

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Харківський національний університет радіоелектроніки
Факультет Комп'ютерних наук
Кафедра програмної інженерії

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

_____ другий (магістерський) _____
(рівень вищої освіти)

_____ Дослідження технологій розпізнавання голосу для веб-додатків _____

Виконав:

студент 2 курсу групи _____ ІПЗм-21-4

_____ Несміян Д.М. _____

(прізвище, ініціали)

121 – Інженерія програмного

Спеціальність _____ Забезпечення _____

Тип програми _____ Освітньо-наукова _____

Керівник _____ доц. Голян Н.В. _____

(посада, прізвище ініціали)

Допускається до захисту

Зав. Кафедри _____

З.В. Дудар

2023 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
Кафедра Програмної інженерії
Рівень вищої освіти другий (магістерський)
Спеціальність 121 – Інженерія програмного забезпечення
(код і повна назва)
Тип програми освітньо-наукова програма
Освітня програма Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«___» _____ 202_ р.

ЗАВДАННЯ**НА КВАЛІФІКАЦІЙНУ РОБОТУ**

студента Несміяну Дмитру Михайловичу
(прізвище, ім'я, по батькові)

1. Тема роботи «Дослідження технологій розпізнавання голосу для веб-додатків» затверджена наказом університету від 10 квітня 2023 р. № 340Ст
2. Термін подання студентом роботи до екзаменаційної комісії 18 травня 2023 р.
3. Вихідні дані до роботи голосові асистенти, технології розпізнавання мови, веб-сайти інтернет-магазинів, інструменти розробки веб-розширень, мова програмування JavaScript
4. Перелік питань, що необхідно опрацювати в роботі мета роботи, аналіз предметної галузі і постановка задачі, аналіз існуючих алгоритмів, формування вимог, опис експерименту, аналіз результатів дослідження

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз предметної галузі	12.04 – 15.04.2023	Виконано
2	Постановка задачі	15.04 – 20.04.2023	Виконано
3	Аналіз існуючих алгоритмів	20.04 – 25.04.2023	Виконано
4	Підготовка до дослідження	25.04 – 28.04.2023	Виконано
5	Проведення дослідження	28.04 – 01.05.2023	Виконано
6	Аналіз результатів дослідження	01.05 – 04.05.2023	Виконано
7	Підготовка пояснювальної записки	04.05 – 07.05.2023	Виконано
8	Підготовка презентації та доповіді	07.05 – 10.05.2023	Виконано
9	Перевірка на плагіат	10.05 – 11.05.2023	Виконано
10	Нормоконтроль	12.05 – 14.05.2023	Виконано
11	Рецензування	14.05 – 16.05.2023	Виконано
12	Занесення диплому в електронний архів	16.05.2023	Виконано
13	Попередній захист	16.05.2023	Виконано
14	Допуск до захисту у зав. кафедри	16.05.2023	Виконано

Дата видачі завдання 12 квітня 2023 р.

Студент _____ Несміян Д.М.
(підпис)

Керівник роботи _____ доц. Голян Н.В.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Кваліфікаційна робота магістра містить: 67 с., 45 рис., 1 табл., 18 джер.

АЛГОРИТМИ ГОЛОСОВОГО РОЗПІЗНАВАННЯ, ВЕБ-САЙТИ, ГОЛОСОВІ ТЕХНОЛОГІЇ ДЛЯ ВЕБ-СТОРИНОК, ГОЛОСОВИЙ АСИСТЕНТ, КОРИСТУВАЦЬКИЙ ДОСВІД

Об'єктом дослідження є голосові технології для веб-сторінок на прикладі голосового асистента.

Метою даної роботи є дослідження методів взаємодії з веб-сторінкою для розробки голосового асистенту, який транслює команди елементам.

Методи дослідження базуються на загальних методах роботи голосового розпізнавання.

Робота ілюструє алгоритм розпізнавання команд користувача, який зменшує кількість рутинних дій під час взаємодії з інтерактивними елементами веб-сторінки. Виконано програмну реалізацію системи з використанням мови програмування JavaScript.

USER EXPERIENCE, VOICE ASSISTANT, VOICE RECOGNITION ALGORITHMS, VOICE TECHNOLOGIES FOR WEB PAGES, WEBSITES

The object of the research is voice technologies for web pages using the voice assistant as an example.

The purpose of this work is to research methods of interacting with a web page for the development of a voice assistant that broadcasts commands to interface elements.

Research methods are based on general methods of voice recognition.

The paper illustrates an algorithm for recognizing user commands, which reduces the number of routine actions during interaction with interactive elements of a web page. The software implementation of the system using the JavaScript programming language was carried out.

Я,

Несміян Дмитро Михайлович

(прізвище, ім'я, по батькові)

Студент групи ППЗм-21-4 здобувач вищої освіти на другому
(магістерському) рівні

Кафедра програмної інженерії,

(повна назва кафедри)

Заявляю: моя кваліфікаційна робота на тему

Дослідження технологій розпізнавання голосу для веб-додатків,

(назва роботи)

Що буде представлена до ЕК для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу EIArKhNURE. Всі 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.....	8
1 Аналіз предметної галузі.....	10
1.1 Аналіз предметної галузі.....	10
1.2 Виявлення проблем та актуалізація рішень	12
1.3 Постановка задачі.....	13
2 Аналіз існуючих алгоритмів	14
2.1 Особливості будови голосових інтерфейсів.....	14
2.2 Опис існуючих алгоритмів.....	17
2.3 Вибір алгоритму	22
3 Опис експерименту	25
3.1 Дані для перевірки	25
3.2 Вибір технологій розробки.....	26
3.3 Апаратно-програмне забезпечення	28
3.4 Процес роботи з даними	30
3.5 Підготовка до експерименту	32
4 Аналіз результатів дослідження	34
4.1 Оцінка результатів	34
4.2 Напрямки подальшої роботи.....	39
Висновки	41
Перелік джерел посилання	42
Додаток А Перелік джерель посилання за науковими напрямками керівника та науковців кафедри програмної інженерії	44
Додаток Б Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ.....	45
Додаток В Слайди презентації.....	46
Додаток Г Апробація результатів роботи.....	54
Додаток Д Матеріали подані до друку.....	65

Додаток Е Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008:2015	67
---	----

ВСТУП

В епоху четвертої промислової революції світ переходить з аналогової парадигми в цифрову. Люди все більше користуються цифровими рішеннями, тому інвестування в сферу інформаційних технологій збільшується. За даними Garther, обсяг інвестицій у сферу програмного забезпечення до кінця 2025 року становитиме більше, ніж 1 мільярд доларів США [1].

Криза, спричинена Covid-19, прискорила технологічне прийняття вже відомих технологічних рішень. Пришвидшення зростання електронної комерції відбулось у світі в середньому на 10 років вперед [2]. Створення інтернет-магазинів набирає популярність не тільки в Україні, а й в усьому світі. Однією з головних причин є карантинні обмеження, але не вони є єдиними. Офлайн бізнес з кожним роком стає все менш прибутковим, а реклама набуває нових форм [3].

Інтернет-магазини мають суттєву перевагу, адже надають можливість споживачам здійснювати замовлення товарів цілодобово, незалежно від їх географічного положення, із наданням найсвіжішої інформації про товар безпосередньо від виробника, маючи можливість порівняння даного товару з аналогічними товарами конкурентів. Крім того, споживачі завжди мають онлайн-підтримку до придбаної продукції. В свою чергу, власники інтернет-магазинів значно економлять на витратах, створюючи електронний каталог товарів, та не витрачають зайвий час, коли інформують клієнтів про нові пропозиції [4].

Перехід великої кількості покупців в інтернет-середовище потребує від розробників створення зручної навігації. Зручна навігація – це один з найважливіших параметрів інтерфейсу комерційного сайту. Вона обов'язково має бути простою, зручною та зрозумілою. Погана навігація може стати причиною підвищення показнику відмов користування сайтом, зменшення часу, проведеного на сайті, а також зниження конверсії. Все це стане наслідком втрати доходів.

Інтернет-магазини, які мають зручну навігацію, можуть стати чудовим середовищем для впровадження голосових технологій, що стануть одним з

ключових факторів підвищення конверсії. Крім того, голосова навігація привабить нових користувачів, для яких голосове супроводження підвищить зручність використання інтернет-магазину.

Метою роботи є отримання концепту алгоритму, що потенційно може бути впровадженій для використання в середовищі інтернет-Магазину. Реалізація взаємодії користувача і веб-сторінки за допомогою голосу. Дослідження проблем при впровадженні голосового розпізнавання на веб-сторінку, а також незручностей, що виникають у користувача при використанні голосового супроводження.

Об'єктом дослідження цієї роботи є аналіз дій користувачів інтернет-магазинів.

Предметом дослідження виступає створення алгоритму голосової взаємодії користувача та веб-сторінки інтернет-магазину та прогнозування підвищення конверсії завдяки впровадженню такої реалізації.

В процесі роботи необхідно проаналізувати принцип роботи, можливості, а також виділити ключові переваги та недоліки існуючих голосових асистентів. Враховуючи отримані результати, необхідно створити власне рішення для веб-сайтів, щоб досягти поставлену мету.

Здійснена робота дозволила отримати голосового асистента для веб-сайтів, який має в собі елементи керування веб-сторінкою. Отриманий результат дозволяє зробити висновки щодо зменшення кількості рутинних операцій під час введення інформації про товари під час купівлі в інтернет-магазинах за допомогою голосу.

За результатами дослідження було опубліковано тези в рамках 27 Міжнародного молодіжного форуму «Радіоелектроніка та молодь у XXI столітті», а також була подана стаття на 18 міжнародну конференцію з ІКТ в освіті, дослідженнях і промислових застосуваннях.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Аналіз предметної галузі

Інтернет-магазин – це місце в інтернеті, де відбувається прямий продаж товарів споживачеві (юридичній або фізичній особі), враховуючи доставку. Розміщення споживацької інформації, замовлення товару і угода відбуваються також всередині мережі (безпосередньо на сайті інтернет-магазину). Головна відмінність інтернет-магазину від звичайного магазину полягає у відсутності потреби фізичних засобів. Інфраструктура інтернет-магазину реалізована програмно [5].

Створення інтернет-магазинів – це тривалий процес, розділений на кілька етапів. Завдяки такому поділу під час створення інтернет-магазину значно легше орієнтуватись на те, які дії чи маніпуляції необхідно проводити в певний період часу. Обов'язковими етапами є вибір ніші для створення інтернет-магазину, створення плану робіт, розробка структури сайту та створення оформлення для сайту.

На етапі вибору ніші приймається рішення, які товари буде продавати інтернет-магазин. Варто враховувати, буде одна група товарів, чи декілька. Кожна група може по-різному викликати зацікавленість у клієнтів. Ніша в інтернет торгівлі може вплинути на безліч факторів, в тому числі й на те, наскільки просто або складно буде просувати створений інтернет-магазин. Важливим показником є конкуренція у ніші. Враховуючи, що кожен підприємець прагне перенести свій бізнес в інтернет, конкуренція у ніші також зростає.

