НПОУ «ЯКУТСКИЙ КОЛЛЕДЖ ИННОВАЦИОННЫХ ТЕХНОЛОГИЙ» (НПОУ «ЯКИТ»)

Отделение информационных технологий и туризма

Специальность 09.02.07 «Информационные системы и программирования»

КУРСОВАЯ РАБОТА

по дисциплине: Языки программирования

по теме:

Создание десктопного приложения для семантического анализа русского языка с использованием библиотеки Hugging Face Transformers на Python

Исполнитель студент гр. КИСП-23(1)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Л.П. Лебедева

подпись, дата

Руководитель,

преподаватель\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Е.С.

подпись, дата

Якутск, 2025

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 3](#_Toc192935699)

[Цели проекта 3](#_Toc192935700)

[Задачи проекта 3](#_Toc192935701)

[Актуальность проекта 4](#_Toc192935702)

[Объект и предмет исследования 4](#_Toc192935703)

[ТЕОРЕТИЧЕСКАЯ ЧАСТЬ 6](#_Toc192935704)

[ПРАКТИЧЕСКАЯ ЧАСТЬ 7](#_Toc192935705)

[1. Архитектура приложения 7](#_Toc192935706)

[2. Реализация приложения 7](#_Toc192935707)

[2.1. Установка необходимых библиотек 7](#_Toc192935708)

[2.2. Код приложения 8](#_Toc192935709)

[ЗАКЛЮЧЕНИЕ 11](#_Toc192935710)

# ВВЕДЕНИЕ

В условиях стремительного роста объемов текстовой информации актуальность автоматизированного анализа текста становится все более заметной. Семантический анализ позволяет извлекать значения и интерпретировать смысл текстовых данных, что находит применение в различных областях, таких как обработка отзывов, анализ настроений и чат-боты. Данный отчет описывает процесс создания десктопного приложения для семантического анализа русского языка с использованием библиотеки Hugging Face Transformers на Python.

Цели проекта

Основная цель проекта заключается в разработке десктопного приложения для семантического анализа русского языка, использующего библиотеку Hugging Face Transformers. Приложение должно обеспечивать пользователям возможность быстро и эффективно анализировать текстовые данные, выявляя их смысл и настроение.

Задачи проекта

1. Исследование теоретических основ семантического анализа: Изучить ключевые аспекты семантического анализа, включая методы и подходы, применяемые в обработке естественного языка.
2. Выбор технологий и инструментов: Определить оптимальный стек технологий для реализации приложения, включая язык программирования, библиотеки и инструменты для создания интерфейса.
3. Разработка интерфейса пользователя: Создать интуитивно понятный графический интерфейс, который позволит пользователям вводить текст и получать результаты анализа.
4. Имплементация модуля обработки текста: Реализовать функциональность семантического анализа с использованием предобученных моделей из библиотеки Hugging Face Transformers.
5. Тестирование и отладка приложения: Провести тестирование приложения, выявить и исправить возможные ошибки, а также убедиться в корректности работы всех функций.
6. Документация и представление результатов: Подготовить документацию по использованию приложения и представить результаты работы проекта заинтересованным сторонам.

Актуальность проекта

Актуальность создания десктопного приложения для семантического анализа текста обусловлена несколькими факторами:

1. Рост объемов текстовой информации: В условиях цифровизации и доступности информации на русском языке возрастает потребность в инструментах, позволяющих эффективно обрабатывать и анализировать текстовые данные.
2. Необходимость в автоматизированных решениях: Бизнес и научные исследования требуют автоматизации процессов анализа текстов, что позволяет сократить время и ресурсы, затрачиваемые на ручной анализ.
3. Развитие технологий обработки естественного языка: Современные модели машинного обучения, такие как трансформеры, значительно увеличили точность и скорость семантического анализа, что делает их применение актуальным и необходимым.

Объект и предмет исследования

Объектом исследования является процесс семантического анализа текстов на русском языке, а также применяемые в этом процессе алгоритмы и технологии обработки естественного языка.

Предметом исследования выступает десктопное приложение, реализующее семантический анализ текста с использованием библиотеки Hugging Face Transformers. В рамках этого исследования рассматриваются:

* Архитектура приложения,
* Алгоритмы обработки текста,
* Эффективность выбранных моделей и методов анализа,
* Пользовательский интерфейс и взаимодействие с пользователем.
* Такое структурирование позволяет более глубоко понять как теоретические, так и практические аспекты семантического анализа на русском языке, а также выявить ключевые факторы, влияющие на успешность реализации проекта.

