

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
Институт филологии и языковой коммуникации  
Кафедра теории германских языков и межкультурной коммуникации

УТВЕРЖДАЮ

Заведующий кафедрой ТГЯиМКК

\_\_\_\_\_ /О.В. Магировская/

« \_\_\_\_\_ » \_\_\_\_\_ 2018 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**ИСПОЛЬЗОВАНИЕ СЕНТИМЕНТ-АНАЛИЗА ДЛЯ  
СИНХРОНИЗАЦИИ САУНДТРЕКА И ЭМОЦИОНАЛЬНОЙ  
ТОНАЛЬНОСТИ ТЕКСТА**

45.04.02 Лингвистика

45.04.02.01 Межкультурная коммуникация и перевод

Магистрант

М.К. Кузлин

Научный руководитель

д-р. филол. наук,  
проф. каф. РЯиПЛ  
А.В. Колмогорова

Нормоконтролер

Э.А. Тарасенко

Красноярск 2018

## СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ .....</b>	<b>3</b>
<b>ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ПРЕДПОСЫЛКИ РАЗРАБОТКИ АЛГОРИТМА СЕНТИМЕНТ-АНАЛИЗА ДЛЯ РАНЖИРОВАННОЙ КЛАССИФИКАЦИИ.....</b>	<b>6</b>
1.1. Сентимент-анализ как направление прикладной лингвистики .....	6
1.2. Принципы и методы формирования обучающей выборки для технологии машинного обучения.....	9
1.3. Художественный текст как особый источник данных для обучающей выборки.....	12
1.4. Классификация эмоций по Лёвхейму как основа выделения эмоциональных классов текстовых данных .....	16
1.5. Дизайн программы-приложения для синхронизации саундтрека и эмоциональной тональности текста .....	18
<b>ВЫВОДЫ ПО ГЛАВЕ 1 .....</b>	<b>21</b>
<b>ГЛАВА 2 ОСОБЕННОСТИ ЭМОЦИОНАЛЬНЫХ КЛАССОВ ДАННЫХ ИЗ ХУДОЖЕСТВЕННЫХ ТЕКСТОВ.....</b>	<b>22</b>
2.1. Формирование обучающей выборки.....	22
2.2. Эмоциональный класс текстовых данных «Страх / Ужас (Fear / Terror), Гнев / Ярость (Anger / Rage)» .....	24
2.3. Эмоциональный класс текстовых данных «Стыд / Унижение (Shame / Humiliation), Грусть / Тоска (Distress / Anguish)».....	30
2.4. Эмоциональный класс текстовых данных «Радость / Счастье (Enjoyment / Joy), Интерес / Возбуждение (Interest / Excitement)» .....	36
2.5. Эмоциональный класс текстовых данных «Презрение / Отвращение (Contempt / Disgust)», «Удивление (Surprise)».....	42
2.6. Реализация программы-приложения для синхронизации саундтрека и эмоциональной тональности текста: основные трудности и результаты .....	48
<b>ВЫВОДЫ ПО ГЛАВЕ 2 .....</b>	<b>52</b>
<b>ЗАКЛЮЧЕНИЕ.....</b>	<b>53</b>
<b>СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....</b>	<b>56</b>
<b>ПРИЛОЖЕНИЕ А.....</b>	<b>62</b>

## ВВЕДЕНИЕ

Процесс достижения определенного эмоционального состояния читателем во время чтения и осмысления художественного текста может получить дополнительную стимуляцию при помощи соответствующего музыкального сопровождения текстовосприятия. В этом случае музыка играет ту же самую роль как саундтрек для фильмов, где звук создает дополнительный эффект к визуальным сценам.

**Актуальность** исследования состоит в том, что на данный момент проектов по созданию специального классификатора, способного автоматически определять точную эмоцию, отражённую в тексте, пока немного, а для русскоязычных данных – нет.

**Гипотеза исследования:** Процесс подбора саундтрека зачастую происходит вручную и основывается на тех эмоциях, которые присутствуют в определенном моменте. В данной работе мы хотим доказать, что подобный процесс возможно автоматизировать, прибегнув к имеющимся современным инструментам компьютерной лингвистики.

**Целью** нашего исследования стала разработка компьютерной программы для синхронизации музыкального сопровождения и эмоциональной тональности текста на электронном носителе на основе использования принципов сентимент-анализа текста и классификации эмоций «Куб Лёвхейма».

### **Задачи исследования:**

- 1) с помощью информантов приписать эмоционально окрашенному фрагменту текста ту или иную эмоцию, согласно Кубу Лёвхейма;
- 2) при помощи корпусного менеджера Sketch Engine выявить языковые маркеры той или иной эмоции;
- 3) разработать теоретические принципы базы данных «языковой маркер – имя эмоции – тег Last.fm»

4) описать алгоритм синхронизации языкового маркера и имени эмоции;

5) проанализировать процесс реализации программы-приложения для синхронизации саундтрека и эмоциональной тональности текста

**Материалом для исследования** послужили: 1) отобранные фрагменты текстов из художественных произведений: Дж. К. Роулинг «Случайная Вакансия», К. Тойбин «Бруклин», С. Чбоски «Хорошо быть тихоней», А. Асиман «Зови меня своим именем»; 2) разработанный текстовый корпус, состоящий из 100 фрагментов, полученный при помощи системы «Sketch Engine».

