МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ИНСТИТУТ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ СИСТЕМ

ФАКУЛЬТЕТ КОМПЬЮТЕРНЫХ И ФИЗИКО-МАТЕМАТИЧЕСКИХ НАУК

КАФЕДРА ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ

Допущен к защите

Заведующий кафедрой ПМИ

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Е. В. Разова

**АСПЕКТНЫЙ АНАЛИЗ ТОНАЛЬНОСТИ ОТЗЫВОВ О ФИЛЬМАХ**

Курсовой проект по дисциплине

«Теория алгоритмов»

Выполнил студент группы ФИб-3301-51-00 \_\_\_\_\_\_\_\_\_\_ М. С. Крупин

Руководитель,

канд. пед. наук, доцент кафедры ПМИ \_\_\_\_\_\_\_\_\_\_ Е. В. Разова

Работа защищена с оценкой \_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_2022 г.

Члены комиссии: \_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

Киров 2022

Оглавление

[Введение 3](#_Toc99296423)

[1 Обзор сайтов с отзывами о фильмах 5](#_Toc99296424)

[1.1 Кинопоиск 5](#_Toc99296425)

[1.2 Ivi 6](#_Toc99296426)

[1.3 Отзовик 6](#_Toc99296427)

[1.4 СпасибоВсем.ру 7](#_Toc99296428)

[1.5 Irecommend.ru 8](#_Toc99296429)

[1.6 Выводы по разделу 9](#_Toc99296430)

[2 Обзор методов аспектного анализа тональности 9](#_Toc99296431)

[2.1 Частотный метод 9](#_Toc99296432)

[2.2 Синтаксический метод 10](#_Toc99296433)

[2.3 Машинное обучение с учителем 10](#_Toc99296434)

[2.4 Машинное обучение без учителя 11](#_Toc99296435)

[Библиографический список 12](#_Toc99296436)

[Приложения 14](#_Toc99296437)

[Приложение А 14](#_Toc99296438)

# Введение

Эпоха информационного общества характеризуется постоянно растущими объёмами данных, которыми делятся пользователи со всего мира. Пользователям Интернета постоянно предлагают поделиться своим мнением, оценкой по поводу какой-либо новости, качеству обслуживания, приобретённого товара. Желание людей узнать мнение о том товаре, который они собираются приобрести, у других людей, которые его уже приобрели, привело к росту количества отзывов практически ко всему. На веб-сайтах, предоставляющих какие-либо услуги или продающих товары, начали появляться специальные места, где пользователи могут оставить свой отзыв. Ведь мнение человека о чём-либо в большой части строится от мнений других людей. Также стали даже появляться сайты, предназначенные только для отзывов по продуктам какой-либо тематики, например: «Кинопоиск», «Отзовик», «IMDb».

С точки зрения производителя любой человек – потенциальный покупатель. Следовательно, знание их симпатий и антипатий может оказать большую помощь в разработке новых продуктов и улучшении существующих. Здесь и пригождается аспектных анализ тональности, который систематизирует отзывы, выделяет в каждом из них плюсы и минусы определённого аспекта товара.

Актуальность данной работы заключается в разработке программы, которая будет проводить аспектных анализ отзывов на русском языке, так как при изучении литературы и Интернета таковой найдено не было. Она может применяться на сайте любой организации, которая чем-либо торгует, и хочет иметь статистику об отзывах о своих продуктах. Она также может быть полезно и для самих пользователей в формировании общей оценки о товаре по всем отзывам, либо быстрому и ясному пониманию оценки товара в конкретном отзыве.

Исходя из актуальности, можно выделить цель данной работы: разработать программу для аспектного анализа тональности отзывов о фильмах.

Для достижения цели были поставлены следующие задачи:

1. Сделать обзор сайтов с отзывами о фильмах;
2. Написать парсер отзывов с выбранного сайта;
3. Сделать обзор методов аспектного анализа тональности;
4. Разработать программу аспектного анализа тональности фильмов;
5. Дать оценку качества разработанной программы.

Курсовой проект состоит из пяти разделов, заключения, библиографического списка и приложений. Содержания разделов соответствуют поставленным задачам.

