

# РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ВЕБ СЕРВІСУ ГОЛОСОВИЙ ПОМІЧНИК

Вінницький національний технічний університет

## Анотація

*Розглянуто актуальність задачі розпізнавання голосу. Розглянуто існуючі методи розв'язання поставленої задачі, та запропоновано використання задач розпізнавання голосу для її вирішення.*

**Ключові слова:** голосовий помічник, web сервіс, нейронні мережі, згортоква нейронна мережа.

## Abstract

*Actuality of voice recognition problem is considered. The existing methods of solving the set task are considered, and the use of voice recognition problems for its solution is proposed.*

**Keywords:** voice assistant, web service, neural networks, convolutional neural network.

## Вступ

На сьогоднішній день важко уявити наше життя без сучасних технологій. Смартфони, комп'ютери, ноутбуки – все це стало невідемним у нашій буденності. Але вони були б даремні без типового програмного оснащення. Сьогодні, за допомогою різноманітного програмного оснащення ми можемо здійснювати управління іншими нашими гаджетами та навіть будинками та автівками. Але чи не здається вам, що типове введення інформації за допомогою введення вручну слів дещо застаріле? Все більшого тренду набуває процес введення тієї чи іншої інформації з меншим числом взаємодії із девайсом. Одним із таких напрямків є голосове введення. Однак, без спеціального програмного оснащення такий тип введення інформації є неможливим.

В сучасних операційних системах, що розповсюджені на смартфонах – Android та iOS є вбудовані голосові помічники, які непогано працюють, але вони доступні тільки на девайсах саме з цими ОС. Типове використання комп'ютерів та мобільних девайсів у сьогоднішній та в майбутньому пов'язані із використанням доступу до мережі інтернет, в якій подібного софту до голосового помічника – немає.

## Підхід щодо підвищення точності розпізнавання

Побудова та оптимізація систем голосового помічника є актуальним та перспективним напрямком, оскільки дозволяє робити взаємодію користувача та пристрою більш зручними, збільшувати популярність, відвідуваність та відсоток повторного повернення користувачів до додатку.

Використання голосового помічника залишається за користувачем, він сам вирішує, як йому зручніше виконувати ті чи інші дії, стандартним, типовим способом чи за допомогою голосових команд. Це породжує два підходи для отримання вищевказаної інформації.

Сьогодні оператори стільникового зв'язку пропонують широкий набір сервісних додатків і послуг для задоволення потреб абонентів. Для підтримки конкурентоспроможності кожен оператор намагається зробити користування такими послугами якомога зручнішим для своїх абонентів.

Щоб їх використовувати, абонентам оператора стільникового зв'язку потрібно запам'ятовувати конкретні запити чи абонентські номери, що створює для них істотні незручності. Найчастіше абоненти зберігають необхідні дані у телефонній книзі з певною асоційованою назвою (наприклад, «погода», «стан рахунку», «тарифний план»), за якою згодом здійснюють пошук та виконують запит. Однак багато запитів до операторів є складеними, тобто містять кілька комбінацій цифр, які необхідно ввести абоненту для того, щоб виконати певний запит. Наприклад, запит \*124\*XXX\*380XXXXXXX# (сума) (номер абонента) уможливіло переказати гроші з одного мобільного рахунку абонента оператора Київстар на інший [8]. На такі запити абоненти витрачають

від 5 секунд до кількох хвилин. Крім того, запити для ідентичних послуг різних операторів стільникового зв'язку України різняться між собою.

Google Voice Search.

Розроблене ПЗ дасть змогу скоротити трудовитрати, збільшити продуктивність систем самообслуговування, покращити зручність та швидкість використання сервісів та послуг.

