

**УДК 004.9**

**Д.А. Некрасов А.Д. Мехоношин**

**D.A. Nekrasov A.D. Mehonoshin**

Пермский государственный национальный исследовательский университет

Perm state national research university

**О ВОЗМОЖНОСТИ СОЗДАНИЯ СИСТЕМЫ КВАЗИРЕФЕРИРОВАНИЯ  
КУРСОВЫХ И ДИПЛОМНЫХ РАБОТ СТУДЕНТОВ**

**ABOUT THE POSSIBILITY OF CREATION OF THE SYSTEM OF  
QUASIREFERATION OF COURSE AND DIPLOMATIC WORKS OF  
STUDENTS**

В данной статье рассматриваются различные методы квазиреферирования. На главный план выходит метод квазиреферирования путем подстановки строк в шаблонную строку. Использование данного метода достаточно в рамках статьи (смотрите приведенные результаты в конце статьи). Решение такой задачи на программном уровне облегчит процесс написания реферата для дипломной или курсовой работы студента.

**Ключевые слова:** квазиреферирование; реферат; Microsoft Word; заголовок; шаблон; подстановка; строка; doc; docx.

In this article considers the various methods of quasirefering. The main plan is the method of quasi-referencing by substituting strings into a template string. The use of this method is sufficient within the framework of the article (see the results at the end of this article). Solving such a problem at the program level will facilitate the process of writing an essay for the student's thesis or course work.

**Keywords:** quasirefering; essay; Microsoft Word; caption; template; substitution; string; doc; docx.

Цель: 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.

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 документ Microsoft Word:

- формат doc или docx;
- название главы – заголовок первого уровня;
- название раздела – заголовок второго уровня.

Квазиреферат — реферат, составленный компьютером [1].

Квазиреферирование основано на экстрагировании фрагментов документов, т. е. выделении наиболее информативных фраз и формировании из них квазирефератов.

Краткое изложение исходного материала основывается на выделении из текстов с помощью методов искусственного интеллекта и специальных информационных языков наиболее существенной информации и порождении новых текстов, содержательно обобщающих первичные документы.

Преимущество методов квазиреферирования заключается в простоте их реализации. Однако выделение текстовых блоков, не учитывающее взаимоотношений между ними или учитывающее поверхностно, часто приводит к формированию бессвязных рефератов. Некоторые предложения могут оказаться пропущены, либо в них могут встречаться слова или фразы, которые невозможно понять без предшествующего пропущенного текста [2].

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

- построение N-грамм;
- построение дерева вывода (разбора);
- построение шаблона реферата и дальнейшая подстановка заголовков на определенные позиции.

В области обработки естественного языка N-граммы применяются для предугадывания каждого последующего слова на основе вероятностных моделей. N-граммная модель рассчитывает вероятность последнего слова N-граммы, если известны все предыдущие. При использовании этого подхода для моделирования языка предполагается, что появление каждого слова зависит только от предыдущих слов [3].

Синтаксическое дерево (вывода или разбора), отражающее синтаксическое подчинение одних синтаксических единиц другим называется деревом подчинения. В узлах дерева подчинения стоят синтаксические единицы, а связи между узлами соответствуют разным видам синтаксического подчинения [4].

В рамках данной работы был выбран способ построения шаблона реферата и дальнейшей подстановки заголовков документа на определенные позиции.

Существует шаблонная строка вида:

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

Строка размечена особым образом для того, чтобы ее можно было представить в виде дерева. Строка включает в себя такие символы как: "\*", "\$", "\*\*", где \* – место для вставки названия главы, \$ – номер главы, \*\* – место для вставки всех разделов этой главы.

На вход системе подается документ формата doc или docx, из которого выделяются заголовки. После выделения заголовков выполняется обход дерева с подстановкой в размеченные места той информации, которая необходима. В результате получается заполненная шаблонная строка, что и будет являться результатом работы системы – рефератом.

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.

### Список литературы

1. Герд А.С. Предмет и основные направления прикладной лингвистики // Прикладное языкознание. СПб. – 1996.
2. Ландэ Д.В. Основы интеграции информационных потоков: Монография // Инжиниринг. – 2006.
3. Jurafsky D., Martin J.H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition // Pearson Prentice Hall – 2009.

4. Синтаксическое дерево – Студопедия // URL: [https://studopedia.ru/12\\_415\\_sintaksicheskoe-derevo.html](https://studopedia.ru/12_415_sintaksicheskoe-derevo.html) (дата обращения 19.12.2017).

**Некрасов Дмитрий Алексеевич** – Пермский государственный национальный исследовательский университет, студент, механико-математический факультет, кафедра информационных технологий, 1 курс магистратуры, гр. ФИТ-17, г. Пермь, e-mail: dinekrasov3@gmail.com.

**Мехоношин Александр Дмитриевич** – Пермский государственный национальный исследовательский университет, студент, механико-математический факультет, кафедра информационных технологий, 1 курс магистратуры, гр. ФИТ-17, г. Пермь, e-mail: infernal\_34@mail.ru.