

ВИКОРИСТАННЯ РЕЗУЛЬТАТІВ АНАЛІЗУ ПРИРОДНОГО МОВЛЕННЯ ДЛЯ ПАРАМЕТРИЗАЦІЇ ПОСТАВЛЕНОЇ ЗАДАЧІ

¹ Вінницький національний технічний університет

Анотація

У роботі проведено дослідження можливості параметризації задач за допомогою використання методів та засобів для аналізу тексту. Запропоновано алгоритм попереднього аналізу та параметризації поставленої задачі, результати роботи якого можуть слугувати підґрунтям для подальшої оцінки та розподілу.

Ключові слова: аналіз тексту, параметризація, задача, розподіл.

Abstract

The paper research methods and tools for parameterization of tasks using text analysis tools. An algorithm for preliminary analysis and parameterization of the given task is proposed, the results of which can serve as a basis for further evaluation and distribution.

Keywords: text analysis, parametrization, problem, distribution.

Вступ

Аналіз тексту може бути розглянутий як процес виявлення та розуміння корисної інформації з наданих текстових даних у форматі природного мовлення із подальшим збереженням результатів аналізу у зручному форматі [1].

Процес аналізу може складатися з різних завдань, як от визначення ключових слів, виявлення сутностей, класифікація тексту за темами, настроями, синтаксичною структурою, тощо.

Результати аналізу тексту можуть бути використані для параметризації задачі шляхом використання отриманої інформації як параметрів для корегування подальшої обробки. Наприклад, ключові слова можуть бути використані як вхідні дані для пошуку відповідних інструкцій у базах знань, а інформація про приналежність тексту до певної теми, настроїв автора чи емоційного забарвлення може служити для визначення рівня пріоритетності задачі [2].

Тому, обрана тема є актуальною на сьогоднішній день, оскільки параметризація поставленої задачі є однією з ключових концепцій в обчислювальних науках та програмуванні, а визначивши властивості та характеристики проблеми, можна виконати корегування обчислювального процесу вирішення у цілому та певної реалізації зокрема.

Результати дослідження

В доповіді здійснено дослідження інструментів для аналізу природного мовлення, серед яких Google Cloud Natural Language API, Amazon Comprehend, IBM Watson Natural Language Understanding, та Azure Text Analytics.

Google Cloud Natural Language API – це інструмент для хмарних обчислень, що надає користувачам інтерфейс програмування застосунків для аналізу тексту. За його допомогою можна виконувати аналіз природного мовлення для виявлення сутностей, визначення настроїв, ключових слів, а також класифікації тексту за темами, тощо [3].

Amazon Comprehend – це інструмент аналізу тексту, що надає готові інструменти для виконання різноманітних завдань, що стосуються аналізу природного мовлення без необхідності розробки складних алгоритмів власноруч з нуля. Amazon Comprehend може виконувати різні типи аналітичних завдань, як от виявлення сутностей, визначення ключових фраз, тощо [4].

IBM Watson Natural Language Understanding – це інструмент для аналізу природного мовлення, що дозволяє вирішувати задачі з виявлення сутностей, визначення ключових фраз та настроїв, тощо [5].

Порівняння вище згаданих інструментів наведено у Таблиці 1.

Таблиця 1 – Порівняння інструментів для аналізу природного мовлення

Інструмент	Google Cloud Natural Language API	Amazon Comprehend	IBM Watson NLU	Azure Text Analytics
Виробник	Google	Amazon	IBM	Microsoft
Мови	Більше 100	Більше 100	Більше 20	25
Функції	Виявлення сутностей, аналіз синтаксису, визначення настроїв, класифікація тексту та інше	Виявлення сутностей, аналіз синтаксису, визначення ключових фраз, класифікація тексту та інше	Виявлення сутностей, аналіз синтаксису, визначення настроїв, класифікація тексту та інше	Виявлення сутностей, аналіз настроїв, визначення ключових фраз, класифікація тексту та інше
Цінова політика	Плата за використання	Плата за використання	Плата за використання	Плата за використання
Безкоштовний тариф	Так	Так	Ні	Так
Інтеграція з іншими сервісами	Частково	Так	Так	Так
Максимальний обсяг тексту за запит	1,000,000 символів	5,000,000 символів	256 kB	5,120 символів
Можливість кастомізації	Обмежена	Ні	Так	Обмежена
Рівень точності	Високий	Високий	Високий	Високий