Під час етапу створення плану робіт проводиться визначення термінів та обсягів роботи по створенню сайту. Створення плану дозволяє ретельно відстежувати кожні дії співробітників команди.

При створенні структури сайту безпосередньо проектується веб-сайт. Створення структури має починатись з макету, який буде включати опис головної та другорядних сторінок. Після цього відбувається перехід до другорядних

сторінок, на яких відбувається розділення товарів на категорії, планується реалізація способу отримання грошей від клієнта та надання йому інформації про доставку товарів. Доцільно використовувати правило чотирьох кліків. Суть даного правила полягає в тому, що в створеному інтернет-магазині клієнт перебуваючи на будь-якій сторінці сайту повинен мати можливість дістатися до будь-якої цільової сторінки за допомогою чотирьох кліків.

Оформлення сайту грає чи не найважливішу роль для клієнта при виборі товарів. Клієнт завжди спочатку звертає увагу на зовнішній вигляд товару, а вже потім читає його опис. Якщо дизайн та оформлення не сподобається клієнту, є ризик, що клієнт одразу покине сторінку інтернет-магазину.

Кожен окремий етап створення інтернет-магазину може вплинути на успішність бізнесу, тому дуже важливо, щоб кожному етапу приділялась максимальна кількість часу, сил та вмінь.

Голосовий інтерфейс може покращити існуючий користувацький інтерфейс, оскільки забезпечує більш зручний та менш обмежений спосіб взаємодії людини та комп'ютера. Якісний голосовий інтерфейс допомагає долати неприйняття технології користувачами, оскільки як для його використання не потрібно опановувати нові навички. Голосовий інтерфейс якісно змінює цей спосіб, а отже і ефективність взаємодії користувача з системою. Голосовий пошук від компанії Google та голосовий асистент Siri від компанії Apple є цьому яскравими прикладами, підтверджуючи нагальну необхідність впровадження мовних технологій, зокрема розпізнавання мовлення та голосових інтерфейсів.

Важливий і практичний аспект цих проблем пов'язані з тим, що голосовий інтерфейс є необхідною складовою, коли йдеться про створення комфортних умов життя для людей з порушеннями опорно-рухового апарату, а також фахівцям, які втратили можливість використовувати стандартні засоби в результаті професійного захворювання, травми або каліцтва.

Розпізнавання мови – технологія, що дозволяє використовувати природний для людини мовленнєвий інтерфейс взаємодії з електронною технікою [6].

Після аналізу предметної галузі, можна зробити висновки, що вже виконуються мінімальні умови для впровадження голосового супроводження на веб-сторінки деяких інтернет-магазинів. Зокрема тих, які мають зручний користувацький інтерфейс.

Використання голосового супроводження не лише спростить користування інтернет-магазинами, а також надасть суттєву перевагу власникам сайтів в конкуренції сфери інтернет-продажів.

1.2 Виявлення проблем та актуалізація рішень

За статистикою, в середньому 74% усіх покупців залишають товари в корзині інтернет-магазину без оформлення замовлення [7]. Найголовнішою причиною є помилки, яких припускаються проєктувальники та розробники при створенні інтернет-магазинів. Серед них можна виділити:

- відсутність функції «замовити в один клік»;
- відсутність можливості замовити додати більше однієї одиниці товару в корзину;
- немає повідомлення про успішне додання товару у корзину;
- немає додаткових опцій;
- немає акценту на головний СТА;
- незручне редагування корзини;
- немає пояснень до полів у формах;
- немає швидкого доступу до консультанта;
- немає інструкції про подальші дії для покупця після оформлення замовлення.
- не автоматизована робота з залишеними корзинами;
- не тестується верстка та функціонал інтернет-магазину.

Виявлені проблеми можуть бути вирішені за допомогою сторонніх засобів, що можуть вільно створюватись розробниками у вигляді розширень для браузеру

[8]. Розширення браузера – це мікропрограма, яка доповнює функціонал браузера. Найпопулярніші розширення для Google Chrome та Mozilla Firefox.

Зокрема для користувачів вирішення деяких з цих проблем може надати користувачу можливість виконувати покупки із задоволенням, а власнику запропонувати додаткові пропозиції для продажу.

1.3 Постановка задачі

В кваліфікаційній роботі необхідно дослідити існуючі рішення голосової взаємодії користувача з веб-стінками сайтів та розробити концепт алгоритму голосової взаємодії користувача та веб-інтерфейсу придатний для використання в середовищі інтернет-магазинів. Первинна розробка концепту має спиратись на взаємодію з інтерфейсом, що реалізовано без допущення критичних помилок.

Розроблений алгоритм повинен:

- враховувати послідовність кроків, яку виконує користувач при купівлі товару;
- чітко розпізнавати команди користувача;
- пропонувати повторне введення інформації в разі невдалого розпізнавання;
- взаємодіяти з найголовнішими елементами, через які проходить користувач;
- мати можливість інтегрування додаткових команд;
- бути сумісним із різними технологіями голосового введення.

Має бути передбачено два сценарії роботи: у якості інтегрованого рішення в структуру інтернет-магазину та у вигляді розширення браузера.

2 АНАЛІЗ ІСНУЮЧИХ АЛГОРИТМІВ

2.1 Особливості будови голосових інтерфейсів

Мова – основний вид передачі думок, ідей та почуттів у людському середовищі. Можливість контролю оточення голосом підштовхнув науковців та експертів передових корпорацій до створення таких програмних додатків, як Siri від Apple, Cortana від Microsoft та інших.

Вирішення задачі розпізнавання голосу дозволив перейти від командного рядка та WIMP (Window, Image, Menu, Pointer) до SILK (Speech, Image, Language, Knowledge) інтерфейсу програмних систем. На відміну від взаємодії користувача з комп'ютерною системою за допомогою клавіатури, миші, джойстика та дисплею мовний інтерфейс має наступні переваги:

- для спілкування з комп'ютером людині немає необхідності мати спеціальні навички або уміння в галузі інформаційних технологій;
- мова знижує психологічну та фізичну відстань між людиною і комп'ютером, а також може бути пов'язана за допомогою системи комунікацій;
- мовний інтерфейс надає оперативність і мобільність спілкування, звільнення рук і розвантаження зорового каналу при отриманні інформації.

Розпізнавання мовлення (speech recognition, speech to text (STT)) – автоматизований процес перетворення мовленнєвого сигналу в текстовий потік. [9]

Для вирішення даної задачі існує велика кількість систем автоматичного розпізнавання мовлення, які перетворюють вхідний мовленнєвий сигнал в повідомлення. При цьому повідомлення може бути представлено як у формі тексту, так і одразу перетворено в зручну для його подальшої обробки форму. Форма може бути застосована для формування відповідної реакції програмної системи.

Системи розпізнавання мовлення класифікуються за такими ознаками:

- розмір словника (обмежений набір слів або великий словник) – чим більший розмір словника, з яким працює система розпізнавання мови, тим більше помилок при розпізнаванні слів. Словник, що складається тільки з цифр, може бути

розпізнаний практично безпомилково, тоді як вірогідність помилок при розпізнаванні словника в сто тисяч слів може досягати 45%.

– залежність від диктора (дикторозалежні або дикторонезалежні системи) – дикторозалежна система призначена для роботи тільки з людиною, яка навчала цю систему, в той час як дикторонезалежна система призначена для роботи з будь-яким диктором. На поточному етапі розвитку систем розпізнавання мовлення вірогідність виникнення помилок в дикторонезалежній системі в 3-5 разів більша, ніж у дикторозалежних;

– тип мови (злитна, роздільна). Роздільна мова – це мова, в якій слова відокремлюються одне від одного проміжком тиші. Злитна мова – це природно вимовлений текст. Розпізнавання злитного мовлення складніше, тому що у вимовлених слів немає чітких меж. Системи розпізнавання мовлення, що працюють з ізольованими словами, досягли високо рівня точності розпізнавання – 95-99%, в той час як задача розпізнавання злитної мови в достатній мірі не вирішена.

– призначення (системи диктування, командні системи) – визначає необхідний рівень абстракції, на якому буде відбуватися розпізнавання мови. Системи голосового набору мобільного телефону, де здійснюється розпізнавання за шаблоном, називаються командними. На відміну від них, система диктування вимагає розпізнавання на базі виділення лексичних елементів. При інтерпретації виголошеної фрази вона буде покладатися не тільки на те, що було виголошено в поточний момент, але і на те, як це співвідноситься з тим, що було вимовлено до цього. Також в таку систему повинен бути вбудований набір граматичних правил. Чим суворіші ці правила, тим простіше реалізовувати систему розпізнавання, але набір слів, які вона зможе розпізнати, буде меншим;

– алгоритм, що використовується. Після того, як мовний сигнал розбивається на певні частини, відбувається імовірнісна оцінка належності цих частин до того чи іншого елемента словника, що здійснюється за допомогою одного з алгоритмів розпізнавання;

– за типом структурної одиниці (фрази, слова, фонем) – системі розпізнавання мовлення, які використовують цілі слова або фрази, називаються шаблонними. Вони, як правило, дикторозалежні, і їх реалізація є простішою, ніж створення систем, які розпізнають мовлення на базі виділення лексичних елементів. У таких системах структурними одиницями мови є лексичні елементи;

– за принципом виділення структурних одиниць. Найпоширеніший підхід виділення структурних одиниць заснований на перетворенні Фур'є, яке переводить вихідний сигнал з амплітудно-часового простору в частотний. Однак аналіз Фур'є має цілу низку недоліків, в результаті яких відбувається втрата інформації стосовно часових характеристик оброблюваних сигналів. У зв'язку з цим для завдання виділення структурних одиниць мови виправдано використання вейвлет-аналізу. Вейвлет – це математична функція, яка дозволяє аналізувати частотні компоненти даних. В загальному випадку, аналіз сигналів проводиться в площині вейвлет-коефіцієнтів – масштаб-час-рівень (Scale-Time-Amplitude). Отримані вейвлет-спектри відрізняються від спектрів Фур'є тим, що дають чітку прив'язку властивостей сигналу до часу. Крім вейвлет і Фур'є-аналізу в системах розпізнавання мовлення використовується кепстральний аналіз, але створення таких систем є трудомістким і вимагає високої кваліфікації розробника.

Таким чином голосовий інтерфейс є максимально зрозумілим рішенням для користувача. Серед перелічених характеристик, розмір його словника має покривати усі необхідні потреби запитів користувача. Він має бути адаптований до сфери, в якій використовується. В даному випадку, середовища інтернет-торгівлі. Система має бути дикторонезалежною та здатною сприймати злитну мову. Після отримання аудіодоріжки має бути можливість розбиття її на окремі слова. Такий підхід дозволить впровадити підтримку як командного режиму, так і режиму диктування. Потрібно також враховувати унікальність слів у словнику. Якщо слова дуже схожі, то похибка розпізнавання збільшується. Інші критерії наразі не є ключовими, тому не потребують детального розгляду в рамках цієї роботи.

