

СИСТЕМА ПРОВЕРКИ ТЕКСТА НА УНИКАЛЬНОСТЬ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА ШИНГЛОВ

Чекан В.О.

ГГТУ им. П.О. Сухого, г. Гомель, Республика Беларусь

В статье рассматривается программный комплекс, реализующий систему проверки текста на уникальность (выявления плагиата).

Ключевые слова: антиплагиат, сравнение, шинглы.

Развитие информационных технологий сегодня во многом упростило существование человека, что колоссально отразилось на его работе: появилась возможность с помощью простых алгоритмов и несложных манипуляций решать различные задачи в максимально короткие сроки. Например, искать нужную информацию. Достаточно воспользоваться Интернетом, составить корректный запрос для поисковой программы, а через секунду уже можно анализировать результаты поиска.

Однако есть и отрицательная сторона: такая простота и доступность получения информации дали возможность недобросовестным людям копировать чужие мысли. Такой вид деятельности называется плагиатом – использование, перефразирование и подведение итогов работы в любой форме без подтверждения ссылками на источники и представление её как своей собственной работы. Свое распространение он получил с развитием информационных технологий среди школьников, студентов, аспирантов, докторантов, журналистов [1].

Для решения проблемы предлагается разработать систему сравнения текстов на уникальность, или антиплагиат-систему. В общем виде антиплагиат представляет собой программно-аппаратный комплекс, способный выявить оригинальность текста путем его сравнения с различными источниками, размещенными в локальной базе данных, сети Интернет, а в отдельных случаях – в закрытых информационных системах отдельных органов власти или организаций (например, высших учебных заведений, оффлайн-библиотек). По итогам анализа система будет формировать отчет, ключевым показателем которого выступает процент оригинальности. Такая система позволит снизить процент заимствования чужих работ и мотивировать пользователей более добросовестно подходить к написанию научных работ [2].

Система выполняет следующие функции:

- настройка параметров для сравнения;
- сравнение двух заданных текстов;
- сравнение заданного текста с имеющимися в локальной базе данных документами;
- использование алгоритма шинглов и статических хеш-функций для сравнения элементов текста [3][4][5];
- вывод результата сравнения текстов в процентном соотношении;

– при сравнении с текстами локальной базы данных указывать документы, в которых были найдены совпадения.

Приложение должно предоставлять графический интерфейс для удобного управления системой. Для хранения данных о работах использовать базу данных. Работа представляет собой несколько полей, среди которых:

- идентификатор (число);
- имя человека, добавившего работу;
- название документа;
- содержимое документа (текст).

Для разработки используется язык программирования C#. Для создания интерфейса пользователя применяется WPF. Получение исполняемого кода и его отладка – MS Visual Studio. Используемая для хранения записей база данных – MySQL. Программа разрабатывается под операционную систему Windows. Для тестирования приложения используется Unit-тестирование.

Список источников

1. Абрамова, Н.Ю. Проблема плагиата в научных работах / Абрамова Н.Ю // Научная периодика: проблемы и решения – 2013 г. – 20 с.
2. Шарапов, Р. В. Система проверки текстов на заимствования из других источников / Шарапов Р. В, Шарапов Е. В – Муром: МИ ВлГУ – 32 с.
3. Cindy A.K.M., Security Analysis of MD5 Algorithm in Password Storage / Cindy A.K.M., Wang Z., Deb Das S. // Applied Mechanics and Materials – 2013. –№2706 – р. 27-29
4. Xuejia, L. Advances in Cryptology. Finding SHA-1 Characteristics: General Results and Applications / Xuejia L., Chen K. – Shanghai: Shanghai, 2006. – 466 p.
5. Koopman, P. Cyclic Redundancy Code (CRC) Polynomial Selection / Koopman P., Chakravarty T. – Pittsburgh: Carnegie Mellon University – 109 p.

The article discusses a software package that implements a system for checking text for uniqueness (plagiarism detecting).

Keywords: anti-plagiarism, comparison, shingles.