozilla, тобто Chrome, Firefox, Safari. Це інтерфейс Web Speech API, який призначений для надання вебзастосункам можливості оброблювати голосові дані [22]. Проект побудований на основі сучасних технологій глибокого навчання, зокрема рекурентних нейронних мережах (RNN) з архітектурою LSTM. Спільнота активно працює над додаванням підтримки різних мов та діалектів. Вже доступна підтримка багатьох мов, зокрема української та англійської. Також Mozilla надає попередньо натреновані моделі, але користувачі можуть тренувати власні моделі, використовуючи свої дані, що забезпечує можливість адаптації системи до специфічних потреб. DeepSpeech може бути додатково конвертовано в TFLite для використання на мобільних пристроях, що робить можливим впровадження розпізнавання мовлення на різних платформах.

Для проведення експерименту було створено вебзастосунок, який містить форму з тестовим набором полів. Її можна заповнити, набираючи текст вручну, але також є можливість натиснути кнопку “Слухати голос”, після чого підключається розпізнавання мовлення, що дозволяє надиктувати вміст для полів (див. рис. 3).

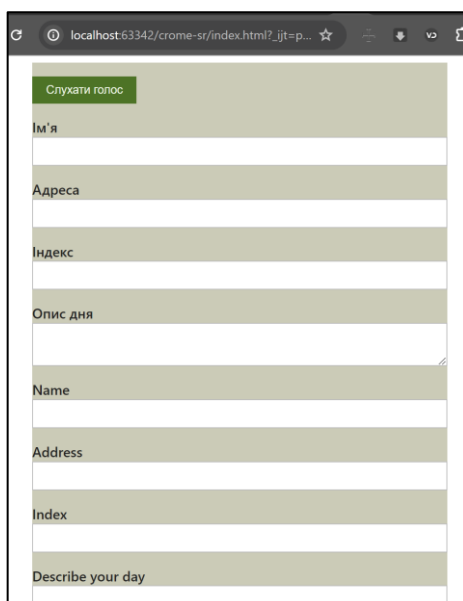


Рисунок 3 – Розроблений застосунок на базі Mozilla Speech Recognition API

Поля продубльовані українською та англійською мовами, оскільки саме ці мови використовуватимуться для перевірки точності розпізнавання мовлення.

Під час проведення тестування модель показала себе набагато краще за Speech Recognition Anywhere. Вона чудово розпізнає англійську мову неносія, й аналогічно добре розпізнає українську. Помилки виявлено у слові “вебдодаток”, але лише коли воно в реченні. В інших випадках модель правильно розпізнає слова нового правопису української мови.

Також спостерігається своєрідна робота з наборами цифр – модель або намагається подати їх у форматі телефону (з дефісами), або ставить між цифрами пробіли (див. табл. 9).

Таблиця 9 – Приклади розпізнавання фраз Mozilla Speech Recognition API

Мова	Очікуваний результат	Отриманий результат
Англійська	I can fill out any form now	I can fill out any form now
	Baker Street, 226	Baker Street, 226
	663897	6 6 3 8 9 7
	Clear box	Clear box
Українська	Я маю чудовий вебдодаток	Я маю чудовий веб-додаток
	Вебдодаток	Вебдодаток
	Вулиця Плеханівська 63/8	Вулиця Плеханівська 63/8
	Етер	Етер
	Катедра	Катедра

В напівавтоматичному режимі застосунок працює наступним чином. Користувач натискає кнопку “Слухати голос”, потім сам фокусується на потрібному полі (за допомогою мишки чи клавіатури) і надиктовує вміст для цього поля.

На відміну від цього, автоматичний режим усуває крок з ручним фокусуванням на полі. Модель Speech Recognition Anywhere, яку було розглянуто в попередньому підрозділі, побудована як повноцінне розширення браузера, і має власну систему для автоматизації заповнення вебформ, а саме – команди для переходу фокусу між полями. Наступні моделі, які розшлянуті в даному дослідженні, не мають таких засобів автоматизації, адже надають лише інструмент для розпізнавання мовлення. Тому надалі було додатково покращено застосунок і реалізовано повну автоматизацію заповнення форм.

