Министерство образования и науки Республики Казахстан

Северо-Казахстанский университет им. М. Козыбаева

СЕМЕСТРОВЫЙ ПРОЕКТ 3

На тему: «»

ИКТ1АПО.01 СП

|  |  |
| --- | --- |
| Выполнили студенты  группы АПО-19 | Бурханов Р.Р.  Вагина О.А.  Семейников А.Н. |
| Научный руководитель  доцент, PhD | Астапенко Н.В. |

Петропавловск, 2022

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 3](#_Toc98280509)

[1 Аналитическая часть 5](#_Toc98280510)

[1.1 Исследование предметной области 5](#_Toc98280511)

[2 Проектирование автоматизированной информационной системы 6](#_Toc98280512)

[2.1 Распределение задач 6](#_Toc98280513)

[2.2 Описание баз данных 7](#_Toc98280514)

[2.3 Разработка архитектуры программного средства 7](#_Toc98280515)

[2.4 Описание используемых алгоритмов 7](#_Toc98280516)

[2.5 Тестирование программного средства 7](#_Toc98280517)

[ЗАКЛЮЧЕНИЕ 8](#_Toc98280518)

[СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ 9](#_Toc98280519)

# ВВЕДЕНИЕ

Современное развитие информационных технологий и глобальной сети Интернет предоставило широким кругам пользователей доступ к огромным массивам информации. Появилось большое число online-библиотек, содержащих художественную и научнотехническую литературу. Стало возможным читать книги, новости и газеты непосредственно с экрана компьютера.

В сети Интернет стало доступно множество методических указаний, курсов лекций, учебников и т.д. Кроме того, появились огромные коллекции рефератов, готовых лабораторных работ, курсовых и дипломных проектов и даже диссертаций. Использование компьютерной техники сильно облегчило задачу поиска и копирования подобной информации. Если раньше для написания реферата или контрольной работы информацию было нужно, по крайней мере, найти в книгах и переписать (вручную, перепечатать или ввести в компьютер с помощью сканера и программ распознавания текстов), то теперь достаточно ввести название темы в поисковую систему и скопировать найденные материалы. Стал распространяться метод написания работ, получивший название «Copy & Paste». Метод заключается в простом копировании информации из одного или нескольких источников с минимальным редактированием получающегося таким образом текста.

Аналогичная ситуация наблюдается с отчетными материалами внутри учебных заведений. В связи с тем, что большое число пояснительных записок по курсовым и дипломным проектам выполняется с использованием компьютеров, происходит их распространение и повторное использование среди учащихся.

В последнее время наблюдается бурный рост использования в учебном процессе подобной заимствованной информации. Ситуация усугубляется тем, что учащиеся иногда не знают (не читают) то, что написано и «их» работах. Плагиат — умышленное присвоение авторства чужого произведения науки или искусства, чужих идей или изобретений.

Как можно убедиться из определения, подобные заимствованные работы можно отнести к разряду плагиата. Задача обнаружения недобросовестного использования заимствованных текстов в учебных и ученых кругах (фактов плагиата) приобретает высокую актуальность.

В статье «Актуальность и средства создания сайтов интернет-магазинов», авторами которой являются Володченко В. С., Ланцова Д.С., Метельницкая Т.А., Бышок К.А., Романов Э.В., Кадуков К.А., указано, что «в настоящее время интернет представляет собой высокоинтеллектуальную сферу, где происходит обмен, хранение и обработка огромного массива информации». Авторы утверждают, что «…Интернет стал эффективным инструментом для осуществления коммерческой деятельности. Создание Интернет-магазина – один из наиболее выгодных и перспективных инструментов онлайн-бизнеса».

*Целью* семестрового проекта является разработка системы-анализатора эссе, обеспечивающая исправление текст по «.,», проверку количества слов эссе на соответствие требованию, проверка количества слов в предложении (<24), анализ тошноты текста.

Преподаватель имеет возможность устанавливает лимит на количество слов в эссе, сохраняет одно/все эссе в отдельные word файлы, проверяет эссе на плагиат. Студент имеет возможность загружать эссе

Для разработки программного средства необходимо решить задачи синтеза и анализа. К задачам анализа относятся:

* анализ предметной области;
* выбор средств проектирования базы данных;
* выбор средств проектирования программного средства.

После решения задач анализа можно приступить к выполнению задач синтеза. К ним относятся:

* проектирование и создание базы данных;
* разработка архитектуры программного средства;
* реализация алгоритмов программного средства;
* отладка и тестирование программного средства.

Практическая значимость семестровой работы заключалась в использовании результатов работы в практической деятельности образовательных учреждений, систем автоматической обработки текста для анализа русскоязычного текста. Разработанные в ходе семестрового проекта алгоритмы и программное обеспечение может использоваться как в качестве самостоятельного продукта позволяющего облегчить анализ информации в больших объемах текстовых данных, так и в качестве основы для разработки узкоспециализированной системы семантической проверки текста.

# 1 Аналитическая часть

## 1.1 Исследование предметной области

Предметной областью семестровой работы является разработка системы анализа текста. Система позволяет студенту вводить текст эссе. Система перед отправкой преподавателю: исправляет текст по «.,», проверяет количество слов эссе на соответствие требованию, проверяет количество слов в предложении (<24), анализирует тошноту текста. Функции допустимые преподавателю в системе анализа текста: устанавливает лимит на количество слов в эссе, сохраняет одно/все эссе в отдельные word-файлы, проверяет эссе на плагиат. Таким образом система должна содержать несколько ролей входа это студент и преподаватель, поскольку функции у каждого из них различны.

