**Н** **И** **Т** **У** **«** **М** **И** **С** **и** **С** **»**

**ОТЧЕТ** по

**ЛАБОРАТОРНОЙ** **РАБОТЕ**

*«Разработка* *демонстрационного* *прототипа* *приложения* *для* *решения* *специализированной* *задачи* *интеллектуальной* *обработки* *и* *анализа* *информации* *с* *использованием* *современных* *ИИ-сервисов*

*(на* *примере* *системы* *IBM* *Watson* *и* *аналогов)»*

Учебная дисциплина «Методы искусственного интеллекта»

**Группа:** БПМ-17-1

**Студенты:** Бердичевская А. Г.,

Куликова Д. М.

**Преподаватель:** доц., к.т.н. Кожаринов А.С.

**Отметка:**

**Дата** **защиты:** 9.11.2020г

**2020** **г.**

**Содержание**

[Введение 3](#_Toc53915503)

[Описание задачи 3](#_Toc53915504)

[Средства разработки 4](#_Toc53915505)

[Примеры работы программы 5](#_Toc53915506)

[Выводы 9](#_Toc53915507)

[Источники 11](#_Toc53915508)

# 

# Введение

Целью выполнения данной работы является выработка устойчивых умений по организации взаимодействия на уровне программного кода между собственным программным обеспечением и функционалом существующих ИИ-сервисов нового поколения при решении некоторой задачи интеллектуальной обработки, а также определение уровня качества, с которым современные ИИ-сервисы IBM Watson и Google Cloud могут выполнять конкретную задачу, и выявление лучшего из использованных ИИ-сервисов для решаемой задачи .

В качестве задачи в данной лабораторной работе требуется разработать программное обеспечение на уровне демонстрационного прототипа, осуществляющее распознавание англоязычной аудиозаписи и перевод текста на русский язык, используя общедоступные ИИ-сервисы, предоставляемые системой IBM Watson и Google Cloud.

# Описание задачи

Основной задачей, которую решает разработанное программное обеспечение, является распознавание англоязычной аудиозаписи и перевод текста на русский язый.

В качестве исходных данных на вход программе пользователь подает mp3 файл с англоязычной записью длительностью не более 1 минуты.

Данный файл передается сервису Speech-to-Text от IBM Watson или Google Cloud (в зависимости от выбранного пользователем значения). На выходе программа получает распознанный текст с метками спикеров. Полученный текст приводится к читаемому виду с обозначением спикеров и передается на второй сервис от IBM Watson или Google Cloud для перевода на русский язык. Распознанный англоязычный текст и его перевод отображаются на экране в соответствующих окнах.

