

УДК 004.942

Павліченко Данило Олегович

студент Інституту комп'ютерних систем

Одеського національного політехнічного університету

Павличенко Даниил Олегович

студент Института компьютерных систем

Одесского национального политехнического университета

Pavlichenko Daniil

Student of the Institute of Computer Systems

Odessa National Polytechnic University

Науковий керівник:

Мартинюк Олександр Миколайович

кандидат технічних наук, доцент кафедри КІСМ

Одеський національний політехнічний університет

**ДОСЛІДЖЕННЯ МОДЕЛЕЙ ТА МЕТОДІВ РОЗПІЗНАВАННЯ МОВИ
У ВЕБ-СЕРВІСІ ДЛЯ ОРГАНІЗАЦІЇ КОРИСТУВАЦЬКОЇ
ІНФОРМАЦІЇ
ИССЛЕДОВАНИЕ МОДЕЛЕЙ И МЕТОДОВ РАСПОЗНАВАНИЯ
РЕЧИ В ВЕБ-СЕРВИСЕ ДЛЯ ОРГАНИЗАЦИИ
ПОЛЬЗОВАТЕЛЬСКОЙ ИНФОРМАЦИИ
RESEARCH OF MODELS AND METHODS OF SPEECH
RECOGNITION IN A WEB SERVICE FOR ORGANIZATION OF USER
INFORMATION**

Анотація. Розглядається побудова веб-сервісу для організації користувацької інформації із можливістю голосового вводу, розпізнавання мови для надання більш зручнішої функції збереження даних та подальшого

використання їх. Були побудовані схема алгоритму голосового вводу та синтезу звуків у текст, схема основних етапів та методів розпізнавання мови, таких як метод на основі скритих Марковських моделей, метод на основі динамічного програмування, побудови нейронних мереж та метод на основі словникового пошуку. Останній був взятий за основу при розробці веб-сервісу, а модель отримала вдосконалення – введено додатковий словник, який будується та заповнюється динамічним шляхом. Даний підхід дозволяє прискорити пошук по мовним словникам за рахунок того, що тепер у запита є можливість оброблюватись за двома існуючими гілками, замість однієї. Також у дослідженні розглядається, основні аналоги розробленого веб-сервісу, їх недоліки та переваги. Представлено фрагмент побудованої мережі Петрі.

Ключові слова: web-сервіс, системи розпізнавання мови, клієнт, сервер, сховище даних, мережа Петрі, граф досяжних маркувань.

Аннотація. Рассматривается построение веб-сервиса для организации пользовательской информации с возможностью голосового ввода, распознавания речи для предоставления более удобной функции хранения данных и дальнейшего использования их. Были построены схема алгоритма голосового ввода и синтеза звуков в текст, схема основных этапов и методов распознавания речи, таких как метод на основе скрытых Марковских моделей, метод на основе динамического программирования, построения нейронных сетей и метод на основе словарного поиска. Последний был взят за основу при разработке веб-сервиса, а модель получила совершенствования - введено дополнительный словарь, который строится и наполняется динамичным путем. Данный подход позволяет ускорить поиск по языковым словарям за счет того, что теперь в запроса есть возможность обрабатываться по двум существующим ветвями, вместо одной. Также в исследовании рассматривается, основные аналоги

разработанного веб-сервиса, их недостатки и преимущества. Представлен фрагмент построенной сети Петри.

Ключевые слова: веб-сервис, системы распознавания речи, клиент, сервер, хранилище данных, сеть Петри, граф достижимых маркировок.

