СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 2](#_Toc194748915)

[ТЕОРИТИЧЕСКАЯ ЧАСТЬ 3](#_Toc194748916)

[1.1 ТЕРМИНОЛОГИЯ ПО ПРОЕКТУ И ГЛОССАРИЙ 3](#_Toc194748917)

[1.2 РАСПРЕДЕЛЕНИЕ РОЛЕЙ И РАБОТЫ 3](#_Toc194748918)

[1.2 СТЕК ТЕХНОЛОГИЙ 4](#_Toc194748919)

# ВВЕДЕНИЕ

В наше время современный рынок электронной коммерции характеризуется высокой конкуренцией, где ключевым фактором успеха становится удовлетворенность клиентов. Одним из основных источников обратной связи являются отзывы покупателей, анализ которых позволяет компаниям улучшать продукты, корректировать маркетинговые стратегии и повышать уровень сервиса.

Одним из самых популярных маркетплейсов является Wildberries который является одним из самых крупнейших маркетплейсов в России, и в некоторых странах СНГ, на котором ежедневно публикуются десятки тысяч отзывов. Ручной анализ такого объема данных будет очень трудно, поэтому программа веб парсинга сайтов является самым удобным способом для анализа большого объема отзывов.

Актуальность темы:

1. Веб парсинг - автоматизированный сбор и структурирование информации с сайтов при помощи программы или сервиса.

2. Сентимент-анализ позволяет быстро оценить общее настроение клиентов и выявить проблемные зоны.

3. Графики после анализа.

4. Интерфейс.

Цель работы:

1. Разработка программы веб парсинг сайта wildbberies на языке программирования Python.

2. Разработка и работа с моделькой (сентимент анализ).

3. Разработка функции графиков после анализа.

4. Разработка интерфейса.

После того как рассмотрели введению перейдем к теоретической части и потом приступим к самой разработки программы.

# ТЕОРИТИЧЕСКАЯ ЧАСТЬ

# ТЕРМИНОЛОГИЯ ПО ПРОЕКТУ И ГЛОССАРИЙ

Веб парсинг – автоматизированное получение информации с веб-сайтов с помощью программ, которые называются парсерами. Парсеры собирают и систематизируют определенный контент с веб-страниц, документов, API или даже исходного кода, например, с помощью веб-парсинга можно собрать отзывы. Парсинг имеет широкий круг применения можно использовать для анализа различных типов контента с веб-сайтов, такие как сведения о конкурентах, отзывы клиентов, новостные статьи, контент в социальных сетях, аудитории групп, частоту поисковых запросов, популярных заголовки и связанные ключевые слова.

Sentiment analysis (по-русски, анализ тональности) – это область компьютерной лингвистики, которая занимается изучением мнений и эмоций в текстовых документах. Целью анализа тональности является нахождение мнение в тексте и определение его свойств. В зависимости от поставленной задачи нас могут интересовать разные свойства, например:

1. Автор – кому принадлежит это мнение (Александр Жуков)
2. Тема – о чем говорится во мнении (выступление российский олимпийцев)
3. Тональность – позиция автора относительно упомянутый темы (положительная)

Интерфейс – это набор инструментов, который позволяет пользователю взаимодействовать с программой. В более широком смысле термин обозначает любые инструменты для соприкосновения между разными системами и сущностями. Если говорить простыми словами, интерфейс – это внешний вид сайта, программы или приложения.

График – вид изобразительного искусства, в котором основными изобразительными средствами являются свойства изобразительной поверхности и тональные отношения линий, штрихов и пятен.

# 1.2 РАСПРЕДЕЛЕНИЕ РОЛЕЙ И РАБОТЫ

Романов Дмитрий Дмитриевич отвечает за веб-парсинг

План по выполнению работы по части веб-парсинга:

1. Установить библиотеку Selenium
2. Написать код на языке Python используя библиотеку Selenium
3. С помощью исходного кода, найти отзывы по типу «<div class = “feedback\_info” </div>»

Антонов Викториан Гаврильевич отвечает за анализ тональности

План по выполнению работы по части анализа точности:

1. Изучение массива
2. Определить позитивные, негативные и нейтральные эмоции

Андрей Мазуров отвечает за график

План по выполнению работы по части графика:

1. Создание графика

Андрей Борисов отвечает за интерфейс

План по выполнению работы по части:

1. Создание интерфейса

# 1.2 СТЕК ТЕХНОЛОГИЙ

Selenium -

Исходный код страницы