СОДЕРЖАНИЕ

Введение1

1. Теоретическая часть3
   1. Терминология по проекту и глоссарий3
   2. Стек технологий5
   3. Поставленные задачи6
2. Практическая часть8
   1. Архитектура ПО8
   2. Разработка проекта9
   3. Результат по поставленным задачам12

Заключение14

Список литературы15

ВВЕДЕНИЕ

Современные технологии коммуникации играют ключевую роль в повседневной жизни, бизнесе и социальном взаимодействии. Однако просто отправки сообщений уже недостаточно - важно понимать эмоциональную окраску передаваемой информации. Сентимент-анализ (анализ тональности текста) позволяет автоматически определять эмоциональную нагрузку сообщений, что может быть полезно для улучшения качества общения, анализа обратной связи и принятия решений на основе данных.

**Актуальность** проекта обусловлена растущим спросом на инструменты, способные не только передавать текстовые сообщения, но и анализировать их эмоциональную составляющую. Такие решения могут применяться в службах поддержки, соцсетях, бизнес-коммуникациях и психологических исследованиях.

**Цель проекта** - разработка десктопного приложения для ввода текста и дальнейшего анализа его тональности с выводом в отдельное поле для результатов, позволяющего пользователям оценивать эмоциональную окраску текста в реальном времени.

**Задачи проекта:**

* Изучить принципы работы сентимент-анализа и существующие решения.
* Разработать архитектуру десктопного приложения с поддержкой отправки сообщений.
* Реализовать алгоритм анализа тональности текста.
* Обеспечить удобный интерфейс для взаимодействия с пользователем.
* Протестировать приложение на точность определения эмоциональной окраски текста.

**Объект исследования** - процесс автоматического анализа тональности текста и его интеграция в системы обмена сообщениями.

**Предмет исследования** - методы и алгоритмы сентимент-анализа, применяемые для оценки эмоциональной окраски текстовых сообщений в реальном времени.

Данный проект представляет собой практическую реализацию современных методов обработки естественного языка и может быть полезен как для индивидуального использования, так и для интеграции в бизнес-процессы, требующие анализа эмоциональной реакции пользователей.

1. Теоретическая часть
   1. Терминология по проекту и глоссарий

Сентимент анализ – метод обработки текстовых данных, направленный на определение эмоциональной окраски высказываний. Он классифицирует сообщения как положительные, отрицательные или нейтральные с использованием алгоритмов обработки естественного языка и методов машинного обучения

* Машинное обучение

Машинное обучение представляет собой область искусственного интеллекта, позволяющую программам самостоятельно находить закономерности в больших объемах данных и делать предсказания. В контексте сентимент анализа алгоритмы обучаются на размеченных текстах, чтобы впоследствии автоматически определять эмоциональную тональность новых сообщений

* NLTK (Natural Language Toolkit)

Это одна из самых старых и наиболее полных библиотек для NLP на Python. Она предоставляет множество инструментов для различных задач NLP, таких как токенизация, стемминг, лемматизация, парсинг и многое другое. NLTK широко используется в академических кругах и является отличным инструментом для обучения основам NLP.

* Токенизация

Разделение текста на слова и предложения. Это первый шаг в большинстве задач NLP, так как позволяет разбить текст на более мелкие компоненты для дальнейшего анализа.

* Стемминг и лемматизация

Приведение слов к их базовой форме. Стемминг удаляет окончания слов, а лемматизация приводит слова к их канонической форме.

* Частеречная разметка

Определение частей речи для каждого слова в тексте. Это важно для понимания грамматической структуры текста.

* Парсинг

Анализ синтаксической структуры предложений. Позволяет понять, как слова в предложении связаны друг с другом.

* Корпусы и лексиконы

Доступ к большим наборам текстов и словарей. Это полезно для обучения моделей и проведения исследований.

* Обработка естественного языка (NLP)

Это область искусственного интеллекта, которая занимается взаимодействием между компьютерами и человеческим языком. NLP позволяет компьютерам анализировать, понимать и генерировать человеческий язык. Это важно для множества приложений, таких как чат-боты, автоматический перевод, анализ тональности, автоматическое резюмирование текстов, и многое другое. В современном мире, где объем текстовой информации растет экспоненциально, способность эффективно обрабатывать и анализировать текстовые данные становится критически важной.