# Средства разработки

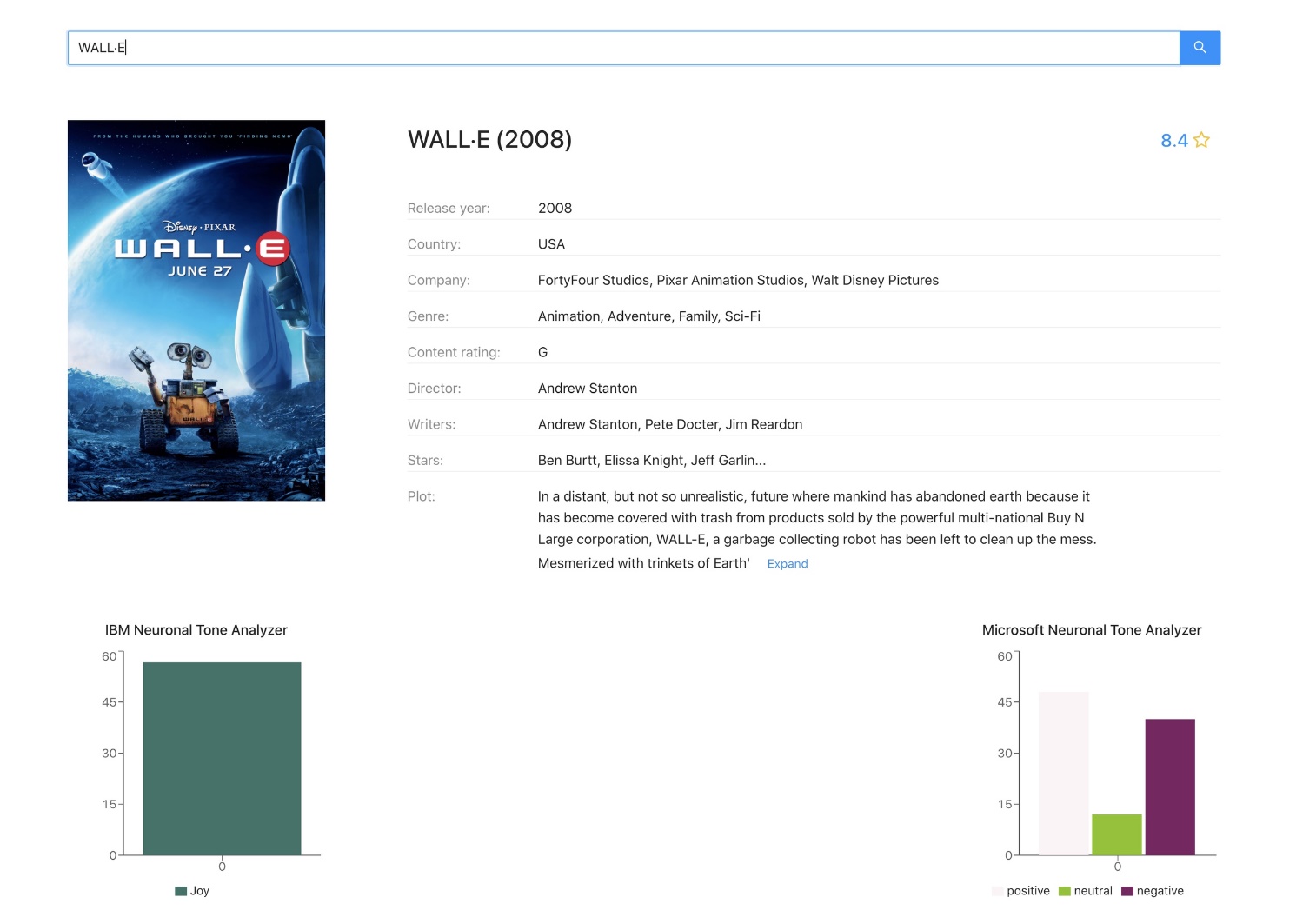
Интерфейс программного обеспечения разработан на языке JavaScript, в частности использована библиотека React. Запуск производится с помощью сервера Node.js Express.

Для получения информации о фильме и субтитрах используются API, которые запрашивают информацию с сервера. API для запроса информации о фильме использует IMDb, а субтитры запрашиваются с ресурса OpenSubtitles.org.

# Примеры работы программы

В данном разделе приведены примеры работы программного обеспечения для различных исходных данных.

В качестве первого примера выступает мультипликационный фильм «Валли» 2008 года выпуска:

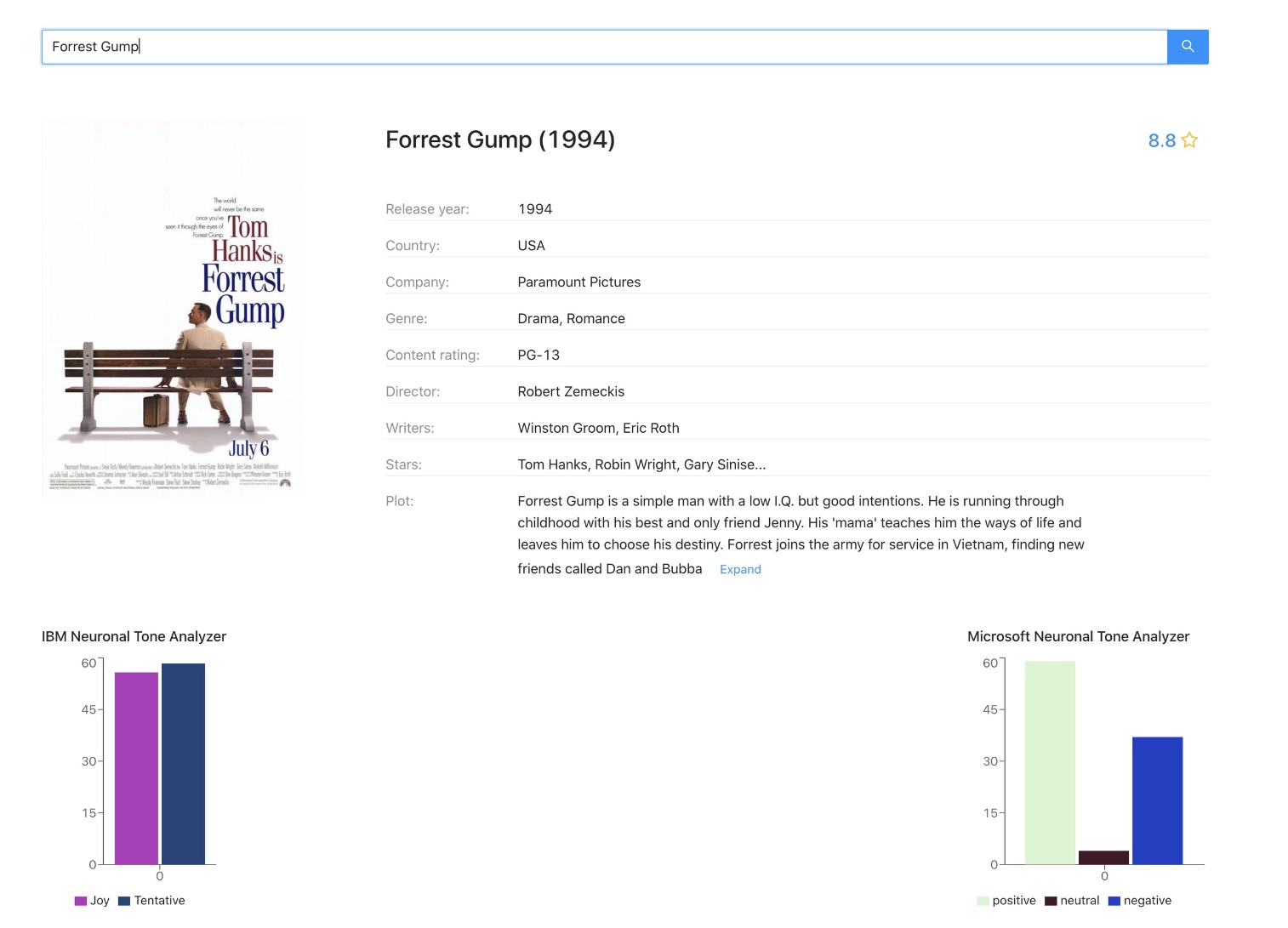


*Рисунок 2 – анализ фильма «WALL-E»*

Из графиков видно, что IBM Watson посчитал, что в 60% всего текста субтитров содержится радость. Других эмоций распознано не было. В какой-то степени это неудивительно, так как в данном мультфильме отсутствует большое количество диалогов и монологов, в большинстве своем два главных героя – Валли и ЕВА – общаются невербально, с помощью жестов и языка тела. Также герои часто произносят лишь имена друг друга, что с точки зрения текста не несет никакой эмоциональной составляющей.

В то же время ИИ-сервис от Microsoft распознал 3 различные эмоциональные окраски речи - позитивную, негативную и нейтральную. Около 45% текста данная нейронная сеть считает позитивно окрашенными, примерно 40% - негативно, а остальные 15% - нейтрально.