2.2 Опис існуючих алгоритмів

Сфера розпізнавання голосу вже давно не нова, проте вона лише зараз набуває широкого застосування на веб-сайтах. Кожне рішення використовує власні технології та засоби, проте жодне з них не є універсальним рішенням, яке б працювало на будь-яких веб-сайтах. Загально усі існуючі рішення базуються на технології голосового розпізнавання [10].

Технологія голосового розпізнавання має два способи взаємодії: режим голосової команди та режим голосового пошуку. Голосовий пошук працює так само, як і звичайний, проте керується голосом.

Під час просування продуктів важливо враховувати і те, і інше — значна частина голосових команд також орієнтована на пошук; тобто люди можуть шукати продукт чи послугу в інтернеті.

Коли здійснюється голосовий запит, технологія перетворює звуковий сигнал у текст. Потім пошукова система або програми обробляють це та виводять відповідний результат.

Обробка голосового запиту здійснюється в чотири етапи:

- фільтрація — потрібна фраза виділяється з хмари шумів. Це необхідно для усунення зовнішніх перешкод.
- цифровізація — звукові хвилі перетворюються в код, який придатний для обробки за допомогою обчислень.
- аналіз — отримані дані обробляються, стискаються та готуються до ідентифікації.
- ідентифікація даних шаблону — запит порівнюється з прикладами з бази даних та історії пошуку, щоб показати конкретні слова та фрази.

Було розглянуто найпопулярніші рішення, що базуються на цій технології.

VoxRow [11] – це інструмент, який дозволяє використовувати силу голосу досить легко і безкоштовно. Для правильної роботи він використовує алгоритми машинного навчання для розпізнавання мови, досягаючи більш ніж 95% точності,

працюючи більше ніж 100 мовами (рис. 1). Для встановлення потрібно лише зареєструватися на сайті та вставити невеликий фрагмент коду в HTML-код або, зокрема, в тег <body>. Клієнту, який відвідує веб-сайт, необхідно дати свою згоду та дозвіл на використання мікрофона, після чого він може використовувати сервіс.

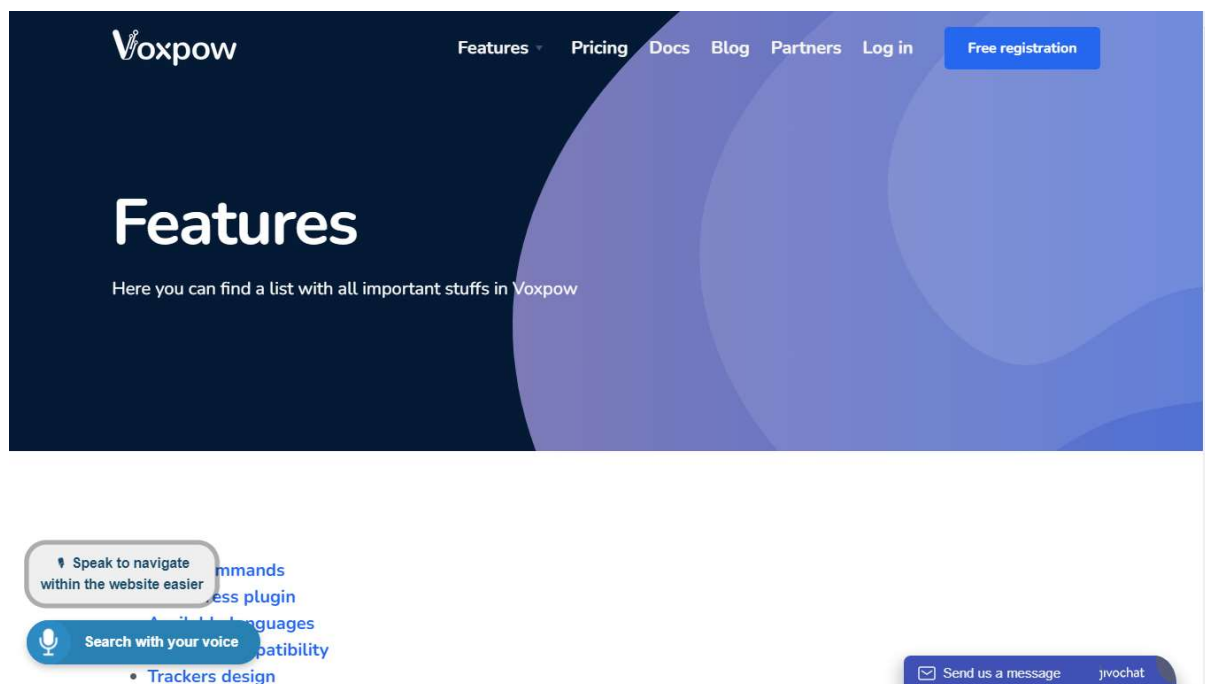


Рисунок 2.1 – Результат запиту на сторінці VoxPow

Демонстраційна версія на офіційному сайті VoxPow працює переважно в режимі кліку по об'єкту на веб-сторінці. Якщо відповідний об'єкт знайдено, команда одразу виконується. В іншому випадку, програма запропонує певні команди, що схожі на об'єкти на сторінці. Серед великої кількості мов немає української.

Handsfree for web [12] – ця програма корисна для навігації в інтернеті без рук. За описом вказано, що підтримуються сотні команд (рис. 2). Функціонал дозволяє зробити голосом усе, що робив за допомогою миші та клавіатури. А саме: переглядати інтернет, відкривати веб-сайти, клікати по об'єктах на веб-сторінці, рухатись, диктувати текст, заповнювати будь-який тип введення, копіювати та вставляти. Також є можливість керувати закладками, завантаженнями, історією та вкладками. Окрема зручна функція – керування аудіо та відео вмістом сторінки.

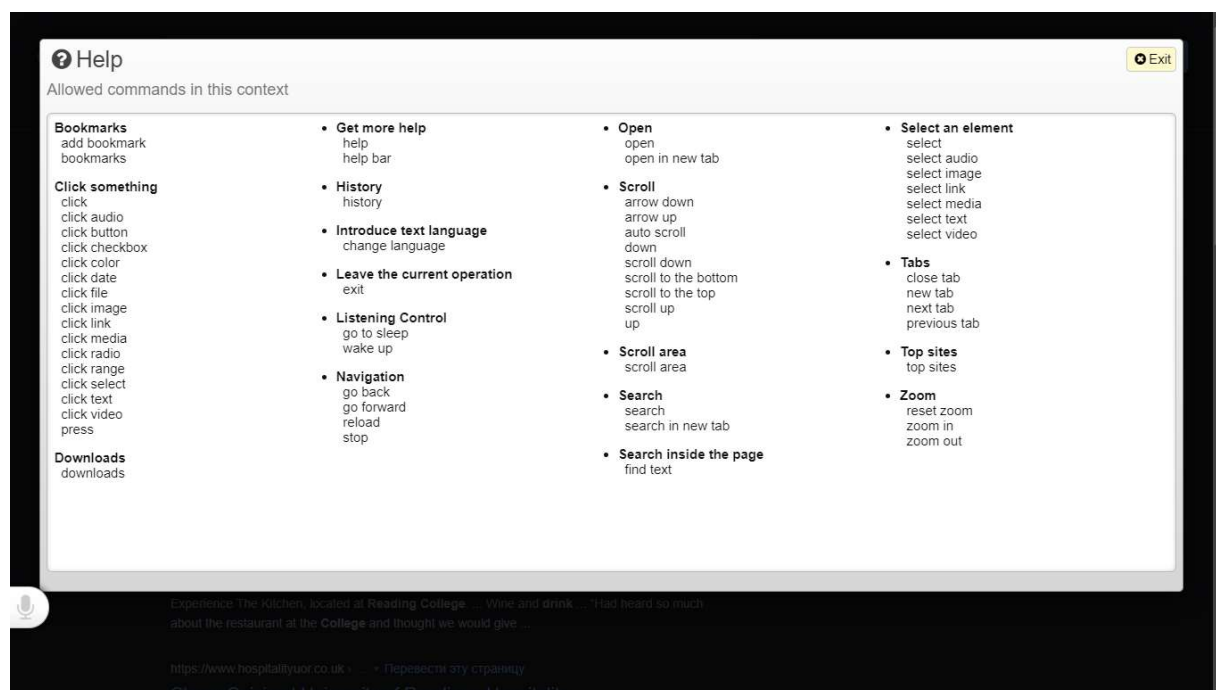


Рисунок 2.2 – Набір команд розширення Handsfree for web

Тим не менше, варто перевіряти кожну команду окремо. Інструмент показує кожну розпізнану команду незалежно від того, чи вона відповідає тому, про що сказав користувач, чи ні. Завдяки цьому було визначено, що не всі з розпізнаних команд були виконані відповідно до їх опису. Зокрема група команд, які відповідають за роботу з текстом на веб-сторінці.

Великою перевагою цього інструменту є взаємодія з іншими сервісами та пошуковими мережами. Деякі запити замість пошукової мережі Google використовують DuckDuckGo. Для пошуку достатньо ввести команду «Search» («Пошук»), після чого необхідно сказати, що саме необхідно знайти. Також є підтримка кліку по об'єктах, проте перед тим як перейти до об'єкта, необхідно спочатку сказати спеціальний номер, за допомогою якого програма знайде об'єкт на сторінці. Команди прокручування сторінки, відкривання та закривання вкладок також присутні.

Якщо казати лише про інструменти голосового розпізнавання без додаткових команд, то варто звернути увагу на Web Speech API [13]. Web Speech API надає дві різні сфери функціональності — розпізнавання мовлення та синтез мовлення

(також відомий як текст у мовлення або TTS), які відкривають нові цікаві можливості для доступності та механізмів керування.

Розпізнавання мовлення передбачає отримання мовлення через мікрофон пристрою, яке потім перевіряється службою розпізнавання мовлення зі списком граматики. Повертається як результат у вигляді текстового рядка, із яким в подальшому можна виконувати подальші дії.

Web Speech API має для цього основний інтерфейс контролера — `SpeechRecognition` — плюс ряд тісно пов'язаних інтерфейсів для представлення граматики, результатів тощо. Як правило, система розпізнавання мовлення, що встановлена на пристрої за замовчуванням, буде використовуватися для розпізнавання мовлення.

На рисунку 3 наведено демонстрацію роботи алгоритму. Для початку потрібно клікнути по простору та назвати будь-який з перелічених кольорів. Після вдалого розпізнавання фон сторінки зафарбовується в обраний колір.

