«Шаблон»

Название файла: **СиТП – Фамилия И.О. № Группы(подгруппа)**

Смолянинов Юрий Александрович

Б9120-09.03.04прогин

Подгруппа 3

**Предметная область:** Обработка естественного языка

**Название системы:** Платформа для работы с естественным языком

**Краткое описание:** Платформа, которая позволяет обрабатывать русский язык. В частности, позволяет реферировать входящий текст.

1. Цели системы

Платформа должна быть гибкой, т.е. чтобы в неё была возможность подключать новые функции, и платформа должна уметь функцию реферирования.

1. Функциональные требования
   1. Система должна осуществлять формирование объектного представления текста в виде списка деревьев разбора, состоящих их размеченных слов;
   2. Система должна предоставлять возможность загрузки текста, объектное представление которого планируется получить;
   3. Система проверять загружаемый текст на корректность (то есть на наличие недопустимых символов);
   4. При формировании объектного представления текста система должна осуществлять формализацию этого текста с использованием анализатора;
   5. При формировании объектного представления текста система должна осуществлять сопоставление входящим в этот текст словам их значение из словаря значений;
2. Диаграмма вариантов использования (UC)
3. Сценарий использования

**Получение объектного представления текста**

**Автор:** Смолянинов Юрий.

**Дата создания:** 13.11.2022.

**Основной актор:** Пользователь.

**Краткое описание:** Пользователь загружает текст в модуль предобработки текста. Модуль предобработки текста выполняет формирование объектного представления текста и предоставляет его пользователю.

**Предварительные условия:**

Словарь значений слов обновлен до последней версии.

**Выходные условия:**

Пользователь получил объектное представление предоставленного им текста.

**Нормальное направление:**

1. Пользователь загружает текст в подсистему формирования объектного представления текста модуля предобработки текста.
2. Подсистема формирования объектного представления текста обращается к подсистеме работы с анализатором и передает ей полученный от пользователя текст.
3. Семантический анализатор выполняет разбор полученного текста и передает результат своей работы подсистеме работы с анализатором.
4. Подсистема работы с анализатором преобразует текстовое представление для каждого из предложений результата работы анализатора в дерево разбора (дерево разбора содержит список слов и соответствующих им тегов) и передает эти деревья подсистеме формирования объектного представления текста.
5. Подсистема формирования объектного представления текста передает подсистеме обработки слов список слов и соответствующих им тегов для каждого из предложений, полученных от подсистемы работы с анализатором.
6. Подсистема формирования объектного представления текста заменяет слова в деревьях разбора на соответствующие размеченные слова, в итоге получая объектное представление текста, которое возвращается пользователю как результат работы модуля предобработки текста.

**Альтернативные направления:**

1а. Пользователь ничего не ввел в поле текст:

1а1. Система сообщает пользователю о неудаче, дает совет и ожидает ввода текста.

3а. Ошибка при анализе текста:

3а1. Система сообщает пользователю о неудаче, дает совет.

**Исключения:**

Загруженный текст некорректен (в тексте найдены недопустимые символы, которые препятствуют его обработке).

**Приоритет:** Высокий.

**Частота использования:** используется в каждом случае работы с системой.

**Реферирование текста**

**Автор:** Смолянинов Георгий.

**Дата создания:** 12.11.2022.

**Основной актор:** Пользователь.

**Краткое описание:** Пользователь через вводит текст, и система проводит реферирование текста, с последующим выводом результата реферирования.

**Предварительные условия:** Отсутствуют.

**Выходные условия:**

POST-1.Пользователь ввел исходный текст.

**Нормальное направление:**

1. Пользователь подает текст для реферирования.
2. Система проводит реферирования.
3. Система выводит результат реферирования пользователю.

**Альтернативные направления:**

1а. Пользователь ничего не ввел в поле текст:

1а1. Система сообщает пользователю о неудаче, дает совет и ожидает ввода текста.

2а. Ошибка при анализе текста

2а1. Система сообщает пользователю о неудаче, дает совет и возвращает на предыдущий шаг

**Исключения:**

Текст пуст.

