

Совмещение проектов по анализу ИТ и интеллектуальным ИТ: создание системы квазиреферирования дипломных и курсовых работ студентов



Выполнили студенты группы ФИТ-17
1 курса магистратуры механико-
математического факультета:

Некрасов Дмитрий Алексеевич
Мехоношин Александр Дмитриевич

Целью создания системы является разработка системы квазиреферирования текста.

Для реализации поставленной цели были сформулированы следующие задачи:

- описать проблемы создания квазиреферата;
- показать возможность выделения заголовков из документа Microsoft Word;
- показать способ, как можно создать квазиреферат, который будет достаточен в учебных целях (пояснения в заключении).

Объектом автоматизации является процесс реферирования курсовых и дипломных работ студентов.

Функции системы:

- загрузка и считывание текстового документа;
- выделение заголовков из документа;
- построение реферата текста.

На вход системе подается текстовый документ:

- формата doc или docx;
- название главы – заголовок первого уровня;
- название раздела – заголовок второго уровня;
- заголовок не является маркированным списком.

Задачу создания текста реферата на основе выделенных заголовков из документа можно решить следующими способами:

- построение N-грамм;
- построение дерева вывода (разбора).

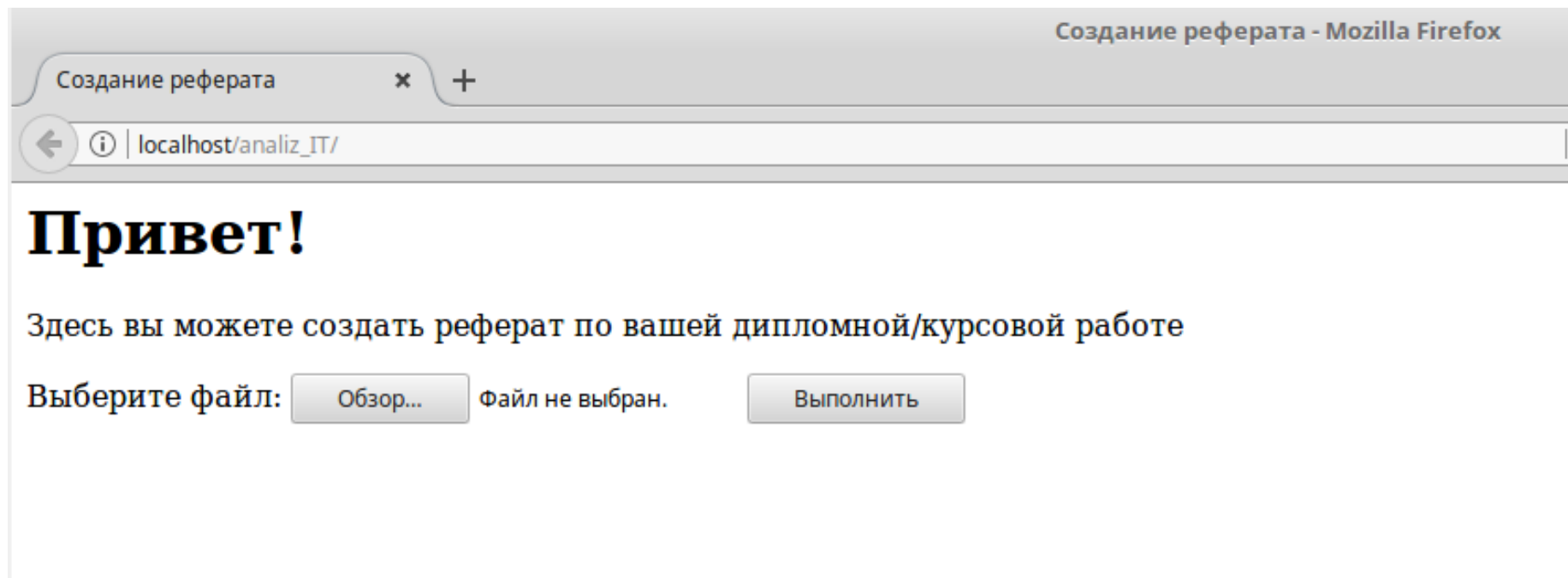
Мы выдвинули гипотезу, что для построения системы квазиреферирования дипломных и курсовых работ студентов (в учебных целях) нам будет достаточно создать шаблонную строку, в которую на определенные позиции будут подставляться заголовки из работ студентов.

Пример такой строки:

```
( В ( главе ( $ ( рассматривается ( * ) ) ) ) ) ( Освещаются  
( такие вопросы ( как: ( ** ) ) ) )
```

где * — место для вставки названия главы, \$ — номер главы, ** — место для вставки всех разделов этой главы.

Пример №1



Пример №1

В главе 1 рассматривается анализ предметной области. Освещаются такие вопросы как: обзор существующих решений, методики реферирования, различие между денотатом и сигнификатом, коллокации, выделение ключевых слов. В главе 2 рассматривается проектирование. Освещаются такие вопросы как: список использованных сторонних библиотек для python, обоснование выбора базы данных, классификация методов извлечения ключевых слов, извлечение коллокаций, автореферирование. В главе 3 рассматриваются результаты. Освещаются такие вопросы как: выделение ключевых слов, выделение коллокаций, рефераты.

Результаты

В ходе тестирования работы системы были случайно выбраны 50 дипломных и курсовых работ, построены по ним квазирефераты. Доля корректных рефератов (без грамматических ошибок, без пунктуационных ошибок, без ошибок по смыслу) равняется 0,48. По каждой второй дипломной или курсовой работе строился корректный реферат. Используемый способ не дает положительных результатов даже для такой несложной задачи. Но он был признан достаточным для работы в учебных целях. Реализовав данную систему мы убедились, что для построения квазирефератов лучше использовать другие способы.

Заключение

В ходе выполнения текущей работы была разработана система квазиреферирования текстов.

Также в ходе выполнения были реализованы следующие задачи:

- были описаны проблемы создания квазиреферата;
- была показана возможность выделения заголовков из документа Microsoft Word;
- был показан способ, как создать квазиреферат, который будет достаточен в учебных целях.

Спасибо за внимание!