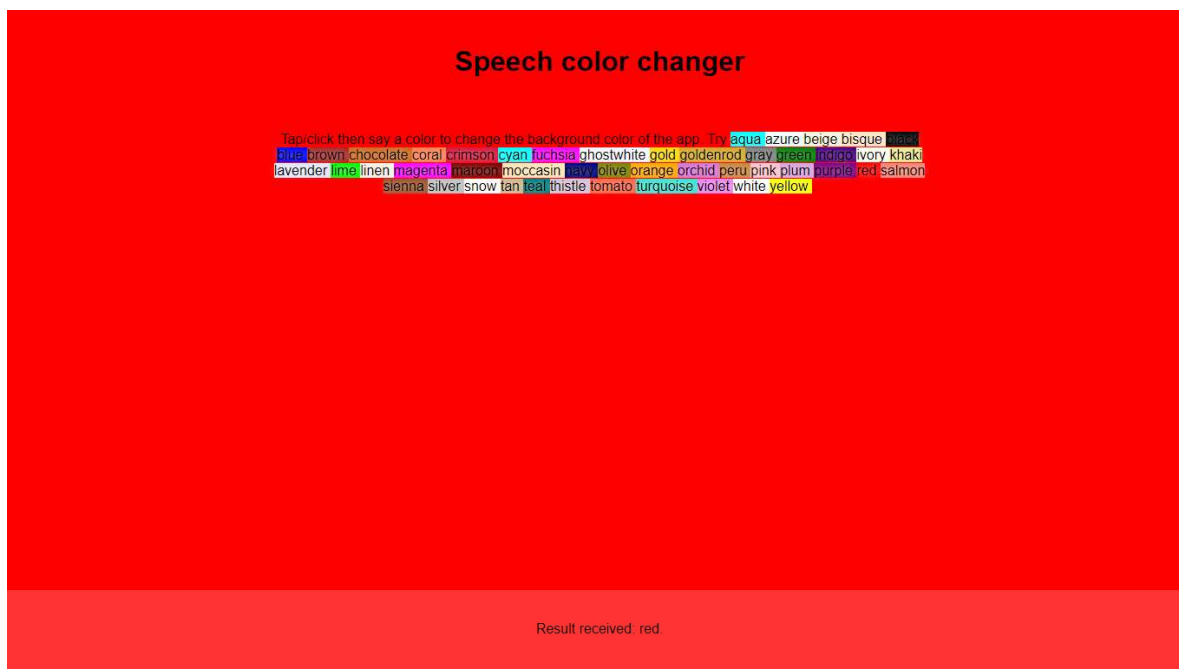


Рисунок 2.3 – Зміна кольору за допомогою Web Speech API

Наразі використання Web Speech API обмежено лише браузером Google Chrome, проте зважаючи на популярність цього браузера [14], рішення може

використовуватись для впровадження на веб-сайти та побудови власного алгоритму. На жаль, воно також не має підтримки української мови.

Крім інструментів голосового розпізнавання було проаналізовано також способи застосування цих інструментів. У дослідженні Мандіпа, Фаргани та Харуни було детально розглянуто проблеми використання голосового розпізнавання. Дослідження охоплювало хмарну систему розпізнавання голосу (SRS) для електронної комерції у якості сценарію використання [15]. Особливу увагу надано незрячим користувачам, які найбільше потребують рішень керування за допомогою голосу. Під час дослідження було визначено, що стандартів, встановлених Консорціумом Всесвітньої павутини (W3C) щодо доступності веб-контенту (WCAG 2.1), недостатньо для вирішення всіх проблем доступності. У дослідженні порівняли роботу сервісів розпізнавання голосу IBM Watson, Google та Amazon на прикладі власного розробленого рішення.

Окремо варто виділити інші підходи, що використовувались для використання незрячими, оскільки впровадження голосового рішення для незрячих суттєво підвищить користувацький досвід, а також стане чудовим ґрунтом для впровадження голосових рішень у інших продуктах. Португальське дослідження 2012 року показало, що лише 62% веб-сайтів було придатним для використання незрячими [16]. Варто зазначити, що в цій країні доступність регулюється на законодавчому рівні та регламентується за W3C (World Wide Web Consortium). Ще одне дослідження свідчить про те, що модель WAI (Web Accessibility Initiative) поступається принципам Accessibility 2.0, оскільки використовує підхід, в якому незрячий, для якого створюють рішення є лише користувачем. Accessibility 2.0 пропонує поставити користувача на етап проектування ПЗ [17].

В результаті проведеного аналізу було вирішено використовувати підхід, що поєднував би у собі підходи порівняння роботи голосових технологій VoxPow, Handsfree for Web та Web Speech API для розробки концепту голосового алгоритму, що був би придатний для використання у веб-додатках інтернет-магазинів. Придатність використання цих сайтів незрячими за допомогою голосу також має бути врахована.

2.3 Вибір алгоритму

Оскільки аналоги не надають повного спектру взаємодії користувача та веб-сторінки інтернет-магазину, стає потреба у створенні власного програмного рішення та алгоритму його роботи.

Головною вимогою алгоритму, який розробляється, є поєднання технологій введення та виведення голосу, тому результуючий алгоритм має поєднувати у собі риси аналогів, а саме:

- голосове введення команд та інформації для обробки;
- озвучення результату та супроводження користувача під час користування веб-сторінками інтернет-магазину;
- масштабування для інтеграції команд керування вмістом сторінки безпосередньо користувачем;
- можливість працювати з різними технологіями голосового розпізнавання;
- взаємодію з найголовнішими елементами інтернет-магазину: пошуком, фільтрацією та сортуванням.

Голосове введення команд безумовно надасть користувачу можливість обрати зручний спосіб взаємодії з елементами інтернет-магазину. Це також дозволить користувачу створити персональну послідовність дій для замовлення товарів або послуг.

Супроводження користувача доцільно реалізувати у вигляді голосового бота, який буде вмикатись за необхідністю та допоможе користувачу вирішити складне для нього питання. Вмикання бота може бути виконано за допомогою спеціальної кнопки. Для того, щоб зацікавити користувача у використанні бота, варто надати кнопці вигляд привітного персонажа.

Масштабування необхідне не тільки для розширення функціоналу та можливостей інструменту, а також для того, щоб впроваджувати користувачам можливість обрати власні алгоритми швидкого замовлення товарів.

Можливість роботи з різними голосовими технологіями не просто необхідна, вона обов'язкова! Перш за все тому, що не можна зосереджувати роботу лише на одній голосовій технології. Кожна технологія має свої ключові можливості, переваги та тарифні плани. Перемикання між ними дозволить не тільки обрати користувачу зручний інструмент, а ще й дозволить заощадити.

Окремо варто приділити увагу питанню роботи з інструментами пошуку, фільтрації та сортування. Найкраща реалізація алгоритму має отримувати голосовий запит та автоматично розкласти його на слова, які вже автоматично будуть розкладатись за фільтрами, і вже тільки після цього буде надіслано запит для обрання товарів. Якщо після введення усіх фільтрів буде кілька варіантів зі швидких підказок, користувачу має випасти поле для швидкого заповнення. Крім того варто залучати взаємодію з категоріями товарів та послуг, якщо вони є у наявності. Фільтрація та сортування товарів залежать перш за все від користувацького інтерфейсу. Велика ймовірність, що для адаптації алгоритму під інтернет-магазини, доведеться використовувати персональний підхід до кожного з них. Особливо уважно потрібно бути з тими магазинами, де оновлення товару відбувається одразу при застосуванні фільтру.

За результатами проведеного аналізу можна визначити, що для замовлення товару алгоритм буде працювати наступним чином:

- користувач переходить на веб-сторінку інтернет-магазину, що оснащений технологією голосового супроводження;
- користувач робить клік по голосовому помічнику, і підтверджує дозвіл на використання мікрофону, якщо це є необхідним;
- користувач каже команду «пошук» для обрання поля пошуку;
- користувач називає необхідний товар та кілька параметрів, що виділять товар серед інших;
- система проводить розділення за фільтрами та надсилає результуючий запит на пошук усіх схожих за запитом товарів;
- користувач бачить товари. За необхідності користувач може використати голосового помічника для озвучення певних параметрів товарів;

- при промовлянні команди для додавання товару до козики, товар буде додано до козики;
- при промовлянні команди для купівлі товару, відбувається перехід на сторінку продажу;
- при введенні даних про замовника товару, помічник може допомогти заповнити деякі поля за допомогою голосу;
- при остаточному заповненні даних, відбувається перехід до сторінки оплати. Помічник очікує завершення оплати, після чого запропонує користувачу купити ще один товар.

Таким чином, алгоритм проаналізований та спроектований.

3 ОПИС ЕКСПЕРИМЕНТУ

3.1 Дані для перевірки

Для проведення експерименту, необхідно обрати дані, які будуть використовуватися для перевірки.

Зважаючи, що смартфон рано чи пізно стає незамінним асистентом користувача для дозвілля та роботи в сучасному цивілізованому світі, то експеримент проводитиметься саме з алгоритмом вибору смартфона. Крім того, програма «Держава в Смартфоні» сприятиме попиту на смартфони.

Шлях користувача має включати дії, вказані в діаграмі активності (рис. 3.1).

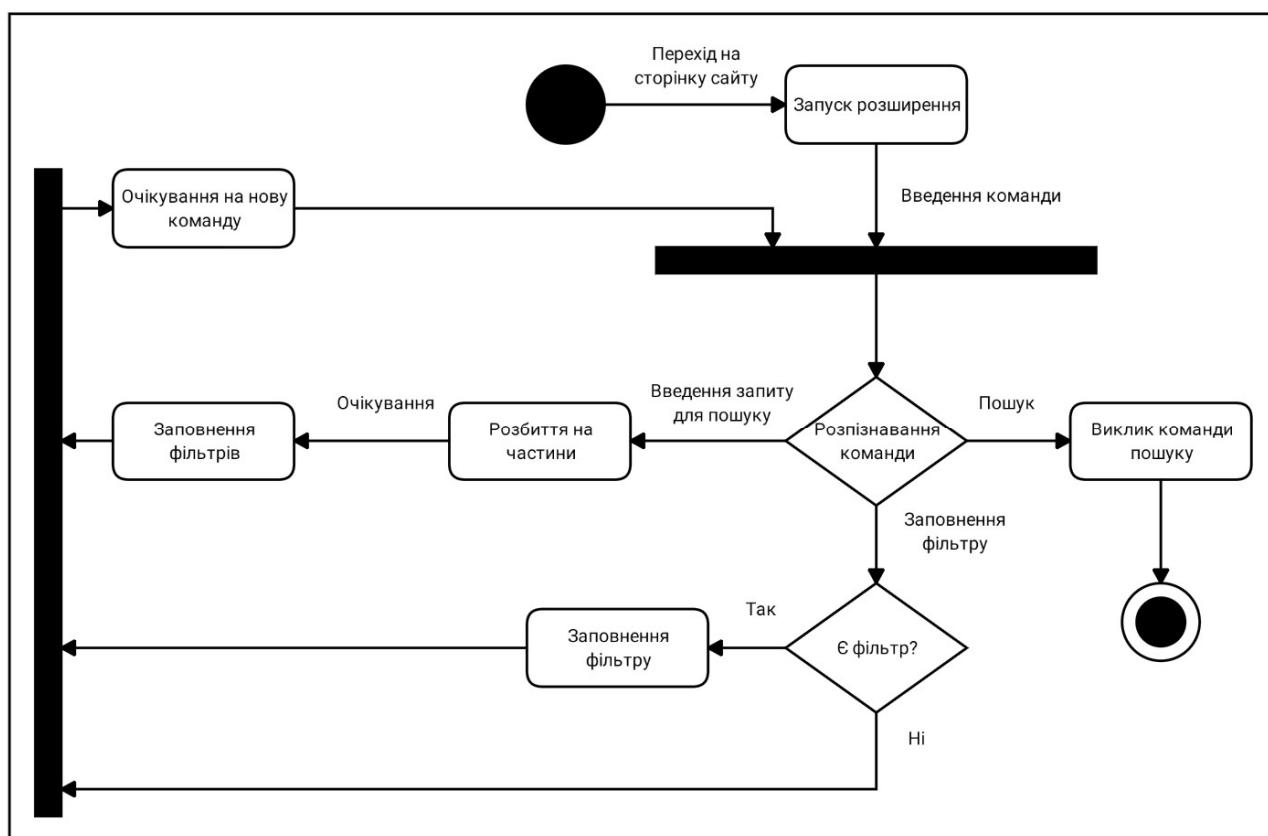


Рисунок 3.1 – Шлях користувача під час пошуку

Важливо врахувати, що деякі з цих дій будуть повторюватися кілька разів. Зокрема процес заповнення фільтрів, адже користувачу може бути недостатньо запропонованих результатів та він захоче побачити більше.

