

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА
ПОЛІТЕХНІКА» ІНСТИТУТ КОМП'ЮТЕРНИХ НАУК ТА
ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Кафедра систем штучного інтелекту



ЗВІТ

про виконання практичної роботи №4
з курсу «Науковий процес та робота з науковими джерелами»

Основи етики наукових досліджень: плагіат та засоби його виявлення

Виконав:

ст. гр. КН-409

Тимошенко Павло Олександрович

Перевірив:

доцент каф. СШІ, к.т.н., Ізонін І. В.

Львів – 2021

1. Постановка завдання

Далі описано пункти, які потрібно виконати у межах цієї практичної роботи.

1. Ознайомитись із поняттям «академічної доброчесності» та видами академічного плагіату.
2. Здійснити перевірку звіту до практики (1 сторінки тексту) із використанням двох довільних програм сервісів. Подати результати перевірки, порівняти результати, отримані двома різними програмами.
3. Описати основні шляхи уникнення плагіату. Подати основні кроки усунення елементів плагіату з аналізованої роботи.
4. Здійснити опис проведеної роботи.

2. Хід роботи

В цьому розділі описано процес виконання практичної роботи та пророблені кроки.

Я не зміг знайти свій звіт з практичної роботи у швидкому доступі, тому у цій практичній роботі аналізував уривок (розміром в одну сторінку) з розділу «Обговорення результатів експериментів» з курсової роботи.

2.1. Аналізований текст

Такі результати, як 81% точності при класифікації слів та 71% точності при маркуванні речень, не можна вважати дуже високими. Втім, у цій роботі ми показово не використовували жодного препроцесингу чи виправлення результатів – це зроблено для демонстрації того, що розглянуті алгоритми є потужними та надійними навіть на сирих необроблених даних. Імплементация та аналіз різних

технік покращення результату виходять за межі задач, поставлених на початку роботи. Однак, найдієвіші з них все одно варто згадати, оскільки це може мати цінність для подальших досліджень:

- для частин мови, склад яких не має тенденції змінюватись з часом (службові частини мови та займенники), можна створити список. Тепер слова, які входитимуть у такі списки можна класифікувати із стовідсотковою точністю;
- при тренуванні моделі ПММ замість того, щоби додавати у матрицю ймовірностей спостережень кожне слово із частиною мови що згадані вище, можна додати спеціальний об'єкт, який позначатиме всю частину мови в цілому. Завдяки цьому зменшиться кількість об'єктів, що зберігаються у моделі, а отже і розмір у пам'яті та час класифікації;
- мітки **X**, **PUNCT** та **SYM**, що відповідають за іноземні слова (або просто випадкових набір символів), пунктуацію та будь-які інші небуквенні символи відповідно можна визначати алгоритмічно;
- у промаркованому реченні можна визначати стійкі помилкові патерни та виправляти їх. Такий підхід також відомий як «маркувальник Бриля».

Окремої уваги заслуговують результати експерименту із прорідженням простору. Їх можна розуміти так, що існує спосіб скоротити обсяг тренувальних даних для задачі класифікації, але при цьому не зменшити точність отриманої моделі. [...]

3. Перевірка сервісом Content-watch

Для того, аби перевірити текст, написаний у Latex-форматі за допомогою сервісу Content-watch[1], я видалив із нього все форматування та вставив у

поле тексту, як показано на Рис. 1. Результат цієї перевірки продемонстровано на Рис. 2

The screenshot shows the Content Watch website. At the top is the logo, which is an eye with a copyright symbol inside, and the text "CONTENT WATCH". Below the logo are five navigation links with icons: "Проверка текста" (document icon), "Проверка сайта" (WWW icon), "Защита сайта" (shield icon), "Магазин подписок" (shopping cart icon), and "API" (gears icon). The main heading is "Проверить текст на уникальность". To the right of the heading, it says "Длина текста: 1673 (без пробелов: 1436)". Below the heading is a large text area containing a disclaimer and several bullet points. The disclaimer states that the results are not highly accurate (81% for words, 71% for sentences) and are for demonstration purposes only. The bullet points discuss the limitations of the model, the possibility of adding a special object for the entire sentence, and the use of symbols like X, PUNCT, and SYM for punctuation and non-alphabetic symbols. At the bottom of the text area, there is a field "Игнорировать сайт:" followed by a checkbox "Запомнить". Below this is a large blue button labeled "Проверить".