Для досягнення такого результату було використано ChatGPT API. Сформовано запит, в який передаються заголовки полів форми та текст, надиктований користувачем, після розпізнання його, в даному випадку, за допомогою Mozilla Speech Recognition API (див. рис. 4).

Є пронумеровані заголовки полів

0. Адреса
1. Введіть ваше ім'я
2. Опис дня

Ось текст, надиктований користувачем, який хоче заповнити ці поля.

Ім'я Андрій Борисович Адреса вулиця Плеханівська, 63
Мій день пройшов чудово

Визначи за назвами полів, який текст призначений для якого поля і сформує json з масивом формату номер_поля: текст. У відповіді надай лише json, без коментарів. Для деяких полів може не бути тексту, не включай такі поля в json.

Рисунок 4 – Приклад запиту для ChatGPT API

Модель ChatGPT аналізує наданий текст і визначає за змістом, яка інформація призначена для якого поля (див. рис. 5). Таким чином, користувач може заповнити всі поля, не фокусуючись на жодному. Замість цього достатньо буде промовити заголовок поля. Наприклад, сказати “Адреса: Харків, вулиця Плеханівська, 63”.

```
json
{
  "0": "вулиця Плеханівська, 63",
  "1": "Андрій Борисович",
  "2": "Мій день пройшов чудово"
}
```

Рисунок 5 – Згенеровані ChatGPT дані для заповнення полів форми

У відповідь ChatGPT API надає рядок у форматі JSON, який можна легко прочитати і програмно заповнити поля. Перевагою такої автоматизації через ChatGPT API є те, що це високо навчена текстова модель, яка дуже добре розуміє зміст тексту. Тому користувач може посилатися на поле майже будь-яким способом і ChatGPT зрозуміє, про яке саме поле йде мова. Немає потреби розробляти команди для фокусування на полі й обмежувати користувача ключовими словами. Натомість, користувач може заповнювати форму максимально природнім мовленням. Можна промовляти заголовок поля, або його порядковий номер, або навіть лише вміст.

Після проведення тестування на точність розпізнавання мовлення, було визначено середні показники часу, необхідного для заповнення тестової вебформи (див. табл. 10).

Показник часу для заповнення вручну було взято з попередньої серії тестів для Speech Recognition Anywhere. Ручне фокусування на полях виконувалось за допомогою мишки.

Таблиця 10 – Середній час заповнення тестової форми з Mozilla Speech Recognition API

Час заповнення вручну, сек.	Час заповнення голосом, сек.	
	Ручне фокусування на полях	Автоматичне фокусування на полях
43,516	22,847	21,142

Маємо кращий результат для автоматичного фокусування, порівняно зі Speech Recognition Anywhere. Цього разу автоматизація виявилась швидшою, ніж ручне фокусування. Це пов'язано з тим, що в реалізації з ChatGPT не потрібно чекати, поки модель розпізнає команду переходу для кожного поля форми. До того ж, чим більше міститься полів у вебформі, тим більший буде виграш часу порівняно з напівавтоматичним режимом, особливо враховуючи те, що користувач може заповнити всю форму за один запит до API ChatGPT, якщо диктуватиме без великих пауз.

3.4 OpenAI Whisper API

Whisper – це відкритий проєкт моделі розпізнавання мовлення, створена компанією OpenAI. До неї надається набір запитів API на платформі для розробників, які й використовувались протягом цього експерименту. Модель підтримує багато мов, зокрема українську та англійську, а також є найбільш гнучкою, бо не залежить від оточення, в якому використовується.