**Приоритет:** Высокий.

**Частота использования:** используется в каждом случае работы с системой.

1. User Story
2. **Как** пользователь, **я хочу** вводить текст, **чтобы** система проводила реферирование, введенного мною текста;
3. **Как** пользователь, **я хочу** сохранять результат реферирования, **чтобы** потом оценить его;
4. **Как** пользователь, **я хочу** видеть удобный и интуитивно понятный интерфейс у системы, **чтобы** было удобнее работать с системой;
5. **Как** пользователь, **я хочу** получать сообщение об ошибке, если введенный мною текст некорректен, **чтобы** я мог исправить ошибку;
6. **Как** разработчик платформы, **я хочу** добавлять новые функции в платформу, **чтобы** улучшать её функционал;
7. **Как** разработчик платформы, **я хочу** изменять текущие функции, **чтобы** повышать качество платформы;
8. **Как** разработчик, **я хочу** добавлять новые словари, **чтобы** улучшать качество функций платформы;
9. **Как** разработчик, **я хочу** изменять текущие словари, **чтобы** улучшать качество функций платформы.
   1. Must do /Should do/Could do

|  |  |  |
| --- | --- | --- |
| **Must do** | **Should do** | **Could do** |
| **Как** пользователь, **я хочу** вводить текст, **чтобы** система проводила реферирование, введенного мною текста | **Как** разработчик, **я хочу** добавлять новые словари, **чтобы** улучшать качество функций платформы |  |
| **Как** пользователь, **я хочу** сохранять результат реферирования, **чтобы** потом оценить его | **Как** разработчик, **я хочу** изменять текущие словари, **чтобы** улучшать качество функций платформы |  |
| **Как** пользователь, **я хочу** видеть удобный и интуитивно понятный интерфейс у системы, **чтобы** было удобнее работать с системой |  |  |
| **Как** разработчик платформы, **я хочу** добавлять новые функции в платформу, **чтобы** улучшать её функционал |  |  |
| **Как** разработчик платформы, **я хочу** изменять текущие функции, **чтобы** повышать качество платформы |  |  |

1. Архитектурно контекстная диаграмма (АКД)

Архитектурно-контекстная диаграмма (АКД) отображает основные информационные и программные компоненты, необходимые для выполнения поставленных задач. Информационные компоненты изображаются цилиндрами, программные компоненты – прямоугольниками, связи по передаче данных и управлению – стрелками с соответствующими подписями.

Архитектурно-контекстная диаграмма платформы для работы с естественным языком с функцией реферирования представлена на рисунке 2. В черной рамке отражен состав модуля реферирования текста.

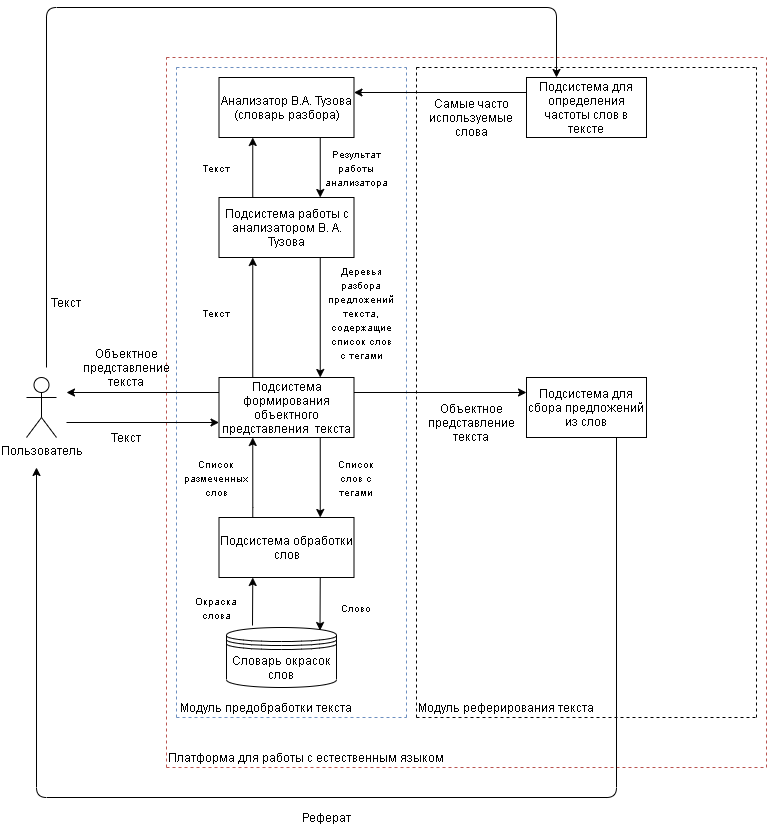


Рисунок 2 - АКД системы

Модуль реферирования текста включает в себя следующие компоненты:

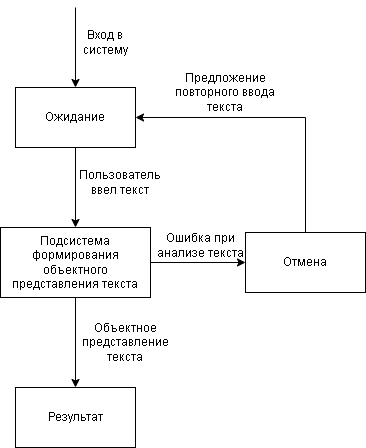
Подсистема для определения частоты слов в тексте.