**Основными методами исследования** в работе являются методы сплошной и специальной выборки, сравнительный метод, анкетирование и корпусное ориентирование.

**Объект исследования:** рассматриваются языковые маркеры эмоций, согласно классификации Лёвхейма, в художественном тексте.

**Предмет исследования:** возможность автоматической синхронизации языковых маркеров эмоции, художественного текста и тегов саундтреков в соответствующей базе данных.

**Степень разработанности:** выбранная тема в отечественной лингвистике исследовалась такими научными деятелями как Азимов Э.Г., Айвазян С.А., Алифанова О.Г., Арнольд И.В., Бабенко Л.Г., Баранов А.Н., Гельгардт Р.Р. Тональность текста и машинное обучение рассматривались такими учёными как Клековкина М.В., Котельников Е.В., Лукашевич Н.В., Четверкин И.И., Меньшиков И.Л., Пазельская А.Г., Соловьев А.Н.

**Новизна** настоящего исследования заключается в попытке создания компьютерной программы-классификатора русскоязычных текстовых данных по критерию класса эмоций, которые вербализуются в тексте.

**Практическая значимость** настоящего исследования заключается в реализации разработанной программы в различных целях: автоматический

подбор подходящего музыкального сопровождения для любого текста на русском языке; создание комфортной среды для обучения; прививание интереса к чтению и так далее.

**Апробация результатов исследования:** Результаты исследования были представлены в форме доклада на международной конференции «Second International Workshop on Language, Music and Computing» (17.04.2017, г. Санкт-Петербург). Была отправлена на публикацию статья: Кузлин М.К. Использование сентимент-анализа для синхронизации саундтрека и эмоциональной тональности текста // Siberia Lingua. Красноярск. 2018.

**Структура работы.** Основная часть диссертации состоит из двух глав. В первой главе «Теоретические предпосылки разработки алгоритма сентимент-анализа для ранжированной классификации» определены основные термины и методы, необходимые для выполнения практической части исследования. Во второй главе «особенности эмоциональных классов данных из художественных текстов» представлен анализ текстовых фрагментов и описан процесс реализации необходимого программного обеспечения. В заключении представлены выводы исследования. В список использованных источников включено 60 источника.

# ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ПРЕДПОСЫЛКИ РАЗРАБОТКИ АЛГОРИТМА СЕНТИМЕНТ-АНАЛИЗА ДЛЯ РАНЖИРОВАННОЙ КЛАССИФИКАЦИИ

## 1.1. Сентимент-анализ как направление прикладной лингвистики

Сентимент-анализ информационных потоков имеет большой потенциал применения для мониторинговых, аналитических и сигнальных систем, для систем документооборота и рекламных платформ, таргетированных по тематике веб-страниц. Текст на естественном языке, помимо информации, может выражать эмоциональную оценку того, о чём сообщается.

Тональность – это эмоциональное отношение автора высказывания к некоторому объекту (объекту реального мира, событию, процессу или их свойствам/атрибутам), выраженное в тексте. Подобный метод анализа можно использовать во многих сферах деятельности. Анализ тональности текста осуществляет определение полярности текста и извлечение аспектов из эмоционально окрашенного текста [Ермаков и др., 2001: 283].

Сентимент-анализ подходит для усиления продаж и улучшения маркетинговых стратегий компании (отслеживание отзывов покупателей и ответов на обзоры), определяя идеологические изменения и анализируя тенденции в планировании политических стратегий на основе сообщений из-за рубежа, финансовых отчетов и сообщений из социальных сетей [Bollen et al., 2011: 2]. Сентимент-анализ или эмоциональный анализ можно рассмотреть как усовершенствованную версию анализа мнений, так как этот анализ стремится к более подробной классификации документов на основе эмоций, которые они выражают [Chatzakou et al., 2013: 323].

Некоторые исследователи «предприняли попытки совместить статистические и лингвистические методы для достижения лучших результатов, однако испытывают при этом затруднения при анализе более длинных и сложных предложений» [Jakob et al., 2010: 1036]. Кроме того,

такие исследователи как Якоб Н. решили не делать различий между объектом оценки (например: смартфон), его составляющими частями (например: экран, фронтальная камера) и его характеристиками (например: удобство применения); и, определяют ближайшую именную группу в качестве основного объекта оценки» [Jakob et al., 2010: 1037].

В современных системах автоматического определения эмоциональной оценки текста чаще всего используется двухмерное эмотивное пространство: позитив или негатив [Aue et al., 2005: 237].

Основной задачей в анализе тональности является классификация полярности данного документа, то есть определение, является ли выраженное мнение в документе или предложении позитивным, негативным или нейтральным [Ceron et al., 2014]. Более развёрнуто, «вне полярности» классификация тональности выражается, например, такими эмоциональными состояниями, как «злой», «грустный» и «счастливый».

