

Автор:

Кривов'яз Владислав Олександрович,
студент 4 курсу, спеціальності 121
«Інженерія програмного забезпечення»

Науковий керівник:

кандидат педагогічних наук, старший викладач
кафедри комп'ютерної та програмної інженерії
Галицький Олександр Вадимович

РОЗРОБКА ДОДАТКУ ДЛЯ ЕКСПРЕС-ПЕРЕВІРКИ НА ПЛАГІАТ ФРАГМЕНТІВ НАУКОВИХ РОБІТ СТУДЕНТІВ

Анотація. Метою дослідження є ознайомлення з варіантами взаємодії з пошуковими системами через програмний код; розглядання і вибір алгоритмів для визначення схожості двох текстів, які можливо застосовувати для (під)фрагментів тексту. Завданням дослідження є процес створення додатку для настільних операційних систем з графічним інтерфейсом для експрес-перевірки на унікальність фрагментів наукових робіт студентів, який буде досить зрозумілим для використання викладачам або студентам для самоперевірки.

Ключові слова: антиплагіат, порівняння тексту, пошукові системи, настільний додаток.

Вступ. У теперішній час, у зв'язку з розвитком технологій пов'язаних з глобальною мережею, значно полегшився процес пошуку, використання і розповсюдження відомостей, відповідь на більшість запитань можна знайти за декілька хвилин скориставшись смартфоном, ноутбуком або настільним ПК при наявності виходу в глобальну мережу.

Але із-за такої простоти зараз все частіше виникають випадки недобросовісного присвоєння чужих робіт, виводів, досліджень, ідей, тощо, в учнівській та студентських роботах, навіть якщо це здійснювалось не усвідомлюючи, це являється плагіатом - привласнення авторства на чужий твір або на чуже відкриття, винахід чи раціоналізаторську пропозицію, а також використання у своїх працях чужого твору без посилання на автора [1].

Актуальність подібної розробки зумовлюється статистичними даними, за якими в 2016 році плагіатом було понад 1/3 студентських робіт в Україні [2]. Тобто зараз виникає необхідність мати простий і безкоштовний спосіб перевіряти відсутність копії певного фрагменту тексту в глобальній мережі, з достатньою точністю і без затримок в декілька хвилин, як на більшості онлайн сервісах.

Мета роботи: розробка додатку для експрес-перевірки на плагіат фрагментів наукових робіт студентів, який дозволить перевірити наявність певного тексту в глобальній мережі з використанням пошукових систем, надавати оцінку приблизної схожості знайденого, подібного, під-фрагменту тексту.

Постановка задачі. Визначити стек технологій для виконання цілей майбутнього додатку для експрес-перевірки на унікальність фрагментів наукових робіт студентів. Визначити засоби взаємодії з пошуковими системами по глобальній мережі і їх придатність для цілей даного проекту. Дослідити види за алгоритми порівняння рядків довільного тексту за їх вихідними результатами, визначити які краща за все підійдуть для поставлених цілей і спробувати реалізувати їх в додатку. Визначити сценарії використання додатку користувачами. Виділити частину параметрів які можливо зробити опціональними (на вибір) для користувачів, для подальшого винесення їх в вікно налаштувань.

Основна частина. Для реалізації додатку було обрано інтегроване середовище розробки JetBrains Rider і крос-платформний графічний каркас AvaloniaUI для платформи dotnet.

Взаємодія з пошуковою системою виконується через SearX (пошукова система з відкритим кодом, що об'єднує результати інших пошукових систем, при цьому не зберігає відомості про своїх користувачів [3]).

Для оцінювання схожості текстових фрагментів використано програмну бібліотеку StringSimilarity [4], яка містить велику кількість алгоритмів для визначення схожості та

відстані текстів. У даному додатку були використані: «Відстань Дамерау – Левенштейна» і «Q-Gram».

Принцип роботи для користувача - при відкритті додатку в головному вікні користувач вставляє текст для перевірки та натискає кнопку для запуску перевірки, додаток, в невидимому для користувача режимі, очищує текст від неістотних символів і розділяє його по реченням, на запити придатні для пошукових систем, надто великі речення повторно розділяються на майже рівні (по кількості слів) частин; далі виконуються запити до пошукової системи по підготовленим запитам, яка віддає у відповідь невеликі фрагменти знайденого тексту на веб-сайтах і посилання на них; отриманих від пошукової системи фрагментів достатньо для оцінювання схожості з реченням для перевірки, після знаходження найбільш схожого фрагменту, додаток додає його в таблицю що відображається користувачу, з колонками: під-фрагмент тексту що перевірявся; знайдений фрагмент в глобальній мережі; оцінка схожості фрагментів; посилання на джерело в глобальній мережі.

І так для кожного речення, в кінці перевірки додаток виводить користувачу вікно з відомостями про загальну схожість текстів і списком найчастіших посилань.

Також було виділено наступні параметри для винесення їх в вікно налаштувань: розмір тексту що відображається в графічному інтерфейсі додатку; мова тексту що перевіряється; перемикач для автоматичного виправлення заміни кирилических літер схожими латинськими, або навпаки перемикач для додаткового відображення більш справедливому, але і більш затратному для вираховування, загального проценту схожості в результуючому вікні.

Висновок. При дослідженні здійснено аналіз різних варіантів взаємодії з пошуковими системами через програмний код і розглянуто деякі алгоритми для визначення схожості двох текстів, в результаті чого отримані знання були застосовані при створенні невеликого додатку, при використанні якого є можливість перевірити наявність такого ж або схожого тексту у глобальній мережі, додаток працює на більшості сучасних операційних системах і потребує для роботи тільки наявності підключення до глобальної мережі.

Список використаної літератури

1. Визначення плагиату [Електронний ресурс]. URL: <https://uk.wikipedia.org/wiki/Плагіат>
2. Статистичні дані про плагиат в наукових роботах студентів за 2016 рік [Електронний ресурс]. URL: <https://web.archive.org/web/20200508145453/https://www.plag.pl/statystyki-plagiatowania>
3. Пошукова система SearX [Електронний ресурс]. URL: <https://searx.github.io/searx>
4. Програмна бібліотека StringSimilarity [Електронний ресурс]. URL: <https://github.com/feature23/StringSimilarity.NET>
5. Nguyen, D. T. T. (2021). University students' understandings, attitudes and experiences on plagiarism . Cypriot Journal of Educational Sciences, 16(4), 1471–1478. <https://doi.org/10.18844/cjes.v16i4.6001>
6. Bard, Gregory V. (2007), "Spelling-error tolerant, order-independent pass-phrases via the Damerau–Levenshtein string-edit distance metric
7. Платформа .NET [Електронний ресурс]. URL: <https://docs.microsoft.com/ru-ru/dotnet>
8. Lapinskyi, V. V., Mykytenko, P. V., Halytskyi, O. V. (2021). DESIGN OF MEDICAL INFORMATION SYSTEMS USER INTERFACE. Information Technologies and Learning Tools, 85(5), 1-13. <https://doi.org/10.33407/itlt.v85i5.4407>