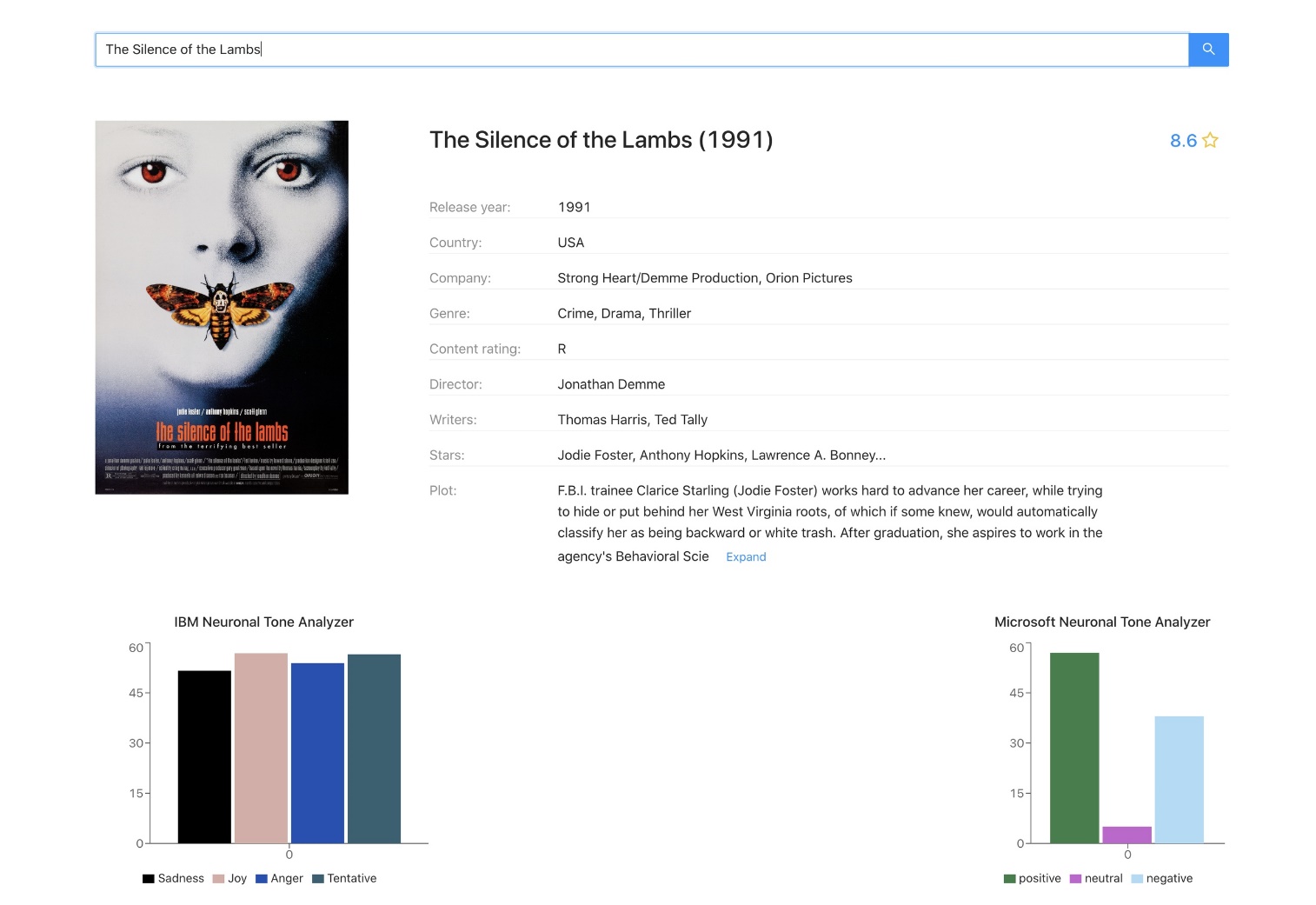
В другом случае для фильма «Форрест Гамп» 1994 года выпуска мы видим следующие результаты работы программы:



*Рисунок 3 – анализ фильма «Forrest Gump»*

IBM Watson считает, что примерно наполовину речь актеров наполнена радостью, а на другую половину – робостью и неуверенностью. При этом искусственный интеллект от Microsoft обнаружил в 35% фильма негативные эмоции, в 60% - позитивные и лишь 5% речи актеров не были окрашены эмоционально. Как бы ни разнились результаты работы двух нейронных сетей в данном конкретном случае, из их графиков очевидно одно сходство – оба анализа находят этот фильм крайне эмоциональны: IBM на 100%, а Microsoft – на 95%.