NLP используется в различных отраслях, включая здравоохранение, финансы, маркетинг и даже правосудие. Например, в здравоохранении NLP помогает анализировать медицинские записи и исследования, чтобы улучшить диагностику и лечение пациентов. В финансах NLP используется для анализа новостей и отчетов, чтобы предсказать рыночные тренды. В маркетинге анализ тональности помогает понять, как клиенты реагируют на продукты и услуги, что позволяет компаниям лучше адаптировать свои предложения.

* Классификация

Классификация – метод распознавания и группировки объектов (в данном случае текстовых сообщений) по предварительно заданным категориям, что в сентимент анализе позволяет определять позитивные, негативные или нейтральные высказывания

* Глубокое обучение

Глубокое обучение – подвид машинного обучения, основанный на использовании нейронных сетей с несколькими слоями, позволяющих выявлять сложные закономерности в данных. Применяется для улучшения точности анализа эмоциональной окраски текста

* API

API (интерфейс программирования приложений) – набор правил и протоколов, с помощью которых обеспечивается обмен данными между различными компонентами приложения. В рассматриваемом решении API обеспечивает интеграцию модулей отправки сообщений и анализа их содержания

* UI и UX

UI (пользовательский интерфейс) относится к визуальному оформлению приложения, а UX (пользовательский опыт) – к удобству использования и удовлетворенности пользователя от взаимодействия с продуктом. Хорошее сочетание UI и UX является критически важным для успешного восприятия приложения конечными пользователями

* Предварительная обработка данных

Предварительная обработка данных (Data Preprocessing) включает очистку, нормализацию текста, удаление стоп-слов, пунктуации и лишних символов. Этот этап помогает подготовить данные к дальнейшему анализу и снижает шум, повышая точность классификации

* Нейронные сети

Нейронные сети – вычислительные модели, имитирующие работу биологических нейронов, применяемые для сложных задач классификации и распознавания образов. В сентимент анализе их используют для выявления скрытых закономерностей в языке

* 1. Стек технологий
* Язык программирования

Использование Python для реализации серверной логики, алгоритмов обработки естественного языка и машинного обучения. Дополнительно можно применять JavaScript (или TypeScript) при создании клиентской части

* Фреймворк для серверной разработки

FastAPI или Flask для разработки REST API, обеспечивающих обмен данными между клиентской частью и сервером. Эти фреймворки позволяют быстро разворачивать и масштабировать приложение

* Библиотеки для обработки естественного языка

NLTK, spaCy или библиотеки от HuggingFace (например, Transformers) – для обработки текстов, токенизации, лемматизации и анализа тональности сообщений. Они позволяют интегрировать готовые модели сентимент анализа или обучить свои

* Фреймворки и библиотеки машинного обучения

scikit-learn для классического машинного обучения (например, наивный Байес, логистическая регрессия) и TensorFlow или PyTorch для глубокого обучения. В зависимости от сложности и требований к точности можно выбирать подходящий инструмент

* Система контроля версий и CI/CD

Git для контроля версий кода, GitLab CI, Jenkins или GitHub Actions для организации процессов непрерывной интеграции и доставки, позволяющих быстро тестировать и деплоить обновления

* Мониторинг и логирование

Инструменты вроде Prometheus для мониторинга, Grafana для визуализации метрик и ELK-стека (Elasticsearch, Logstash, Kibana) для централизованного логирования, что поможет оперативно обнаруживать и устранять проблемы.

* 1. Поставленные задачи

Разработать проект в Microsoft Visual Studio Code на языке программирования Python, суть проекта состоит в разработке десктопного приложения для анализа тональности текста. Использование библиотек для определения тональности текста.

Разработать интерфейс для приложения, расположение, шрифт и размер текста, поле для ввода текста, поле для отображения результата анализа текста.

Разработка метода анализа тональности текста через библиотеку TextBlob для обработки текстовых данных на языке Python. Она предоставляет простой API для решения задач обработки естественного языка (NLP).

1. Практическая часть
   1. Архитектура ПО

Microsoft Visual Studio Code редактор исходного кода, разработанный Microsoft для Windows, Linux и macOS. Позиционируется как «лёгкий» редактор кода для кроссплатформенной разработки веб- и облачных приложений.

