

Реализация локального антиплагиата с помощью объектного приложения языка C++

Ленкин Алексей Викторович

*Приамурский государственный университет имени Шолом-Алейхема
студент*

Колесников Алексей Александрович

*Приамурский государственный университет имени Шолом-Алейхема
студент*

Лучанинов Дмитрий Васильевич

*Приамурский государственный университет имени Шолом-Алейхема
старший преподаватель кафедры информационных систем, математики и
методик преподавания*

Аннотация

В данной статье рассмотрена проблема присвоения текстовой информации, пути ее решения с помощью программных средств. Описана программная реализация локального антиплагиата с помощью объектной среды языка программирования C++.

Ключевые слова: Антиплагиат, язык программирования C++, информация, уникальность.

The implementation of local Anti-plagiarism using C++ object language application

Lenkin Aleksei Viktorovich

*Sholom-Aleichem Priamursky State University
student*

Kolesnikov Aleksey Aleksandrovich,

*Sholom-Aleichem Priamursky State University
student*

Luchaninov Dmitry Vasilyevich

*Sholom-Aleichem Priamursky State University
Senior lecturer of the Department of Information Systems, Mathematics and
teaching methods*

Abstract

The problem of assigning textual information is considered, ways to solve the problem with the help of software is described. The implementation of the local Anti-plagiarism software using object-oriented programming language C ++ is described.

Keywords: Anti-plagiarism, the programming language C ++, information unique.

В современное время свободный доступ к сети Интернет есть почти у каждого. Как правило, это делает любую информацию доступной и открытой общественности, но также приносит и новое понятие – информационное хищение, когда материалы, созданные одним человеком, используются без его согласия другими. В целях предотвращения данного процесса были созданы специальные программы для поиска информационного плагиата, которые, используя собственную базу источников, проверяют, сколько в проверяемых материалах заимствованных данных и выводят результат, опираясь на который можно делать вывод об уникальности. В данной статье описано исследование процесса поиска плагиата и разработка собственной локальной версии проверки электронных документов.

На данный момент создано уже довольно много программ поиска плагиата в электронных документах, наиболее часто используемая программа принадлежит ЗАО «Анти-Плагиат». Данный продукт широко используется во всех образовательных организациях высшего образования страны для проверки работ студентов, а также научных работ. Алгоритм проверки плагиата хранится компанией в секрете, но общая методика проверки известна [1]:

1. Система собирает информацию из различных источников: загружает из сети Интернет и обрабатывает сайты, находящиеся в открытом доступе, базы научных статей и рефератов. Загруженные документы проходят процедуру фильтрации, в результате которой отбрасывается бесполезная (с точки зрения потенциального цитирования) информация.

2. На следующем этапе каждый из полученных таким образом текстов определённым образом форматируется и заносится в системную базу данных.

3. Все пользовательские документы, загружаемые для проверки, ставятся в очередь на обработку.

4. После проверки документа, пользователь получает доступ к отчёту, в котором представляются результаты.

Данная система во многом и хороша и используется уже достаточный промежуток времени, но она имеет некоторые недостатки [2]:

1. С каждым улучшением программы студенты придумывают новые способы обмана системы. К примеру, раньше было возможно заменить буквы кириллицы на аналогичные латиницы и программа считала такие слова разными.

2. Системе обязательно нужен выход в сеть Интернет для проверки документа, чтобы сверить его с базой данных.

3. Даже если выход в сеть Интернет есть, проверка электронного документа может затянуться на долгое время, причем дело не в размере проверяемого файла, в обычной (бесплатной) версии на пользователя накладываются некоторые ограничения, а именно: 1 документ на 6 минут и низкий приоритет в очереди на проверку.

Частично решить данную проблему может программа для поиска плагиата в электронных документах, реализованная локально, то есть без подключения к сети Интернет. Именно, в рамках написания данной статьи было разработано приложение, которое, в отличие от программы ЗАО «Анти-Плагиат», использует локальную (оффлайн) базу документов.

Данное приложение разработано с помощью среды программирования Visual Studio 2008. Оно имеет минималистичный дизайн, лишенный ненужных опций и предназначено для проверки документов формата doc. Интерфейс программы очень прост: 4 пункта меню, 3 поля вывода и одна кнопка. После добавления необходимых файлов на проверку, а также заполнения базы эталонных файлов, они соответственно отобразятся в соответствующих полях вывода. Перед началом работы с программой необходимо удалить все пробелы из названий проверяемых и эталонных файлов.

Опишем процесс работы с программой. В комплекте программы существует два модуля «Antiword» и «Antiplagiat». До непосредственной работы с программой папки с данными модулями необходимо поместить в корень системного диска (см. рис. 1).

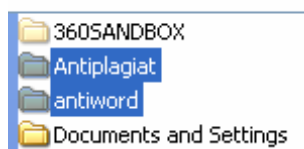


Рисунок 1 – Папки с модулями программы

Сама программа запускается с помощью файла `antiplagiat.exe` из папки `Antiplagiat`.