В основе поиска совпадений по системе антиплагиата лежит семантический анализ текста. То есть, первоначальное разбиение его на отдельные части, дробление их на составляющие отрезки, а затем уже программа осуществляет по каждому отрезку поиск совпадений в сети интернете, отслеживание ссылок на соответствующие документы, подсвечивание их для автора и выдачу полного отчета по обработанному отрывку. Антиплагиат предоставляет информацию об общем уровне уникальности документа в процентах. При нахождении схожих или идентичных отрывков текста в сети, программа антиплагиат помечает эти отрывки цветным маркером, указывая после общего анализа ссылки на источники плагиата. Кроме подсветки не уникальных фрагментов, антиплагиат предоставляет информацию об общем уровне уникальности документа в процентах. Он высчитывается по установленному алгоритму. Абсолютная уникальность текста по антиплагиату — 100 % это величина достаточно условная. В любом случае, рассчитанная величина не является абсолютной истиной, она всего лишь свидетельствует об уровне владения языком автором текста, его умением обрабатывать и оригинально предоставлять информацию и не использовать чужие тексты в своих работах в первоначальном виде.

Помимо общего уровня антиплагиата система указывает на ошибки в тексте, исправляет текст по возможности. Алгоритм проверки на ошибки осуществляет анализ каждого слова на несовпадение с правильным написанием, а после подсвечивает ошибки и выдает список, которой следует проверить и вручную исправить написание. Сравнивает длину эссе с данными, которые ранее были введены преподавателем и при несоответствии требованиям запрещает отправку текста до тех пор, пока они не выполнены.

Таким образом, для работы автоматизированной информационной системы необходимо разработать следующие алгоритмы: исправление текста по «.,», проверка эссе на соответствие требованиям, анализ текста на тошноту, проверка на плагиат. После завершения исследования предметной области следует приступить к проектированию системы анализа текста.

# 2 Проектирование автоматизированной информационной системы

## 2.1 Распределение задач

Группа, участвующая в семестровом проекте на тему: «Разработка АИС Wildberries», состоит из 4 студентов. Техническое задание представляет собой создание макетов проекта, исследование предметной области, проектирование и подключение баз данных, создание форм, алгоритмов, архитектуры проекта, создание EXE-файла. Распределение технического задания представлено в таблице 1.

Таблица 1. Распределение обязанностей технического задания

|  |  |  |  |
| --- | --- | --- | --- |
| Неделя  Студент | 7 неделя | 8 неделя | 9 неделя |
| Семейников Артем |  |  |  |
| Вагина Оксана |  |  |  |
| Бурханов Руслан |  |  |  |

Составление отчета включает в себя написание введения, аналитической части, проектной части, заключения, а также добавления списка использованной литературы. Аналитическая часть включает в себя исследование предметной области, проектная часть – описание баз данных, архитектуры и алгоритмов проекта, тестирование программного средства. Распределение обязанностей по выполнению отчета представлено в таблице 2.

Таблица 2. Распределение обязанностей выполнения отчета

|  |  |  |  |
| --- | --- | --- | --- |
| Неделя  Студент | 7 неделя | 8 неделя | 9 неделя |
| Серик |  |  |  |
| Вагина Оксана |  |  |  |
| Бурханов Руслан |  |  |  |

После распределения работ можно приступить к выполнению задания в следующем порядке: первым этапом является проектирование баз данных. Спроектированные базы данных в дальнейшем должны быть описаны в архитектуре программного средства. Следующим этапом является разработка общих алгоритмов проекта, а также алгоритмов обработки данных о пользователях и товарах. После подключения баз данных и сборки алгоритмов и архитектуры проекта, проводится тестирование программного средства.

## 2.2 Описание баз данных

При разработке приложения, одним из основополагающих моментов, является ответ на

## 2.3 Разработка архитектуры программного средства

Архитектура программного средства (ПС) – это его строение как оно видно (или должно быть видно) извне его, т.е. представление ПС как системы, состоящей из некоторой совокупности взаимодействующих подсистем. В качестве таких подсистем выступают отдельные файлы. Разработка архитектуры является первым этапом борьбы со сложностью ПС, на котором реализуется принцип выделения относительно независимых компонентов.

Файл EXE объединяет в себе файлы

## 2.4 Описание используемых алгоритмов

Алгоритм - это система точных и понятных предписаний о содержании и последовательности выполнения конечного числа действий, необходимых для решения любой задачи данного типа.

В разработанной программе используются различные алгоритмы. Одним из таких алгоритмов является алгоритм регистрации в АИС. Алгоритм происходит следующим образом:

* Начало алгоритма;

## 2.5 Тестирование программного средства

После сбора архитектуры и создания EXE-файла можно приступить к тестированию программного средства. При тестировании приложения Wildberries можно убедиться в работоспособности кнопок для перехода между страницами, заполнении полей со страницы регистрации в базу данных names.csv. При запуске EXE-файла приложения открывается страница входа с функционалом заполнения логина и пароля, дальнейшего входа по ним, либо регистрации в приложении. Страница входа представлена на рисунке 11.

# ЗАКЛЮЧЕНИЕ

Таким образом, разработанная система «» проверки текстов на плагиат показала достаточно хорошие результаты. Использование фильтрации текста, стемминга и преобразования символов, позволило системе находить заимствованные тексты даже при их незначительной модификации. Особенностью системы является возможность проведения проверки, как по внутренней базе источников, так и по источникам сети Интернет. Это делает систему достаточно универсальным средством проверки текстов и выгодно отличает ее от существующих систем. Выдаваемые системой отчеты позволяют оценивать правомерность найденных заимствований текстов. Система может использоваться для проверки уникальности студенческих работ (курсовых и дипломных проектов, рефератов и контрольных работ). Еще одной областью применения может служить использование системы для проверки докладов, представляемых на студенческие и молодежные научные конференции.

# СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