Одним из минусов данного подхода является то, что эмоциональную составляющую документа не всегда можно однозначно определить, то есть документ может содержать как признаки позитивной оценки, так и признаки негативной. Ранние работы в этой области включают в себя различные методы распознавания полярности обзоров товара и отзывов о фильмах соответственно. Это пример работы на уровне документа [Maas et al., 2011].

Можно классифицировать полярность документа по многополосной шкале. Такие исследователи как Б. Снайдер и Р. Барзилай расширили основные задачи классификации киноотзывов от оценки «положительный или отрицательный» и ушли в сторону прогнозирования рейтинга по 3-х или 4-балльной шкале [Snyder et al., 2007].

Другим методом определения тональности является использование систем шкалирования, посредством чего словам, обычно связанным с отрицательными, нейтральными или позитивными тональностями, ставятся в соответствие числа по шкале от  $-10$  до  $10$  (от самого отрицательного к

самому положительному) [Banea et al., 2010]. Вначале фрагмент неструктурированного текста исследуется с помощью инструментов и алгоритмов обработки естественного языка, а затем выделенные из этого текста объекты и термины анализируются с целью понимания значения этих слов. [Лукашевич и др., 2011: 76]. Проблема иногда может быть более сложной, чем классификация полярности: субъективность слов и фраз может зависеть от их контекста, а объективный документ может содержать в себе субъективные предложения (например, новостной источник, цитирующий и интерпретирующий мнение человека) [Voüy et al., 2009].

Модель более подробного анализа называется анализом на основе функции (аспекта). Эта модель ссылается на определение мнений или настроений, выраженных различными функциями или аспектами сущностей, например, у сотового телефона, цифровой камеры или банка. Аспект – это атрибут или компонент сущности, исследуемой на предмет эмоциональной тональности например, экран сотового телефона или же качество съемки камеры [Jakob et al., 2010]. Эта проблема требует решения ряда задач, например, у идентификации актуальных сущностей, извлечения их функций/аспектов и определения, является ли мнение, высказанное по каждой функции/аспекту, положительным, отрицательным или нейтральным [Клековкина и др., 2012].

Ежедневно миллионы людей выражают свое мнение о продуктах, услугах и предложениях, среди других, используя платформы онлайн, такие как социальные сети, блоги, вики, форумы, и так далее. Естественно, автоматическое извлечение выраженных мнений или подразумеваемых чувств самым точным способом стало очень важным для компаний, маркетологов и исследователей. [Пазельская и др., 2001] Сентимент-анализ относится к процессу идентификации нетривиальной, субъективной информации из коллекции исходных материалов, которые содержат скрытую информацию. Такой процесс может быть применен на различных текстовых



источниках, например, в пределах одного документа или, даже, на конкретный фрагмент в тексте [Пазельская и др., 2011].

Сентимент-анализ подходит для «усиления продаж и улучшения маркетинговых стратегий компании (отслеживание отзывов покупателей и ответов на обзоры), определяя идеологические изменения и анализируя тенденции в планировании политических стратегий на основе сообщений из-за рубежа, финансовых отчетов и сообщений из социальных сетей» [Bollen et al., 2011: 3].

Анализы, основанные на словарях, приводят к вполне удачным результатам, но они часто страдают от низкого количества слов, то есть встречаются документы, которые не содержат ни одного из слов словаря. Это особенно заметно на коротких текстовых отрывках, которые используются в коммуникации пользователей в интернете [Witten et al., 2005]. Кроме того, такие подходы не захватывают более скрытые проявления чувств и эмоций, так как они не рассматривают контекст, в котором люди выражаются.

## 1.2. Принципы и методы формирования обучающей выборки для технологии машинного обучения

Формирование большого количества обучающей выборки обладает существенно важным значением для успешного решения задач машинного обучения. Зачастую проблемы машинного обучения сводятся именно к правильному формированию обучающего множества. Погрешности в создании обучающего множества обычно оказываются критичными и способны свести на нет эффективность самих алгоритмов обучения. В машинном обучении считается, что наличие хороших обучающих данных гораздо важнее качества алгоритма обучения. В то же время, в современной литературе по машинному обучению вопросам формирования обучающего множества уделяется недостаточное внимание, зачастую данные вопросы

«полностью игнорируются, недостаточно развита теоретическая база, объясняющая явления, возникающие в процессе формирования множества обучающих данных» [Антонова и др., 2013].

Точность sentiment-анализа и дальнейшее машинное обучение определяется выбранным набором средств классификации обучающей выборки. Выборка – это некоторая часть объектов, которая выступает в качестве объектов непосредственного изучения. В рамках нашего исследования, нам потребуется составить первую выборку, которая называется «обучающая выборка» (training sample), по которой происходит изначальные настройки и дальнейший процесс машинного обучения.

«Машинное обучение (machine learning) изучает методы построения алгоритмов, способных самостоятельно обучаться. Различают два типа обучения: обучение по прецедентам, или индуктивное обучение, основано на выявлении общих закономерностей по частным эмпирическим данным и дедуктивное обучение, которое предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний» [Воронцов, 2011].

