

Розробка програмного забезпечення голосового інтерфейсу на основі бібліотек Python

Вінницький національний технічний університет

***Анотація.** Запропоновано програмну модель інтелектуального помічника з голосовим інтерфейсом на основі бібліотек Python. Розглянуто основні існуючі бібліотеки для розпізнавання голосу, взаємодії з операційною системою та інтернет браузером. Розроблено розмовну функцію для взаємодії голосового помічника з користувачем. Розглянуто технологію DialogFlow, що є сервісом взаємодії людини з комп'ютером, заснованим на комунікації природною мовою із можливістю навчатись і розмовляти з людиною. Досліджено задачі, критерії та обмеження розуміння природної мови.*

Ключові слова: голосовий помічник, python, siri, голосовий інтерфейс, машинне навчання

***Abstract** A software model for the intelligent helper with a voice interface based on Python libraries is proposed. The main existing libraries for voice recognition, interaction with the operating system and the Internet browser are considered. Conversational function for voice assistant interaction with user has been developed. DialogFlow technology, which is a human-computer interaction service based on natural language communication with the ability to learn and speak with a person, is discussed. The tasks, criteria and limitations of natural language comprehension are investigated.*

Keywords: voice assistant, python, siri, voice interface, machine learning

Вступ

Голосовий помічник або інтелектуальний особистий помічник - це програмний агент, який може виконувати завдання або послуги для людини на основі словесних команд, тобто шляхом інтерпретації людської мови і відповіді за допомогою синтезованих голосів. Користувачі можуть задавати питання своєму помічнику, за допомогою усних команд, управляти пристроями домашньої автоматизації і відтворенням мультимедіа за допомогою голосу, а також керувати іншими основними завданнями, такими як електронна пошта, списки справ, відкривати або закривати будь-який додаток [1].

Голосові помічники можуть включати прилади, відповідати на питання, відтворювати музику, розміщувати онлайн-замовлення і виконувати більшість інтелектуальних задач.

Прототипом голосового помічника є Briana (Brain Artificial), що є інтелектуальним особистим помічником, інтерфейсом на людській мові, програмним забезпеченням для автоматизації та розпізнавання голосу для ПК з Windows. Briana - це багатофункціональне програмне забезпечення для додатків штучного інтелекту, яке дозволяє взаємодіяти з комп'ютером за допомогою голосових команд на більшості мов світу. Briana точно перетворює мову в текст для більш ніж 100 мов світу [2].

Результати дослідження

Для розробки голосового помічника існують вже готові рішення у вигляді безкоштовних бібліотек мови програмування Python. Для реалізації поставлених задач було обрано наступні бібліотеки для роботи з розпізнаванням голосу та його обробки: SpeechRecognition, PyAudio pywin32, pyriwin32, gTTS.

Для взаємодії з операційною системою було обрано модулі `os` та `sys`, також для взаємодії з браузером використано наступні бібліотеки: `requests`, `BeautifulSoup4`, `Selenium` [3].

Для виконання голосовим помічником розглянути вище функції, необхідно закодувати логіку кожної з них єдиним способом. Перший крок - розробити метод, який буде інтерпретувати голосову відповідь користувача. Наступним кроком є розробка методу, який буде перетворювати текст в мову.

Спочатку створюється цикл, що підтримує виконання голосових команд. Для реалізації головного методу голосова команда передається для користувача в якості параметра. Потім створюється кілька операторів `if`, відповідних кожній функції відповіді. Отже, для кожної голосової команди створюються змістові модулі всередині оператора `if`. Таким чином можна описати основний алгоритм розмовної функції та створення користувацьких команд.

`Dialogflow` - це онлайн сервіс від компанії `Google`, який дозволяє створювати власні чат-боти або функції для голосового помічника. `Dialogflow` зіставляє фразу користувача природною мовою з деяким «немовним» значенням (званим `intent`) і дає відповідь. `Dialogflow` тримає контекст діалогу, щоб на наступну фразу можна було зреагувати в контексті попередніх реплік [4].

Розроблений інтелектуальний помічник дозволить спростити комунікацію користувача з операційною системою, програмами та роботою в інтернет браузері. З набором зазначених технологій впровадження розробки можливо в різних галузях, причому задачі конкретної галузі визначають вимоги до навчання помічника.

Одним із викликів, що виникає в процесі обробки природної мови, є проблема синонімії, в результаті якої одне поняття може бути вираженим кількома різними словами. Як наслідок, релевантні документи, в яких використано синоніми понять, що було вказано користувачем у запиті, можуть бути не визначені системою. Вплив синонімії є особливо відчутним при створенні систем машинного перекладу. Сьогодні програм, які оперують з усіма типами неоднозначностей у широкому спектрі галузей не існує, проте є програми, що можуть коректно реагувати на неоднозначності у вузьких сферах [5].

Отже метою є провести аналіз існуючих реалізацій та базуючись на дослідженнях розробити власний голосовий помічник, націлений на україномовного користувача.

Висновки

Загальною тенденцією є інтелектуалізація інтерфейсів, тобто призначені для користувача інтерфейси стають все більш природними у використанні. Останні досягнення покладаються на доповнену реальність, мову жестів і голосових команд. Інтерфейси переважно реалізуються шляхом постановки запитань і обробки відповідей, ніж шляхом введення деталей в онлайн-формі.

В роботі розроблено розмовну функцію для взаємодії голосового помічника з користувачем. Чим більше людина взаємодіє з пристроями, що активуються голосом, тим більше трендів і патернів, які система ідентифікує на основі одержуваної інформації. Для навчання помічника розглянуто основні існуючі бібліотеки розпізнавання голосу. Потім ці дані можна використовувати для визначення користувацьких переваг і смаків, що є довгостроковою перевагою для того, щоб зробити додаток інтелектуальним. Так, у перспективі `Google` і `Amazon` прагнуть інтегрувати голосовий штучний інтелект, здатний аналізувати людські емоції і реагувати на них.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Віртуальний помічник [Електронний ресурс]. Режим доступу до матеріалу: https://uk.wikipedia.org/wiki/Віртуальний_помічник
2. Briana - voice assistant [Електронний ресурс]. Режим доступу до матеріалу: <https://en.wikipedia.org/wiki/Briana>
3. Посібник зі створення голосового помічника [Електронний ресурс]. Режим доступу до матеріалу: <https://news.myseldon.com/ru/news/index/210459979>
4. DialogFlow [Електронний ресурс]. Режим доступу до матеріалу: <https://dialogflow.com/>
5. Обробка природної мови [Електронний ресурс]. Режим доступу до матеріалу: https://uk.wikipedia.org/wiki/Обробка_природної_мови

Мясніков Ігор Ігорович — студент групи ІПІ-18мс, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, email: ihormiasnikov@gmail.com

Ракитянська Ганна Борисівна — к.т.н, доцент, доцент кафедри програмного забезпечення, Вінницький національний технічний університет, м. Вінниця

Miasnikov Ihor I. — Department Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email: ihormiasnikov@gmail.com

Rakytianska Hanna B. — PhD (Eng.), Associate Professor of Department for Software Engineering, Vinnytsia National Technical University, Vinnytsia.