Это система разбирает текст на слова, после чего подсчитывает их количество и сортирует в порядке убывания.

Подсистема для сбора предложений из слов.

На основе объектного представления текста собирает из слов предложения, учитывая их падежную форму, склонения и т.д. На выходе подсистемы получается понятный для человека текст на естественном языке.

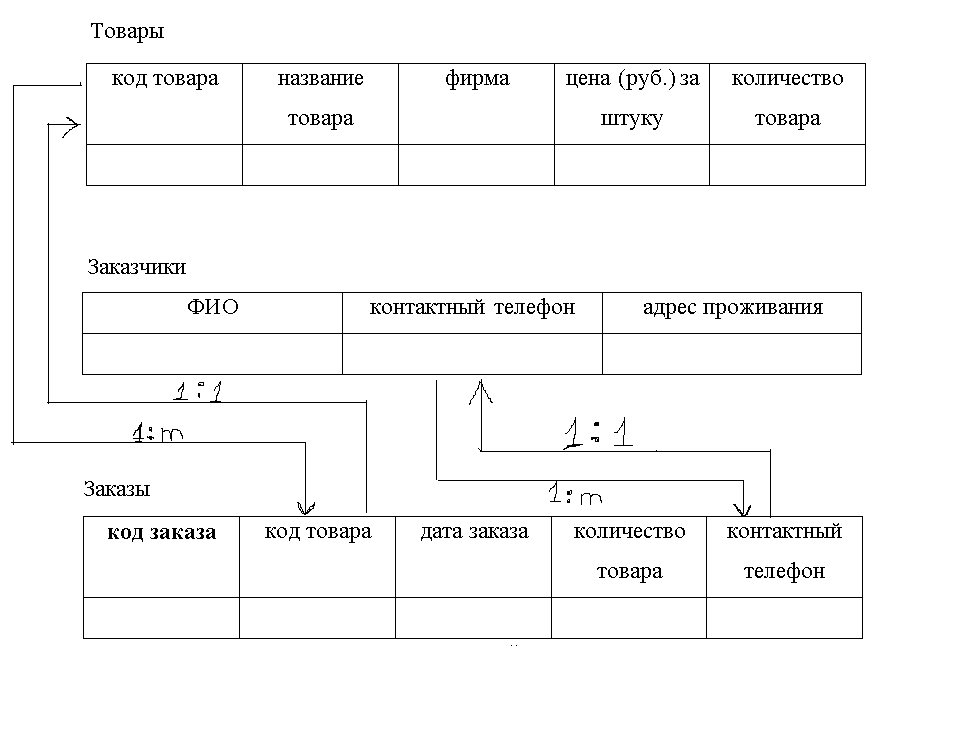
1. Диаграмма перехода из состояния в состояние (ДПСС)



* 1. Таблица состояний.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Состояния  **Состояния** | **Ожидание** | **Подсистема формирования объектного представления текста** | **Отмена** | **Результат** |
| **Ожидание** | false | false | true | false |
| **Подсистема формирования объектного представления текста** | true | false | false | false |
| **Отмена** | false | true | false | false |
| **Результат** | false | true | false | false |

1. Модель данных. ER-модель (фрагмент)



1. Требования к входным данным. Требования к выходным данным

**Требования в входным данным**

Входными данными модуля предобработки текста является строковое представления обрабатываемого текста.

Требования к входным данным:

Текст написан на русском языке;

Текст содержит не менее одного предложения;

Количество предложений текста не превышает 5;

Длина текста не превышает 140 символов;

Каждое предложение текста начинается с большой буквы и завершается точкой.

**Требования к выходным данным**

Выходными данными модуля предобработки текста является объектное представление текста, полученного в качестве входных данных модуля.

Объектное представление представляет собой древовидную структуру, состоящую из листьев-размеченных слов. Каждое размеченное слово содержит следующую информацию, которая может быть использована для дальнейшей обработки текста (в частности модулем определения тональности текста):

* непосредственно слово;
* теги слова;
  + номер альтернативы слова в словаре;
  + морфологические признаки;
  + связи слова с указанием их типов и номеров присоединяемых слов.

1. Макет интерфейса