Для добавления проверяемых файлов нужно нажать пункт меню «Файл» – «Выбрать файлы для проверки» (см. рис. 2).

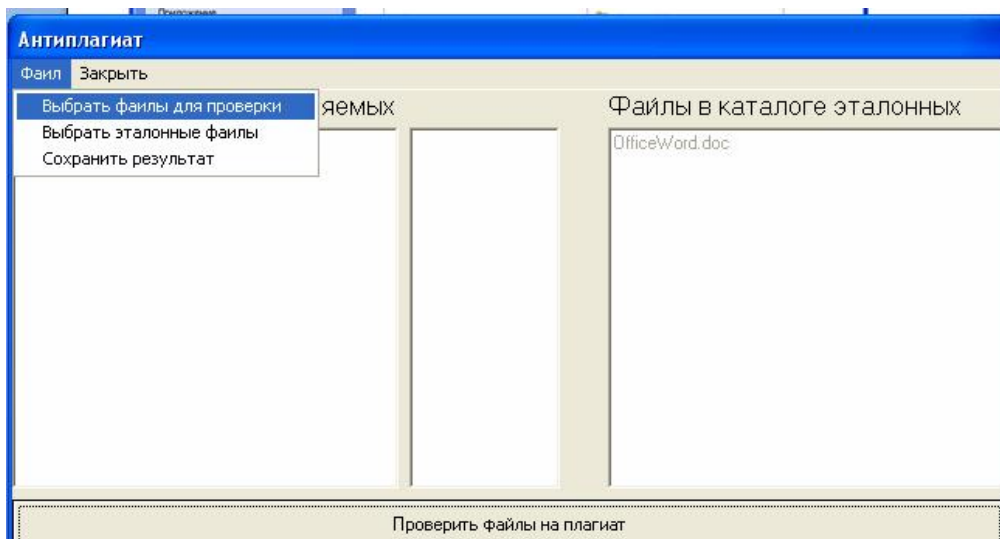


Рисунок 2 – Добавление проверяемых файлов

База текстов (эталонные файлы) добавляется через пункт меню «Файл» – «Выбрать эталонные файлы» (см. рис. 3).

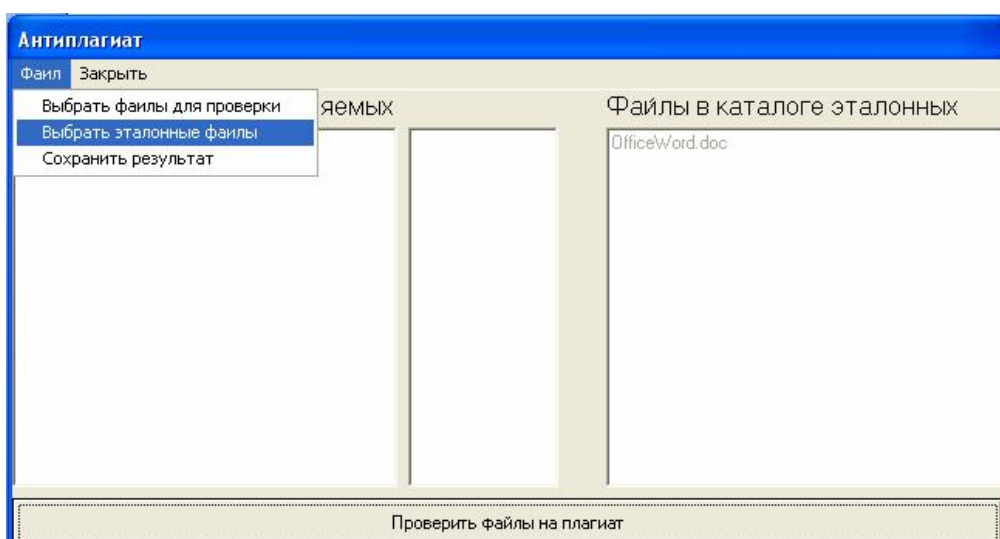


Рисунок 3 – Добавление базы текстов

После этого необходимо:

1. Убедиться, что все добавленные файлы отобразились в боковых полях.

2. Нажать кнопку «Проверить файлы на плагиат».

В результате работы программы во второй колонке отобразится процент плагиата напротив каждого проверяемого документа (см. рис. 4). При этом, для громоздких текстов выполнение программы может заметно увеличиться вследствие особенностей алгоритма.

