УДК 004.8

ДОСЛІДЖЕННЯ ТЕХНОЛОГІЙ РОЗПІЗНАВАННЯ МОВНИХ КОМАНД TELEGRAM-БОТОМ

Павленко А.Ю.[1]; Михеєнко Б.О.[2]
[1,2]Національний університет кораблебудування імені адмірала Макарова,
[1,2]Україна, Миколаїв
[1]pavlenko.a.yu@gmail.com; [2]dizexen@gmail.com

Анотація. *Досліджено технології автоматичного розпізнавання мовлення (ASR) та їх інтеграцію в Telegram-бот. Проведено аналіз фреймворків і розроблено прототип системи. Тестування підтвердило її ефективність і придатність для використання. Рішення має прикладне значення для автоматизації бізнес-процесів через голосові інтерфейси.*

Ключові слова: *розпізнавання мовлення; ASR; Telegram-бот; голосові команди; Whisper; Google; Vosk; Java Spring Boot; голосовий інтерфейс; автоматизація.*

Вступна частина. У сучасному цифровому середовищі стрімко зростає потреба у створенні інтелектуальних систем, що забезпечують зручну та ефективну взаємодію між користувачем і технічними засобами. Одним із перспективних напрямів розвитку таких систем є впровадження голосових інтерфейсів, які використовують технології розпізнавання мовлення. Особливої актуальності набуває їх застосування у сфері електронної комерції, обслуговування клієнтів і логістики, де важливо забезпечити швидке й точне оформлення замовлень. У цьому контексті Telegram-боти виступають як зручна платформа для реалізації голосових сервісів.

Особливістю роботи є поєднання мовних технологій і системної інтеграції у вже існуючу архітектуру бот-системи.

Метою роботи є підвищення ефективності автоматичного розпізнавання мовлення (ASR) за допомогою розробки Telegram-бота, який здатен приймати голосові замовлення, трансформувати їх у текстовий формат і автоматично обробляти за заданими бізнес-правилами.

Основна частина. У межах дослідження було реалізовано прототип Telegram-бота з голосовим інтерфейсом, що інтегрує технології автоматичного розпізнавання мовлення (ASR) для обробки користувацьких запитів. На початковому етапі проведено порівняльний аналіз сучасних ASR-рішень, зокрема Whisper (OpenAI), Google Speech-to-Text, Vosk та Mozilla DeepSpeech. Whisper продемонстрував найвищу точність розпізнавання для української мови, а також здатність до локального розгортання без залежності від хмарних сервісів. У дослідженні [1] підтверджено ефективність Whisper у багатомовному середовищі, зокрема для мов з обмеженим ресурсом, таких як українська. Модель стабільно працює з різними акцентами та типами мовлення, що є критично важливим для практичного застосування в бізнес-середовищі [2].

Для оцінки конкурентних рішень було використано результати порівняльного дослідження [3], яке охоплює як комерційні, так і відкриті ASR-сервіси. Google Speech-to-Text, хоча й демонструє високу якість транскрипції, вимагає постійного підключення до хмарної інфраструктури, що обмежує його використання в автономних системах. Whisper, навпаки, забезпечує локальну обробку аудіо, що відповідає вимогам до безпеки даних і швидкодії.

Архітектура Telegram-бота побудована на основі фреймворку Java Spring Boot, який забезпечує модульність, масштабованість і підтримку REST-сервісів. Spring Boot є ефективним інструментом для побудови мікросервісних систем, особливо в контексті автоматизації бізнес-процесів [4]. У реалізованій системі передбачено багатоступеневу

обробку голосових повідомлень: приймання аудіо через Telegram API, конвертацію у формат .wav, транскрибування за допомогою Whisper, аналіз тексту відповідно до бізнес-логіки, а також збереження результатів у базі даних PostgreSQL.

Особливу увагу приділено моделюванню діалогової взаємодії між користувачем і ботом. Запропоновано механізм контекстного керування станами, що дозволяє уточнювати запити, обробляти неструктуровані фрази та забезпечувати природну комунікацію. Для підтримки адміністративного контролю реалізовано веб-інтерфейс, який дозволяє переглядати транскрипції, змінювати статуси замовлень та аналізувати історію взаємодії. Досвід впровадження Telegram-ботів у дослідницьких середовищах підтверджує ефективність API як механізму автоматизації інформаційної взаємодії [5].

У процесі тестування було змодельовано типові сценарії взаємодії, зокрема оформлення замовлення, уточнення деталей та скасування запиту. Розроблене рішення має високу адаптивність і може бути застосоване в різних галузях, зокрема у сфері громадського харчування, логістики, електронної комерції та реєстрації користувачів. Перспективними напрямами розвитку є впровадження багатомовності, використання моделей машинного навчання для класифікації запитів, інтеграція з платіжними сервісами та розширення функціоналу адміністративної панелі.

Висновки. На основі аналізу обрано оптимальний варіант для реалізації – Whisper, як відкритий, точний та локалізований інструмент із підтримкою української мови. Для реалізації логіки бота використано фреймворк Java Spring Boot, що дозволяє забезпечити модульність, масштабованість і підтримку REST-сервісів. За результатами тестів середня точність розпізнавання українського мовлення досягла 87%, що є прийнятним для комерційного застосування. Додатково оцінено швидкість обробки повідомлень (середній час від надсилання голосу до відповіді – 2.4 секунди).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Amorese, T., Greco, C., Cuciniello, M., Milo, R., Sheveleva, O., & Glackin, N. (2024). Automatic speech recognition (ASR) with Whisper: Testing performances in different languages. *CEUR Workshop Proceedings, 3574.* http://ceur-ws.org/Vol-3574/paper1.pdf

[2] Graham, C., & Roll, N. (2024). Evaluating OpenAI's Whisper ASR: Performance analysis across diverse accents and speaker traits. *JASA Express Letters, 4(2), 025206*. https://doi.org/10.1121/10.0024876

[3] Ferraro, A., Galli, A., La Gatta, V., & Postiglione, M. (2023). Benchmarking open source and paid services for speech to text: An analysis of quality and input variety. *Frontiers in Big Data, 6, Article 1210559.* https://doi.org/10.3389/fdata.2023.1210559

[4] Ahmad, N. H. (2024). Architectural patterns and challenges in Spring Boot for microservices: Evaluating automation strategies for scaling, monitoring, and deployment in complex software ecosystems. *International Journal of Information Technologies and Artificial Intelligence, 8(8).* https://www.researchgate.net/publication/385686967

[5] Abdillah, A., Rahmatulloh, A., Widiyasono, N., & Rizal, R. (2024). Designing a Telegram bot with web scraping capabilities for automating scientific article downloads. *Journal of Advanced Computing Technology and Application, 6(1).* https://doi.org/10.54554/jacta.2024.06.01.004

Pavlenko A.[1], Mykheienko B.[2]

Research of Language Command Recognition Technologies by Telegram-Bot

Abstract. *Automatic speech recognition (ASR) technologies and their integration into Telegram bot are studied. Frameworks are analyzed and a prototype of the system is developed. Testing has confirmed its effectiveness and usability. The solution has applied value for business process automation via voice interfaces.*

Keywords: *speech recognition; ASR; Telegram bot; voice commands; Whisper; Google; Vosk; Java Spring Boot; voice interface; automation.*