Результати роботи будь-якого з вище згаданих інструментів для аналізу природного мовлення можуть слугувати важливими джерелами структурованої інформації. Отримані характеристики можуть слугувати підґрунтям для побудови процесу автоматизованого прийняття рішення.

Запропоновано алгоритм пошуку попередньої інформації, що може відігравати ключову роль у підвищенні швидкодії вирішення поставленої задачі шляхом пошуку відповідності між результатами аналізу та відомими записами бази знань (див. рисунок 1).

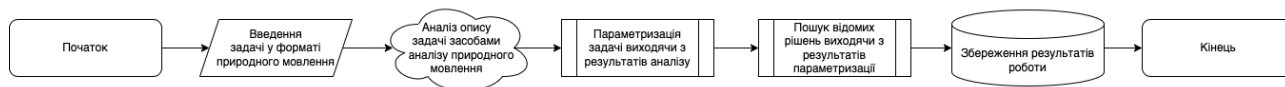


Рис. 1. Алгоритм пошуку відомих рішень виходячи з результатів параметризації задачі

Використання запропонованого алгоритму дозволяє уникнути дублювання виконаної роботи, застосувати найкращі практики та підходи, які були випробувані в схожих умовах, зменшити час вирішення проблеми, забезпечити консистентність та стандартизацію процесу прийняття подальших рішень, а також врахувати минулий досвід.

Висновки

Проведено дослідження дослідження інструментів для аналізу природного мовлення, серед яких Google Cloud Natural Language API, Amazon Comprehend, IBM Watson Natural Language Understanding, та Azure Text Analytics, зведено результати аналізу у порівняльну таблицю.

Запропоновано алгоритм пошуку попередньої інформації, що може відігравати ключову роль у підвищенні швидкодії вирішення поставленої задачі шляхом пошуку відповідності між результатами аналізу та відомими записами бази знань.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. МОДЕЛІ ГЛИБОКОГО НАВЧАННЯ ДЛЯ ВИРІШЕННЯ ЗАДАЧІ КЛАСИФІКАЦІЇ ТЕКСТОВОЇ ІНФОРМАЦІЇ [Електронний Ресурс] – Режим доступу: <https://itce.vntu.edu.ua/index.php/itce/article/view/901>
2. Аналіз даних [Електронний Ресурс] – Режим доступу: https://uk.wikipedia.org/wiki/Аналіз_даних
3. Natural Language AI [Електронний Ресурс] – Режим доступу: <https://cloud.google.com/natural-language>
4. Amazon Comprehend [Електронний Ресурс] – Режим доступу: <https://aws.amazon.com/comprehend/>
5. IBM Watson Natural Language Understanding [Електронний Ресурс] – Режим доступу: <https://www.ibm.com/products/natural-language-understanding>

Слободян Роман Віталійович — аспірант кафедри автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, м. Вінниця, e-mail: romich.prof@gmail.com.

Богач Ілона Віталіївна – к.т.н., доцент кафедри автоматизації та інтелектуальних інформаційних технологій, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м.Вінниця, e-mail: ilona.bogach@gmail.com

Slobodian Roman V. – Postgraduate of the faculty of Automation and Intelligent Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: romich.prof@gmail.com.

Bogach Ilona Vitaliivna – Ph.D., Associate Professor of Automation and Intelligent Information Technologies, Faculty of Computer Systems and Automatics Vinnytsia National Technical University, Vinnytsia, e-mail: ilona.bogach@gmail.com.