Во время sentiment-анализа текстовых ресурсов могут возникнуть проблемы при классификации документов, поэтому, одним из важных этапов создания обучающей выборки является правильный выбор эмоционального классификатора и источник данных для анализа [Антонова и др., 2013].

Машинное обучение находится на стыке математической статистики, методов оптимизации и классических математических дисциплин, но имеет также и собственную специфику, связанную с проблемами вычислительной эффективности и переобучения [Усталов, 2012]. Многие методы индуктивного обучения разрабатывались как альтернатива классическим подходам. Многие методы тесно связаны с извлечением информации и интеллектуальным анализом данных (Data Mining). Наиболее теоретические разделы машинного обучения объединены в отдельное направление, теорию

вычислительного обучения (Computational Learning Theory и COLT) [Hall, 1999; Wiebe et al., 2005].

Машинное обучение является не только математической, но и практической, инженерной дисциплиной. Использование теории не сразу приводит к методам и алгоритмам, которые можно применить на практике. Чтобы успешно запустить процесс обучения, приходится применять дополнительные настройки, компенсирующие несоответствие сделанных в теории предположений условиям реальных задач. Практически ни одно исследование в машинном обучении не обходится без экспериментов на обучающих данных или данных для тестирования [Айвазян и др., 1983].

Для технической реализации sentiment-анализа работе использовано машинное обучение с учителем. Обучение с учителем – один из способов машинного обучения, в ходе которого испытуемая система принудительно обучается с помощью примеров «стимул-реакция». Реализуется оно по следующему алгоритму: сбор коллекции документов, необходимых для обучения классификатора; представление каждого документа в виде вектора признаков; определение для каждого документа типа тональности для обучения классификатора; обучение классификатора по выбранному алгоритму классификации; использование модели для анализа [Pang et al., 2002; 2004].

На основе этих данных требуется «восстановить зависимость (построить модель отношений стимул-реакция, пригодных для прогнозирования), то есть построить алгоритм, способный для любого объекта выдать достаточно точный ответ. Для измерения точности ответов, так же как и в обучении на примерах, может вводиться функционал качества» [Lu et al., 2011: 351; Четвёркин, 2013: 101].

Таким образом, мы убедились, что на успешность проведения sentiment-анализа влияют такие факторы как правильно составленная обучающая выборка и выбор способа машинного обучения. Только результат

работы программы зависит не только от объема обучающей выборки, но и от того, какой конкретно текст мы вводим для тестирования.

### 1.3. Художественный текст как особый источник данных для обучающей выборки

Ввиду использования в нашем исследовании определенной классификации эмоций (о которой мы расскажем в нашем исследовании в параграфе 1.5) было принято решение использовать *художественный текст* как источник данных для составления обучающей выборки.

Главным моментом успешного машинного обучения с учителем, является использование корректно составленной обучающей выборки, поэтому, одним из факторов выбора художественных текстов как материала для составления обучающей выборки и дальнейшего анализа стало то, что художественный стиль имеет воздействующую функцию на читателя (влияние с помощью образов на мысли и чувства) [Меньшиков, 2012].

В первую очередь, как и любой другой вид текста, Л.Г. Бабенко и Ю.В. Казарин определяют художественный текст как «словесное речевое произведение, сложный языковой знак, внутри которого реализованы языковые единицы всех уровней (от фонемы до предложения)» [Бабенко и др., 2009: 30], следовательно, наличие языковых единиц всех уровней позволит нам найти подходящие языковые маркеры для каждой эмоциональной группы.

В художественном тексте разнообразные средства языка составляют единую, стилистическую и эстетическую систему, к которой неприменимы нормативные оценки, прилагаемые к отдельным функциональным стилям литературного языка [Гельгардт, 2011].

То, как в художественном тексте сочетаются разнообразные языковые средства, какие стилистические приемы использует писатель, как он

переводит понятия в образы, составляет предмет стилистики художественной речи [Голуб, 2001: 365].

«...Понятие «стиля» в применении к языку художественной литературы наполняется иным содержанием, чем, например, в отношении стилей делового или канцелярского и даже стилей публицистического и научного. Язык художественной литературы не вполне соотносителен с другими стилями, он использует их, включает их в себя, но в своеобразных комбинациях и в преобразованном виде...» [Виноградов, 1959: 365].

«Для художественного произведения характерны восприятие посредством чувств и перевоссоздание действительности, автор стремится передать прежде всего свой личный опыт, свое понимание и осмысление того или иного явления» [Голованова и др., 2008: 29].

Лексический состав и функционирование слов в художественном стиле речи имеют свои особенности. В число слов, составляющих основу и создающих образность этого стиля, прежде всего, входят образные средства русского литературного языка, а также слова, реализующие в контексте свое значение [Болотнова, 2009: 312]. Это слова широкой сферы употребления. Узкоспециальные слова используются в незначительной степени, только для создания художественной достоверности при описании определенных сторон жизни [Земская и др., 2006: 147].

В художественном стиле речи очень широко используется речевая многозначность слов, что открывает в нем дополнительные смыслы и смысловые оттенки, а также синонимия на всех языковых уровнях, благодаря чему появляется возможность подчеркнуть тончайшие оттенки значений и эмоций. Автор использует все богатства языка, создаёт свой неповторимый язык и стиль, к яркому, выразительному, образному тексту [Мерков, 2001; Баранов, 2007].