Нехай потенційні користувачі будуть шукати смартфони, які будуть відрізнятися за кольором, обсягом оперативної пам'яті, вбудованим сховищем та наявністю модулю бездротової передачі даних малого радіусу дії (Near Field Communication – NFC). Також в рамках експерименту припустимо, що цим користувачам буде достатньо обрати будь-який смартфон з перших запропонованих на сторінці. Час на прокрутку вони витратити не будуть.

Варіанти характеристик для пошуку наведено в таблиці 3.1

Таблиця 3.1 – Приклади характеристик пошуку

№	Колір	Обсяг ОЗП	Обсяг сховища	Наявність NFC
1	Червоний	4	64	Так
2	Чорний	4	64	Так
3	Чорний	6	128	Ні
4	Чорний	6	128	Так
5	Чорний	8	128	Так
6	Червоний	8	128	Так
7	Чорний	4	128	Ні

Таким чином дані для пошуку було обрано

3.2 Вибір технологій розробки

Розроблений алгоритм дуже просто інтегрується в клієнтські сторінки сайтів, адже розроблений за допомогою стандартних веб-інструментів. Серед цих інструментів мова розмітки веб-сторінок (HTML), каскадна таблиця стилів (CSS)

та мова програмування JavaScript. Ці засоби є базовими для побудови веб-розширень.

Оскільки функціонал прототипу реалізований у вигляді веб-розширення, окремий вузол з базою даних для зберігання даних користувача не є необхідним. Тим не менше, розширення браузеру потребує файлу, який буде зберігати поточні налаштування користувача для взаємодії з інструментом.

Для розпізнавання мови при тестуванні використовувався Web Speech API з англійським мовним пакетом. Тестування в наявних українських інтернет-магазинах не є можливим, адже вони не оснащені англійською локалізацією. З цієї причини тестування відбулося в середовищі тестової веб-сторінки, яка імітує поведінку інтернет-магазину (рис. 3.2). Для створення прототипу використовувалось середовище розробки Visual Studio Code версії 1.78.

Розширення та прототип тестувалися в середовищі браузера Google Chrome, але в майбутньому можуть бути вдосконалені для використання в інших браузерах, побудованих на ядрі Chromium. Серед цих браузерів в першу чергу можуть бути Mozilla Firefox, а також Microsoft Edge.

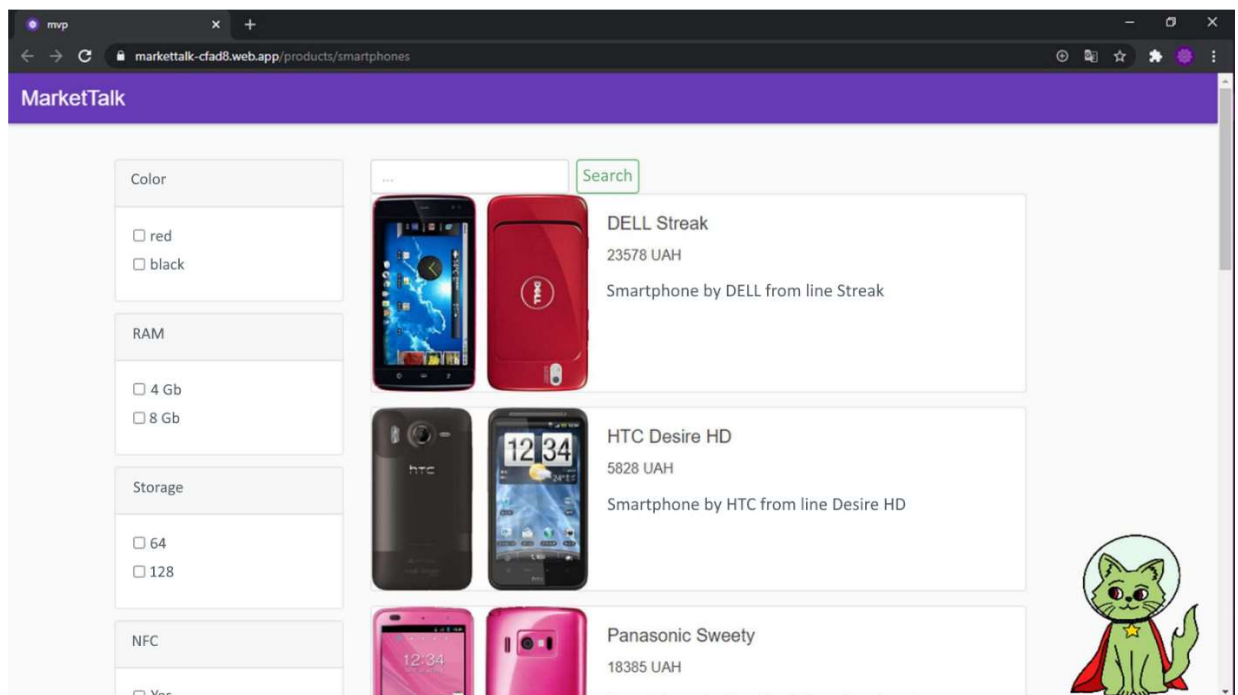


Рисунок 3.2 – Сторінка прототипу

Головний інтерфейс прототипу має поле пошуку, область фільтрів та зону переліку товарів.

Кнопка асистента в наданому прикладі реалізована у вигляді kota. Вона візуально відображує, що розширення працює, і з ним можна взаємодіяти. Для запуску голосової команди можна натиснути на зображення або натиснути комбінацію гарячих клавіш. Після відпускання клавіш або зображення, відбувається перехід до розпізнавання введеної команди.

3.3 Апаратно-програмне забезпечення

Для роботи голосового розпізнавання велику роль грають компоненти апаратного забезпечення.

Загальні вузли, які приймають участь в обробці голосового запиту користувача відображені на діаграмі розміщення (рис. 3.3).

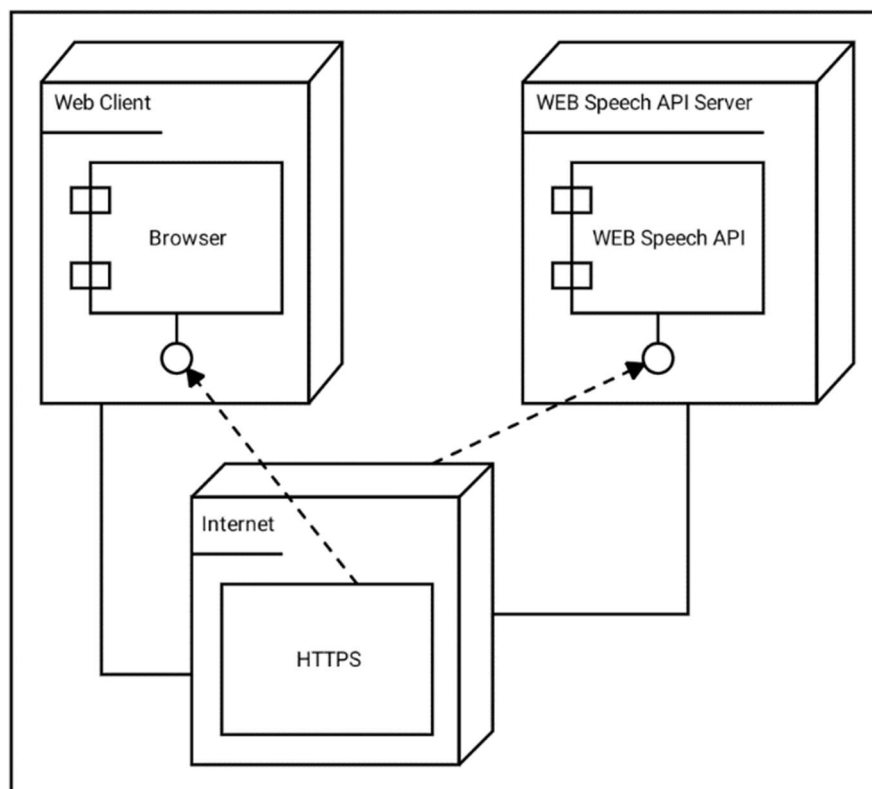


Рисунок 3.3 – Діаграма розгортання

Зокрема звук від користувача захоплюється мікрофоном та передається для опрацювання в операційну систему, де розкодовуються драйверами пристрою. Оброблений звук приймається інструментами браузеру та передається на подальшу обробку до серверів WEB Speech API. Результат повертається у вигляді текстової команди, з якою в подальшому працює логіка веб-розширення.

З'єднання інтернет є каналом комунікації між застосунком та сервером розпізнавання, тому воно грає не менш важливу роль в процесі передачі даних. З'єднання має бути не тільки надійним, але й швидким. На щастя, рівень розвитку сучасних технологій надає можливість розпізнавання даних на швидкості до 100 Мбіт/с, що є цілком доступним для потенційного користувача. Крім того, поступовий прогрес все частіше змушує користувачів переходити до цього стандарту.

Будь-який комп'ютер, який працює з голосовими технологіями, має бути оснащений процесором з двома ядрами, SSD-накопичувачем та щонайменше 4 гігабайтами оперативної пам'яті. Ці вимоги встановлюються не скільки рішенням, скільки операційною системою. Ресурсів має бути достатньо для підтримки працездатності системи разом з браузером та веб-розширенням.

Варто також враховувати тип клавіатури, з якою буде відбуватися порівняння. Мембранні клавіші мають більш плавний хід, що додатково уповільнює процес набору тексту в порівнянні з механічними клавішами.

Враховуючи досвід тестування попередньої версії голосового алгоритму, для підвищення точності захоплення звуку бажано додавати до системи звукову карту, яка буде розділяти звук мікрофону та динаміків або навушників [18]. Самі навушники на якість розпізнавання не впливають, і є лише зручним інструментом взаємодії користувача та комп'ютера. Вони можуть бути обрані користувачем за його особистими вподобаннями.