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Семантический анализ текста включает несколько ключевых этапов:

1. Извлечение значений: Определение значений слов и выражений в зависимости от контекста. Это особенно важно для русского языка, где слова могут иметь разные значения.
2. Анализ отношений: Выявление связей между словами, таких как синонимы, антонимы и гиперонимы.
3. Контекстуальное понимание: Учет контекста, в котором используются слова, для более точной интерпретации их значения.
4. Методы анализа:
   * Семантические сети: Графические представления концепций и их взаимосвязей.
   * Модели на основе трансформеров: Использование предобученных моделей, таких как BERT и его русскоязычные версии, для анализа текста.

Стек технологий

Для реализации десктопного приложения были выбраны следующие технологии:

1. Языки программирования:
   * Python: Основной язык для разработки приложения, обладающий богатой экосистемой библиотек для обработки текста.
2. Библиотеки:
   * Hugging Face Transformers: Библиотека, предоставляющая доступ к предобученным моделям для семантического анализа.
   * Torch: Библиотека для работы с нейронными сетями, необходимая для работы с моделями Transformers.
   * Tkinter: Библиотека для создания графического пользовательского интерфейса (GUI).

ПРАКТИЧЕСКАЯ ЧАСТЬ

Архитектура программного обеспечения

В данной секции подробно описывается процесс разработки десктопного приложения для семантического анализа текста на русском языке, включая архитектуру, реализацию, тестирование и результаты.

1. Архитектура приложения

Приложение состоит из нескольких ключевых компонентов, каждый из которых отвечает за определенную функциональность:

* Графический пользовательский интерфейс (GUI): Позволяет пользователю вводить текст, запускать анализ и получать результаты.
* Модуль обработки текста: Использует модель из библиотеки Hugging Face Transformers для выполнения семантического анализа.
* Модуль управления: Обрабатывает взаимодействие между интерфейсом и модулем обработки, управляет передачей данных.

Схема архитектуры приложения

┌──────────────────────────┐

│ Графический интерфейс │

│ (Tkinter, кнопки и текст)|

└───────────────┬──────────┘

│

▼

┌──────────────────────────┐

│ Модуль управления │

│ (логика взаимодействия) │

└───────────────┬──────────┘

│

▼

┌──────────────────────────┐

│ Модуль обработки текста │

│ (Hugging Face Transformers)|

└──────────────────────────┘

2. Реализация приложения

2.1. Установка необходимых библиотек

Перед началом разработки необходимо установить необходимые библиотеки. Для этого выполните следующую команду:

pip install transformers torch tkinter

2.2. Код приложения

Ниже представлен полный код приложения, реализующего семантический анализ текста:

2.3. Описание кода

1. Импорт библиотек: Импортируем необходимые модули из tkinter для создания интерфейса и transformers для выполнения семантического анализа.
2. Класс SemanticAnalysisApp: Основной класс приложения, который инициализирует интерфейс и обрабатывает ввод пользователя.
   * Метод \_\_init\_\_: Создает элементы интерфейса, такие как метка (label), текстовая область (text area) и кнопка (button).
   * Метод analyze\_text: Вызывается при нажатии кнопки. Извлекает текст из текстовой области и передает его методу perform\_semantic\_analysis.
   * Метод perform\_semantic\_analysis: Использует предобученную модель для анализа текста и возвращает результаты.
3. Запуск приложения: В конце скрипта создается экземпляр Tkinter и запускается основной цикл приложения.

3. Тестирование приложения

Тестирование приложения включает следующие этапы:

1. Функциональное тестирование:
   * Ввод различных текстов для анализа (положительные, отрицательные и нейтральные).
   * Проверка корректности результатов, возвращаемых моделью.
2. Пользовательское тестирование:
   * Оценка удобства интерфейса пользователями.
   * Сбор отзывов о функциональности и возможных улучшениях.
3. Нагрузочное тестирование:
   * Проверка производительности приложения при вводе больших объемов текста.

4. Результаты

После успешной реализации и тестирования приложение позволяет пользователю:

* Вводить текст для анализа.
* Получать результаты семантического анализа на русском языке в виде меток, отражающих настроение текста (например, положительное, отрицательное, нейтральное).
* Интуитивно взаимодействовать с приложением благодаря простому и понятному интерфейсу.

ЗАКЛЮЧЕНИЕ

Создание десктопного приложения для семантического анализа текста на русском языке с использованием библиотеки Hugging Face Transformers представляет собой эффективный способ автоматизации анализа текстовой информации. Использование современных методов обработки естественного языка позволяет значительно улучшить качество анализа и расширить возможности взаимодействия с пользователями. Разработка такого приложения может быть полезной в различных областях, включая маркетинг, социологические исследования и автоматизацию клиентского обслуживания.