

УДК 004.6

Поліщук В.О.¹, Куликовська Н.А.²

¹ студ. гр. КНТ-511м, НУ «Запорізька політехніка»

² старш.викл. НУ «Запорізька політехніка»

МЕТОДИ ОБРОБКИ ПРИРОДНОЇ МОВИ

Обробка природної мови (Natural Language Processing, NLP) – загальний напрямок штучного інтелекту та математичної лінгвістики. Вона вивчає проблеми комп'ютерного аналізу та синтезу природних мов. Що стосується штучного інтелекту, аналіз визначає розуміння мови, а синтез – створення

грамотного тексту. Вирішення цих проблем буде означати створення зручнішої форми взаємодії комп'ютера та людини.

Розпізнавання людської мови можна розділити декілька напрямів. Вони є різними етапами, які згодом використовуються для загального розпізнавання тексту: розпізнавання мови; класифікація окремих слів та речень; одержання граматичної інформації, такої як основні форми визначення функцій окремих слів у реченні (суб'єкт, дієслово, об'єкт, прикметник тощо); інтерпретація сенсу (часткових) речень; розуміння контекстів речень та зв'язків між ними.

NLP найчастіше використовує методи машинного навчання для отримання інформації та роботи з нею. Алгоритм машинного навчання працює на основі навчання під час виконання. NLP також продовжує вчитися на основі даного вкладу. Ось чому, незважаючи на помилки та поширені стилі мови, NLP прагне правильно передбачити, що хоче сказати користувач. Нижче наведено кроки обробки NLP:

- лексичний аналіз означає поділ цілого сегменту тексту на речення, слова та аналіз їх значення;
- синтаксичний аналіз передбачає зв'язок слів з іншими словами, їх розташування в реченні та їх відносне значення;
- семантичний аналіз перевіряє перевіряється осмисленість речення;
- інтеграція дискурсу. Цей метод аналізує відносне значення речень та їх зв'язок з іншими реченнями;
- прагматичний аналіз. Цей метод стосується реального світового сенсу речення.

Оскільки людські мови постійно і стихійно розвиваються, а комп'ютеру потрібні чіткі та структуровані дані, при обробці виникають певні проблеми та страждає на точність. Крім того, методи аналізу текстів сильно залежать від мови, жанру, теми - завжди потрібне додаткове налаштування. Однак сьогодні багато завдань обробки природної мови все ж таки вирішуються із застосуванням глибокого навчання нейронних мереж.