Враховуючи вимоги до апаратного забезпечення, серед наявних комп'ютерів, тестування відбувалося на машині з наступними характеристиками:

- процесор Intel Xeon e3-1240 v2, 3,4 GHz, 4 cores (8 Threads);

- ОЗП 16 Gb з частотою 1600 MHz;
- накопичувач SSD Kingston A400 240 Gb;
- звукова карта V8;
- навушники з мікрофоном Razer Kraken Essential;
- механічна клавіатура MARVO K917.

Поєднуються комплектуючі за допомогою системної плати ASRock H61M-VG4. Пороговий чипсет визначається підтримкою інструкцій в процесорі та сумісністю з операційною системою, що є індивідуальним для кожної системи.

Середня швидкість інтернету під час тестування складала 94 Мбіт/с.

3.4 Процес роботи з даними

Як вже було зазначено, голосове налаштування прототипу підключається за допомогою Web Speech API. Дані зберігаються в змінну, яка за необхідності буде змінювати значення для підключення іншого API.

```

window.SpeechRecognition = window.webkitSpeechRecognition ||
window.SpeechRecognition;
let recognition = new window.SpeechRecognition();

```

Для прикладу використано невеликий набір даних, з якого збираються товари для демонстрації роботи рішення.

```

let db = {
  categories: {
    "smartphones": {
      id: "smartphones",
      filters: {
        color: {
          name: "color",
          options: ["red", "black"]
        }
      }
    }
  },
  products: {
    "0": {
      id: "0",

```

```

    categoryId: "smartphones",
    title: "iPhone SE 2020",
    description: "Apple iPhone SE is a continuation of the line of compact smartphones. Like the iPhone 5S, the body of the SE model is made in a similar design and dimensions, as well as materials. The size and type of the display remained unchanged – it is an IPS matrix with a diagonal of 4 inches and a dot density of 326 pixels per inch.",
    brand: "Apple",
    price: "16999",
    discount: "13",
    currency: "UAH",
    img:
"https://i.citrus.ua/imgcache/size_180/uploads/shop/e/5/e5e54fb94120a0e5f5084440c08fbd8a.jpg",
    filter: {
      color: "red"
    }
  }
}

```

При купівлі користувач називає продукт, який бажає купити та детальні характеристики до нього. Веб-сайт отримує голосову команду від користувача та розбирає її по словам. За допомогою слів заповнюються відповідні фільтри та відбувається автоматичне натискання кнопки пошуку.

```

export default class Search extends HTMLElement {
  _productsService = new ProductsService();

  constructor() {
    super();
  }

  get productsService() {
    return this._productsService;
  }

  updateSearchInput = ({userInput}) => {
    console.log("updateSearchInput", userInput);
    this.querySelector(".search__input").value = userInput;
  }

  find = async ({userInput}) => {
    console.log("find", userInput);
    let result = await this.productsService.getProductByName(userInput)
    console.log("find", result);
  }
}

```

Результат надсилається користувачу, і користувач може надати команду з озвучення товарів у переліку.

3.5 Підготовка до експерименту

Перед початком покупки, користувач має налаштувати алгоритм під власне використання. Для цього передбачено окрему сторінку налаштування розширення.

Діаграма активності користувача та реакції системи в процесі налаштування роботи розширення, відображена на рисунку 3.4.

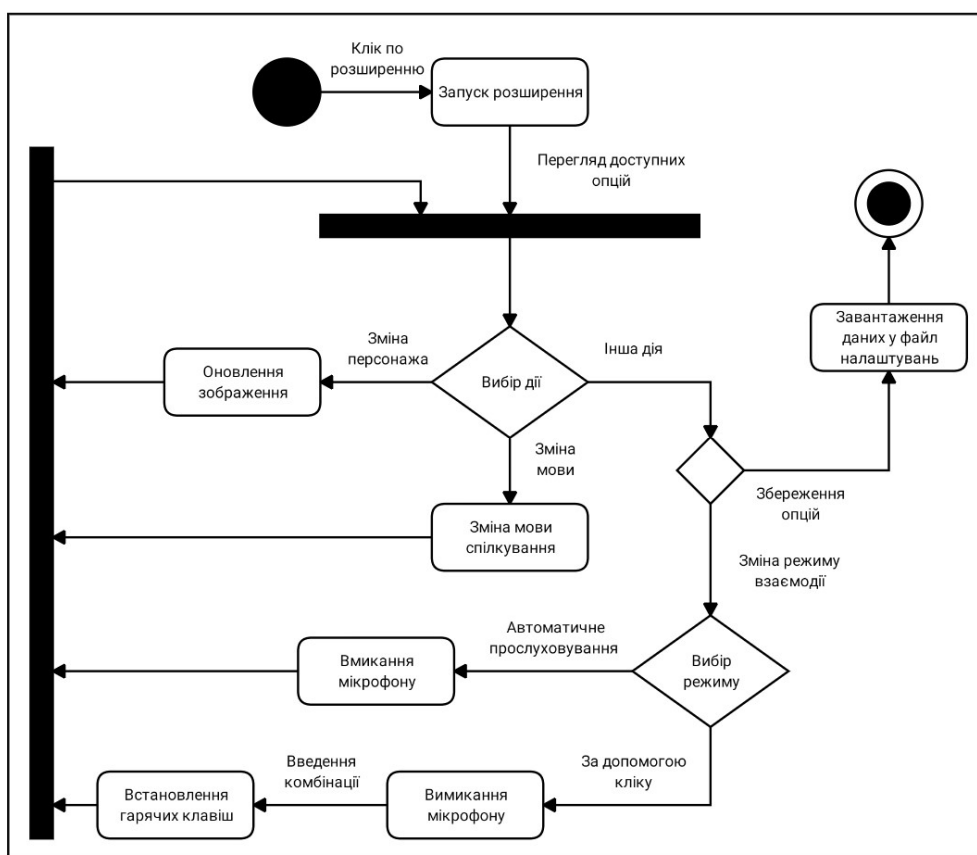


Рисунок 3.4 – Дії користувача перед початком налаштування розширення

Після завершення налаштування користувач починає роботу з алгоритмом. Після переходу на сторінку, з'являється зображення персонажа, який супроводжує користувача. За умовчанням обрано зображення кота.

Під час першого запуску розширення користувач надає доступ на використання мікрофону. Ця дія виконана до проведення експерименту та збережена для всіх подальших ітерацій експерименту.

Вимірювання часу починається від моменту, коли користувач перейшов на сторінку інтернет-магазину, яка повністю завантажується. Для зручності користувача, комбінація клавіш для активації голосового асистента була запрограмована за допомогою макросу клавіатури. Це дозволяє користувачу натискати одну клавішу замість трьох. Користувач натискає гарячу клавішу та промовляє команду «Search», чим активує голосове розпізнавання. Під час наступного натискання клавіші, користувач промовляє пошуковий запит, який будується на прикладі обраних даних. Нехай для прикладу це буде «Smartphone black 4 Gb RAM 64 Gb NFC».

Система обробляє запит користувача, трансформує аудіо в текст та заповнює поле пошуку та необхідні фільтри. В даному випадку, мають заповнюватися ті фільтри, які зазначені в пошуковому запиті. Зважаючи на те, що перегляд опису товару є однаковою процесом як під час заповнення з клавіатури, так і під час голосового керування, то ним можна знехтувати.

Наступним кроком користувач натискає гарячу клавішу для активації голосового пошуку і промовляє команду «Add to cart», щоб додати товар до кошика.

Подальші дії включають в себе перехід на сторінку оформлення товару, заповнення інформації про товар та переміщення між полями пошуку товару. В рамках експерименту передбачено, що користувач вже замовляв товари на цьому сайті, отже його облікові дані збережені та доступні для заповнення. Такий підхід дозволяє уникнути проблем з розпізнаванням мовлення, адже імена, прізвища, найменування вулиць та назви деяких міст не є словниковими словами. Користувач має кілька адрес доставки, тому йому пропонується зміна адреси доставки та заповнення номеру відділення або індексу. Процес оплати потребує введення даних банківської картки. В цьому випадку користувачу потрібно продиктувати цифри відповідно до полів, вказаних на сторінці.

Таким чином експеримент було підготовлено та проведено.

4 АНАЛІЗ РЕЗУЛЬТАТІВ ДОСЛІДЖЕННЯ

4.1 Оцінка результатів

За результатами 200 ітерацій для введення як з клавіатури, так і за допомогою голосового асистента, отримано дані, які можна представити у вигляді порівняльних графіків. Кожен графік ілюструє окремий етап дій, які виконували користувачі в процесі вибору своїх товарів. Графік завантаження сторінки без голосового асистенту наведено на рисунку 4.1.



Рисунок 4.1 – Завантаження сторінки без голосового асистента

Як видно з графіку, перше завантаження проходить найдовше, бо перед цим відбувалося очищення даних cookies, і браузеру необхідно більше часу для завантаження сторінки. В подальших ітераціях завантаження сторінки відбувається приблизно з однаковим часом.

Графік завантаження сторінки з голосовим асистентом вже не має таких великих просідань, і час завантаження сторінки вже коливається в межах від 2 секунд до 3,5 секунд (рис. 4.2).

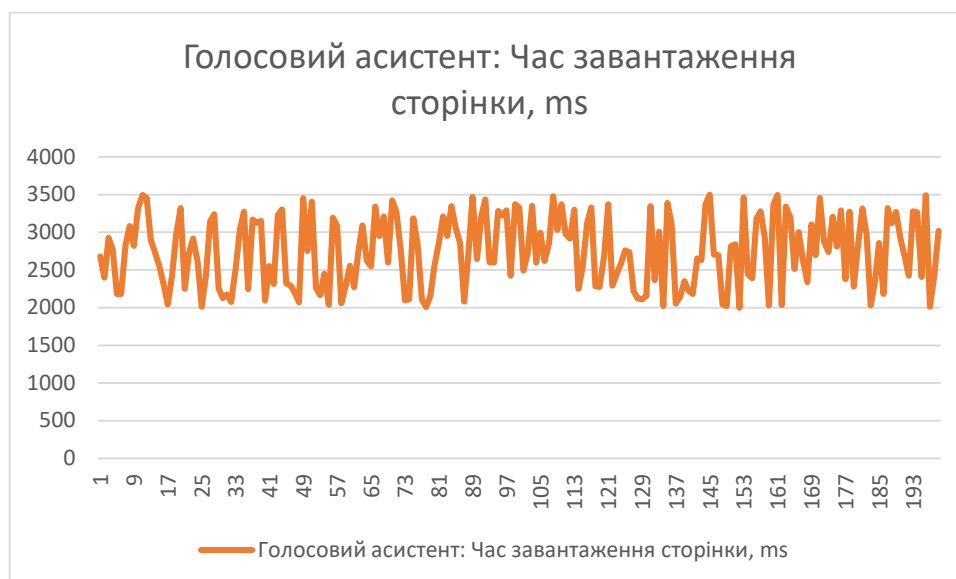


Рисунок 4.2 – Завантаження сторінки з голосовим асистентом

В цьому випадку грає роль швидкість Інтернету та завантаженість комп'ютера фоновими процесами системи, які впливають на час обробки сторінки та розширення.