Summary. The construction of a web service for organizing user information with the possibility of voice input, speech recognition for more convenient data storage and further use is considered. The study also examines the construction of models and methods of speech recognition systems, the main analogues of the developed web service, their shortcomings and advantages. Among the disadvantages were identified inconvenient and incomprehensible interface, the inability to structure the information into categories, the system of voice input, information retrieval and synthesis into text is not developed. The general structural scheme of the service was built, which shows the connection of client, server and data warehouse, the scheme of algorithm of voice input and synthesis of sounds into text, the scheme of the main stages and methods of speech recognition, such as the method based on the hidden Markov models, the method based on dynamic programming, the construction of neural networks and the method based on dictionary search. The latter was taken as a basis for web service development. The web service is divided into three main parts - client, server and data warehouse. The user is also able to store not only text information, but also audio and video files, personal images and pictures. Another advantage of the developed service is the ability to structure information by categories and notebooks, and an intelligent and convenient text editor. There is also the construction of a Petri network and a graph of achievable markings. Petri network shows data exchange between client, server parts and data storage, time of data processing and thanks to statistics shows correct work of the system as a whole. Statistics shows how much data in the form of queries was processed by the system, minimum, maximum, average number of queries, and time of

processing them. On its basis, the decision is made to refine the service, i.e. to simplify it in order to speed up its work.

Key words: *web-service, language recognition systems, client, server, data storage Petri net, graph of achievable markings.*

Постановка проблеми. Завданням розуміння в системах діалогу є переклад послідовності слів, отриманої з виходу розпізнавача слів, в смислове уявлення. Складність завдання розуміння пов'язана з великою невизначеністю в розпізнаній послідовності слів. Причинами спотворень можуть бути: помилки розпізнавання слів, акустичний шум, сторонні звуки, придих, плямкання губами, кашель, озвучені паузи (eee ..., ммм ..., і т.п.), повторення, переривання, пропуски слів, додаткові слова і т. д. Таким чином, основним завданням розуміння мови є нейтралізація перерахованих вище відхилень.

Розроблюваний веб-сервіс надає можливості зберігати користувацьку інформацію, та при необхідності редагувати, видаляти її. Для кожного користувача створюється особистий персональний запис у системі для того, щоб інші користувачі не мали доступу до чужої інформації. Прикладом користувацької інформації можуть бути важливі записи, замітки до якоїсь справи, фотографії, відео. Всі дані у розроблюваному сервісі зберігаються у вигляді заміток, які можна формувати у відповідних блокнотах. А інформацію можна заповняти за допомогою розпізнавання мови та синтезу у текст [1, с.136].

Аналіз останніх досліджень та публікацій. Серед основних аналогів розроблюваного веб-сервісу можна виділити наступні:

- Сервіс «Evernote»;
- сервіс «OneNote»;
- сервіс «Google Keep»;

Серед головних плюсів першого сервісу можна виділити можливість створювати нотатки як друкованим текстом, так і рукописним. Будь-яку інформацію - як графічну, так і текстову - можна зберегти в Evernote скріншотом. При безкоштовному використанні Evernote місце під зберігання заміток не може перевищувати 60 Мб на місяць. Плюс до цього, кожне вкладення в замітку - документ, картинка, відео і т.д. - не може перевищувати 25 Мб.

Одним із головних мінусів цього сервісу можна вважати малий об'єм місця збереження заміток. Також до мінусів можна віднести відносно високу вартість платних функцій сервісу.

До плюсів другого сервісу можна віднести:

- по-перше, це інтеграція (причому двостороння) з MS Office. Наприклад, в разі Outlook можна як відправляти листи в OneNote, так і передавати завдання безпосередньо із записної книжки в календар Outlook, а в разі Word - з'єднувати з документами;
- по-друге, є можливість відправляти замітки на адресу `me@onenote.com`, і вони автоматично будуть додані в записну книжку.
- по-третє, інтеграція з різними сканерами, як «залізними» (Epson, наприклад), так і «софтверними» (Genius Scan), а також зовсім екзотичними, типу «розумних» ручок. По-четверте, після установки OneNote, в «Пристрої та принтери» з'являється новий принтер «Надіслати до OneNote», після чого будь-який додаток, яке вміє друкувати, може відправляти інформацію в записну книжку.

Серед недоліків можна виділити неможливість спільної роботи, відсутність синхронізації між різними пристроями (за винятком пристроїв на базі Windows Mobile), ризик втрати даних у разі неочікуваної поломки комп'ютеру і т.п. [4, с. 345].

