

УДК 004.032.26

Модель розпізнавання мовлення для вирішення задачі керування елементами веб-сайту

Литвин Юлія Олександрівна¹, Строкань Оксана Вікторівна²

*Таврійський державний агротехнологічний університет,
м. Мелітополь, Україна*

¹ асистент кафедри комп'ютерних наук
orcid.org/0000-0001-5797-3151, litwin.yu1994@gmail.com

² кандидат технічних наук, доцент, доцент кафедри комп'ютерних наук
orcid.org/0000-0002-6937-3548, oksana.strokan@tsatu.edu.ua

Анотація. В статті розглядається метод розпізнавання мовлення, а саме можливість використання цієї технології для мовного керування елементами веб-сайтів. У зв'язку з широким впровадження таких технологій у життя людини, постає задача створити голосовий додаток, який би покращив зручність їх використання. Особливістю запропонованої моделі є виконання розпізнавання мовлення не на сервісі або за допомогою будь-яких програмних модулів ([Запрягаев, & Коновалов, 2009](#)), як це відбувається у більшості випадків, а на самому пристрої із застосуванням лише мікрофону. У вже існуючу конструкцію можна легко додавати будь-які команди. Розпізнавання мови реалізовано на сторінці веб-сайту за допомогою мови програмування Javascript. Робота скрипта базується на використанні інтерфейсів SpeechRecognition Web Speech API та SpeechGrammarList API Web Speech ([Черный, 2014](#)).

Ключові слова: розпізнавання; веб-сайт; SpeechGrammarList API Web Speech; скрипт; інтерфейс.

Speech recognition model for solution of website element management tasks

Yulia Lytvyn¹, Oksana Strokan²

*Tavria State Agrotechnological University,
Melitopol, Ukraine,*

¹Assistant of the Department of Computer Science
orcid.org/0000-0001-5797-3151, litwin.yu1994@gmail.com

²Candidate of Engineering Sciences, Associate professor,
Associate professor of the Department of Computer Science
orcid.org/0000-0002-6937-3548, oksana.strokan@tsatu.edu.ua,

Abstract. The article deals with the method of speech recognition, namely, the possibility of using this technology for the language control of website elements. Due to the widespread introduction of such technologies into human lives, the task is to create a voice application that would improve the usability. The feature of the proposed model is the implementation of speech recognition not in the service, as it happens in most cases, but in the device itself, using only a microphone. In the existing design users can easily add any commands. Language recognition is implemented on the website page using the JavaScript programming language. The script work is based on the use of the SpeechRecognition Web Speech APIs and the SpeechGrammarList API for Web Speech. The current direction of the use of speech in the process of interaction between the technical system and the user is the application of such technology for creating comfortable living conditions for people who have a violation of the musculoskeletal system and who have lost the opportunity to use traditional means and methods of dialogue with the system. It analyzed the basic principles of the website and the ability to control it using voice control. To operate the proposed speech recognition model, two interfaces are used to solve the problem of managing elements of a website: SpeechRecognition Web Speech API and SpeechGrammarList API Web Speech. In order to manage the elements of the website, a model is proposed, the implementation of which is possible through the use of a microphone on the user's desktop only. The feature of the proposed model is that it is easy to add any commands to an already existing structure. Such application provides a great perspective for building new web interfaces in combination with artificial intelligence.

Keywords: recognition; website; SpeechGrammarList API Web Speech; script; interface.

ВСТУП

Постановка проблеми. Людство завжди прагнуло до більш універсальної та комфортної взаємодії зі світом комп'ютерних технологій. В області інформаційних технологій взаємодія користувача з технічною системою відбувається різними способами і методами ([Сорокин, 2008](#)). На сьогоднішній день найпопулярнішим способом такої взаємодії є мовлення. Голосовий інтерфейс надає можливість управляти системою без опанування спеціальних навичок і тим самим підвищує якість та ефективність взаємодії користувача з системою.

Актуальним напрямком застосування мовлення у процесі взаємодії технічної системи та користувача є застосування такої технології для створення комфортних умов життя для людей, які мають порушення опорно-рухового апарату та які втратили можливість використання традиційних засобів і способів діалогу із системою.

Яскравим прикладом засобів, які реалізують голосовий пошук, є продукти від компанії Google та голосовий асистент Siri від компанії Apple, які підтверджують необхідність розроблення і впровадження нових мовленнєвих технологій у життя сучасної людини.