Крайне интересными являются результаты работы двух ИИ-сервисов при передаче им субтитров фильма «Молчание ягнят»:

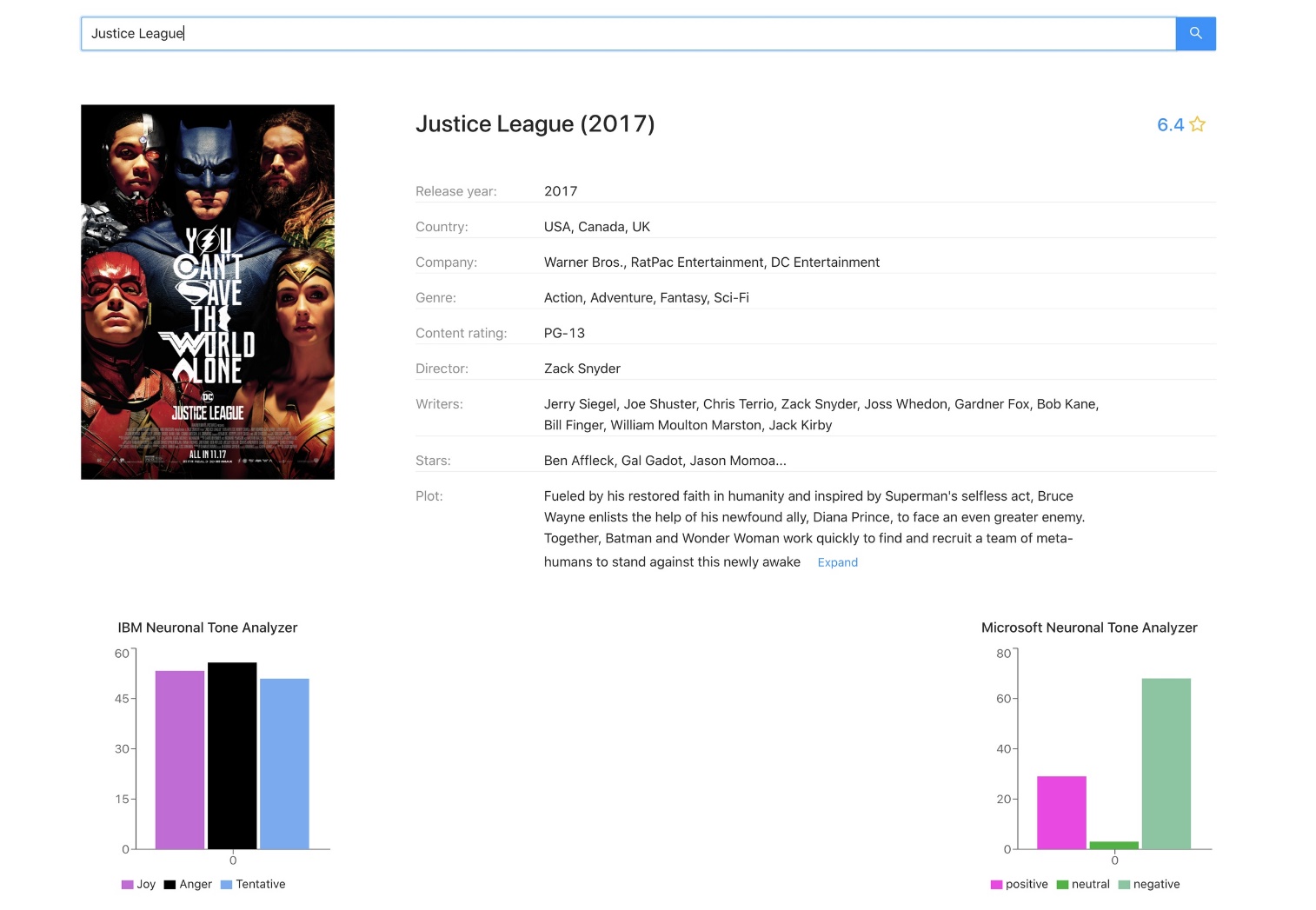


*Рисунок 4 – анализ фильма «The Silence of the Lambs»*

ИИ от IBM распознал 4 разных эмоции: от 50 до 60 процентов всей речи наполнено грустью, злобой, а также робостью. При этом почти 60% текста считаются выражением радости. По такой противоречивой оценке ясно, что в фильме все же преобладают