В художественном тексте на первый план выходят эмоциональность и экспрессивность изображения. Многие слова, которые в научной речи

выступают как четко определенные абстрактные понятия, в газетно-публицистической речи – как социально-обобщенные понятия, в художественной речи несут конкретно-чувственные представления [Кожин, 1982]. Таким образом, стили функционально дополняют друг друга. Поэтому в художественной речи важную роль играют словосочетания, которые создают некое образное представление.

Литературное произведение воздействует на читателя эмоционально. Воздействие художественного текста определяется тем, что он содержит так называемую художественную, или эстетическую, информацию. Эта художественная информация реализуется только в пределах индивидуальной художественной структуры, т.е. конкретного художественного текста. Носителями художественной информации в тексте могут быть любые его элементы, от отдельных слов до целых предложений.

Очень широко в художественном стиле речи используется речевая многозначность слова, которая способствует созданию дополнительных смыслов и смысловых оттенков. Как правило, автор художественного текста не ограничен в выразительных средствах нормами стиля, поэтому может использовать не только кодифицированную лексику литературного языка, но так же семантические единицы разговорной речи, просторечья [Денисенко и др., 2008].

Часто авторы художественных текстов прибегают к синтаксическим средствам, например, к инверсии, которая позволяет усилить семантическую значимость конкретного слова или придать всей фразе особую стилистическую окраску [Гвишиани, 2008: 125].

В художественной речи возможны и намеренные отклонения от структурных норм, необходимые для выделения конкретной мысли, идеи, черты, важной для смыслового наполнения произведения. Они могут быть выражены в нарушении фонетических, лексических, морфологических и других языковых норм [Азимов и др., 2009]. Часто подобные приемы

используются для создания комического эффекта или яркого, выразительного художественного образа.

В действительности характерной чертой художественных текстов является использование всевозможных средств экспрессивности и образности. Это не только единицы функциональных разновидностей литературного языка, но и элементы просторечия, социальных и профессиональных жаргонов, диалектов [Кожина, 1987]. Отбор и употребление этих средств автор подчиняет эстетическим целям, к которым он стремится при создании своего произведения.

Мир художественной литературы – это «перевоссозданный» мир, изображаемая действительность представляет собой в определенной степени авторский вымысел, а значит, в художественном стиле речи главнейшую роль играет субъективный момент. Вся окружающая действительность представлена через видение автора [Розенталь, 1997]. Но в художественном тексте мы видим не только мир писателя, но и писателя в этом мире: его предпочтения, осуждения, восхищение, неприятие. С этим связаны эмоциональность и экспрессивность, метафоричность, содержательная многоплановость художественного стиля речи.

Синтаксический строй художественной речи отражает поток образно-эмоциональных авторских впечатлений, поэтому здесь можно встретить все разнообразие синтаксических структур. Каждый автор подчиняет языковые средства выполнению своих идейно-эстетических задач.

Подводя промежуточные итоги, мы можем сказать, что художественный текст эмоционально окрашен и содержит в себе большое количество языковых средств, что доказывает правильность выбора художественного текста как особого источника данных для обучающей выборки.

#### 1.4. Классификация эмоций по Лёвхейму как основа выделения эмоциональных классов текстовых данных

Представляя собой группу очень сложных и многообразных психических явлений, эмоции с трудом поддаются совокупному анализу. В связи с этим в современной науке существует разное количество классификаций эмоций. Классификации эмоций, доступные в последних исследовательских работах по тональности текста, содержат, как правило, ограниченное количество эмоций (положительная, отрицательная и нейтральная).

Шведский биолог Гуго Левхейм установил, что, эмоциональные состояния, являясь функцией от адаптивных систем человеческого организма, порождаются в лимбической системе и миндалевидном теле головного мозга, но дальнейший сигнал об эмоции активируется и распространяется на другие отделы головного мозга благодаря действию трех моноаминов: серотонина, допамина и норадреналина.

Взяв за основу восьмичленную классификацию С. Томкинса, выявленную на основе анализа типов выражений лица, Г. Левхейм установил корреляцию каждой из восьми эмоций со специфической комбинацией уровней трех названных выше моноаминов [Lövhelm, 2012: 343].

Для нашего исследования за основу было решено взять классификацию эмоций Гуго Лёвхейма. Таким образом, были сформированы восемь базовых эмоций:

Страх / Ужас (Fear / Terror) – высокий уровень допамина при низком – серотонина и норадреналина.

Гнев / Ярость (Anger / Rage) – низкий уровень серотонина при высоком уровне других моноаминов.

Стыд / Унижение (Shame / Humiliation) – низкий уровень всех трех моноаминов.



Грусть / Тоска (Distress / Anguish) – низкий уровень серотонина и допамина, высокий – норадреналина.

Радость / Счастье (Enjoyment / Joy) – высокий уровень серотонина и допамина в сочетании с низким уровнем норадреналина.

Интерес / Возбуждение (Interest / Excitement) – высокий уровень всех трех моноаминов.