CONTENT WATCH

Проверка текста Проверка сайта Защита сайта Магазин подписок API

Проверить текст на уникальность

Длина текста: 1673 (без пробелов: 1436)

Такі результати, як 81% точності при класифікації слів та 71% точності при маркуванні речень, не можна вважати дуже високими. Втім, у цій роботі ми показово не використовували жодного препроцесингу чи виправлення результатів - це зроблено для демонстрації того, що розглянуті алгоритми є потужними та надійними навіть на сирих необроблених даних. Імплементация та аналіз різних технік покращення результату виходять за межі задач, поставлених на початку роботи. Однак, найдієвіші з них все одно варто згадати, оскільки це може мати цінність для подальших досліджень:

- * для частин мови, склад яких не має тенденції змінюватись з часом (службові частини мови та займенники), можна створити список. Тепер слова, які входять у такі списки можна класифікувати із стовідсотковою точністю;
- * при тренуванні моделі ПММ замість того, щоби додавати у матрицю ймовірностей спостережень кожне слово із частиною мови що згадані вище, можна додати спеціальний об'єкт, який позначатиме всю частину мови в цілому. Завдяки цьому зменшиться кількість об'єктів, що зберігаються у моделі, а отже і розмір у пам'яті та час класифікації;
- * мітки X, PUNCT та SYM, що відповідають за іноземні слова (або просто випадкових набір символів), пунктуацію та будь-які інші неbukвенні символи відповідно можна визначати алгоритмічно;
- * у промаркованому реченні можна визначати стійкі помилкові патерни та виправляти їх. Такий підхід

Игнорировать сайт: ☐ Запомнить

Проверить

Рис. 1: Перевірка тексту на плагіат з використанням сервісу Content-watch.

говують результати експерименту із прорідженням простору. їх м
ютити обсяг тренувальних даних для задачі класифікації, але при і
моделі.

Унікальність текста: 100.0%

Править этот текст

Новая проверка

Рис. 2: Результат перевірки тексту сервісом Content-watch.

4. Перевірка сервісом Etxt.ru

Використовуючи сервіс etxt.ru[2], я виконав аналогічні дії. Процес перевірки та результат показано на Рис. 3 та Рис. 4 відповідно.

