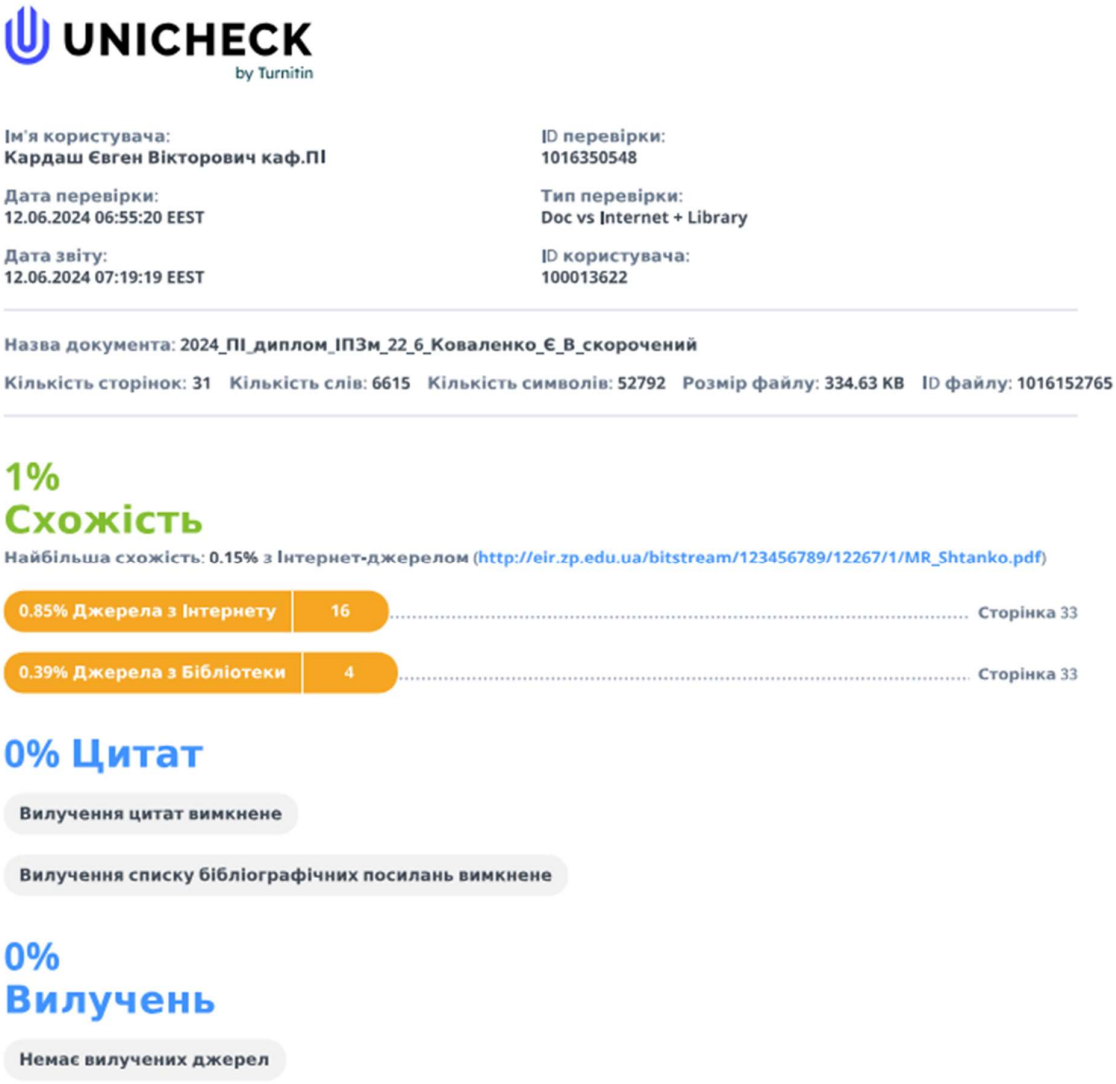
ДОДАТОК А Використані наукові праці викладачів кафедри

1. Шубін І., Снісар С., Литвин С. Формалізація та застосування алгебраїчних методів у автоматизованих інтелектуальних системах. В: Матеріали 8-ї Міжнародної конференції з проблем інфокомунікацій науки та техніки (PIC S and T 2021), Харків, Україна, 2021 (с. 67-70).

