Федеральное государственное бюджетное образовательное учреждение высшего образования "МИРЭА - Российский технологический университет"

Детский технопарк "Альтаир"

Сервис для определения основных

тематик текста

Леванова Ксения Сергеевна

10 класс ГБОУ г. Москвы №1564

Руководитель: Русаков Алексей Михайлович

Преподаватель детского технопарка «Альтаир»

**Москва, 2022**

# Оглавление

[Оглавление 2](#_Toc96373936)

[Введение 3](#_Toc96373937)

[Цель 3](#_Toc96373938)

[Задачи 3](#_Toc96373939)

[Методы и этапы 4](#_Toc96373940)

[Актуальность 4](#_Toc96373941)

[Использованные программного решения 4](#_Toc96373942)

[Обзор проектного решения 5](#_Toc96373943)

[Итоги и перспективы 6](#_Toc96373944)

[Список литературы 6](#_Toc96373945)

## Введение

В наше время все больше люди получают электронные файлы, документы и другие источники большого скопления информации, в которых зачастую сложно ориентироваться, а так как через каждого человека проходит ежедневно большое количество информации очень сложно отследить, о чем именно и когда велся диалог. Для этого используются различные технологии поиска и сортировки таких данных. В данной работе предлагается система, позволяющая проанализировать заданную информацию, для легкого ориентирования с помощью сортировки данных.

## Цель

Разработать дополнительный сервис для определения темы, чтобы облегчить и усовершенствовать его использование с помощью определения основных тематик или содержания диалога

Также сделать интереснее и познавательнее времяпрепровождение пользователя в сфере, как различных диалогов, так и получение конкретной информации, о чем был выделенный текст.

## Задачи

* Определить основной функционал программного сервиса
* Провести обзор и анализ современных технологий и выбрать наиболее пригодны для реализации проекта
* Разработать алгоритм для извлечения основных тематик из текста.
* Разработать прототип программного средства
* Разработать алгоритм визуализации основных тематик из текста.
* Определить перспективы развития проекта

## Методы и этапы

|  |  |
| --- | --- |
| Изучение материалов по данной теме | * Знакомство с имеющимися технологиями * Опросы специалистов * Личный опыт |
| Программирование | * Изучение фреймворка Quasar * Изучение языка Python * Изучение платформы VS Code |
| Внедрение | * Выбор самого благоприятного хостинга для реализации проекта |

## Актуальность

Проблема управления и хранения большого количества информации в огромном количестве диалогов в последнее время всё более актуальна. Для упрощения и удобного применения всей информации используются ряд технологий, как поиск по слову и разделения на разные типы вложений диалога. Технологии хорошо зарекомендованы и стандартизированы. Предлагаемая в данной работе технология развивает идею сортировки и упорядочения по основным направлениям:

1. возможность использовать данные для характеристики своего общения
2. возможность визуально увидеть на экране конкретную информацию в сжатом виде, выбранную из огромного количества информации

## Использованные программного решения

Проект создавался на языке программирования Python. Код писался в интерпретаторе Visual Studio. Создавался на платформе Quasar. После нужно выбрать нужный текст и нажать на кнопку.

## Обзор проектного решения

Выбираем текст для определения 4-ех важнейших тем и вставляем в поле. Нажимаем на кнопку «Определить тематики текста»