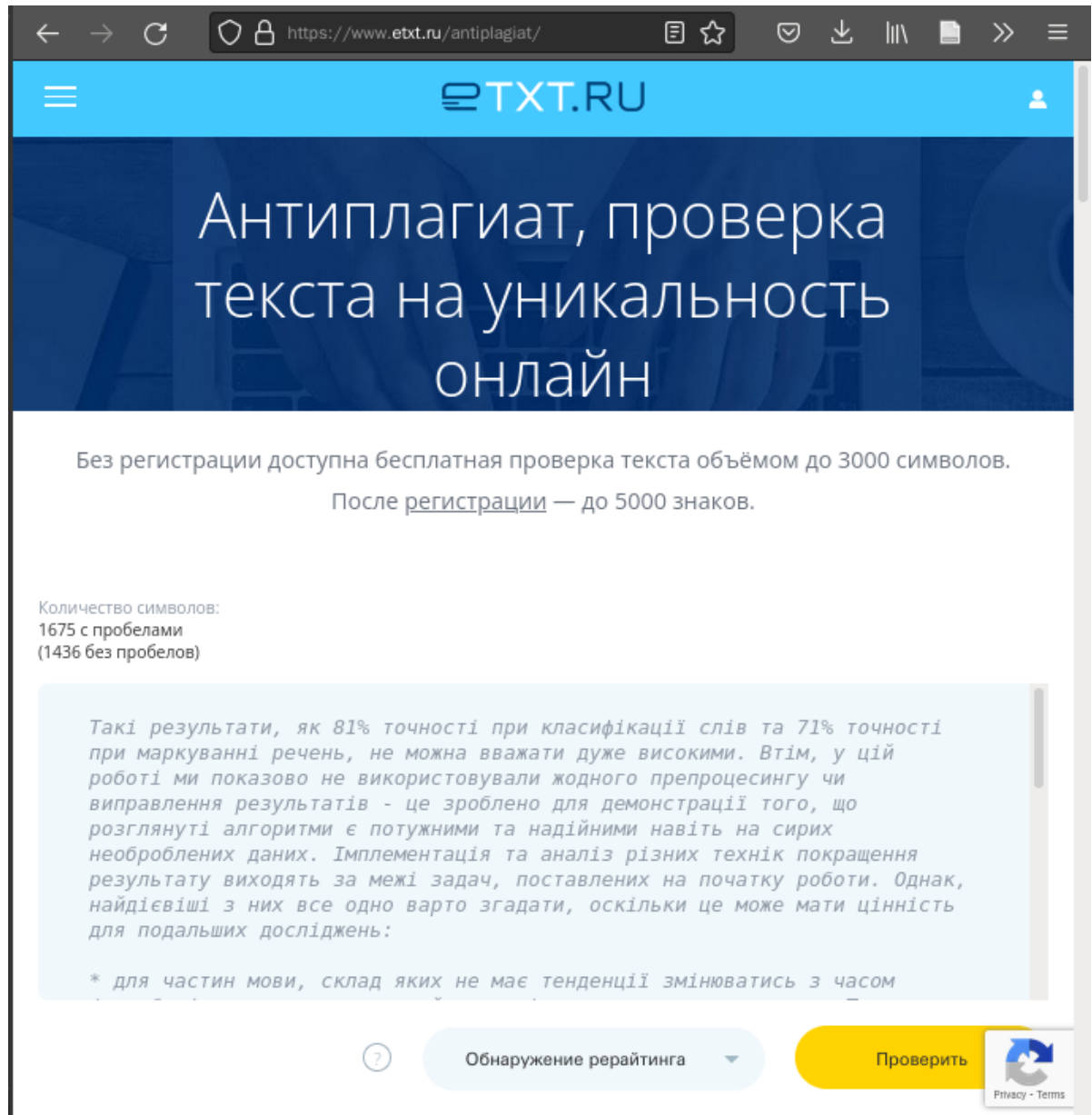


Рис. 3: Перевірка тексту на плагіат з використанням сервісу etxt.ru.

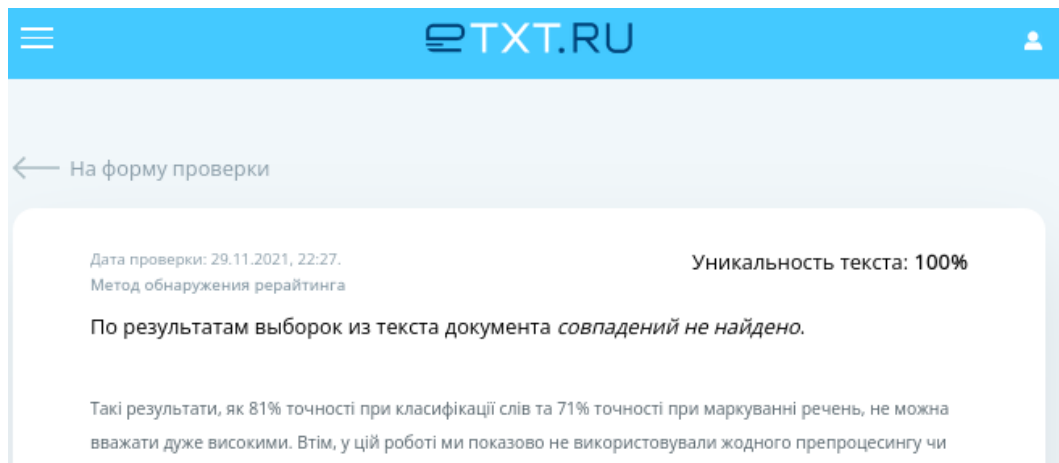


Рис. 4: Результат перевірки тексту сервісом etxt.ru.

5. Результати та висновки

Обидва сервіси показали, що мій текст має 100-відсоткову унікальність. Такий результат був неочікуваним, втім, я ним задоволений!

Це не обов'язково означає, що ніде у світі не існує більше роботи, вислови у якій, є схожими на мої. Цілком ймовірно, що подібні праці просто відсутні в базі пошуку абоощо. При перевірці всього обсягу роботи ми навряд досягнемо 100-відсоткової унікальності, тому що навіть усталені вирази та схожі висновки (особливо коли текст не несе змістового навантаження, а є лише «водою») можна вважати плагіатом. Тож для того, аби уникнути можливого плагіату, я сформулював для себе деякі принципи, які також міг би порадити і комусь іншому.

1. Перш за все потрібно користуватись здоровим глуздом. Якщо адекватно і свідомо описувати результати власної роботи, з власними висновками та міркуваннями, то вірогідність того, що хтось уже використав такі самі висловлювання, на мою думку, є дуже малою.
2. Потрібно уникати сталих беззмістовних виразів та тексту-«води». Наприклад: *у цій роботі ми здійснювали дослідження на предмет відповідно-*

сті досліджуваного об'єкта очікуваним властивостям, сформульованим у меті.

3. При огляді літератури потрібно занотовувати цікаві думки, аби за потреби процитувати їх із правильним вказанням джерела. В іншому разі це може бути плагіатом.

Досвід, отриманий в результаті виконання цієї роботи, я зможу використати при написанні диплому задля уникнення помилкового плагіату.

6. Список використаної літератури

- [1] Антиплагиат онлайн - перевірить текст на унікальність без реєстрації [Електронний ресурс] // Защита контента сайта от копирования. – Режим доступу: <https://content-watch.ru/text/> (дата звернення: 29.11.2021). – Назва з екрана.
- [2] Антиплагиат, перевірить текст на унікальність онлайн. [Електронний ресурс] // Биржа копирайтинга eTXT. – Режим доступу: <https://www.etxt.ru/antiplagiat/> (дата звернення: 29.11.2021). – Назва з екрана.