1 Обзор сайтов с отзывами о фильмах

Для аспектного анализа отзывов о фильмах необходимо сформировать базу этих отзывов. Желательно, чтобы у каждого отзыва уже была его оценка по различным аспектам, так как так будет легче проверять правильность работы программы аспектного анализа. Также было бы плюсом, если бы отзыв делился на составляющие, по которым будет проще его анализировать. Например: достоинства, недостатки, пожелания.

1.1 Кинопоиск

На данном сайте отзывы бывают трёх видов:

1. Положительные, помечаются зелёным фоном;
2. Отрицательные, помечаются красным фоном;
3. Нейтральные, помечаются серым фоном;

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 1.1 – Пример отзыва с сайта «Кинопоиск»

Информации о содержании отзыва достаточно мало. Для аспектного анализа недостаточно знать эмоциональную окраску отзыва. Сам отзыв не делится на составляющие, только если сам автор этого не сделает.

Для Кинопоиска существует API, с помощью которого можно получать всю информацию об отзыве, не прибегая к парсингу.

1.2 Ivi

На данном сайте об отзыве не предоставляется никакой информации. Однако есть общий аспектный анализ отзывов о фильмах, состоящий из режиссуры, зрелищности, игры актёров и сюжета. Каждый из аспектов оценивается по стобалльной шкале.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 1.2 – Пример отзыва с сайта «Ivi»

1.3 Отзовик

Каждый отзыв сайта «Отзовик» делится на пункты: «Достоинства», «Недостатки» и основной раздел. В конце отзыва приводится аспектная оценка фильма, составленная пользователем, которая состоит из сюжета, зрелищности, игры актёров, оригинальности, музыки и звука. Также автор ставит оценку от 1 до 5 и рекомендует или не рекомендует этот фильм. По данным причинам этот сайт хорошо подходит для составления размеченного корпуса отзывов о фильмах.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 1.3 – Пример отзыва с сайта «Отзовик»

1.4 СпасибоВсем.ру

Структура отзыва похожа на структуру отзыва с сайта «Отзовик», однако в оценке отсутствует шкала «Оригинальность». Сам сайт менее популярен по сравнению с вышеупомянутом с сайтом, поэтому отзывов на нём меньше, и собрать их достаточное количество будет труднее.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 1.4 – Пример отзыва с сайта «СпасибоВсем.ру»

1.5 Irecommend.ru

На этом сайте пользователь ставит оценку фильму от 1 до 5, пишет отзыв и по желанию отмечает достоинства и недостатки. То, что достоинства и недостатки есть не в каждом отзыве, значительный минус.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 1.5 – Пример отзыва с сайта «Irecommend.ru»

1.6 Выводы по разделу

В данном разделе был проведён обзор пяти сайтов с отзывами о фильмах. В каждом были выделены достоинства, которые бы лучше всего поспособствовали построению модели анализа аспектов фильма. В результате для дальнейшего использования и создания на его основе размеченного корпуса отзывов был выбран наилучший сайт – «Отзовик», на котором отзыв является структурированным (достоинства, недостатки, основной раздел), даётся оценка по пятибалльной шкале, общая рекомендация, а самое главное – оценка фильма по аспектам: «Сюжет», «Зрелищность», «Игра актёров», «Оригинальность», «Музыка и звук».

[сделать ссылки на литературу]

2 Обзор методов аспектного анализа тональности

2.1 Частотный метод

Данный метод заключается в выявлении наиболее часто встречающихся существительных и их сочетаний, которые и будут являться аспектами. Далее для выявления отношения к ним выбираются прилагательные и другие эмоционально окрашенные слова, находящиеся неподалёку.

Плюсы:

1. Простой, но достаточно эффективный.

Минусы:

1. Выявляет много неаспектов;
2. Пропускает низкочастотные аспекты;
3. Требует ручной настройки: какие существительные и с какой частотой будут считаться аспектами.

2.2 Синтаксический метод

Данный метод находит аспекты на основе синтаксических отношений, в которых они находятся. Очень частым отношением является отношение существительного и прилагательного, например, словосочетание «вкусная еда». Здесь чаще всего поиск происходит по прилагательным, так как они почти всегда принадлежат к какому-то существительному.

