

РЕФЕРАТ

Актуальність теми: для спрощення пошуку необхідної інформації серед наукових публікацій в Україні використовується бібліотечна класифікація. Проте наразі ця система є недосконалою, адже при класифікації допускаються помилки, а в деяких випадках вона виконується для збірника загалом, що призводить до часткової невідповідності для деяких статей, що в нього входять. Також виконання класифікації сторонньою людиною (наприклад, бібліотекарем чи редактором) вимагає багато часу. Вирішенням цієї проблеми є автоматизація процесу класифікації. За рахунок використання машинного навчання можна створити автоматичний класифікатор, який дозволить покращити точність класифікації порівняно з ручною та прискорити класифікацію нових надходжень.

Мета дослідження: створення класифікатора наукових статей за категоріями УДК на основі машинного навчання.

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

- систематизація існуючих алгоритмів класифікації текстових даних;
- збір достатньої навчальних даних, розробка класифікатора на основі машинного навчання;
- тестування та аналіз ефективності отриманого алгоритму;
- визначення подальшого напрямку досліджень.

Об'єкт дослідження: бібліотечна класифікація наукових статей.

Предмет дослідження: алгоритми класифікації текстових даних.

Методи дослідження: для розв'язання поставленої задачі використовувались наївний баєсів класифікатор, нейронні мережі, алгоритм зворотного поширення помилки.

Наукова новизна: найбільш суттєвими науковими результатами магістерської дисертації є дослідження можливостей автоматизації класифікації наукових текстів; пошуку помилок у вже класифікованих текстах; створення алгоритмів класифікації для розрізнення категорій у текстів близьких тематик.

Практичне значення отриманих результатів визначається тим, що запропонований алгоритм дозволяє досягти точності бібліотечної класифікації в

86%, що дозволяє використовувати його для пошуку і виправлення помилок у класифікації текстів, а також як допоміжного засобу при класифікації нових надходжень.

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» в рамках теми «Математичні моделі та технології в СППР». Державний реєстраційний номер 0117U000914

Апробація: основні положення роботи доповідались і обговорювались на XII науково-практичній конференції магістрантів та аспірантів «Прикладна математика та комп'ютинг» (ПМК-2019), а також на третій всеукраїнській науково-практичній конференції молодих вчених та студентів «Інформаційні системи та технології управління» (ІСТУ-2019).

Ключові слова: класифікація текстів, нейронні мережі, УДК, машинне навчання.