

2. Руденко О.Г., Бодяньський С.В. Штучні нейронні мережі - Навчальний посібник для студентів вищих навчальних закладів – К: Компанія СМІТ, 2006, 404 с.

3. Dettmers T. How to Parallelize Deep Learning on GPUs Part 1/2: DataParallelism [Електронний ресурс] / Tim Dettmers – Режим доступу до ресурсу: <http://timdettmers.com/2014/10/09/deep-learning-data-parallelism/>.

4. Згорткова нейронна мережа [Електронний ресурс] – Режим доступу до ресурсу: https://uk.wikipedia.org/wiki/Згорткова_нейронна_мережа.

АСТИТОВА Т. І., ГЛУЩЕНКО В.В.

ДОСЛІДЖЕННЯ СЕРВІСІВ ДЛЯ АНАЛІЗУ ІНФОРМАЦІЇ НА УНІКАЛЬНІСТЬ

ASTISTOVA T. I., GLUSHCHENKO V.V.

RESEARCH SERVICES FOR ANALYSIS OF INFORMATION FOR UNIQUENESS

Annotation. With the development of the Internet, the need for the ability to check the text for uniqueness has grown rapidly. Checking texts for uniqueness is necessary both for scientists, graduate students, and students to identify the similarity of parts of the text borrowed in their work from the Internet, as well as for the owners of various publishing houses and the authors of the texts themselves to make sure of the absolute originality of his work.

The development of information technology has given rise to a rapid increase in the development of software applications, websites and other software products in all areas of our lives. The increase in the number of creation of software codes has given rise to another problem in this direction - this is the problem of plagiarism in software code.

Plagiarism in programming threatens the quality and reliability of software products, violates copyright, and can have serious consequences for developers and organizations.

The goal of the work is to investigate existing services for comparing text and software code, to develop options that would adopt the advantages and take into account the shortcomings of existing methods and algorithms for finding borrowed fragments for comparison.

Keywords: web-oriented systems, HTTP, PostgreSQL, URL , HTML, CSS, Thxt-based, Metrics-based, Token-based

Вступ

З розвитком Інтернету стрімко зростає потреба в можливості перевірки тексту на унікальність. Перевіряти тексти на унікальність потрібно як науковцям, аспірантам, студентам, щоб виявити схожість частин тексту запозичених в своїй роботі з Інтернету, так і власникам різних видавництв і самим авторам текстів, щоб переконатися в абсолютній оригінальності своєї роботи.

Розвиток інформаційних технологій породив стрімке зростання розробок програмних додатків, веб-сайтів та інших програмних продуктів у всіх галузях нашого життя. Зростання кількості створення програмних кодів породило ще одну проблему в цьому напрямку - це проблема

плагіату в програмному коді. Плагіат в програмуванні загрожує якості та надійності програмних продуктів, порушує авторські права та може мати серйозні наслідки для розробників і організацій.

Дослідити існуючі сервіси порівняння тексту і програмного коду, розробити варіанти, які б перейняли переваги та врахували недоліки існуючих методів та алгоритмів пошуку запозичених фрагментів до порівняння є метою роботи.

Основний розділ

На сьогодні можна виділити найбільш поширені програми та сервіси для аналізу текстів, такі, як: eTXT Антиплагіат (зараз AntiPlagiarism.Net), інтернет -сервіс «Advego Plagiatus», онлайн сервіс EduBirdie.

Більшість цих сервісів вміють отримати текст, зрівняти його зі своєю базою текстів або ж з текстами в мережі Інтернет, зробити висновок про його унікальність чи показати відсоток співпадінь з іншим текстом. Деякі сервіси можуть показати які саме частини тексту є плагіатом, також деякі вміють робити регулярну перевірку збережених раніше текстів і зображень, що знаходяться на комп'ютері чи в мережі Інтернет, а ще деякі — перевірку цілих вебсторінок на унікальність та перевіряти унікальність вказаного URL.

Можна виділити основні недоліками цих систем:

1. Хибні спрацювання — шойно написаний власноруч текст буде унікальним всього на 50%;
2. Обмеження по кількості символів та технічні збої;

Алгоритм роботи сервісів можна описати наступним чином:

- вилучити службові дані;
- викинути всі символи, що не входять в діапазон ASCII;
- отримуємо чистий текст;
- пошук співпадінь в локальній базі текстів. [4]

Всі файли в локальній базі текстів завантажуються через мережу Інтернет. Для доступу до файлів сайтів через протокол HTTP можна використовувати бібліотеку CURL. Щоб з файлів вилучити тільки потрібний текст, треба написати власний HTML парсер за допомогою регулярних виразів (regular expressions). Зберігати всі тексти краще у файлової системі додатку, а для просто пошуку та каталогізації документів можна використовувати СУБД PostgreSQL, як база, яка поєднує величезний функціонал, простоту та користування.

Всі описані функції можуть бути реалізовані через скриптові застосунки, написані на мові PHP. Вони можуть легко і швидко доповнюватись та виправлятись.