Плюсы:

1. Находит малочастотные аспекты;
2. Эмоционально окрашенные слова известны или их легко найти.

Минусы:

1. Выделяется много неаспектов;
2. Необходимо ручное составление шаблонов, по которым находятся пары «аспект – отношение к нему».

2.3 Машинное обучение с учителем

В данном случае потребуется размеченный набор данных с заранее известными аспектами и их тональностями, так как на каждом этапе обучения с учителем полученный ответ сравнивается с правильным.

Плюсы:

1. Не требует ручной настройки параметров;
2. Не пропускает низкочастотные аспекты;
3. Не выделяет лишние аспекты.

Минусы:

1. Необходимы размеченные данные.

2.4 Машинное обучение без учителя

В машинном обучении без учителя размеченного набора данных не требуется, алгоритм самостоятельно будет пытаться выявить аспекты с тональностями по каким-либо признакам.

Плюсы:

1. Размеченные данные не нужны;
2. Не пропускает низкочастотные аспекты;
3. Не выделяет лишние аспекты.

Минусы:

1. Требует ручной настройки параметров.
2. Обучение дольше, чем с учителем, так как машина должна сама определять закономерности тональностей и аспектов в отзывах.

# Библиографический список

1. Андрианов И. А. Современные методы аспектно-ориентированного анализа эмоциональной окраски / И. А. Андрианов, В. Д. Майоров, Д. Ю. Турдаков // Труды Института системного программирования РАН. – 2015. – Том 27, вып. 5. – сс. 5-21.

2. IEEE 2018 Fourteenth International Conference on Information Processing, Dec. 21st – Dec. 23rd, 2018: Aspect Based Sentiment Analysis on Product Reviews / A. P. Rodrigues, N. N. Chiplunkar. – Nitte, 2018 – 6 p.

3. Международная конференция «Диалог», 27.05.2015 – 30.05.2015: Извлечение аспектов, тональности и категорий аспектов на основании отзывов пользователей о ресторанах и автомобилях / В. В. Иванов, Е. В. Тутубалина, Н. Р. Мингазов, И. С. Алимова. – КФУ, 2015 – сс. 46-57.

4. Международная конференция «Диалог», 27.05.2015 – 30.05.2015: Высокоточный метод извлечения аспектных терминов для русского языка / В. Майоров, В. Аванесов, И. Андрианов, Н. Астраханцев. – Москва, 2015 – сс. 58-67.

5. Международная конференция «Диалог», 27.05.2015 – 30.05.2015: Sentirueval: Тестирование систем анализа тональности текстов на русском языке по отношению к заданному объекту / Н. В. Лукашевич, П. Д. Блинов, Е. В. Котельников, Ю. В. Рубцова, В. В. Иванов, Е. В. Тутубалина. – Москва, 2015 – сс. 12-24.

6. Международная конференция «Диалог», 27.05.2015 – 30.05.2015: Sentirueval: Анализ тональности твитов о телекоммуникациях и банках на основе метода машинного обучения в рамках Sentirueval / Е. В. Тутубалина, М. А. Загулова, В. В. Иванов, В. А. Малых. – Москва, 2015 – сс. 89-99.

6. Byrne. S. Aspect Identification and Sentiment Analysis in Text-Based Reviews / S. Byrne // Lehigh Preserve Institutional Repository. – 2017. – Lehigh University – 90 p.

7. Chakraborty S. Aspect-based Sentiment Analysis of Scientific Reviews / S. Chakraborty, P. Goyal, A. Mukherjee – Electronic article. – 2020. – Access mode: https://arxiv.org/pdf/2006.03257.pdf (accessed 27 March, 2022).

8. 7th International Conference on Advances in Computing & Communications, Aug. 22nd – Aug. 24th, 2017: Aspect based Sentiment Oriented Summarization of Hotel Reviews / A. P. Rodrigues, N. N. Chiplunkar. – Cochin, 2017 – 9 p.

9. Schouten K. Survey on Aspect-Level Sentiment Analysis / K. Schouten, F. Frasincar // IEEE Transactions on Knowledge and Data Engineering. – 2016. – Vol. 28, Issue 3. – pp. 813-830.

# Приложения

Приложение А