Керування ресурсом за допомогою голосу не є новою. Проте загальнодоступні інструменти (керування ПК, введення пошукових слів) доволі обмежені в області своїх можливостей. Так, наприклад керування комп'ютером (стандартна утиліта Windows) дозволяє пересуватись файловою системою і запускати програми, але не більше того. Siri - персональний помічник і питально-відповідна система, адаптована для iOS. Цей додаток використовує обробку природної мови, щоб відповідати на питання і давати рекомендації. Siri пристосовується до кожного користувача індивідуально, вивчаючи його переваги протягом довгого часу, проте є власністю Apple і доступний лише в смартфонах і планшетах цієї компанії. Google voice service – сервіс компанії Google що дозволяє перетворювати голосовий цифровий сигнал в текст, недоліком є висока чутливість (для прийняттого результату необхідна чітка дикція і розмірений темп голосу) На даний момент не існує інструменту для здійснення навігації веб ресурсами за допомогою голосу, хоча веб стає основним середовищем користувача. Пропонується створити інструментарій для веб навігації за допомогою голосу. Практично всі веб сторінки відображаються за допомогою HTML коду і клієнтських сценаріїв на javascript. Це дозволяє отримати список елементів навігації і взаємодії з користувачем використовуючи стандартні патерни пошуку елементів в DOM дереві за тегами, а також перевіркою всіх тригерів (`—listener`) javascript в користувацьких елементах, оскільки їх побудова також має певні прийняті стандарти. Однією з найважливіших проблем при розпізнаванні голосу є визначення імовірності помилки у визначенні букви в контексті слова чи речення. Одним з варіантів є побудова самодоповнюючої бібліотеки слів і їх контекстів, та визначення порогового значення помилки (ідентифікація помилки в розрізі можливої нової послідовності букв)[1]. Оскільки побудова такої бібліотеки (чи аналогічного інструменту) являє собою задачу високої складності, тому вирішено використати вже існуючий сервіс перетворення голосового сигналу в текстове повідомлення. Роботу інструменту можна декомпонувати на декілька модулів, кожен з яких відповідатиме за певну частину функціональності (як і показано на рисунку 1). Дана концепція відображає модифіковану загальноприйнятую трьох ланкову конструкцію веб застосунків (клієнт – сервер - СКБД) [2]. Модифікація в тому, що при відсутності необхідного значення в базі даних - серверна сторона звертається до Google voice service як до фабрики об'єктів зв'язок (бінарна послідовність – текстова репрезентація) після чого проходить ще один цикл визначення команди (описано нижче).

При завантаженні Веб сторінки Javascript module знаходить і зв'язує всі елементи навігації та вводу на сторінці, перевіряє Local storage і у випадку актуальності даних зв'язує елементи з відповідними командами, в іншому випадку робить запит до веб – сервісу і завантажує вищеописані дані в Local storage для даного сайту конкретної сторінки [4]. Об'єкт listener працює в паралельному потоці, запускаючи процес обробки команди, яку вимовить користувач після ключового слова (для зменшення імовірності помилкової реакції) При отриманні команди об'єкт listener зчитує бінарну послідовність , що є репрезентацією акустичного сигналу, який поступає на вхід користувацького пристрою вводу. Послідовність передається у внутрішній об'єкт – proceeder. В proceeder передана послідовність порівнюється з уже існуючими з урахуванням певного відхилення і вираховує вірогідність співпадиння (потрапляння кожного біта у відповідний допустимий коридор значень) введеної команди з уже існуючими. При знаходженні відповідної команди (вірогідність співпадиння вища порогового значення) - виконується команда, зв'язана з даною бінарною послідовністю (перехід по зв'язаному навігаційному елементі чи активація іншого активного елемента, тощо). При відсутності відповідного значення аналізатор передає дану послідовність на обробку менеджера даних (міститься в javascript модулі). Javascript module даних відправляє асинхронний запит на веб сервіс для обробки.