негативные эмоции и, что более важно, эти эмоции являются сложными – некоторые слова субтитров нейронная сеть отнесла сразу к нескольким эмоциям.

Что касается работы ИИ от Microsoft, то противоположно IBM он считает, что фильм почти на 60% позитивный, на 35% - негативный и на 5% нейтральный.

Ну и наконец еще один интересный пример работы двух нейронных сетей:



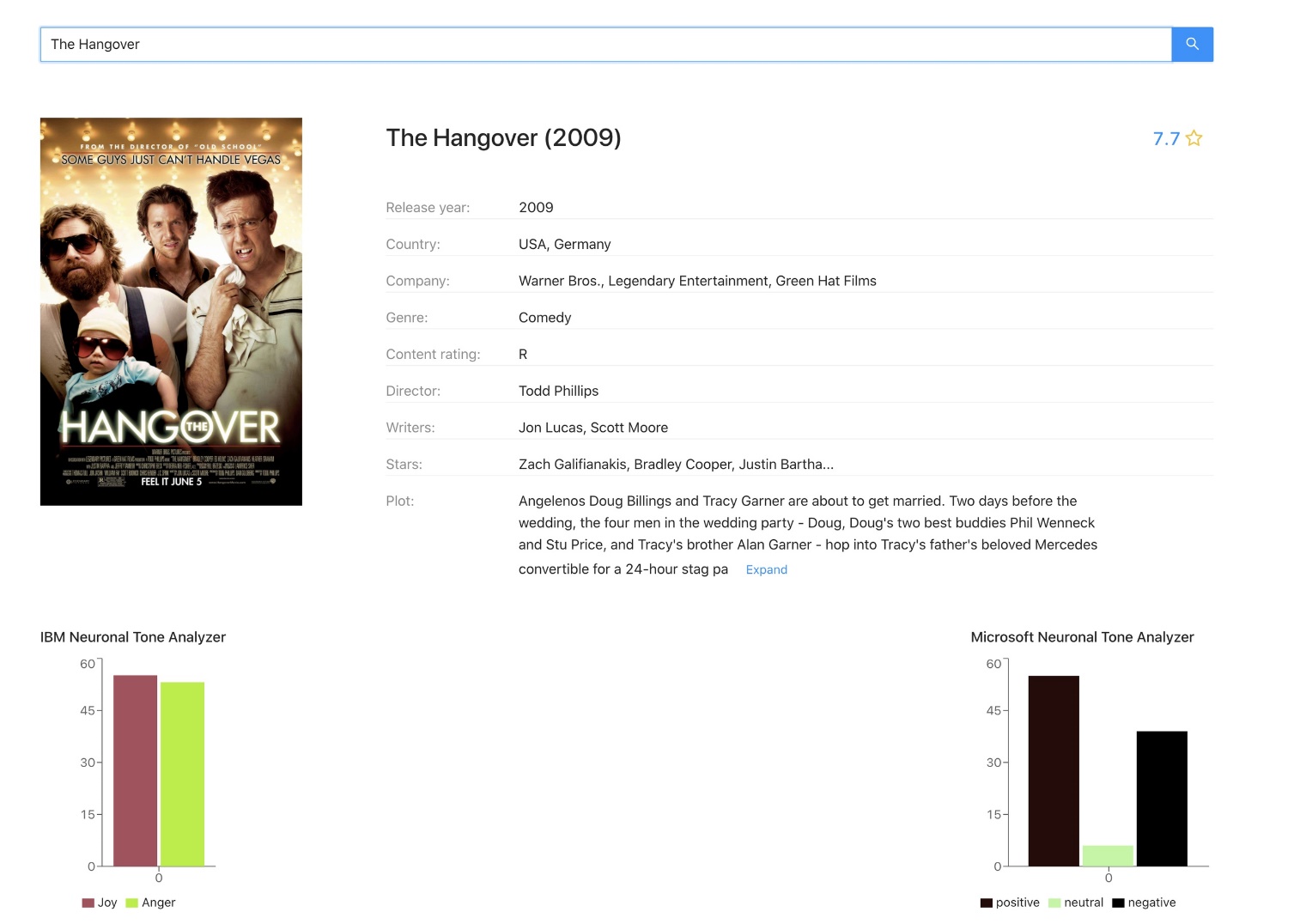
*Рисунок 5 – анализ фильма «Justice League»*