Розглянемо другу проблему плагіату на теперішній час, це проблема плагіату в програмному коді.

Сучасні методи аналізу програмного коду включають порівняння текстових фрагментів, виявлення структурних схожостей та інших ознак плагіату.

Для реалізації аналізу програмних кодів частіше використовують три основні методи: text-based метод, metrics-based та token-based методи.

Короткий опис методів реалізації:

- Text-based метод – в цьому методі не звертається увага на структуру програми та її синтаксичні особливості. Вихідні коди програм показані у виді текстового рядка. Тобто є читання наданих фрагментів коду та сканування за допомогою функціоналу javascript лише словесних символів в обох текстах.
- Metrics-based метод - вихідний код програми не зазнає жодних змін. У якості відправної точки для виявлення плагіату, вибираються деякі кількісні характеристики програми. Це можуть бути змінні або функції. Метод проводить очищення коду від символів відступу та пробільних символів, а для виявлення ключових слів проводиться посимвольний аналіз коду для ідентифікації символів.
- Token-based метод - весь текст перетворюється в послідовність лексем, яка потім сканується для пошуку дублікатів. Всі деталі вихідного тексту, які легко піддаються зміні (імена змінних, функцій, класів, коментарі) будуть ігноруватися. Метод полягає в перетворенні вхідного тексту програмного коду в новий текст, який не має в собі справжніх слів, а являється набором уніфікованих значень, токенів, які в подальшому і використовуються для аналізу тексту.

Аналізуючи web-орієнтовані системи-аналоги було виявлено декілька аналогічних систем з різними рівнями виявлення плагіату, які використовують сукупність різних схем та алгоритмів. Це:

1. Codequiry. Особливістю цієї системи є можливість одночасного аналізу і порівняння декількох груп вхідних кодів.
2. Copyleaks. Система надає послуги по аналізу коду та пошуку в ньому фрагментів співпадаючих частин, порівнюючи з іншими типовими зразками коду.
3. Codegrade. Codegrade є одним із найнадійніших засобів перевірки схожості вихідного коду. Це безкоштовний вебінструмент, який перевіряє подібність для різних мов програмування, таких як Python, Android, iOS, JAVA, PHP, JavaScript тощо. Цей веб-додаток є дуже

корисним для студентів, оскільки в нього вбудоване виявлення плагіату, розроблене для коду, з чудово проникливим онлайн-інтерфейсом і функцією експорту.

Висновки

Аналіз існуючих систем перевірки на схожість тексту та програмних кодів, показав необхідність розробки індивідуальних систем, які допоможуть користувачам отримати достатню кількість інформації для оцінки наявності плагіату. Система сприятиме розвитку етичного програмування та підвищенню якості програмного продукту шляхом запобігання поширенню плагіату у текстах та програмних кодах. Створений продукт буде корисний для різних груп цільової аудиторії, таких як працівників сфери ІТ, компаній, викладачів та інших.

Література

1. Ланде Д. В. Елементи комп'ютерної лінгвістики в правовій інформатиці. // Д.В.Ланде //— К.: НДПП НАПрН України, 2014. — 351 с. — ISBN 978-966-2344-33-2
2. Berkeley Extension [Електронний ресурс] – Режим доступу до ресурсу: <https://bootcamp.berkeley.edu/resources/coding/learn-css/how-does-css-work/>
3. Пасічника В. В. Математична лінгвістика. Квантитативна лінгвістика / В. В. Пасічник, Ю. М. Щербина, В. А. Висоцька, Т. В. Шестакевич / навч. посіб. Кн.1 :: Новий Світ, 2016. – 359 с.
4. Астісова Т.І Розробка автоматизованої системи аналізу текстів «Антиплагіат» / Т.І. Астісова, В.О.Керіб. //Інформаційні технології в науці, виробництві та підприємстві: зб. наук. праць молодих вчених, аспірантів, магістрів кафедри інформаційних технологій проектуванн. - К.: КНУТД, 2017. С. 118-121 – ISBN 978-966.

АСТІСОВА Т.І., КАЛІНЕНКОВ В. І.

РОЗРОБКА WEB-ДОДАТКУ ДЛЯ АВТОМАТИЗАЦІЇ ВНУТРІШНЬО УНІВЕРСИТЕТСЬКИХ ПРОЦЕСІВ

ASTISTOVA T.I., KALINENKOV V.I.
DEVELOPMENT OF A WEB APPLICATION FOR AUTOMATING INTERNAL
UNIVERSITY PROCESSES

Annotation. The article considers aspects of the use of digital technologies in the field of education on the example of the development of software for automation of intra-university processes, namely in the dormitory.

Based on the analysis and research of the process of settlement and registration of students in the dormitory, a system of identification and electronic settlement of students.

The system is intended for local use on the computer of the person responsible for registration of students in dormitories and students living in dormitories. The program stores information about the university dormitories, students who live in them and the occupancy of rooms. Each student has a history of events that affect him, so information about the student is always at hand during the settlement. All this will completely transfer some processes that