Для тестування було взято попередній застосунок з тестовою вебформою і замінено модель для розпізнавання мовлення на Whisper large-v2. Аналогічно попереднім випадкам, тестування проводилося українською та англійською мовами. Як і Mozilla Speech Recognition API, дана модель гарно виконує своє завдання, відсутні складнощі з розпізнаванням обох мов. Залишаються проблеми з цифрами поштового індексу – Whisper або ставить пробіли між цифрами, або розбиває на числа і розділяє їх комами, залежно від стилю вимови користувача. Також модель правильно розпізнає слова нового правопису української мови, і це

перша з протестованих моделей, яка правильно оброблює слово “вебдодаток” (див. табл. 11).

Таблиця 11 – Приклади розпізнавання фраз моделі Whisper

Мова	Очікуваний результат	Отриманий результат
Англійська	I can fill out any form now	I can fill out any form now
	Baker Street, 226	Baker Street, 226
	663897	66, 78, 97
		6 6 3 8 9 7
	Clear box	Clear box
Українська	Я маю чудовий вебдодаток	Я маю чудовий вебдодаток
	Вебдодаток	Вебдодаток
	Вулиця Плеханівська 63/8	Вулиця Плеханівська 63/8
	663897	6 6 3 8 9 7
	Етер	Етер
	Катедра	Катедра
	Ефір	Ефір
	Кафедра	Кафедра

Після проведення тестування на точність розпізнавання мовлення, було визначено середні показники часу, необхідного для заповнення тестової вебформи (див. табл. 12).

Таблиця 12 – Середній час заповнення тестової форми з OpenAI Whisper

Час заповнення вручну, сек.	Час заповнення голосом, сек.	
	Ручне фокусування на полях	Автоматичне фокусування на полях
43,516	25,756	23,337

Показник часу для заповнення вручну було взято з попередньої серії тестів. Ручне фокусування на полях виконувалось за допомогою мишки.

Маємо схожі результати з моделлю Mozilla Speech Recognition API, де автоматизація за допомогою ChatGPT є швидшою, ніж ручне фокусування і, передбачувано, голосове заповнення в будь-якому випадку швидше, ніж набирання тексту.

4 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ

Проведення дослідження дозволило вивчити та порівняти інструменти розпізнавання мовлення, що підходять для заповнення вебформ за допомогою голосу. В даному розділі підбито підсумок для кожного з розглянутих інструментів.

На початку дослідження було відібрано багато способів, які забезпечують розпізнавання мовлення, зокрема голосові помічники, менеджери нотаток, нейронні моделі. Однак не всі вони підходять для реалізації голосового заповнення вебформ. Так, було відсіяно наступні інструменти, адже вони дозволяють надиктувати текст, але не у вебформу (наприклад, лише у власне поле пошуку або нотатку), або принаймні не забезпечують автоматичне заповнення вебформи на основі отриманого тексту:

- Dragon Anywhere;
- Google Voice Typing;
- Zapier;
- Siri;
- Google Assistant;
- Alexa;
- Cortana.

Далі було впроваджено детальне порівняння залишившихся моделей:

- Speechnotes;
- Speech Recognition Anywhere;
- Alan AI;
- Mozilla Speech Recognition API;
- OpenAI Whisper.

В ході аналізу було відібрано три інструменти для практичного експерименту: Speech Recognition Anywhere, Mozilla Speech Recognition API та OpenAI Whisper.