Аналіз останніх досліджень і публікацій. Розпізнавання мови за останні роки зробило великий крок у просуненні. Проведений аналіз інформації з питання розпізнавання мови з метою управління інформаційними системами підтвердив актуальність теми. Сучасний ринок інформаційних технологій насичений багатьма мовленнєвими інтерфейсами. Зокрема, такі компанії як Google та Apple є головними розробниками таких інтерфейсів. Сервіс Google CLOUD SPEECH API дає обмежений трафік та безкоштовне користування лише на рік. Сервіс Mozilla Web Speech API

вбудований у всі сучасні веб-браузери, він застосовує сторонні ресурси, тим самим не створює зайве навантаження на веб-ресурс.

В основі актуальної версії голосового пошуку Google лежить покращений алгоритм для навчання нейронних мереж, створений спеціально для аналізу і розпізнавання акустичних моделей. Але недоліком такого алгоритму є низька швидкість розпізнавання мовлення у порівнянні з сервісом Mozilla Web Speech API, який застосовується для розпізнавання невеликих фраз.

Мета роботи. У зв'язку з вищевикладеним стає актуальним питанням створення простого алгоритму для розпізнавання мовлення. Не всі існуючі задачі цього напрямку інформаційних технологій на сьогодні можна вважати вирішеними. Проблема розробки засобів голосового управління вимагає створення простих і зручних інтерфейсів, які б не вимагали витрати великої кількості ресурсів і мали малий час відгуку на запит користувача. Тому метою статті є представлення моделі розпізнавання мовлення для вирішення задачі керування елементами веб-сайту без застосування важких сторонніх ресурсів.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Використання голосового інтерфейсу у повсякденному житті людини набуває широкого розповсюдження. Голосовий інтерфейс має ряд переваг у порівнянні зі звичайним керуванням елементами веб-сторінок:

- відсутність необхідності у користувача спеціальних навичок;
- максимально наближена природність спілкування з технічною системою;
- можливість керування системою в обмежених умовах: занятість рук, погана видимість тощо;
- миттєве реагування системи на зміну мови спілкування і т.ін.

На сьогоднішній день керування за допомогою голосового мовлення може застосовуватися майже на різноманітних пристроях: на стаціонарних комп'ютерах, мобільних пристроях, годинниках, безпілотних літальних апаратах тощо ([Davies, Biddulph, & Balashek, 1952](#)). Збільшення обчислювальних потужностей таких пристроїв дозволило для них удосконалювати вже існуючі та створювати нові програми з функцією розпізнавання мовлення.

Для розпізнавання голосового мовлення у веб-браузері існує багато сервісів, серед яких: Google, Mozilla Web Speech API, Siri та багато інших. В основі актуальної версії голосового пошуку Google лежить покращений алгоритм для навчання нейронних мереж, створений спеціально для аналізу і розпізнавання акустичних моделей ([Klass, 1962](#)). Тобто звичайний алгоритм покращений за допомогою нейронних мереж, він дає більш точне розпізнавання голосового мовлення, але при цьому втрачає швидкість розпізнавання.

В нашому випадку необхідно розпізнавати невеликі фрази при мінімальній витраті ресурсів. Для вирішення даної задачі найкраще підходить розпізнавання на основі обмеженого словника ([Jurafsky, & Martin, 2009](#)).

Слід відмітити те, що веб-браузери не дозволяють веб-сайтам користуватись камерою користувача, мікрофоном та іншими приладами без згоди на те користувача. Крім того, потрібно, щоб на веб-сайті стояв ssl-сертифікат, який створює захищене з'єднання по протоколу https. Якщо такий сертифікат є, з'являється повідомлення, яке

запитує у користувача дозвіл на використання його пристроїв для роботи на цьому веб-сайті.

У методі, який пропонується, розпізнавання голосового мовлення реалізовано на сторінці веб-сайту за допомогою мови програмування Javascript. Був доданий скрипт, який при натисканні на іконку з мікрофоном на робочому столі користувача починає записувати мовну команду. Це дозволить підвищити точність та швидкість розпізнавання саме для конкретного набору слів та дає можливість використання алгоритму на будь-якому веб-сайті. Було вирішено використовувати відкриту технологію розпізнавання мовлення Mozilla Web Speech API (рис. 1).