Наступний графік відображує безпосередньо час пошуку товарів користувачами (рис. 4.3). Варто зазначити, що користувачі не були проінформовані про інші варіанти товарів, і шукали тільки той товар, який був призначений саме для них.

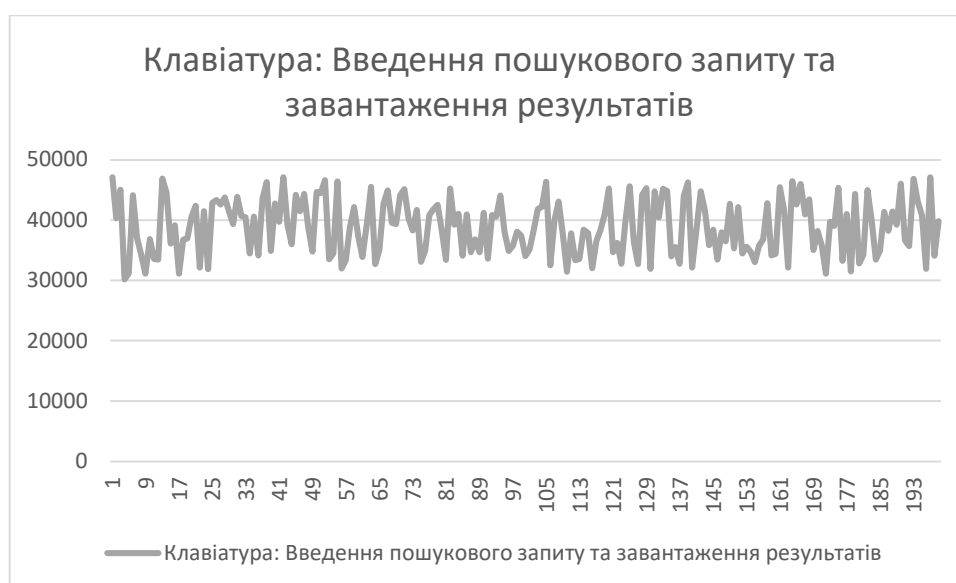


Рисунок 4.3 – Пошук товару без голосового асистента

Цей етап дуже сильно залежить від навичок користувача. Серед них як швидкість набору тексту, так і досвід роботи з інтернет-магазинами. Зважаючи на те, що користувачі бачать сторінку вперше, то й результати від 30 секунд на ознайомлення з її вмістом та друкування пошукового запиту це підтверджують.

У випадку з пошуком за допомогою голосового асистента, можна сказати, що користувачам також необхідно час на знайомство з інтерфейсом перш ніж переходити до пошуку (рис. 4.4).

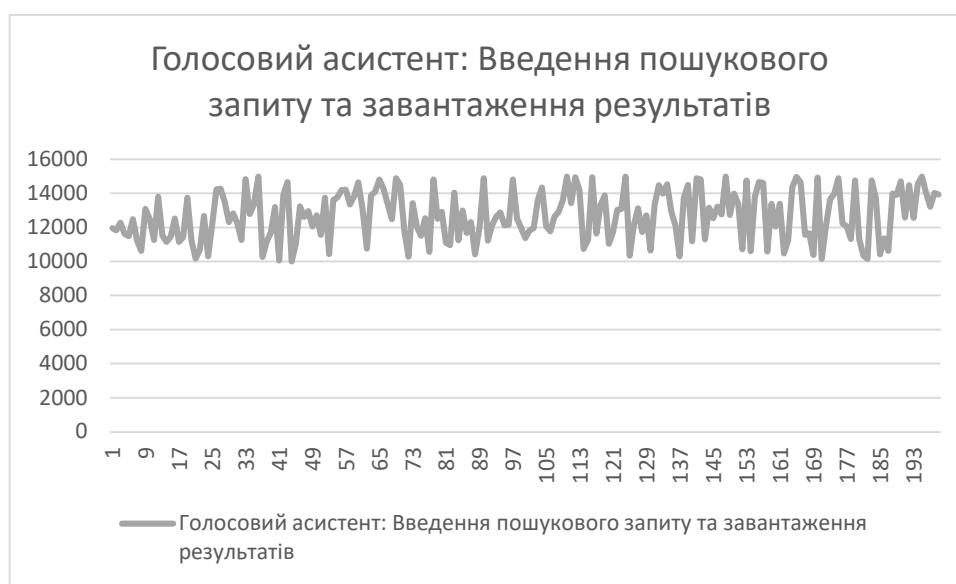


Рисунок 4.4 – Пошук товару з голосовим асистентом

Зважаючи на те, що користувач вже ознайомлений з роботою голосового асистента, він витрачає приблизно такий самий час для ознайомлення з вмістом веб-сторінки, але значно швидше вводить запит за допомогою голосу.

Додавання товару до корзини є інтуїтивно-зрозумілим, тому цей етап для користувача проходить досить швидко (рис. 4.5). Можуть бути певні відмінності з іншими інтернет-площадками, адже дизайн сторінок відрізняється. Зважаючи на те, що в цьому прототипі кнопка додавання товару реалізована лише на сторінці товару, то час оновлення сторінки та вибору товару теж враховано. В інших платформах кнопка додавання в корзину може бути реалізована безпосередньо на сторінці вибору товарів, і це може вплинути на результати.

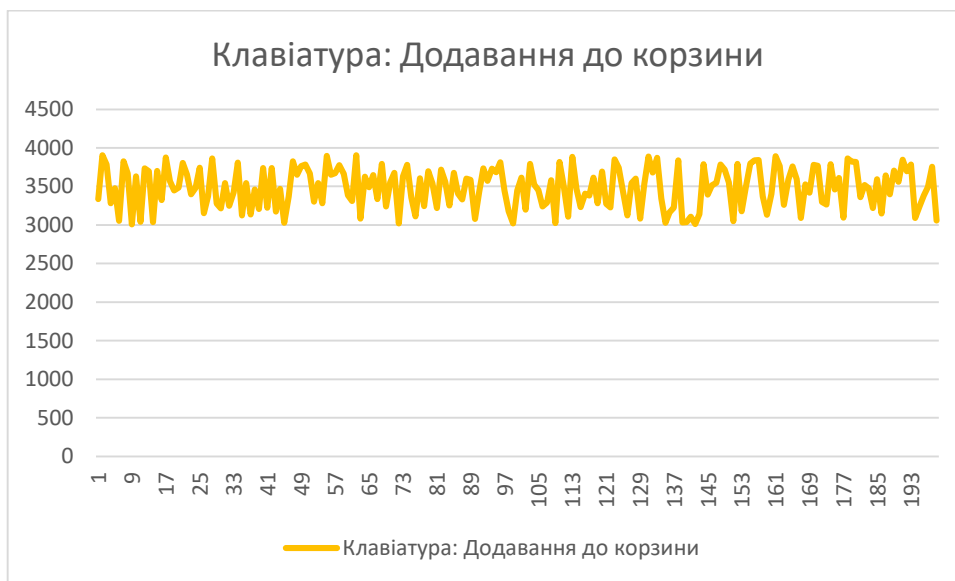


Рисунок 4.5 – Додавання товару до корзини без голосового асистента

Голосовий помічник на цьому етапі трохи програє введенню з клавіатури, адже потребує часу на перевантаження та обробку запиту користувача (рис. 4.6). Тим не менше, цей програш коливається в межах двох секунд, тому не є критичним.

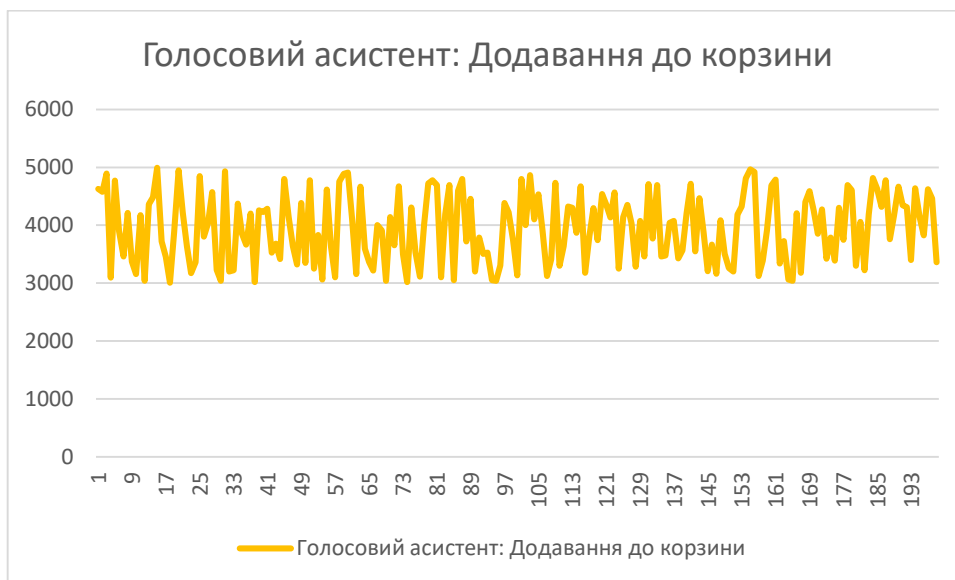


Рисунок 4.6 – Додавання товару до корзини з голосовим асистентом

Процес підтвердження купівлі товару потребує найбільше часу не зважаючи на те, що все максимально спрощено для користувача (рис. 4.7). Потрібно певний

час на авторизацію під час купівлі, адже після кожної ітерації експерименту відбувається вихід з облікового запису та ретельна перевірка введених даних.

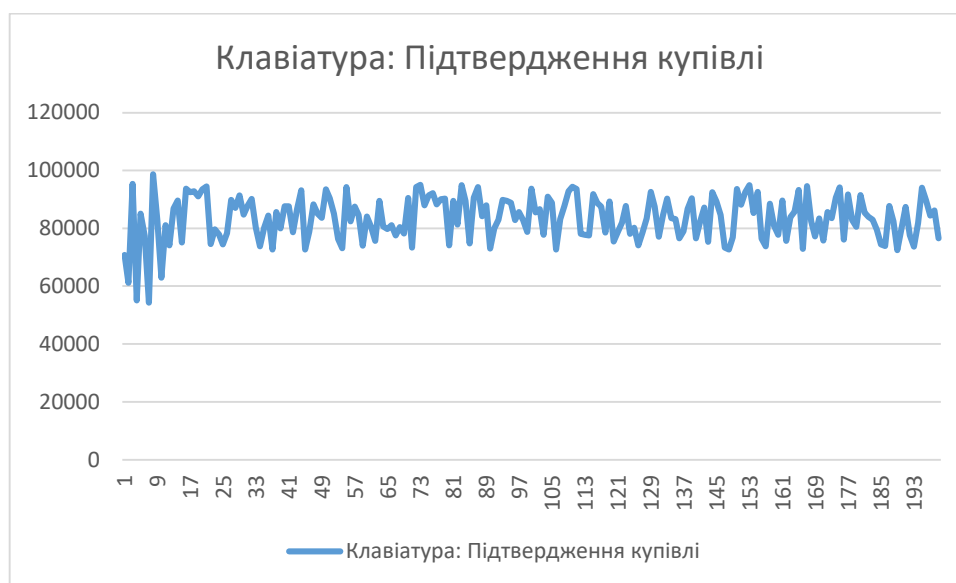


Рисунок 4.7 – Підтвердження купівлі без голосового асистента

Результати купівлі з голосовим асистентом трошки швидші (рис. 4.8). Суттєвий вигрaш йде на етапі роботи з числами, адже голосове розпізнавання працює з ним без проблем, і дозволяє користувачу диктувати цифри не відриваючи очей від картки.