Speech Recognition Anywhere є розширенням для браузера, що одразу створює низку переваг та недоліків. Недоліком є підтримка лише браузерів

Chrome та Edge, що є руйнівним з точки зору платформонезалежності. Оскільки це готове рішення, немає можливості його змінити під потреби конкретних користувачів та/або вебсайтів. Також до очевидних мінусів Speech Recognition Anywhere відноситься те, що його не можна закріпити за певним сайтом. Це інструмент, який користувач має встановити сам. Але, з іншого боку, таким чином голосове керування з'явиться на всіх формах, з якими може стикнутися користувач в межах підтримуваних браузерів. Це робить даний інструмент ідеальним готовим рішенням для персонального використання, максимально простим в освоєнні та налаштуванні. Він вимагає мінімальних зусиль для інтеграції і підходить для користувачів без технічних знань завдяки простому інтерфейсу. Однак, його не можна назвати універсальним, через підтримку лише двох браузерів, а також посередній рівень роботи з українською мовою. Однією з переваг Speech Recognition Anywhere є наявність команд для керування браузером і, зокрема, переходами між полями вебформи. Але ці команди розпізнаються лише англійською. Деякі розпізнаються і коли, наприклад, користувач промовляє англійську команду при обраній українській мові, але не всі. Також модель не оновлено відповідно до нового українського правопису. З розпізнаванням англійської мови теж є проблеми – модель плутає співзвучні слова, принаймні якщо їх промовляє не носій. Помітно, що пріоритет розширення зміщений на англійськомовних користувачів, хоча і є підтримка багатьох інших мов. Таким чином, найкращий досвід взаємодії з інструментом отримують саме такі користувачі, й особливо ті, що володіють англійською мовою на рівні носія. Тож, якщо користувач є прихильником браузерів Chrome або Edge, добре знає англійську і має необхідність отримати заповнення вебформ голосом найпростішим способом, Speech Recognition Anywhere – найкраще рішення. В якості бонусу, функціонал голосового керування не обмежується вебформами, а також включає в себе всі елементи браузера.

Набір API Mozilla Speech Recognition вимагає більше зусиль для налаштування та інтеграції, оскільки потребує створення вебзастосунку. Таке рішення є хорошим варіантом для власників вебсайту, які хочуть додати голосове

керування для форм на своєму сайті. Для такого завдання й слід обирати гнучке рішення, готове до адаптації під специфічні потреби. Модель набагато краще розпізнає мовлення, ніж Speech Recognition Anywhere, це стосується як української, так й англійської мов. Але трапляються помилки з оновленим правописом української мови. Подібно до інших моделей, що розглядаються в даному дослідженні, є підтримка широкого спектру інших мов.

Хоча налаштування інтеграції вимагає певних технічних знань, в той же час воно максимально просте. На жаль, ця простота обумовлена обмеженнями щодо платформ, на яких доступне використання, а саме – виключно браузері Mozilla, тобто Chrome, Edge та Safari.

На відміну від Speech Recognition Anywhere, у даної моделі немає готового продукту для використання, тож окрім створення застосунку-оболонки, який дозволяє заповнювати вебформи голосом, додатково необхідно робити реалізацію повної автоматизації – коли застосунок бере на себе задачу розпізнавання, для якого поля призначений надиктований текст. Це ще більше спростить користувачам заповнення вебформ, а також прискорить цей процес, особливо для великих форм. Одним з найзручніших способів досягнення такого функціоналу є передача надиктованого тексту після його розпізнання в ChatGPT з відповідним запитом визначити, який текст призначений для якого поля. Перевагами такого способу є мовнезалежність і висока спроможність ChatGPT до аналізу тексту. Це дозволяє не обмежувати користувачів ключовими словами чи командами, роблячи заповнення вебформи максимально природнім процесом. Користувачу достатньо промовляти заголовок поля, а в деяких випадках навіть це не потрібно. Наприклад, ChatGPT легко визначить, з яким полем треба співставити надиктовані адресу та ім'я. Отож, для реалізації такої задачі було обрано саме API ChatGPT.

Загалом, інструмент займає збалансовану позицію між простотою Speech Recognition Anywhere та комплексною складністю OpenAI Whisper. Якщо для поставленої задачі не є проблемою можливість застосування лише у браузерах Mozilla (які є одними з найпопулярніших), вебзастосунок на основі Mozilla Speech

Recognition API є чудовим вибором, завдяки невибагливості до процесу інтеграції та підтримки.

