

# Разработка программного обеспечения для визуализации тезауруса



ПОДГОТОВИЛИ ПРОЕКТ:

ШЕВЦОВ ВЛАДИСЛАВ СЕРГЕЕВИЧ

ХРУЛЁВ АНАТОЛИЙ СЕРГЕЕВИЧ

10 КЛАСС "А" ШКОЛЫ №1542.

РУКОВОДИТЕЛЬ ПРОЕКТА:

РУСАКОВ АЛЕКСЕЙ МИХАЙЛОВИЧ

ПРЕПОДАВАТЕЛЬ ДЕТСКОГО ТЕХНОПАРКА

«АЛЬТАИР»

# Цель работы

Разработать программное средство для визуализации тезауруса. Для создания данного программного продукта существует большое количество различных математических средств, которые можно скомбинировать для более точного извлечения ключевых слов, формирования их в тезаурус и последующей их визуализации. Визуализация данных предоставит более понятный и информативный вид, который поможет для более углубленного и интенсивного восприятия полученной информации.

Тезаурус — словарь, собрание сведений, корпус или свод, полномерно охватывающие понятия, определения и термины специальной области знаний или сферы деятельности.

# Задачи

- Обзор и анализ современных математических методов обработки текстовой информации.
- Обзор и анализ имеющегося программного обеспечения для обработки текстовой информации;
- Выбор и обоснование инструментальных средств для решения поставленной задачи;
- Разработка и реализация алгоритма визуализации тезауруса;
- Разработки структуры программного обеспечения;
- Анализ экономической эффективности программного обеспечения, подготовка документации.

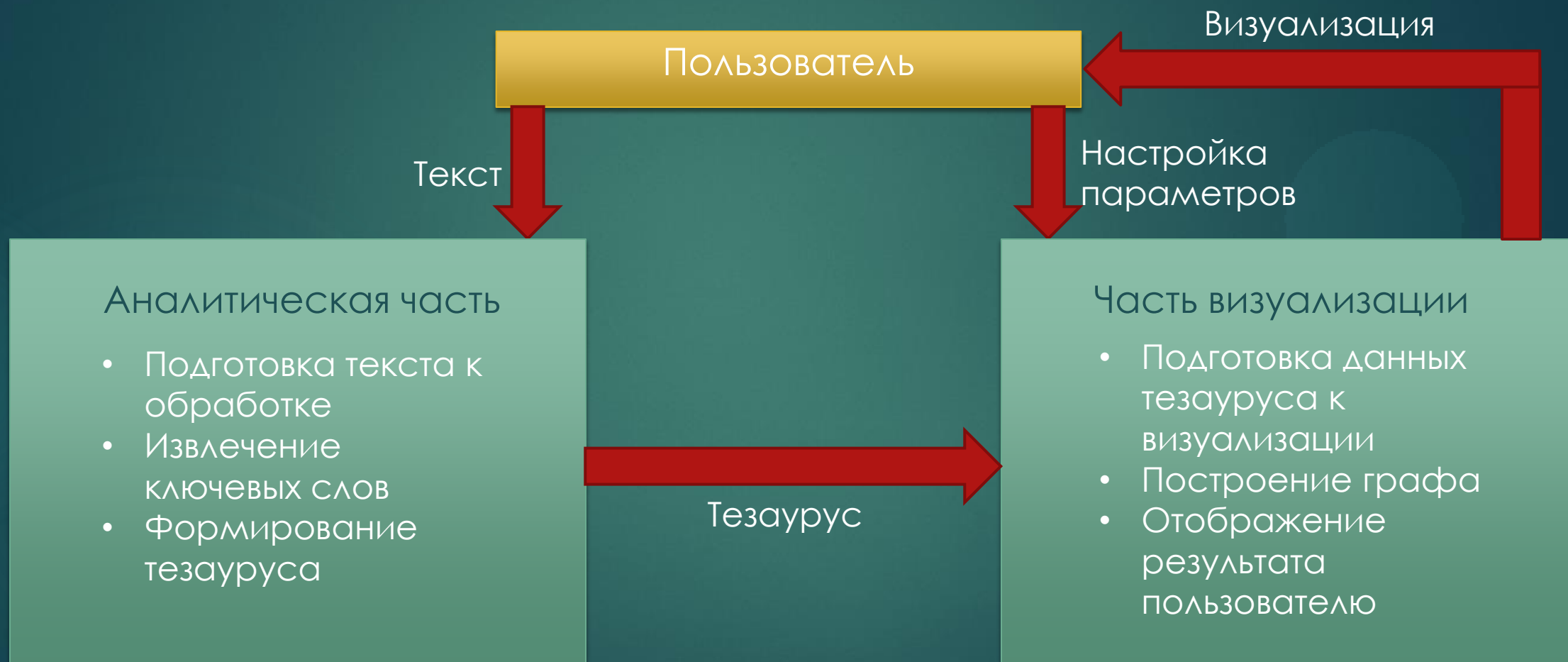
# Этапы работы

- Определение характеристики предметной области
- Обзор и анализ программных решений, пригодных для анализа и визуализации текстовой информации
- Обзор и анализ современных методов анализа текста
- Анализ моделей извлечения ключевых слов
- Анализ моделей визуализации тезауруса
- Разработка структуры программного комплекса
- Обоснование выбора инструментальных средств для разработки ПО
- Описание основных алгоритмов работы ПО
- Разработка и отладка программного обеспечения

# Оборудование

Программное обеспечение (Среда разработки – Visual Code, Языки программирования – Python и JavaScript, JavaScript-библиотеки jQuery и Django), компьютер под управлением OS Windows. Для получения наилучшего результата было произведено ознакомление с технической частью проектов (к примеру, морфологический анализатор pymorphy).

# Структурная схема программного обеспечения



# Исправление ошибок

В ходе разработки остались невыясненными некоторые вопросы, связанные с алгоритмом удаления ненужных для анализа слов, что будет дальнейшим направлением для исследования. В настоящий момент, код, использующий эти функции, временно не используется.

# Результаты

В ходе работы было разработано программное обеспечение для визуализации тезауруса. Также было проведено тестирование разработанной программы, которое показало работоспособность данного программного обеспечения и соответствие поставленным задачам в рамках индивидуального проекта продвинутого уровня. В результате проделанной работы, было разработано программное обеспечение, позволяющее проанализировать текст, извлечь ключевые слова, определить их параметры, сформировать тезаурус, а также визуализировать их в виде графа.



# Обзор и анализ аналогичных программных решений

Критерии сравнения	Thinkmap Visual Thesaurus	Visuwords	Gephi	Advego.com	Наш проект
Платформа	Веб приложение	Веб приложение	Windows/Linux	Веб приложение	Веб приложение
Открытость	Нет	Нет	Да	Нет	Да
Язык словаря	Английский	Английский	Нет	Русский	Русский
Анализ текста	Нет	Нет	Нет	Да	Да
Визуализация информации	Да	Да	Да	Нет	Да
Простота освоения	Нет	Да	Нет	Да	Да
Тип лицензии	Платно	Бесплатно	Бесплатно	Бесплатно	Бесплатно

# Перспективы проекта

Данный проект, в силу расширяемости своего функционала, пригодится в любом маркетинговом сервисе. С помощью этого приложения пользователь может эффективно организовывать поиск и извлечение ключевых слов из текста, формировать из них тезаурус и выполнять визуализацию в виде графа.

Спасибо за внимание!