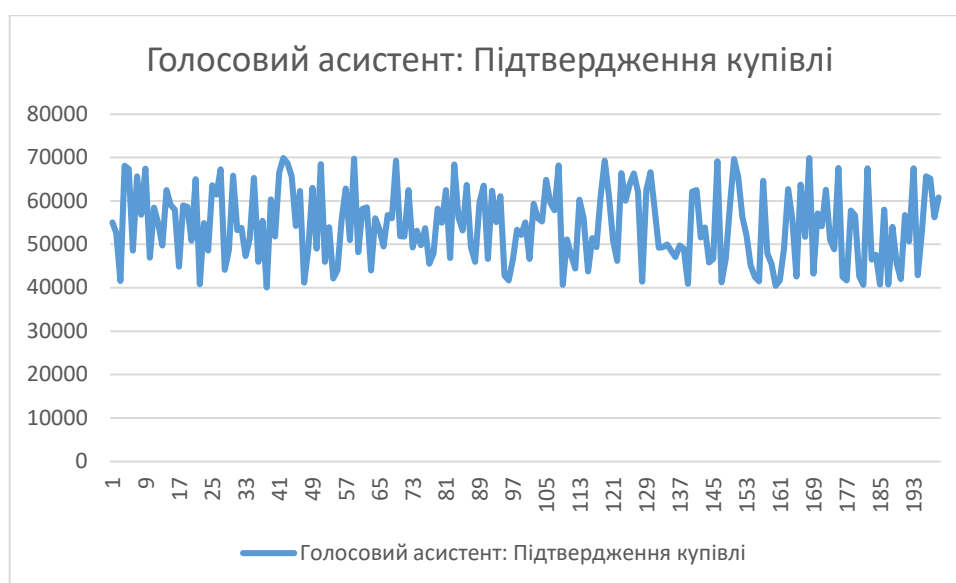


Рисунок 4.8 – Підтвердження купівлі з голосовим асистентом

Отже за проведеними дослідженнями можна сказати, що рішення суттєво пришвидшує передачу даних від користувача веб-сторінці, коли ведеться заповнення довгих рядків даних, а також при роботі з заповненням чисел. Тим не менше, алгоритм уповільнює процес швидких або дій одного кліку (наприклад, під час додавання товару до кошика). Також алгоритм не є працездатним при роботі з несловниковими даними.

В середньому розроблена система пришвидшує отримання даних користувачем на 30%.

Також система спрощує взаємодію користувача з веб-інтерфейсами, чим підвищує конверсію, адже користувач має можливість купити більше товарів за той самий час. Завдяки цьому з'являються умови для підвищення попиту на ринок голосових рішень. Покращення голосових рішень дозволить користувачам з особливими потребами спростити процес купівлі в інтернеті.

4.2 Напрямок подальшої роботи

Алгоритм дійсно є робочим рішенням, проте потребує вдосконалення і поки не може бути застосований в комерційних умовах. Потрібно провести роботу із покращення точності розпізнавання, а також впровадження рішення не лише у браузері Google Chrome, а також в інших браузерах.

Варто також зробити перевірку в інших операційних системах та інших комплектаціях комп'ютера, що значно вплине на швидкість реагування розробленого рішення. Користувач точно не буде задоволений, якщо розпізнавання буде працювати занадто повільно.

Крім того, варто дослідити ключові відмінності інтерфейсів інтернет-магазинів, для адаптації рішення під кожен з них. Розташування кнопок та елементів введення на сторінках сайтів впливає на сценарій поведінки користувача, і відповідно впливає на використання голосового асистента та запити до нього. Потрібно адаптувати рішення під врахування цих сценаріїв.

Важливою перепорою для впровадження також є відсутність підтримки української локалізації, що тимчасово не дозволяє впровадити рішення в простір українських інтернет-магазинів та інших рішень, зокрема в сфері освіти.

Також варто передбачити можливі нестандартні дії користувача, які можуть негативно вплинути не тільки на роботу алгоритму голосового супроводження, а також інтернет-магазину, на якому буде використовуватись алгоритм. Варто додати захист від спаму, а також перевірку на слова, які можуть нашкодити репутації магазину. Якщо врахувати, що рішення здатне заповнювати текстові поля, користувач може спробувати використати його для написання відгуку. В такому випадку, можна додати додатковий шар перевірки даних, які вводить користувач.

Враховуючи появу нових технологій та зростання популярності штучного інтелекту, можна налагодити взаємодію асистента з рішенням для озвучення додаткової інформації про обраний товар. Це дозволить користувачу дізнаватися інформацію про товар з інтернету без потреби переходу на іншу веб-сторінку тією мовою, яка йому зручна для розуміння. Крім того, це стане зручним доповненням для користувачів з особливими потребами, які потребують додаткових рішень для користування веб-сайтами.

ВИСНОВКИ

Під час виконання кваліфікаційної роботи були розглянуті популярні рішення взаємодії користувача та веб-сторінок із використанням голосових технологій. Було вивчено детальний алгоритм їх роботи та потенційні сфери використання на прикладі інтернет-магазинів. Після детального аналізу були знайдені суттєві недоліки конкурентів в існуючих рішеннях, які можуть бути виправлені створеним рішенням.

Було розроблено власний алгоритм та створено на його основі прототип, який в подальшому може стати робочим рішенням. Він може стати продовженням впровадження інтернету, що керується голосом. Після додаткового аналізу та детальної оцінки ризиків під час впровадження рішення, робота над реалізацією повноцінного продукту буде можливою.

Дослідження показало, що впровадження голосового асистента пришвидшує введення даних на сторінку інтернет-магазину на 30%, що дає користувачу можливість придбати більше товарів. Це в свою чергу підвищить конверсію. Програмно реалізувати голосового Асистента в тому вигляді, щоб він охоплював більшість популярних сайтів досить складно, адже необхідно враховувати особливості побудови сайтів. Тим не менше, навіть у вигляді прототипу він вже дозволяє спростити рутинні операції, з якими доводиться стикатися кожному користувачу в процесі купівлі товарів.

Застосування цього алгоритму на стаціонарних ПК потребує додаткового обладнання у вигляді мікрофону. Процес купівлі мікрофону навряд чи потішить користувача. Але найбільшою проблемою залишається точність розпізнавання та підтримка локалізації. Крім незнання користувачами мови, з якою може взаємодіяти сервіс, серйознішою проблемою є відсутність підтримки цих мов веб-сайтами. Якщо не буде підтримуватися необхідна мова, то інструмент не буде користуватися попитом в країні, де планується його впровадження. Поширення мовних пакетів голосового розпізнавання вирішить цю проблему.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Top 9 software investment trends to watch in 2023, <https://www.gartner.com/en/digital-markets/insights/top-9-software-investment-trends-to-watch-in-2023> (дата звернення: 01.04.2023);
2. Постковідний світ та технології: як пандемія вплинула на стартап-екосистему | Економічна правда – URL: <https://www.epravda.com.ua/columns/2021/01/19/670136/> (дата звернення: 27.01.2021);
3. Створення інтернет-магазину: Підводні камені й етапи розробки – URL: <https://glyanec.net/ua/blog/stvorennya-internet-mahazynu-pidvodni-kameni-y-etapy-rozrobky> (дата звернення 29.01.2023);
4. Терещенко Е. Ю. Розвиток інтернет-торгівлі в умовах сучасного бізнес-середовища. Ефективна економіка. 2019. № 9;
5. Типи сайтів. Частина 3. Інтернет-магазин | Web-Lighthouse – URL: <https://web-lighthouse.com/news/типи-сайтів-частина-3-інтернет-магазин/> (дата звернення: 01.02.2023);
6. Розпізнавання мови. – URL: <https://www.cybermova.com/speech/розпізнавання-мови.html> (дата звернення: 02.02.2023).
7. THE REMARKETING REPORT – URL: <https://d34w0339mx0ifp.cloudfront.net/global/downloads/ig/2016-The%20Remarketing%20Report-Q3.pdf> (дата звернення: 20.02.2023);
8. Розширення для браузерів. Їх типи, користь, застосування. – URL: <https://evergreens.com.ua/ua/articles/browser-extensions.html> (дата звернення: 13.03.2023);
9. Shopynskiy M., Golian N., Afanasieva I. Long short-term memory model appliance for generating music compositions. 2020 IEEE international conference on problems of info-communications. Science and technology, IEEE (2020), pp. 239-242.

10. Speech recognition and voice commands for your site | by Voxpow | Voice Tech Podcast | Medium – <https://medium.com/voice-tech-podcast/speech-recognition-and-voice-commands-for-your-site-e1931d60ba03> (дата звернення: 25.03.2023);
11. Speech Recognition and Voice Commands for websites — Voxpow.com – URL: Speech Recognition and Voice Commands for websites — <https://voxpow.com/> (дата звернення 29.03.2023);
12. Handsfree for Web – Voice Control – Browse using your voice – URL: <https://www.handsfreeforweb.com/en/> (дата звернення: 04.04.2023);
13. SpeechRecognition – Web APIs | MDN – URL: <https://developer.mozilla.org/en-US/docs/Web/API/SpeechRecognition> (дата звернення: 07.04.2023);
14. Найпопулярніші браузери, якими користувалися українці у 2020 році – Pingvin.Pro – URL: <https://pingvin.pro/gadgets/article-gadget/najpopulyarnishi-brauzery-yakymy-korystuvalysya-ukrayinczi-u-2020-roczy.html> (дата звернення: 18.04.2023).
15. Mandeep Singh Kandhari, Farhana Zulkernie, Haruna Isaah. A Voice Controlled E-Commerce Web Application – URL: <https://arxiv.org/ftp/arxiv/papers/1811/1811.09688.pdf> (дата звернення 22.04.2023).
16. Amelia Silva, Maria Jose, Sandrina Teixeira, Anabela Martins Silva, Telma Maia. Accountability & Web Accessibility of the Portuguese official municipal Websites. Vision 2020: Innovation Management, Development Sustainability, and Competitive Economic Growth.
17. Kelly, B., Sloan, D., Brown, S., Seale, J., Lauke, P., Ball, S. & Smith, S., 2009. Accessibility 2.0: Next Steps For Web Accessibility. Journal of Access Services, 6 (1 & 2), pp. 265-294.
18. Несміян Дмитро. Атестаційна робота: Веб-портал для взаємодії різних категорій учасників навчального процесу. ХНУРЕ, Харків, 2020 – 80 с.