

РЕФЕРАТ

Актуальність теми: задача визначення власних назв у тексті є проміжним етапом для багатьох інших задач. Результат її роботи – визначення відношень між власними назвами та категоріями, до яких вони відносяться. Далі ці відношення широко використовуються для автоматизації побудови ієрархій та категорій неструктурованих документів, наприклад новинних статей. Також вони мають застосування для реалізації пошуку за тегами, що дозволяє уникнути повнотекстового пошуку для економії обчислювальних ресурсів та часу. Інше актуальне застосування – реалізація сентиментального аналізу, коли серед вхідних даних довільної структури, наприклад даних з соціальних мереж, визначаються власні назви, та емоційне забарвлення тексту, що стосується їх.

Іншою актуальною проблемою є надзвичайно мала увага, що приділяється до задач обробки природньої мови у контексті української мови, і задача визначення власних назв не є виключенням.

Мета дослідження: створення моделі визначення власних назв у тексті на основі машинного навчання.

Для реалізації поставленої мети були сформульовані **наступні завдання:**

- систематизація існуючих підходів визначення власних назв;
- підготовка даних для навчання;
- розробка моделі на основі машинного навчання;
- тренування та тестування моделі;
- аналіз ефективності отриманого алгоритму;
- пошук подальшого напрямку досліджень.

Об'єкт дослідження: визначення власних назв у тексті українською мовою.

Предмет дослідження: алгоритми визначення власних назв у тексті.

Методи дослідження: для вирішення поставленого завдання використовувались нейронні мережі, мережі довгої короткочасної пам'яті, метод випадкових умовних полів.

Наукова новизна: найважливішими науковими результатами магістерської дисертації є дослідження можливостей визначення власних назв у текстах українською мовою при відносно малих розмірах тренувальних даних;

дослідження залежностей впливу кількості вхідних даних на результати визначення власних назв у текстах українською та англійською мовах; дослідження впливу нормалізації текстових даних на результати навчання моделей визначення власних назв.

Практичне значення отриманих результатів визначається тим, що запропонована архітектура моделі дозволяє досягти значення середнього гармонійного влучності та повноти в 92% для текстів англійською мовою, та в 72% для текстів українською мовою, що дозволяє використовувати її для визначення власних назв у текстах довільної тематики.

Зв'язок роботи з науковими програмами, планами, темами: робота виконувалась на кафедрі автоматизованих систем обробки інформації та управління Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського».

Апробація: основні положення роботи доповідались і обговорювались на VI Всеукраїнській науково-практичній конференції молодих вчених та студентів «Інформаційні системи та технології управління»(ІСТУ-2021).

Ключові слова: виділення власних назв, нейронні мережі, машинне навчання, мережа довгої короткочасної пам'яті, умовні випадкові поля