Языке программирования Python мультипарадигмальный высокоуровневый язык программирования общего назначения с динамической строгой типизацией и автоматическим управлением памятью, ориентированный на повышение производительности разработчика, читаемости кода и его качества, а также на обеспечение переносимости написанных на нём программ.

Библиотека TextBlob это библиотека Python, которую можно использовать для обработки текстовых данных. Среди задач, которые можно решать с ее помощью: анализ тональности, токенизация, исправление орфографии и многие другие задачи обработки естественного языка. В этой статье я покажу вам TextBlob в Python.

Библиотека tkinter пакет для языка Python, нужный для работы со средствами Tk. Библиотека Tk написана на языке программирования Tcl и содержит в себе компоненты GUI. Входит в стандартную библиотеку Python.

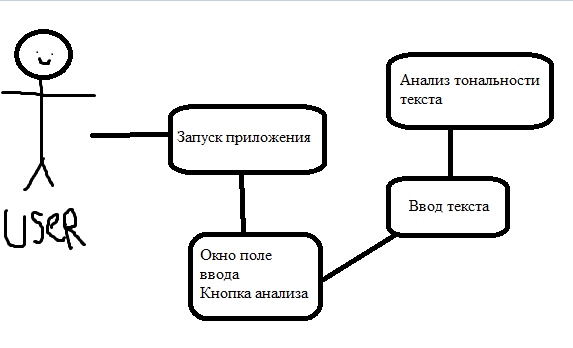


Рис 1. Use case diagram.

* 1. Разработка проекта
* Скачиваем библиотеку TextBlob через терминал.

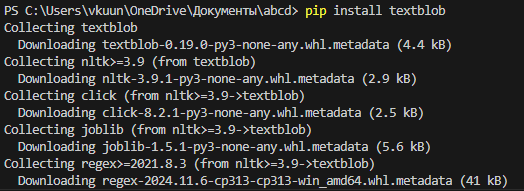


Рис 2. Библиотека textblob.

* Импортируем наши библиотеки.

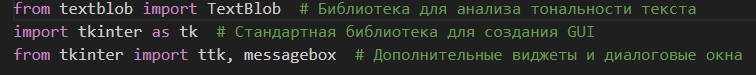


Рис 3. Импорт библиотек.

* Создаем основной класс нашего приложения.

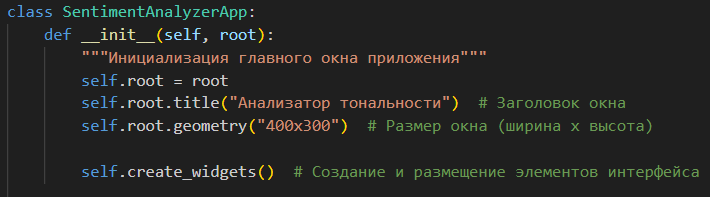


Рис 4. Инициализатор главного окна.

* Пишем код для нашего минималестичного интерфейса.

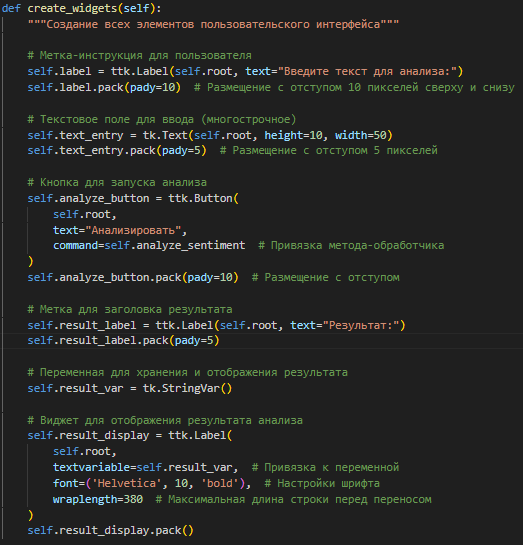


Рис 5. Создание элементов интерфейса.

* Делаем основной метод анализа тональности текста.