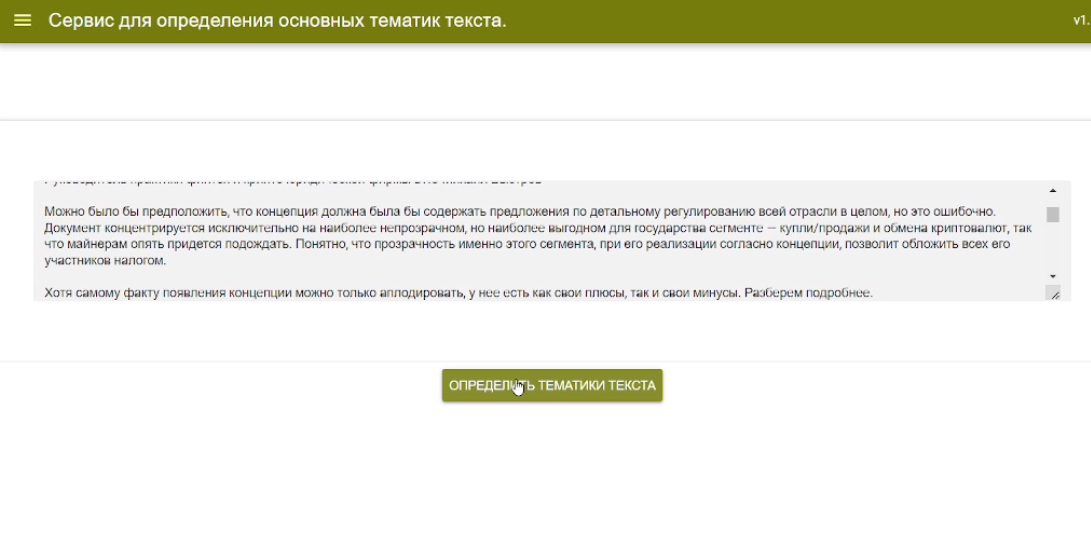


Рис. 1. Исходные данные

После обработки данных мы получаем 4 ряда слов, из которых мы можем сформулировать темы. Самые большие слова – самые значимые в данной теме.

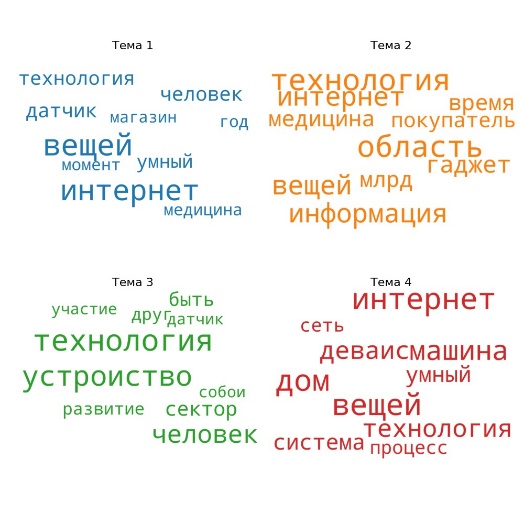


Рис. 2. Темы, полученные путем анализа текста

## Итоги и перспективы

В итоге, в проекте было реализована главная идея и задача:

* Изучены основы создания принципиальных программ;
* Описан функционал разрабатываемого сервиса;
* Разработан прототип программы и ее алгоритмы

Следующими направлениями разработки данного приложения могут быть:

* разработка многопользовательского доступа
* использование сервиса для других проектов
* разработка новых функций, еще более совершенствующих и помогающих в решение данной проблемы

## Список литературы

[Учебник] Федоров Д. Ю. Программирование на языке высокого уровня Python. Изд-во Питер. 2019.

[Учебник]. Бэрри П. Изучаем программирование на Python. Изд-во Litres, 2019.

[Электронный ресурс] Руководство по фреймворку quasar. Дата обновления 01.02.2022. URL: <https://quasar.dev/>

[Учебник] Заяц А. М., Васильев Н. П. Проектирование и разработка WEB-приложений. Введение в frontend и backend разработку на JavaScript и node. js. 2019.

[Учебник] Браун И. Веб-разработка с применением Node и Express. Полноценное использование стека JavaScript. 2-е издание. – Издательский дом Питер, 2021.

[Учебник] Бенджамин Б., Ребекка Б., Тони О. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. – " Издательский дом"" Питер""", 2018.

[Электронный ресурс] Библиотека Gensim Дата обновления 01.02.2022. URL: https://radimrehurek.com/gensim/