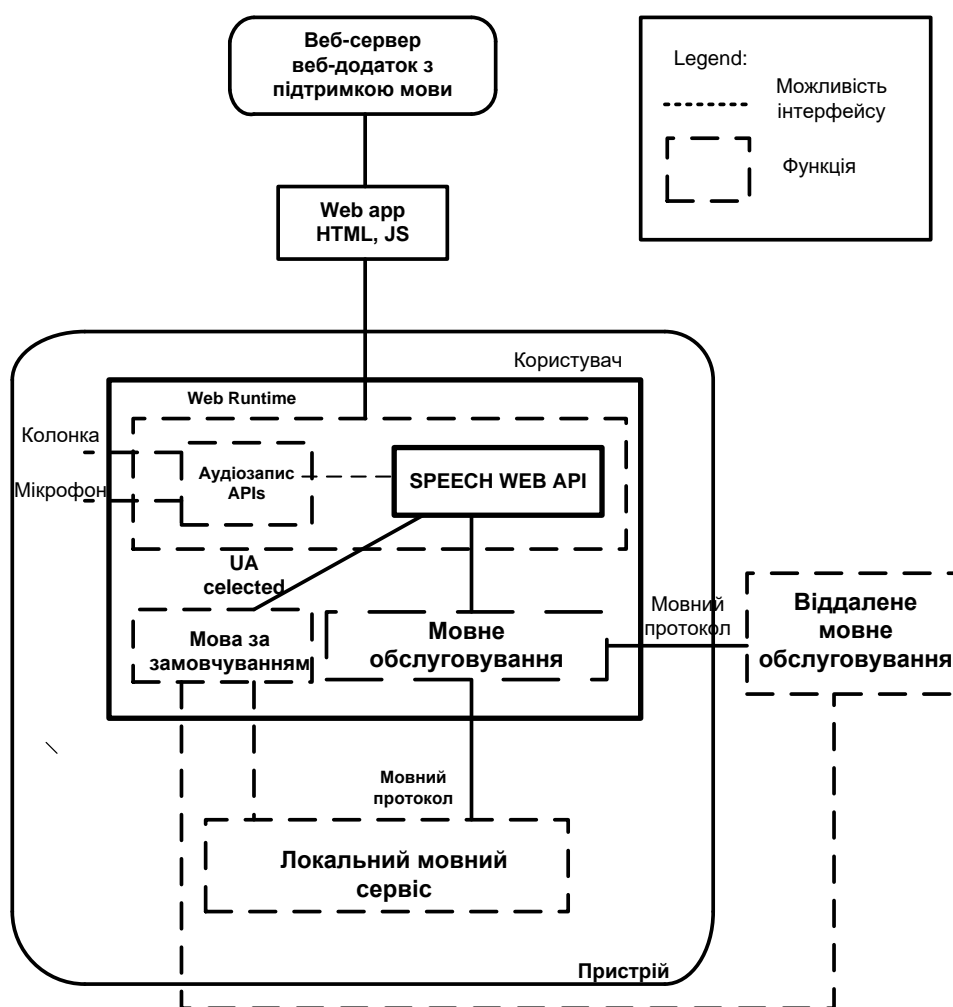


Рис. 1. Схема роботи Mozilla Web Speech API

Для роботи запропонованої моделі розпізнавання мовлення для вирішення задачі керування елементами веб-сайту використано два інтерфейси: SpeechRecognition Web Speech API та SpeechGrammarList API Web Speech.

Інтерфейс SpeechRecognition Web Speech API є контролером для сервісу розпізнавання мови (Як працює, 2018). Створити новий клас recognition можна так:

```
var recognition = new SpeechRecognition();
```

Інтерфейс SpeechGrammarList API Web Speech має структуру списку об'єктів SpeechGrammar, які містять слова та фрази, які необхідно розпізнати.

Граматика визначається з використанням формату JSpeech (JSGF.)

```
var grammar = '#JSGF V1.0; grammar phrase; public <phrase> = ' + phrase +';';  
var speechRecognitionList = new SpeechGrammarList();  
speechRecognitionList.addFromString(grammar, 1);
```

Для роботи інтерфейсу SpeechRecognition потрібно встановити базове налаштування (зв'язати граматику з інтерфейсом та вказати мову роботи):

```
recognition.grammars = speechRecognitionList;  
recognition.lang = 'ua-UA';
```

Для того, щоби почати розпізнавання мовлення необхідно створити подію і натиснути на ікону мікрофону. Розпізнавання мовлення почнеться за допомогою функції:

```
recognition.start();
```

Для отримання даних після розпізнавання мови необхідно творити подію наступним чином:

```
recognition.onresult = function(event) {  
    //дії з отриманими даними  
}
```