# Выводы

Разработанное программное обеспечение можно модульно встраивать в более сложные системы, например, по формированию подбора фильмов для пользователя под его настроение. Это, безусловно, увеличивает интерес к подобного рода программам.

Что касается сравнения качества и корректности работы двух ИИ-сервисов – от IBM и от Microsoft – то очевидного преимущества не продемонстрировала ни одна нейронная сеть. В качестве достоинств IBM Watson можно выделить широкий спектр эмоций, которые способна распознавать сеть, что означает ее умения больше точно распознавать эмоции, а не разделять их только на позитивные и негативные, как это делает Microsoft Text Analytics.

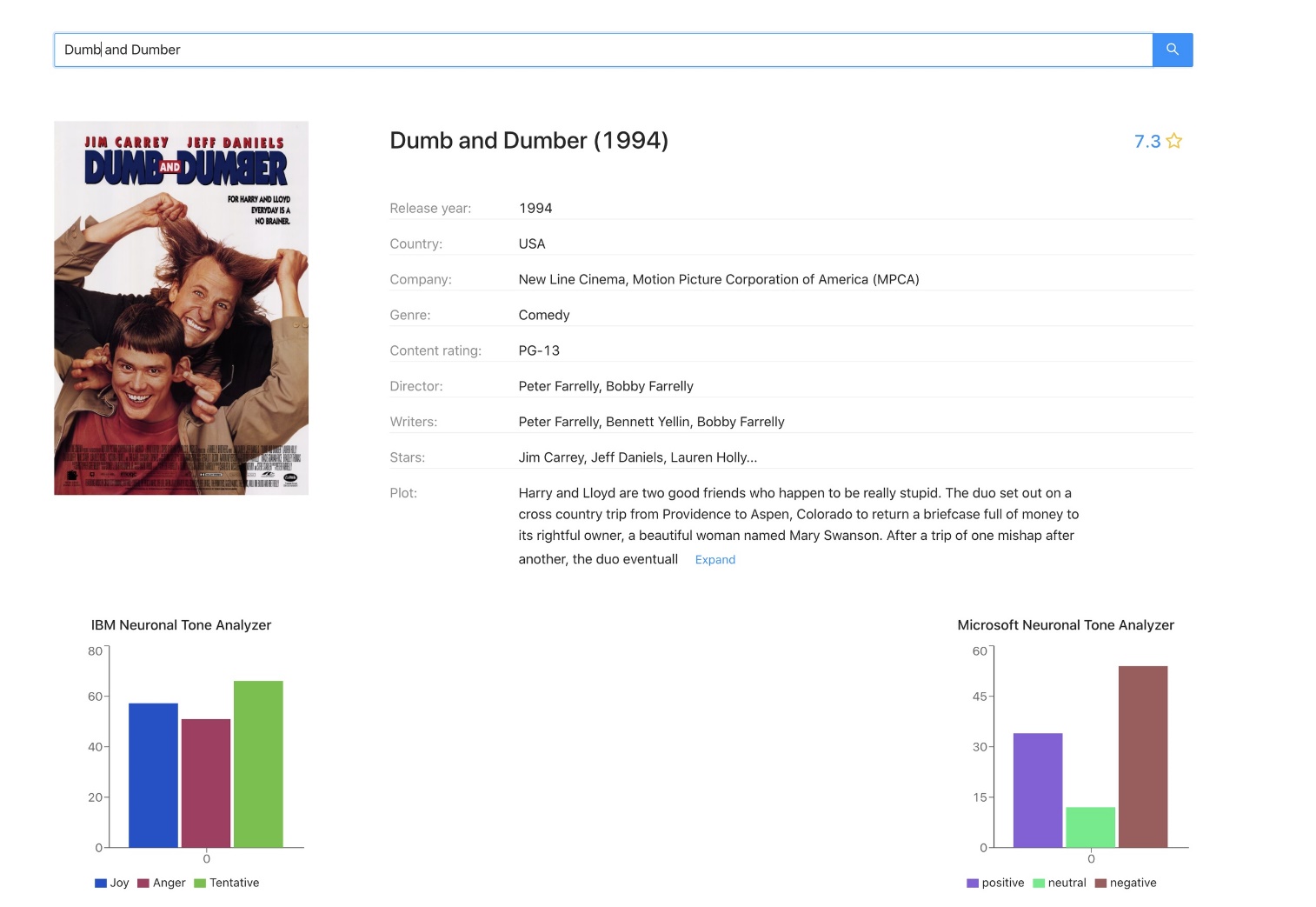
С другой стороны, у IBM результаты зачастую получаются прямо противоположные, так что такой анализ дает очень мало релевантной информации. Например, по результатам работы данной нейро-сети над субтитрами фильма «Мальчишник в Вегасе» нельзя точно сказать, радостный это фильм или злой:



*Рисунок 6 – анализ фильма «The Hangover»*

В таком случае этот фильм попадет как в категорию злых фильмов, так и радостных с одинаковыми показателями содержания этих эмоций в речи актеров, хотя эти эмоции считаются противоположными.