OpenAI Whisper є найбільш навченою та гнучкою моделлю в даному дослідженні. Вона, подібно до Mozilla Speech Recognition, надає API для проведення запитів. Аналогічно, навколо неї треба будувати вебзастосунок і робити автоматизацію за допомогою ChatGPT, якщо цього вимагає поставлене завдання.

Перевагами Whisper є точність розпізнавання та відсутність прив'язки до платформи. Модель відмінно розпізнає англійську мову, їй не заважає неідеальна вимова й навіть помірний фоновий шум. Рівноцінна ситуація й з українською мовою. Це єдина з протестованих моделей, яка безпомилково визначає слова нового правопису. В той же час, перевазі високої точності OpenAI Whisper протиставляється недолік у вигляді трохи більшого часу, потрібного на розпізнавання мовлення, порівняно з іншими моделями, розглянутими в дослідженні.

За результатами тестування, OpenAI Whisper виглядає найбільш підходящим інструментом для створення універсального застосунку голосового заповнення вебформ завдяки високій точності розпізнавання, відсутності обмежень щодо платформ, на яких такий застосунок зможе працювати, широкій підтримці мов та діалектів. За це, очікувано, модель платить підвищеною складністю реалізації. Але така універсальність дозволяє закривати набагато більший перелік задач та вимог. Наприклад, висока точність навіть за складних умов і підтримка багатьох мов дозволяють будувати багатомовні міжнародні застосунки.

На основі проведеного вивчення моделей та застосунків, які побудовані або можуть бути побудованими на їх основі, можемо робити їх порівняння між собою (див. табл. 13).

Таблиця 13 – Порівняння моделей відносно одна одної

Категорія	Speech Recognition Anywhere	Mozilla Speech Recognition API	OpenAI Whisper
Підтримувані мови	+	+	+
Простота інтеграції	+++	++	+
Швидкість роботи	++	+++	+
Точність розпізнавання української мови	+	++	+++
Точність розпізнавання англійської мови	+	++	+++
Платформонезалежність	+	++	+++
Разом	9	12	13

Для порівняння було обрано просту шкалу, яка оцінює показники кожної моделі відносно інших. Оскільки моделей три, найкращою оцінкою є три очки, найгіршою – одне. Бачимо, що за загальними показниками виграє OpenAI Whisper. Її платформонезалежність, точність та загальна універсальність переважають недоліки, пов'язані зі складністю реалізації. Все ж, в конкретному випадку може бути більш доцільно використання інших моделей.

Окремим етапом дослідження було створення такої автоматизації, щоб користувач міг зазначити голосом, яку поле він заповнює. Це було реалізовано за допомогою API ChatGPT. Результати замірів часу заповнення тестової вебформи трьома способами зведено до таблиці 14.

Таблиця 14 – Середній час заповнення тестової форми

Інструмент	Час заповнення вручну, сек.	Час заповнення голосом, сек.	
		Ручне фокусування на полях	Автоматичне фокусування на полях
Speech Recognition Anywhere	43,516	24,233	26,572
Mozilla Speech Recognition API		22,847	21,142
OpenAI Whisper		25,756	23,337

Представлені показники часу є середніми значеннями десяти спроб заповнення вебформи однаковими даними. Далі було розраховано відсоток прискорення голосового заповнення вебформи з автоматизацією та без відносно заповнення ручним набором тексту, а також заповнення голосом з автоматизацією відносно заповнення без автоматизації (див. табл. 15).