Серед основних плюсів третього сервісу можна виділити:

- загальний доступ до даних. Можна запланувати сімейне свято або якийсь проект. Друзям і колегам буде доступна вся інформація, внесена в замітку;
- прискорений пошук і виділення записів кольором допоможуть швидко підібрати необхідні записи;
- синхронізація між пристроями дозволить працювати з нотатками у будь-який час і будь-якому місці;

Але, не дивлячись на це, Google Keep являється частиною «екосистеми» Microsoft, що є одним із самих великих плюсів і одночасно з цього витікає багато мінусів:

- сервіс стає недоступним при оновленні сервісів Google;
- так як сервіс являється розробкою компанії Google, до прохань та пропозицій користувачів у більшості випадків навіть не звертають уваги; користувачі всеодно не зможуть відмовитись від використання продуктів цієї компанії, так як вони являються частиною системи Google.

Відокремлення невирішених раніше частин загальної проблеми.

На даний момент існує багато сервісів та веб-додатків, котрі надають змогу зберігати користувацьку інформацію, але є також один вагомий мінус – людина не завжди має змоги зупинитись та вводити інформацію «від руки». Тому питання створення можливості вводу та зберігання інформації у голосовому вигляді та синтезі її у текст є відкритим і наразі [2, с.125].

Мета статті. Розробка структури веб-сервісу із застосуванням нового методу розпізнавання мови – на основі динамічного створення додаткового мовного словника, який дозволить прискорити швидкість роботи розпізнавання, та проведення її аналізу.

Викладення основного матеріалу. На рисунку 1 показані основні етапи та методи розпізнавання мови. До них відноситься аналіз Фур'є, голосовий аналіз, метод побудови скритих Марковських моделей, побудова

нейронних мереж, динамічне програмування, та метод на основі словникового аналізу [3, с. 250].

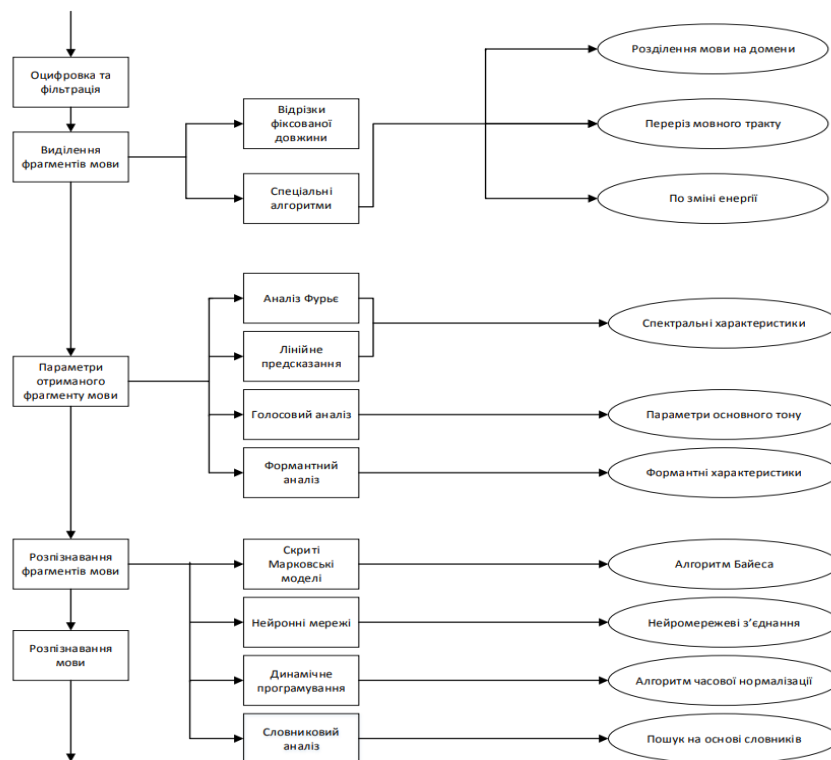


Рис. 1. Методи розпізнавання мови

Джерело: розробка автора

Розроблювана мережа Петрі поділяється на наступні основні частини:

- генератор та клієнтська, яка генерує 7 типів запитів (реєстрація, автентифікація, створення блокнотів та заміток, голосовий ввід, розпізнавання мови та синтез у текст, виправлення даних, пошук);
- серверна, у якій проводиться обробка всіх типів запитів;
- сховище даних, де зберігаються дані;
- рендерна, яка відображає на клієнтській частині веб-сервісу потрібну інформацію;
- статистична, у якій показується середній час обробки даних, їх мінімальна та максимальна кількість та сумарний час обробки.