В работе нейронной сети от Microsoft же почти во всех тестах есть перевес в сторону положительных или отрицательных эмоций, что помогает до просмотра фильма сформировать свое представление о настроении, которое подарит кинолента. Однако случаются ситуации как, например, с фильмом «Тупой и еще тупее», где нейронная сеть Microsoft выделила большое число негативных эмоций, хотя это известная всем комедия:



*Рисунок 7 – анализ фильма «Dumb and Dumber»*

В таком случае, вероятнее всего, работает правило об отсутствии чувства юмора у нейронной сети, что мешает ей распознавать сарказм и иронию, а также черный юмор, который, безусловно, присутствует в данном фильме. Поэтому результаты работы ИИ-сервиса не совпадают со среднестатистической эмоциональной оценкой людей для этого фильма, относящегося к жанру комедии.

Таким образом, не рекомендуется использовать анализ текста речи актеров нейронными сетями с целью получения эмоциональной оценки фильма, так как на нее оказывает огромное влияние еще и интонация, с которой говорится та или иная фраза. Поэтому вероятнее всего работа нейронной сети по анализу эмоциональной окраски фильма на основе аудио-дорожки киноленты выдаст более корректные результаты, так как оценит не только текст, но и настроение, с которым его произносит актер.

# Источники

1. <https://www.ibm.com/cloud/watson-tone-analyzer>
2. <https://www.npmjs.com/package/opensubtitles-api>
3. <https://www.opensubtitles.org/ru>
4. <https://www.kinopoisk.ru/>
5. <https://azure.microsoft.com/en-us/services/cognitive-services/text-analytics/>
6. <https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/>