

УНІВЕРСИТЕТ ІМЕНІ АЛЬФРЕДА НОБЕЛЯ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА
МАГІСТРА

на тему

*“ Мультимовний класифікатор змістового
наповнення текстових документів ”*

Виконав: здобувач 4 курсу, групи КН-22м

Спеціальність 122 ”Комп’ютерні науки”

Алексеев Р. Л.

Керівник: Рижков Ігор Вікторович

Доктор технічних наук, доцент

м. Дніпро

2023 р.

АНОТАЦІЯ *

Родіон А.С. Мультимовний класифікатор змістового наповнення текстових документів

У кваліфікаційній роботі було досліджено докладний аналіз існуючих методів машинного навчання та обробки природної мови (NLP) для вирішення задачі класифікації текстів. Запропонований мультимовний класифікатор заснований на глибокому навчанні та використовує сучасні нейронні архітектури, такі як рекурентні нейронні мережі (RNN) та трансформери. Особлива увага приділяється проблемі перекладу текстів різними мовами та представленню текстових даних у мультимовних векторних просторах.

У результаті експериментальних досліджень проводиться оцінка продуктивності розробленого класифікатора на багатомовних корпусах текстових даних. Результати показують, що запропонований метод здатний ефективно класифікувати тексти різними мовами і перевищує існуючі підходи у цій галузі. Ця робота може бути корисною для фахівців у галузі обробки текстів та машинного навчання, а також для розробників систем автоматичного оброблення інформації, що працюють з мультимовними даними.

Ця робота може бути корисною для фахівців у галузі обробки текстів та машинного навчання, а також для розробників систем автоматичного оброблення інформації, що працюють з мультимовними даними.

Ключові слова: мультимовний класифікатор, текстові документи, обробка природної мови (NLP), RNN, автоматична обробка тексту

ЗМІСТ

| | |
|--|----|
| ВСТУП..... | 7 |
| 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ | |
| 1.1 Вивчення об'єкту дослідження | 9 |
| 1.1.1 Аналіз існуючої інформаційної системи..... | 11 |
| 1.1.2 Обґрунтування важливості мультязичного класифікатора | 13 |
| 1.2 Визначення основних завдань і об'єкта дослідження..... | 14 |
| 1.3 Аналіз сучасних рішень у сфері мультязичного аналізу текстів | 15 |
| 1.4 Постановка задачі..... | 17 |
| 2 ПРОЕКТУВАННЯ ТА РЕАЛІЗАЦІЯ WEB-ДОДАТКУ | |
| 2.1 Визначення варіантів використання веб-додатку | 18 |
| 2.2 Опис алгоритмів роботи системи | 23 |
| 2.2.1 Розробка алгоритмів для класифікації текстових документів..... | 28 |
| 2.3 Створення бази даних для зберігання текстів класифікації..... | 30 |
| 2.4 Розробка інтерфейсу користувача та навігаційної структури | 32 |
| 3 ТЕСТУВАННЯ WEB-ДОДАТКУ | |
| 3.1 Інтерфейс та порядок роботи з системою | 33 |
| 3.2 Реалізація механізмів авторизації та реєстрації | 39 |
| 3.3 Тестування роботи програмного забезпечення | 52 |
| ВИСНОВКИ..... | 62 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... | 63 |
| ДОДАТОК А,Б..... | 64 |