Презрение / Отвращение (Contempt / Disgust) – высокий уровень серотонина при низком уровне двух других моноаминов.

Удивление (Surprise) – высокий уровень серотонина и норадреналина в сочетании с низким уровнем допамина.

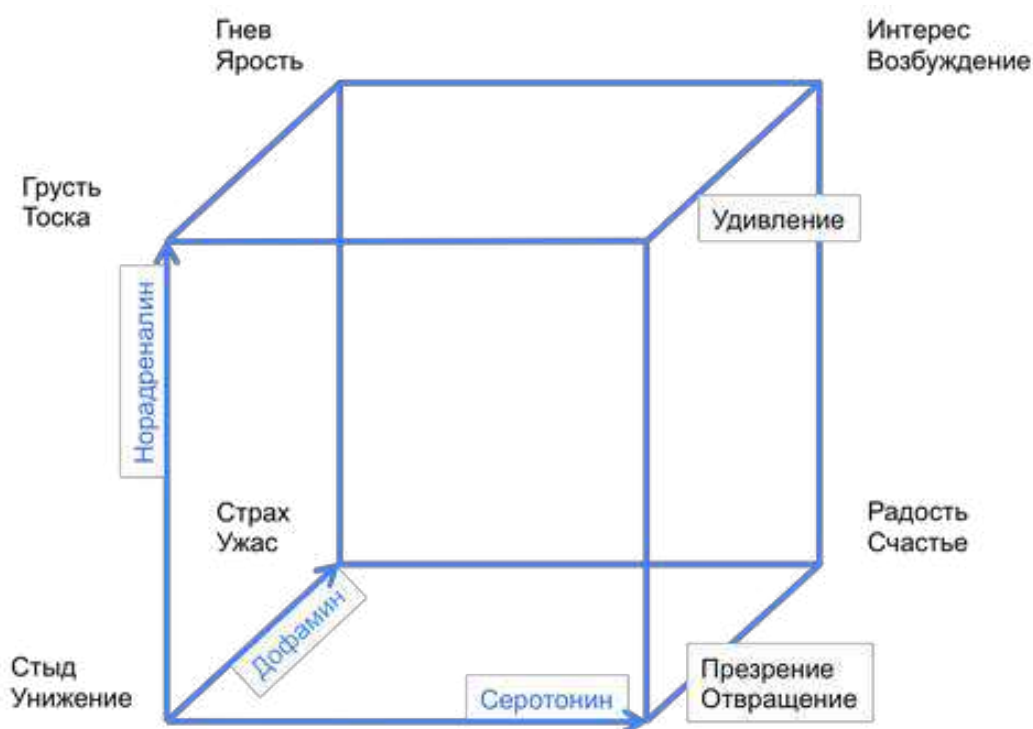


Рисунок 1. Куб Левхейма

Поскольку одной из целей является создание классификатора, то в качестве подхода для создания такой системы мы выбрали парадигму машинного обучения с учителем (*обучение по прецедентам*). В процессе функционирования классификатора на основе большого массива эмпирических данных строится статистическая модель, где на ввод подаются собственно данные, представленные в числовом виде, а в качестве вывода мы имеем один из возможных восьми. Каждой паре «ввод-вывод» в получившейся модели соответствует вероятность принадлежности определенной конфигурации данных на входе определенному классу на выходе. Исходный массив данных с уже проставленными экспертом / экспертами классами, который используется для создания статистической модели (тренировки классификатора), называется обучающей выборкой. Статистические выводы, полученные на качественно составленной обучающей выборке, можно распространить на всю совокупность подобных данных [Колмогорова и др., 2018].

Побочным продуктом работы классификатора является множество различных промежуточных вычислений, которые могут быть использованы для анализа взаимосвязей внутри выборки, поиска закономерностей и релевантных признаков.

#### 1.5. Дизайн программы-приложения для синхронизации саундтрека и эмоциональной тональности текста

Достижение чувств определенных эмоций при прочтении художественных книг может быть облегчено путем прослушивания определенной музыки, которая соответствует определенному эмоциональному состоянию, которое было предложено автором. Чтение (визуальная модальность) дополняется прослушиванием музыки (аудиальная модальность). В этом случае музыка играет ту же роль, что и саундтрек к

фильмам, где звук создает дополнительный эффект для визуальных сцен. Но также можно выбрать музыку для дополнения визуальных моделей и изображений, которые появляются в сознании, когда вы читаете историю, чтобы увеличить эмоциональный эффект книги.

Процесс создания саундтрека для текста состоит из следующих шагов:

1. Разделение текста на несколько фрагментов;
2. Определите настроение (эмоции) каждого фрагмента;
3. Поиск музыки, которая соответствует определенному настроению;
4. Прикрепление музыки к текстовым фрагментам с соответствующей эмоцией.

Но для выбора необходимой музыки появится необходимость решить несколько проблем:

1. Какую музыку мы должны выбрать для данной извлеченной эмоции;
2. Как выровнять время воспроизведения музыкальных дорожек со скоростью чтения.