На рисунку 2 наведено фрагмент мережі Петрі, що відповідає основним етапам розпізнавання мови на основі методу використання мовних словників.

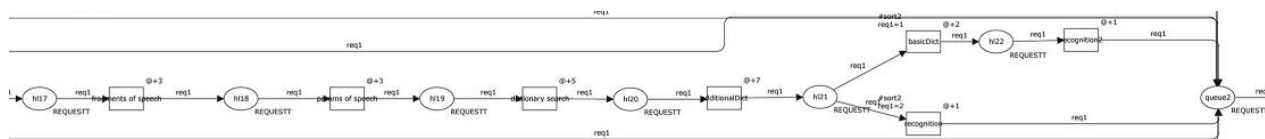


Рис. 2. Система розпізнавання мови у вигляді мережі Петрі

Джерело: розробка автора

Фрагмент, представлений вище показує такі етапи розпізнавання мови, як Фільтрація звукового сигналу, виділення фрагментів та параметрів мови, на основі яких проводиться словниковий пошук. Пошук здійснюється за допомогою основного словника та додаткового, який створюється динамічно на основі часто використовуваних слів.

Удосконалення заключається у введенні позицій $P = \{hl_{17...22}\}$ й переходів:

$$T = \{fragments\ of\ speech, params\ of\ speech, dictionary\ search, additionalDict, basicDict, recognition, recognition2\}$$

та зв'язків:

$$E_1 = \{(hl_{15} \times encryption), (hl_{16} \times coincidence), (hl_{17} \times fragments\ of\ speech), (hl_{18} \times params\ of\ speech), (hl_{19} \times dictionary\ search), (hl_{20} \times additionalDict), (hl_{21} \times basicDict), (hl_{21} \times recognition1), (hl_{22} \times recognition2)\}$$

й:

$$E_2 = \{(data\ processing_{1...2} \times hl_{15...16})\}.$$

Нові зв'язки дають змогу оброблюватись запиту на розпізнавання мови швидше, ніж за допомогою одного мовного словника. Такий підхід дає перевагу у часі обробці перед аналогічними сервісами.

Процес голосового вводу показаний нижче на рис.3. Для того, щоб почати процес голосового вводу, користувач має натиснути кнопку, яка

вмикає мікрофон. У цей момент починається процес розпізнання звуків. Кожен момент часу звук відправляється на сервер для обробки розпізнання звуків, де на сервері відбувається процес розпізнання звуків та синтез у текст.

У випадку, якщо користувач мовчить або на сервері не було розпізнано звуки – процес голосового вводу вважається невдалим, та користувач має знову повторити процес вводу [5 с. 48].