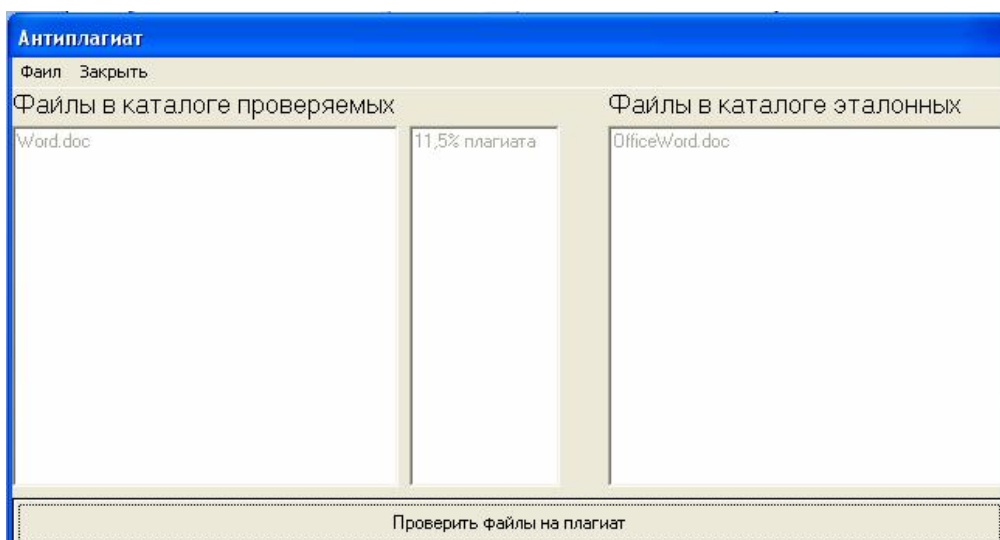


Рисунок 4 – Результат работы программы

При желании полученную информацию можно сохранить, нажав пункт меню «Файл» – «Сохранить результат» и в диалоговом окне выбрать папку сохранения.

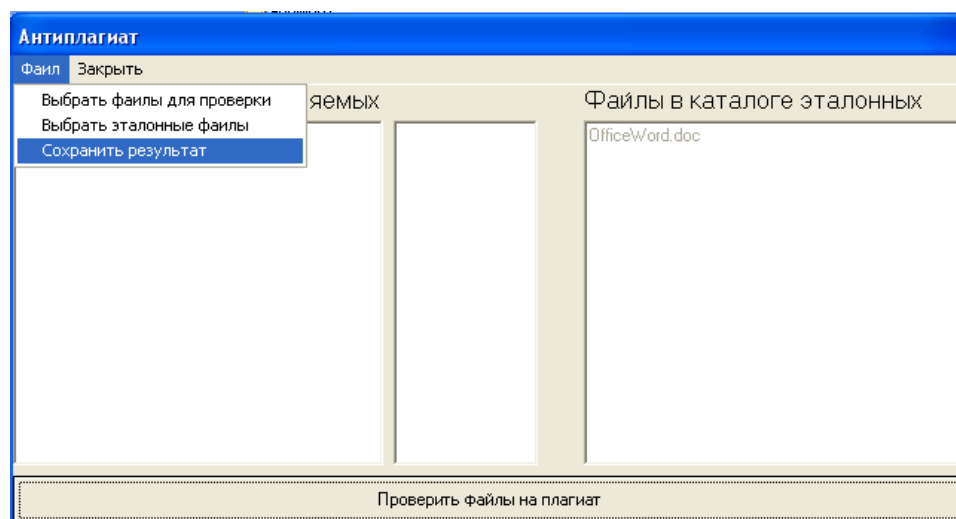


Рисунок 5 – Сохранение результатов работы программы

Программа использует простой алгоритм проверки и состоит из нескольких функций, но основными являются:

1. Zerofikator – производит форматирование входного файла, удаляя из него ненужные для проверки знаки препинания и цифры, а также переводит все буквы в нижний регистр.

2. Plagiater – анализирует два входных файла на процент плагиата, процесс следующий: считывается слово из проверяемого файла и производится поиск его в эталонном файле, если слово найдено, прибавляет счётчик к количеству повторившихся слов и повторяет это, сравнивая следующие, пока не найдутся различные, тогда, если набранное количество слов больше 3 – отнимает это количество повторов из счётчика и дальше ищет другие повторы. В результате работы программы, процент плагиата

вычисляется по следующей формуле («количество повторившихся слов» / «общее число слов в документе»)*100 %.

3. Программа использует готовый парсер “Antiword” для перевода документов формата doc в понятный программе формат txt.

В результате исследования можно описать преимущества и недостатки разработанного программного продукта.

Преимущества программы:

1. Не требуется подключение к сети Интернет для работоспособности.
2. Быстрота проверки текста (основное время занимает не поиск плагиата, а стандартизация текста для проверки).
3. Возможность сохранения результата в документ формата doc.
4. Проверка популярного формата электронных документов doc.
5. Минималистический дизайн.

Недостатки программы:

1. Долгое время стандартизации текста.
2. Возможны ложные срабатывания на частые фразы из 5 и более слов (к примеру, «В данной статье рассмотрена проблема...»).
3. Нет проверки изображений и формул, они попросту игнорируются при проверке.
4. Необходимость отсутствия в пути размещения программы и имени файлов пробелов.

Библиографический список

1. Против рейдерства в науке [Электронный ресурс]. URL: http://www.chaskor.ru/article/protiv_rejderstva_v_nauke_32043 (дата обращения 17.12.2015).
2. Система Антиплагиат [Электронный ресурс]. URL: <http://www.antiplagiat.ru/> (дата обращения 17.12.2015).