После выбора эмоционального набора мы должны определить механизм маркировки определенного текста с эмоциями из Куба Лёвхейма. Такая цель может быть достигнута за счет использования автоматических инструментов (например, применяя методы анализа текущих настроений) или ручную (маркировка текстовых дополнений через опрос).

Из-за отсутствия необходимых данных и из-за того, что наша цель была всего лишь концептом, мы решили выполнить следующие шаги.

1. Выбрать русскоязычное художественное произведение;
2. Разделить текст на более мелкие фрагменты, которые могут быть эмоционально окрашены;
3. Промаркировать фрагменты с нужной эмоцией.

После разделения текста на эмоциональные выдержки нам нужно было обозначить его большим количеством оценщиков. Мы решили использовать платформу краудсорсинга для быстрого получения результатов, большого охвата данных и рандомизации задач.

Мы попросили экспертов прочитать текстовые фрагменты и выбрать эмоцию из восьми возможных эмоций, которые в основном соответствуют настроению текста, или если ни эмоции не подходят, ни текст был очень нейтральным, а эксперт должен был выбрать «нейтральный» вариант.

## ВЫВОДЫ ПО ГЛАВЕ 1

В результате проведенного исследования мы выяснили, что сентимент-анализ информационных потоков имеет большой потенциал применения для мониторинговых, аналитических и сигнальных систем.

Для реализации системы распознавания тональности текста был использован метод машинного обучения с учителем. Суть метода состоит в обучении машинного классификатора на коллекции заранее размеченных текстов.

Для данного исследования был выбран художественный текст как особый источник данных для обучающей выборки. Одним из главных факторов выбора художественных текстов как материала для нашего анализа стало то, что художественный стиль имеет воздействующую функцию на читателя (влияние с помощью образов на мысли и чувства).

Для нашего исследования за основу было решено взять классификацию эмоций Г. Лёвхейма – биохимика, установившего, что каждая из восьми возможных эмоций проявляется со специфической комбинацией уровней трех: допамина, норадреналина и серотонина.

Достижение определенного эмоционального состояния при прочтении художественного текста может быть облегчено путем прослушивания музыки одновременно с чтением при условии, именно для этого, для нашего исследования был разработан дизайн программы-приложения для синхронизации саундтрека и эмоциональной тональности текста.

## ГЛАВА 2 ОСОБЕННОСТИ ЭМОЦИОНАЛЬНЫХ КЛАССОВ ДАННЫХ ИЗ ХУДОЖЕСТВЕННЫХ ТЕКСТОВ

### 2.1. Формирование обучающей выборки

Выбор источников первичных данных является важным этапом для нашего исследования значение для достоверности обучающей выборки.

В нашем случае, мы выбрали художественное произведение – роман Дж. К. Роулинг «Случайная Вакансия», К. Тойбин «Бруклин», С. Чбоски «Хорошо быть тихоней», А. Асиман «Зови меня своим именем». Выбор был обусловлен тем, что после просмотра рецензий на данные книги, мы выяснили, что текст вызывает сильный эмоциональный отклик у читателей, создавая эмоциональное состояние практически изоморфное состоянию героев: «Напишу лишь, что эта книга побуждает чувствовать. Именно чувствовать, а не обдумывать, осмыслять и переваривать. Она читается за один миг и ты как «не губка, но фильтр» пропускаешь её через себя будто становишься героем данного произведения. Чувствуешь, что человек может быть бесконечен!».

Затем, основываясь на данной восьмеричной классификации эмоций мы предложили 35 экспертам оценить коллекцию текстовых фрагментов, приписав каждому из них определенную эмоцию из восьми предложенных. Задание было сформулировано следующим образом:

*Прочитайте фрагмент текста. Подумайте, какое эмоциональное переживание отражено в нем. Соотнесите это переживание с одной из восьми эмоций. Если, читая текстовый фрагмент, вы не чувствуете в нем никакой эмоции – выберите «нейтрально».*

Для примера были приведены некоторые фрагменты текста уже классифицированные экспертами как передающие определенную эмоцию. Решая проблему объема предлагаемого для оценки текстового фрагмента, мы пришли к выводу, что одного предложения явно недостаточно информантам,

поэтому через процедуру экспериментальных проб остановились на объеме в 90–250 слов, которые организованы в сверхфразовое единство. Представляем наиболее показательный пример:

*«Боль пушечным ядром взорвала мозг. Колени ударились о холодный асфальт, но Барри этого не почувствовал; череп захлестнуло волной огня и крови; агония была столь мучительной, что терпеть её не было сил, но он терпел, потому что от вечности его отделяла ровно минута. Мэри вскрикнула – и уже не умолкала. Из бара выскочили какие-то люди. Один тут же ринулся назад, вспомнив, что среди членов клуба есть двое врачей на пенсии. Знакомая супружеская пара, услышав из ресторана крики, даже не прикоснулась к еде и побежала на стоянку, чтобы оказать посильную помощь. Муж набрал по мобильному 999. «Скорая» двадцать пять минут ехала из Ярвила – это был ближайший крупный город. Когда мигалка осветила пульсирующей синевой место действия, Барри лежал в луже рвоты, без движения и без признаков жизни; Мэри в разорванных колготках стояла рядом на коленях и сжимала его руку, сотрясаясь от рыданий и повторяя его имя» [Роулинг, 2013].*

Данный фрагмент из произведения Дж. К. Роулинг «Случайная Вакансия» был оценен следующим образом: 33 эксперта при прочтении выбрали вариант «Страх \ Ужас», 2 эксперта при прочтении выбрали вариант «Грусть \ Тоска».