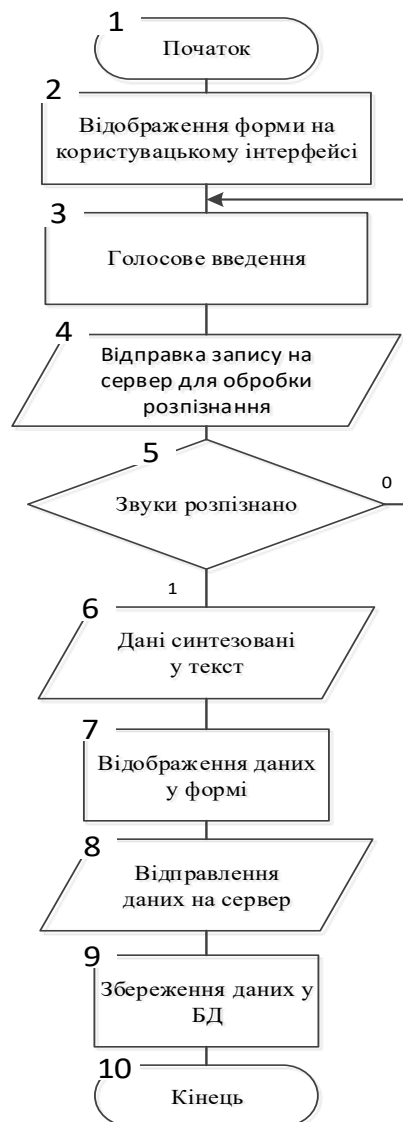


Рис. 3. Схема алгоритму голосового вводу та синтезу звуків у текст

Джерело: розробка автора

У результаті успішної обробки, сервер відправляє відповідь у вигляді результату, синтезованого у текст. На протязі голосового вводу, сервер

відправляє проміжні результати обробки. Фінальний результат відправляється, коли користувач завершає процес голосового вводу.

Наступним етапом дані, синтезовані у текст відправляються на клієнтську частину сервісу, після чого відображаються на користувацькому інтерфейсі у формі для вводу даних. Після чого дані з клієнтської частини сервісу відправляються на сервер, де проходять подальшу обробку. Сервер у свою чергу посилає запит до бази даних на збереження даних [5, с. 342]. У випадку вдалої операції збереження даних, база даних відправляє відповідь про це до сервера.

На цьому процес голосового вводу та синтезу звуків у текст вважається завершеним.

Нижче, на рисунку 4 представлені графіки залежності середнього часу обробки запитів від їх кількості для семи типів запитів. Моделювання проводилося п'ять разів із кроками від 10000 до 50000 запитів.



Рис. 4. Графіки залежності середнього часу обробки запитів від їх кількості

Джерело: розробка автора

З графіків вище видно, що найбільше всього запитів поступало по сьомому ланцюгу, який відповідає пошуку у веб-сервісі. Час обробки цього запиту завжди мінімальний, бо несе найменше навантаження на сервіс. Один із найскладніших у обробці являється запит на розпізнавання мови та синтез мови у текст, бо складається з вводу голосової інформації, очистки та

фільтрації від шумів, виділення частин мови, розпізнання мови на основі словникового пошуку, синтезу у текст. Також це видно із часу обробки даного типу запиту.

Висновки і пропозиції. Розроблюваний web-сервіс повинен надавати користувачам можливість спрощувати процес збереження та структурування особистої інформації. Основними функціями являються реєстрація, автентифікація, створення блокнотів для структурування даних, створення заміток із можливістю вводу голосової інформації, відео-файлів, функція голосового вводу, розпізнавання мови та синтезу у текст, редагування даних, пошук та збереження особистої інформації у сховищі даних.

У дослідженні представлені побудова моделей та методів систем розпізнавання мови, схема алгоритму голосового вводу та синтезу звуків у текст, схема основних етапів та методів розпізнавання мови, таких як метод на основі скритих Марковських моделей, метод на основі динамічного програмування, побудови нейронних мереж та метод на основі словникового пошуку. Останній був взятий за основу при розробці веб-сервісу. Також представлено побудову фрагменту мережі Петрі, на основі яких був зроблений висновок щодо задовольняючого процесу роботи веб-сервісу.

Література

1. Моделі, методи, алгоритми та архітектури систем розпізнавання мови: [збірник] / Російська акад. наук, Обчислювальний центр ім. А. А. Дородніцина; [Відп. ред. В. В. Рязанов]. Москва: Обчислювальний центр ім. А. А. Дородніцина РАН, 2006. 136 с.
2. Фонетичний аналіз зливої промови на основі нейронної мережі / Б. А. Абдієв. М.: ВЦ АН СРСР, 1990. 306 с.

3. Метьюз Марті Динамічне веб-програмування / Марті Метьюз, Джон КРОНАУ. М .: Ексмо, 2015. 384 с.
4. Нізамутдінов Марсель Тактика захисту і нападу на Web-додатки / Марсель Нізамутдінов. М .: БХВ-Петербург, 2017. 432 с.
5. Петін Віктор АРІ Яндекс, Google і інших популярних веб-сервісів. Готові рішення для вашого сайту. М .: БХВ-Петербург, 2015. 480 с.