ДОДАТОК Б Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ



UNICHECK
by Turnitin

Ім'я користувача:
Кардаш Євген Вікторович каф.ПІ

ID перевірки:
1016350548

Дата перевірки:
12.06.2024 06:55:20 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
12.06.2024 07:19:19 EEST

ID користувача:
100013622

Назва документа: 2024_ПІ_диплом_ІПЗм_22_6_Коваленко_Є_В_скорочений

Кількість сторінок: 31 Кількість слів: 6615 Кількість символів: 52792 Розмір файлу: 334.63 KB ID файлу: 1016152765

1%
Схожість

Найбільша схожість: 0.15% з Інтернет-джерелом (http://eir.zp.edu.ua/bitstream/123456789/12267/1/MR_Shtanko.pdf)

0.85% Джерела з Інтернету 16 ..... Сторінка 33

0.39% Джерела з Бібліотеки 4 ..... Сторінка 33

0% Цитат

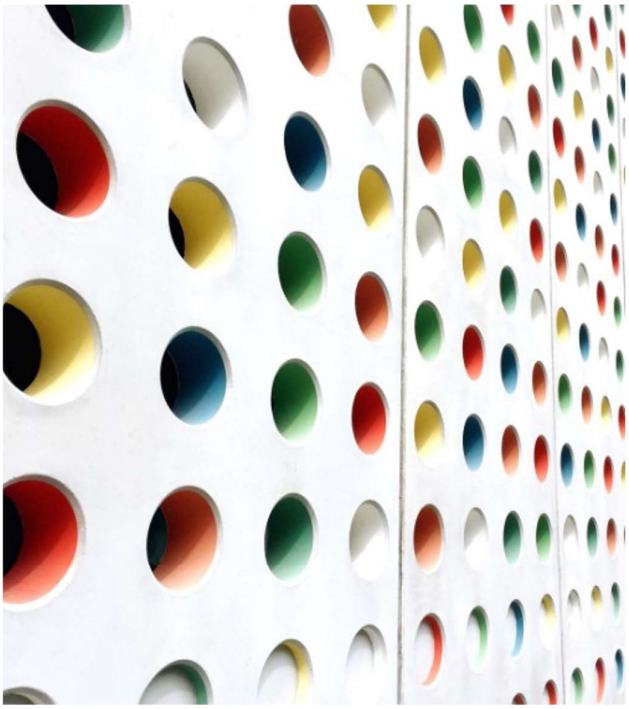
Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0%
Вилучень

Немає вилучених джерел

ДОДАТОК В Слайди презентації



Дослідження методів розробки лінгвістичних ресурсів перетворення корпусів текстів

ВИКОНАВ :
СТ. ГР. ІПЗМ-22-6 КОВАЛЕНКО Є.В.

КЕРІВНИК :
ПРОФ. КАФ. ПІ ШУБІН І.Ю.

Актуальність дослідження

- **Величезні обсяги текстових даних**: Швидке зростання текстової інформації вимагає ефективних методів обробки та аналізу
- **Роль лінгвістичних ресурсів**: Корпуси текстів, словники, і граматики є ключовими для розробки сучасних технологій.
- **Нові технології**: Штучний інтелект і машинне навчання значно покращують можливості аналізу тексту.
- **Адаптація до викликів**: Потрібні нові методи для швидкої та точної обробки складних текстових даних.

Аналіз предметної галузі

- Лінгвістичні ресурси мають багату історію починаючи з друкованих словників і граматику XV століття. З розвитком технологій вони еволюціонували до електронних ресурсів, таких як цифрові словники та корпуси текстів, що дозволяють глибоко аналізувати мовні явища.