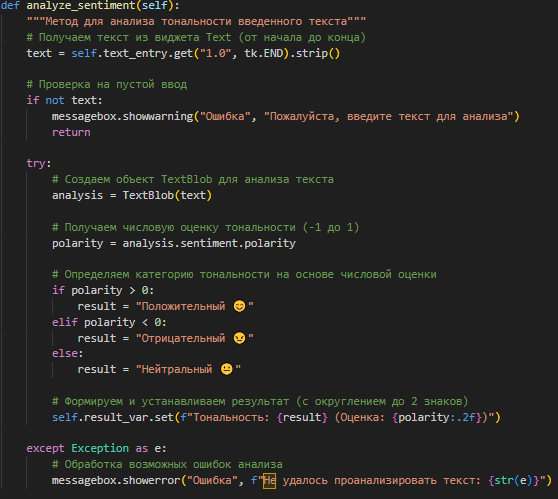


Рис 6. Метод сентимент анализа.

* Основной код для работы приложения.

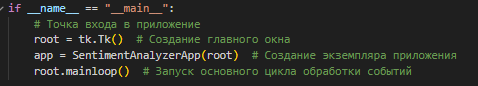


Рис 7. Запуск приложения.

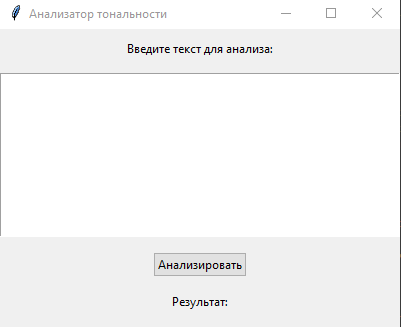


Рис 8. Основное окно приложения.

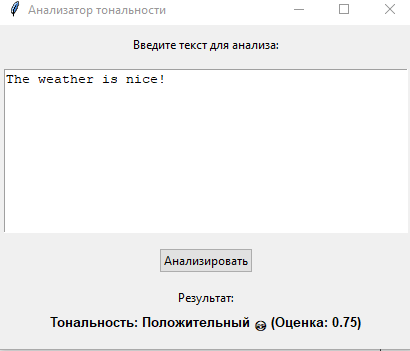


Рис 9. Пример работы приложения.

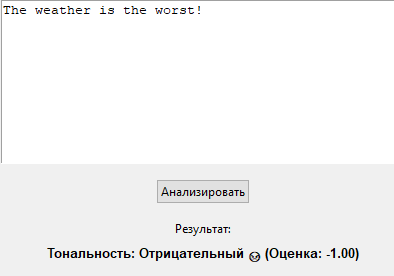


Рис 10. Пример работы с отрицательным результатом.

* 1. Результаты по поставленным задачам

|  |  |  |
| --- | --- | --- |
| Характеристика | Текущий проект | Альтернативный проект |
| Поддержка языков | Только английский | 10+ языков |
| Интерфейс | Минималистичный | Dashboard с графиками и историей запросов |
| Точность анализа | Базовый алгоритм TextBlob | Ансамбли моделей (BERT + LSTM, точность 85-90%) |
| Визуализация | Текст + эмодзи | Интерактивные графики |
| Доп. функции | — | Экспорт отчетов, API для интеграций |
| Производительность | Мгновенный результат | Задержка 2-3 сек (сложные модели) |
| Системные требования | Любой компьютер | Требуется GPU для обработки больших текстов |
| Стек технологий | Python + Tkinter + TextBlob | Python + FastAPI + Transformers |

Заключение

В ходе выполненной практической работы было выполнено создание приложения для анализа тональности текста. Во время работы были изучены и освоены практические навыки программирования.

Список литературы

[Сентимент-анализ. Определяем эмоциональные сообщения на Хабре / Хабр](https://habr.com/ru/companies/simbirsoft/articles/539508/)

[Сентимент анализ текста / Хабр](https://habr.com/ru/companies/brandanalytics/articles/262595/)

[Анализ тональности в русскоязычных текстах, часть 1: введение / Хабр](https://habr.com/ru/companies/vk/articles/516214/)

[Анализ тональности в русскоязычных текстах, часть 2: основные исследования / Хабр](https://habr.com/ru/companies/vk/articles/516726/)

[Анализ тональности в русскоязычных текстах, часть 3: вызовы и перспективы / Хабр](https://habr.com/ru/companies/vk/articles/516730/)

[Анализ тональности текстов с помощью сверточных нейронных сетей / Хабр](https://habr.com/ru/companies/vk/articles/417767/)

[Анализ текстовых данных с помощью NLTK и Python / Хабр](https://habr.com/ru/companies/otus/articles/774498/)

[Библиотеки для NLP на Python](https://sky.pro/wiki/python/biblioteki-dlya-nlp-na-python/)