На веб сервісі дана послідовність передається на Google voice service і у відповідь отримується текстова репрезентація послідовності, що в свою чергу передається на клієнт. На клієнті Javascript module даних зв'язує дану послідовність з командою, що асоціюється з отриманим текстовим значенням (Наприклад, на сторінці присутнє посилання «Новини», тоді голосова команда «новини» буде відкривати дане посилання), і відсилає знайдену зв'язку на сервер, де вона записується в БД для даного сайту і сторінки. Навігатор виконує необхідну дію, що прив'язана до вхідної послідовності. Такий підхід має декілька до кінця неопрацьованих пунктів і обмежень. Серед яких: робота навігатора передбачає використання HTML 5 сумісного браузеру; На даний момент не визначений коефіцієнт допустимого відхилення при порівнянні бінарних послідовностей; Робота

навігатора сильно залежить від якості зчитувального пристрою користувача і чіткості його мовлення. Аналіз роботи створених прототипів показав, що дані про команди для веб сторінок слід категорувати не лише за конкретними сторінками, а і профілювати за місцем розташування та загальними показниками голосу (за одним комп'ютером можуть працювати декілька осіб, зазвичай ці особи не міняються (наприклад в сім'ї), тому можна визначити значення відхилення для кожного користувача), таким чином, дозволяючи визначити найоптимальніший рівень відхилення для кожного користувача, що в свою чергу забезпечить підвищення якості роботи інструменту. Такий підхід дозволить уніфікувати процес визначення відхилення незалежно від мовної приналежності чи особливостей мовлення кожного окремого користувача. Інструмент буде являти собою клієнт – серверний додаток (клієнт - javascript модуль що може бути доданий до будь-якої веб сторінки) з використанням асинхронних запитів і паралельного потоку HTML5, для мінімізації впливу клієнтського коду на загальну поведінку веб сторінки [5]. На даний момент інструмент реалізований в розділених прототипах, не зв'язаних між собою. Проте планується побудова першого повнофункціонального прототипу, і як тестове середовище планується використати портал Національного Антарктичного центру, оскільки ним користуються різні категорії населення (за віковими і мовними ознаками), а також присутність великої кількості послань та елементів управління, робота яких може бути переведена на голосове управління.

### Висновки

Розглянуто типи опрацювання та синтезу голосу людини. Проаналізовано технології голосових систем.

Розглянуто класифікації та математичні моделі алгоритмів та структур даних, що використовуються при роботі з голосовими командами на базі яких будуються голосові помічники.

Проведено комбінування моделей оскільки голосовий помічник – це комплексне рішення, яке потребує складніших структур даних ніж базові завдання.

Обґрунтовано вибір моделі визначення голосових команд голосовим помічником. Наведено переваги і недоліки нечіткої логіки.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Обробка зображень, розпізнавання образів. [Електронний ресурс]. – Режим доступу: <http://ua.textreferat.com/referat-8342-1.html>
2. Viola, P. Robust Real-Time Object Detection: tech. report / P. Viola, M. Jones. – (city: Cambridge) Cambridge, 2001. – 320 p. 2. Как работает детектирование лиц [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://www.digital-sky.ru/point-3/artcateg-17/article-10.html>, свободный – Загл. с экрана.
3. opencv-extension-library [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://opencv-extensionlibrary.googlecode.com/svn/trunk/QtOpenCV/example/facedetect/facedetect.c>, свободный – Загл. с экрана.
4. Graves, Alex; and Schmidhuber, Jürgen; Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks, in Bengio, Yoshua; Schuurmans, Dale; Lafferty, John; Williams, Chris K. I.; and Culotta, Aron (eds.), Advances in Neural Information Processing Systems 22 (NIPS'22), December 7th–10th, 2009, Vancouver, BC, Neural Information Processing Systems (NIPS) Foundation, 2009, pp. 545–552.

**Чопко Андрій Ігорович** — студент групи 1КН-17м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: [ripiuk96@gmail.com](mailto:ripiuk96@gmail.com)

Науковий керівник – **Озеранський Володимир Сергійович** — ст.викл. кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

**Andrii I. Chopko** – student of Information Technologies and Computer Engineering Department, 1CS-17m, Vinnytsia National Technical University, Vinnytsia, e-mail: [ripiuk96@gmail.com](mailto:ripiuk96@gmail.com)

Supervisor – **Ozeransky S. Volodymyr** — senior teacher Vinnytsia National Technical University, Vinnytsia.