Аналіз предметної галузі

- **Словники:** Забезпечують значення, синоніми та приклади використання слів. Важливі для пошуку точних значень і розширення лексикону.
- **Граматики:** Описують правила побудови речень, синтаксичні та морфологічні структури мови. Використовуються для навчання мови та аналізу її структури.
- **Корпуси текстів:** Великі збірки текстів, які використовуються для аналізу мовних даних у реальних умовах. Вони містять тексти з різних джерел і жанрів, що дозволяє досліджувати мовні закономірності.

Постановка задачі

•**Ціль дослідження**: Розробката оптимізаціяметодіві технологійдля ефективногостворення перетвореннята аналізутекстових корпусів використовуючисучасні лінгвістичніресурсита інструменти

•**Завдання дослідження**:

- Аналізіснуючих методіврозробкилінгвістичнихресурсів
- Вибір та розробканових методів
- Програмнареалізація
- Розробкарекомендацій та методичнихвказівок

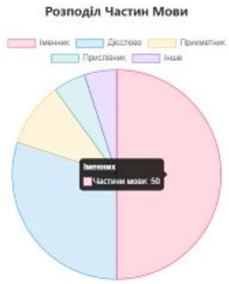
Технічні рішення

- Програмне забезпечення побудоване на модульній архітектурі, що полегшує додавання нових функцій та забезпечує гнучкість в підтримці системи.
- Кожен модуль відповідає за окремі функції, такі як обробка даних, анотація текстів та інтерфейс користувача
- HTML, CSS, JavaScript: Було використано для створення інтерфейсу користувача
- Asp Net Core, C#, REST API, Stanford NLP для створення серверної частини та обробки тексту

Функціонал програмного рішення

Визначення частини мови (Part-of-Speech Tagging)

Цей процес визначає граматичні категорії (іменник, дієслово, прикметник тощо) для кожного слова. Це важливо для аналізу граматичної структури текстів і може бути використано для більш глибокого лінгвістичного аналізу.



Функціонал програмного рішення

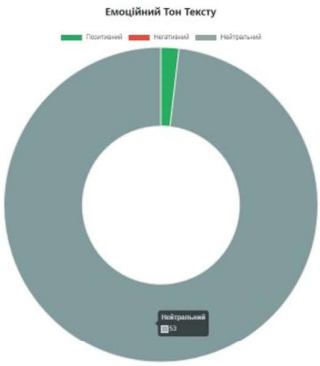
Визначення іменованих сутностей (Named Entity Recognition, NER)

NER дозволяє виявляти та класифікувати іменовані сутності в тексті, такі як імена людей, організації, місця тощо. Це корисно для вилучення конкретної інформації з текстів і аналізу зв'язків між сутностями

Функціонал програмного рішення

Сентимент-аналіз (Sentiment Analysis)

Сентимент-аналіз оцінює емоційний тон тексту, що дозволяє визначити, чи є текст позитивним, негативним або нейтральним. Це може бути корисно для аналізу відгуків або інших текстів, де емоційний тон є важливим.



Висновки

- Корпуси текстів є фундаментальними ресурсами для лінгвістичних досліджень і розробки NLP систем. Вони забезпечують реальний контекст для аналізу мовних явищ та навчання моделей машинного навчання, що є критично важливим для підвищення точності та надійності обробки текстів.
- Створення, оновлення та підтримка якісних текстових корпусів вимагає ретельного підходу, включаючи забезпечення репрезентативності, точності анотацій та дотримання етичних стандартів.
- Інтеграція сучасних NLP інструментів значно спрощує та прискорює процеси обробки та аналізу текстових корпусів, забезпечуючи високу точність та надійність результатів.

ДОДАТОК Г Апробація результатів роботи

Intelligent Systems Workshop on 8th International Conference on Computational Linguistics and Intelligent Systems (CoLInS-2024), сторінки 105-121.

ДОДАТОК Д Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008:2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент (посада) | програмної інженерії (кафедра) | ІПЗм-22-6 (група)

Коваленко Євгеній Владиславович
(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

зауважень немає

Експерт _____ (підпис) Олена ОЛІЙНИК (прізвище, ініціали)

14.06.2024