Таблиця 15 – Відносне прискорення різних методів заповнення вебформи

	Голосове з ручним фокусуванням на полях відносно ручного набору тексту, %	Голосове з автоматичним фокусуванням на полях відносно ручного набору тексту, %	Голосове з автоматичним фокусуванням на полях відносно голосового з ручним фокусуванням на полях, %
Speech Recognition Anywhere	44,31	38,93	-9,65

Кінець таблиці 15

Mozilla Speech Recognition API	47,5	51,42	7,46
OpenAI Whisper	40,81	46,37	9,39
Середній рівень прискорення	44,2	45,57	8,43 (*)

Показники прискорення були обчислені за наступною формулою 3.

$$P = ((T_1 - T_2) / T_1) * 100 \quad (3)$$

де P – відсоток прискорення,

T₁ – початковий час,

T₂ – кінцевий час.

За результатами бачимо, що заповнення вебформи голосом забезпечує велику економію часу порівняно з набиранням тексту. У випадку додаткової автоматизації середній відсоток прискорення з розглянутими моделями складає 45,57%. Без автоматизації 44,2%. Прискорення автоматизації відносно звичайного заповнення голосом складає 8,43%. Цей показник був обчислений без урахування Speech Recognition Anywhere, оскільки через особливості цього інструменту і його власну реалізацію автоматизації, вона призводить до більших витрат часу.

ВИСНОВКИ

В ході проведення дослідження було проведено вивчення предметної області аналізу неструктурованих текстів в контексті автоматичного заповнення вебформ, поставлено задачу дослідження, яке включатиме в себе визначення методів аналізу, які підходять для заповнення вебформ голосом, здійснення їх порівняння, тестування їх впровадження в застосунку та розробку такого застосунку, проведено визначення та порівняння методів аналізу неструктурованих текстів та ступеня їх придатності для автоматизації заповнення вебформ. Було виділено низку методів та інструментів, і для їх порівняння було використано метод експертної оцінки, лінійну адитивну згортку з ваговими коефіцієнтами та аналіз за принципом Парето. Було отримано перелік інструментів та методів, які будуть тестуватися шляхом їх впровадження в застосунок, спеціально розроблений для цього, а саме – Speech Recognition Anywhere, Mozilla Speech Recognition API та OpenAI Whisper.

Було сплановано та проведено експеримент з обраними інструментами. Експеримент включав в себе детальне вивчення кожної моделі, тестування готових застосунків на їх основі та/або створення власного, який надавав можливість заповнювати вебформи голосом. Було визначено наступні критерії моделей:

- підтримувані мови;
- підтримувані браузері;
- точність розпізнавання мовлення;
- найкращі сценарії застосування;
- найкращі способи реалізації.

Також було реалізовано автоматизацію у вигляді визначення за допомогою ChatGPT API, для якого поля користувач надиктовує текст, що дозволило ще більше спростити процес заповнення вебформи та зменшити необхідний для цього час.

Було проведено заміри часу для трьох методів заповнення тестової вебформи:

- ручне введення тексту;
- голосове заповнення з ручним фокусуванням на полі перед початком диктування;
- голосове заповнення без необхідності фокусування на полях.

