

РЕФЕРАТ

Розмір пояснювальної записки – 143 аркушів, містить 33 ілюстрацій, 32 таблиць, 8 додатків, 65 посилань на джерела.

Актуальність теми. У роботі досліджено проблему обробки великої кількості текстових даних у контексті виявлення плагіату із застосуванням Big Data рішень. Проаналізовано основні особливості сучасних платформ для виявлення запозичень, визначено їхні переваги та недоліки. Виявлено необхідність удосконалення методів обробки природної мови, що застосовуються до текстових даних, з метою підвищення ефективності та точності виявлення плагіату.

Мета дослідження. Основною метою є підвищення швидкості роботи та надійності програмного забезпечення для виявлення плагіату на надвеликих масивах даних.

Об'єкт дослідження: програмне забезпечення для виявлення плагіату.

Предмет дослідження: методи, програмна архітектура, засоби створення програмного забезпечення для виявлення плагіату на великому об'ємі даних.

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

- аналіз наявних рішень;
- підбір та підготовка словника українських та англійських синонімів;
- створення покращеного методу виявлення плагіату з використанням словників синонімів;
- розробка програмної архітектури та реалізація програмного забезпечення з використанням big data рішень;
- актуалізація морфологічної бібліотеки для української мови;
- оцінка ефективності запропонованого рішення.

Наукова новизна результатів магістерської дисертації полягає в тому, що запропоновано удосконалений метод перевірки на плагіат за допомогою словника синонімів та додавання підтримки масштабованості та обробки надвеликих масивів

даних. Результат досягнутий шляхом розробки модернізованої архітектури програмного забезпечення.

Практичне значення отриманих результатів полягає в тому, що розроблено модифікований метод та програмне забезпечення, що може використовуватися в наукових закладах, та конференціях для забезпечення академічної доброчесності та зниження використання запозичених текстів.

Зв'язок з науковими програмами, планами, темами. Робота виконувалась на кафедрі інформатики та програмної інженерії Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського" в рамках теми «Методи та технології високопродуктивних обчислень та обробки надвеликих масивів даних». Державний реєстраційний номер 0117U000924.

Апробація. Наукові положення дисертації пройшли апробацію на Всеукраїнському конкурсі студентських наукових робіт знань і спеціальностей 2023 рік, Київ.

Публікації. Наукові положення дисертації опубліковані в:

- 1) Halaiko D.O., Oliynyk Y.O. SEARCHING TEXT SIMILARITY PARALLEL METHOD. SoftTech-2022. 2022. №004.414.
- 2) Yurii Oliinyk, Danylo Halaiko, Iryna Mukha, Oleksandr Ocheretianyi. Plagiarism detecting hash-based parallel method. Proceedings of the 7th international conference, COLINS-2023. 2023. Vol. IV. P. 131–143.
- 3) Галайко Д.О., Олійник Ю.О. Застосування сховищ даних для виявлення плагіату в текстових документах. Адаптивні системи автоматичного Управління. 2024. Том 2 №45.

Ключові слова: ЗАПОЗИЧЕННЯ, ПЛАГІАТ, ТЕКСТ, ВЕЛИКІ ДАНІ, BIG DATA, DYNAMODB, AWS, FIREBASE.