

ДОСЛІДЖЕННЯ СЕРВІСІВ AWS ДЛЯ ОБРОБКИ ПРИРОДНОЇ МОВИ

Дубок В. Ю.

Науковий керівник – к.т.н., доц. Кравець Н. С.

Харківський національний університет радіоелектроніки, каф. ПІ

м. Харків, Україна

e-mail: vladyslav.dubok@nure.ua

This work explores the current state and prospects of natural language processing (NLP). The main objectives of NLP, such as understanding, generation, and processing of text, have been described. The text scrutinizes the challenges inherent in grappling with the inherent ambiguity of human language, as well as the methodologies adopted to mitigate such complexities. The purpose of this work is to compare AWS services and identify the best one for processing user feedback. AWS has been discussed as a cloud platform offering a wide range of NLP services, such as Amazon Lex, Amazon Polly, Amazon Transcribe, Amazon Translate, and Amazon Comprehend.

Сучасне суспільство стрімко розвивається у галузі інформаційних технологій та штучного інтелекту. Завдяки швидкому зростанню обчислювальної потужності, доступності великих обсягів даних і новаторським методам в галузі алгоритмів, технології розвиваються надзвичайно швидко, впливаючи на всі сфери життя. Однією з ключових сфер, яка відіграє важливу роль у цьому процесі, є обробка природної мови.

Обробка природної мови або NLP (Natural Language Processing) це напрям лінгвістики інформаційних технологій та штучного інтелекту, спрямований на вивчення проблем комп'ютерного аналізу та синтезу природної мови. Основними цілями NLP є: розуміння – створення алгоритмів для розпізнавання речень, фраз, слів та граматичні структури, а також зрозуміти значення тексту, генерація – можливість створення текстів які будуть граматично правильними та зрозумілими для людей, обробка — виконання різних задач з текстом таких як: переклад, пошук.

Розуміння природної мови іноді вважають AI-повною задачею, тому що розпізнавання живої мови потребує величезних знань системи про навколишнє середовище та можливості взаємодіяти з ним. Проблема розуміння людського мовлення полягає саме у його неоднозначності. Одні і ті ж самі фрази та слова можуть мати різні значення в залежності від контексту речення. Під час генерації тексту важливо передати не тільки сенс, а й врахувати особливості природної мови. Вплив вище перелічених явищ є особливо відчутним при створенні систем машинного перекладу.

Для вирішення задач NLP використовується широкий спектр методів, таких як:

Статистичні методи: ґрунтуються на аналізі статистичних закономірностей мови.

Методи машинного навчання: використовують алгоритми машинного навчання для вивчення мовних закономірностей.

Лінгвістичні методи: ґрунтуються на знаннях про лінгвістичну структуру мови.

Наразі обробка природної мови одна з найпоширеніших тем яку зараз досліджують та розвивають протягом останніх років. Можливість аналізувати, обчислювати та розуміти просту людську мову за допомогою комп'ютерів і машинного навчання була впроваджена в різні програми та платформи. Таким чином зараз стало можливо автоматизувати деякі процеси які є більш ефективнішими ніж використання людей. Одним з таких процесів є обробка відгуків користувачів.

В останні десятиліття зростає значення відгуків користувачів як важливого джерела інформації для прийняття рішень. Аналіз цих відгуків став необхідним завданням для компаній, які прагнуть зберегти конкурентоспроможність і задовольнити потреби своїх клієнтів. Однак, разом зі зростанням обсягу відгуків, з'являються і складності. Проблема полягає в тому, щоб ефективно аналізувати велику кількість даних, що надходять у різних форматах та від різних джерел. При цьому необхідно враховувати не лише текстовий зміст відгуків, але і їх контекст, тон, емоційне забарвлення та інші аспекти, що можуть впливати на їх значущість. В даній роботі розглядається вирішення цих проблем за допомогою сервісів AWS.

Метою роботи є порівняння сервісів AWS з визначенням найкращого з них для обробки відгуків користувачів.

AWS пропонує широкий спектр сервісів для NLP, які дозволяють розробникам швидко та легко створювати інтелектуальні додатки.

Amazon Lex – платформа для розробки чат-ботів та віртуальних помічників, які розуміють та генерують людську мову. Вони можуть вести діалог з користувачами, відповідати на питання, надавати інформацію та виконувати завдання. Завдяки таким широким можливостям їх можна застосовувати у створенні системи особистого віртуального помічника.

Amazon Polly – сервіс для перетворення тексту в мову з реалістичним звучанням з можливістю налаштування темпу, гучності, інтонації і підтримкою десятків мов та голосів.

Amazon Transcribe – сервіс для перетворення мови в текст. Забезпечує високу точність розпізнавання мови навіть при наявності фонових шумів або присутності акценту.

Amazon Comprehend – сервіс для аналізу тексту, який допомагає отримувати цінні знання текстових даних. Він використовує машинне навчання для автоматичного аналізу тексту та виявлення ключових інформаційних блоків. У результаті аналізу тексту, отримуємо список

сутностей, які поділені та розпізнані по категоріям (імена людей, місця, організації, дати, тощо). Також є можливість отримати настрої проаналізованого тексту, і сутностей у ньому. Наразі не всі мови підтримуються для аналізу, але є можливість виявити якою мовою написаний текст.

Дані сервіси будуть порівнянні за наступними критеріями: мова — яку кількість мов підтримує сервіс для обробки тексту, відповідність функціональності — чи достатньо функцій сервісу для вирішення питання обробки відгуків користувачів, моделі — чи підтримує сервіс можливість додавати власні моделі для обробки запитів і які є обмеження.

Для тестування даних сервісів буде створено веб додаток, за допомогою якого можна буде перевірити чи придатний сервіс AWS для обробки відгуків користувачів. Така схема допоможе дізнатися обмеження кожного сервісу та можливості комбінувати функціонал між собою, вирішуючи недоліки один одного.

За результатами порівняння сервісів за критеріями очікується отримання комплексної оцінки за допомогою якої можна буде вирішити які сервіси підходять або не підходять для вирішення проблеми обробки відгуків користувачів.

Отже, обробка природної мови — це область яка динамічно розвивається і має широкий спектр застосувань. Сервіси AWS для NLP дають можливість розробникам створювати інтелектуальні додатки, які розуміють, генерують та обробляють текст, подібно до того, як це роблять люди. В рамках даної роботи було розглянуто можливість використання сервісів AWS для обробки тексту відгуків користувачів. За результатами аналізу текстів з допомогою Amazon Comprehend отримаємо необхідні дані для: формування загальної оцінку продукту, автоматичної класифікації відгуків на позитивні та негативні та виявлення відношення користувачів до сутностей продукту для типізації відгуків, що підвищить розуміння сильних та слабких сторін продукту.

Список використаних джерел

1. What is Natural Language Processing NLP? : вебсайт URL: https://aws.amazon.com/what-is/nlp/?nc1=h_ls (дата звернення: 03.03.2024).
2. Liddy, E.D. 2001. Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc.
3. Vyrodov K., Chupryna A., Kotelnikov R. Detecting of Anti-Ukrainian Trolling Tweets //Proceedings <http://ceur-ws.org> ISSN. — 2023. — Т. 1613. — С. 0073.