В результаті дослідження було отримано докладні характеристики досліджуваних інструментів, сценарії використання, в яких найбільш виправдане застосування кожного з них, а також таблиці з їх порівняльними характеристиками та розрахунками часу, необхідного для заповнення вебформи різними методами та відсотки прискорення кожного з них. На основі отриманих даних було зроблено висновок, що будь-який спосіб голосового заповнення вебформи є швидшим за набирання тексту, при цьому є можливість реалізувати автоматизацію і ще більше прискорити процес заповнення.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Работягов В. А. Метод ідентифікації людини на основі індивідуального мовного коду / Работягов В. А // Автореферат дисертації на здобуття наукового ступеня кандидата технічних наук – Харків: Навчально-науковий видавничо-поліграфічний центр ХНУРЕ – С. 17.
2. Text Mining: Discovering Patterns in Unstructured Data [Електронний ресурс]. – 2022. – Режим доступу до ресурсу: <https://medium.com/@bharani.blogs/text-mining-discovering-patterns-in-unstructured-data-a53d281c4dbd> (дата звернення: 15.12.2023).
3. Работягов В. А. Люліна К. П. Дослідження методів розпізнавання іменованих сутностей у неструктурованому тексті / К. П. Люліна, О. П. Турута // Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління : тез. доп. дванадцятої міжнародної науково-технічної конференції, 27–28 квітня 2022 р. – Т.
4. Турута О. П. Використання моделі dall-e для створення у програмній системі на основі штучного інтелекту для поштового клієнту GMAIL / О. П. Турута, Ж. В. Дейнеко, І. Є. Мічурін // Поліграфічні, мультимедійні та web-технології : тези доп. ІХ Міжнар. наук.-техн. конф., 14-18 травня 2024 р. – Т. 1. – Харків: ТОВ «Друкарня Мадрид», 2024. – С. 95-96.
5. Speechnotes [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://speechnotes.co> (дата звернення: 15.12.2023).
6. Speech Recognition Anywhere [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://chromewebstore.google.com/detail/speech-recognition-anywhere/kdnnmhpmcakdilnofmllgcigkibjonof?pli=1> (дата звернення: 15.12.2023).
7. Alan AI [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://www.actionableai.com> (дата звернення: 15.12.2023).
8. Mozilla Speech Recognition API [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://developer.mozilla.org/en-US/docs/Web/API/SpeechRecognition> (дата звернення: 15.12.2023).

9. Whisper AI [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://openai.com/research/whisper> (дата звернення: 15.12.2023).

10. Dragon Anywhere [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://www.nuance.com/dragon/dragon-anywhere.html> (дата звернення: 15.12.2023).

11. Google Voice Typing [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://support.google.com/docs/answer/4492226?hl=en> (дата звернення: 15.12.2023).

12. Zapier integrations with Temi [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://zapier.com/blog/updates/1691/temi-integrations> (дата звернення: 15.12.2023).

13. Apple Siri [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://www.apple.com/siri/> (дата звернення: 15.12.2023).

14. Google Assistant [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: https://assistant.google.com/intl/ru_ru/ (дата звернення: 15.12.2023).

15. Amazon Alexa [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://www.alexa.com/> (дата звернення: 15.12.2023).

16. Microsoft Cortana [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://support.microsoft.com/uk-ua/topic/cortana-%D1%82%D0%B0-%D0%BA%D0%BE%D0%BD%D1%84%D1%96%D0%B4%D0%B5%D0%BD%D1%86%D1%96%D0%B9%D0%BD%D1%96%D1%81%D1%82%D1%8C-47e5856e-3680-d930-22e1-71ec6cdde231> (дата звернення: 15.12.2023).

17. Метод експертних оцінок [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_%D0%B5%D0%BA%D1%81%D0%BF%D0%B5%D1%80%D1%82%D0%BD%D0%B8%D1%85_%D0%BE%D1%86%D1%96%D0%BD%D0%BE%D0%BA (дата звернення: 15.12.2023).

18. Методи розв'язання багатокритеріальних задач оптимізації [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <http://um.co.ua/3/3-2/3-22610.html> (дата звернення: 15.12.2023).

19. Pareto Analysis [Електронний ресурс]. – 2023. – Режим доступу до ресурсу: <https://www.mindtools.com/afzbk2y/pareto-analysis> (дата звернення: 15.12.2023).

20. Typing Speed Test – Livechat [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: <https://www.livechat.com/typing-speed-test/#/> (дата звернення: 02.05.2024).

21. Computerised assessment of handwriting and typing speed [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: https://www.researchgate.net/profile/Joanna-Horne/publication/272148120_Computerised_assessment_of_handwriting_and_typing_speed/links/54db62d70cf233119bc6123e/Computerised-assessment-of-handwriting-and-typing-speed.pdf (дата звернення: 02.05.2024).

22. Web Speech API – MDN Web Docs [Електронний ресурс]. – 2024. – Режим доступу до ресурсу: https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API (дата звернення: 02.05.2024).