Для отримання першої фрази можна використати наступну конструкцію:

```
var speechResult = event.results[0][0].transcript;
```

Отриману фразу необхідно перевести у нижній регістр та за допомогою конструкції switch визначити чи відповідає фраза, яка була вимовлена користувачем, тим фразам, які виконують певну дію на сайті:

```
var word = speechResult.toLowerCase();  
switch(word){  
    case 'запам'ятати':  
        check(true);  
        break;  
    case 'не запам'ятовувати':  
        check(false);  
        break;  
    ...}
```

Тобто у case можна записати будь-яку фразу для виконання первинної операції. Функції можуть бути будь-якими. Найпоширенішими є дії з DOM-елементами сторінки.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У даній роботі проведенні дослідження в області застосування розпізнавання голосового мовлення показали його важливість та необхідність в процесі діалогу людини і технічної системи. Проаналізовано основні принципи роботи веб-сайту і можливість керування ним за допомогою голосового керування. З метою керування роботою елементами веб-сайт запропонована модель, реалізація якої можлива за рахунок використання лише мікрофону на робочому столі користувача. Особливістю запропонованої моделі є те, що у вже існуючу конструкцію можна легко додавати будь-які команди. Таке застосування дає велику перспективу для будівництва нових веб-інтерфейсів у поєднанні зі штучним інтелектом ([Макаров, 2017](#)).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- Запрыгаев, С. А., & Коновалов, А. Ю. (2009). Распознавание речевых сигналов. *Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии*, 2, 39-48.
- Макаров, В. (2017). Как устроен искусственный интеллект: распознавание речи. Взято с <https://www.popmech.ru/technologies/392382-kak-ustroen-iskusstvennyy-intellekt-raspoznavanie-rechi/>.
- Сорокин, В. Н. (2008). *Синтез речи*. Москва: Наука.
- Черный, Д. В. (2014). Сверхбыстрое распознавание речи без серверов на реальном примере. Взято с <https://habrahabr.ru/post/237589>.
- Як працює команда «Ok Google» на пристрої Android – Пристрій Android – Пошук Google Довідка. (2018). *Support.google.com*. Взято з <https://support.google.com/websearch/answer/6031948?hl=uk&co=GENIE.Platform%3DAndroid>.
- Davies, K. H., Biddulph, R., & Balashek, S. (1952). Automatic Speech Recognition of Spoken Digits. *Journal of the Acoustical Society of America*, 24 (6), 637-642.
- Jurafsky, D. & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall.
- Klass, P. J. (1962). Fiber Optic Device Recognizes Signals. *Aviation Week & Space Technology*, 77 (20), 94-101. N.Y.: McGraw-Hill. Retrieved from https://archive.org/stream/Aviation_Week_1962-11-12#page/n46/mode/1up.

REFERENCES (TRANSLATED AND TRANSLITERATED)

- Zapryagaev, S. A., & Kononov, A. Yu. (2009). Recognition of speech signals. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyj analiz i informacionnye tekhnologii*, 2, 39-48. (in Russian)
- Makarov, V. (2017). How is the artificial intelligence: speech recognition. Retrieved from <https://www.popmech.ru/technologies/392382-kak-ustroen-iskusstvennyy-intellekt-raspoznavanie-rechi/>. (in Russian)
- Sorokin, V. N. (2008). *Synthesis of Speech*. Moscow: Nauka. (in Russian)
- Cherny, D. V. (2014). Super-fast speech recognition without servers on a real example. Retrieved from <https://habrahabr.ru/post/237589>. (in Russian)
- How the "Ok Google" command works on the Android device – Android Device – Google Help Search. (2018). *Support.google.com*. Retrieved from <https://support.google.com/websearch/answer/6031948?hl=uk&co=GENIE.Platform%3DAndroid>. (in Ukrainian)
- Davies, K. H., Biddulph, R., & Balashek, S. (1952). Automatic Speech Recognition of Spoken Digits. *Journal of the Acoustical Society of America*, 24 (6), 637-642. (in English)

Jurafsky, D. & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall. (in English)

Klass, P. J. (1962). Fiber Optic Device Recognizes Signals. *Aviation Week & Space Technology*, 77(20), 94-101. N.Y.: McGraw-Hill. Retrieved from https://archive.org/stream/Aviation_Week_1962-11-12#page/n46/mode/1up. (in English)

Матеріал надійшов до редакції 15.05.2018