Далее в практической части нашей работы мы разберем фрагменты, проанализируем имеющиеся языковые маркеры для каждой эмоции, и каким образом они помогут для последующего автоматического анализа с помощью классификатора.

## 2.2. Эмоциональный класс текстовых данных «Страх / Ужас (Fear / Terror), Гнев / Ярость (Anger / Rage)»

Первым эмоциональным классом для анализа стала группа «Страх / Ужас». Далее мы рассмотрим несколько текстовых фрагментов и их дальнейший анализ.

**Фрагмент №1:** *«Тесса поднялась и вошла в комнату Пупса с испуганным и озабоченным видом. Не произнося ни слова, Колин взял стичечный коробок и показал ей содержимое.*

*– Ох, – слабо произнесла Тесса.*

*– Он сказал, что идёт сегодня куда то с Эндрю Прайсом, – напомнил Колин. Тессу напугало, что у него на щеках перекатываются желваки. – Я только что проходил мимо этого нового кафе на площади. Эндрю Прайс сейчас там – работает, убирает со столов. Где же, спрашивается, Стюарт?*

*На протяжении нескольких недель Тесса делала вид, будто верит Пупсу, когда тот говорил, что отправляется куда то с Эндрю. Не один день она убеждала себя, что Сухвиндер, должно быть, ошибается, считая, что Пупс «гуляет» (то есть становится вровень) с Кристал Уидон» [Роулинг, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Страх / Ужас». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Страх / Ужас»: «испуганным», «озабоченным», «напугало».

**Фрагмент №2:** *«Бог ты мой, – в третий раз протянул Говард, не отвечая на приглушённые расспросы Ширли: его вниманием полностью завладел Майлз. – Рухнул замертво прямо на стоянке? – Да-да, – подтвердил Майлз. – Я как увидел, сразу понял: ему конец. Это была его первая ложь, и он отвёл взгляд от жены. Она прекрасно помнила, как он своей ручищей покровительственно обхватил вздрагивающие плечи Мэри: «Его поставят на ноги... поставят на ноги...» «Но в конце-то концов, – думала Саманта,*



*оправдывая мужа, – кто мог знать, как оно обернётся, если медики тут же нацепили маски и забежали со шприцами?» Похоже, они надеялись его спасти – не понимали, что это без толку, пока в больнице к Мэри не подошла молоденькая докторша. У Саманты перед глазами до сих пор с ужасающей чёткостью стояла эта картина: и беззащитное, застывшее лицо Мэри, и вид этой гладко причёсанной, очкастенькой девушки в белом халате, собранной и немного настороженной... В сериалах такое показывают что ни день, но когда взаправду... – Ровным счётом ничего» [Роулинг, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Страх / Ужас». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Страх / Ужас»: «замертво», «ему конец», «ужасающей».

**Фрагмент №3:** *«Но, позволяя этим мыслям бежать со всей доступной им быстротой, Эйлиш все-таки останавливалась, когда ее сознание подходило к настоящему страху, если не ужасу или чему-то похуже, к размышлениям о том, что ей предстоит навсегда потерять ее мир, что она никогда больше не сможет провести привычный день в этом привычном доме, что остаток ее жизни обратится в борьбу с неведомым. Впрочем, сходя вниз, к Роуз и маме, она разговаривала с ними о делах практических и сохраняла веселость» [Тойбин, 2017].*

По результатам опроса, в данном фрагменте содержится эмоция «Страх / Ужас». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Страх / Ужас»: «страху», «ужасу», «навсегда потерять».

**Фрагмент №4:** *«Дорогой на нее напал страх перед предстоящим днем. А немного позже, когда она сидела, глядя в меню, у стойки закуской, ей вспомнились обрывки другого сна, который ко времени пробуждения наполовину стерся из памяти. В нем она тихим днем летела, словно на воздушном шаре, над спокойным морем. Под собой она видела утесы Куш-Гэна и мягкий песок Балликоннигара. Ветер отнес ее к Блэкуотеру, потом к Баллаху, к Монагиру, а оттуда к Винегар-Хиллу и Эннискортти. Она так ушла*

в воспоминания об этом сне, что официант из-за стойки спросил, все ли у нее в порядке» [Тойбин, 2017].

По результатам опроса, в данном фрагменте содержится эмоция «Страх / Ужас». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Страх / Ужас»: «*страх*».

Вторым эмоциональным классом для анализа стала группа «Гнев / Ярость». Далее мы рассмотрим несколько текстовых фрагментов и их дальнейший анализ.

**Фрагмент №5:** *«Он возмущался, как будто медицина в который раз показала свою никчёмность, не совершив самого простого и очевидного. Эндрю внутренне злорадствовал. Он заметил, что в последнее время отец, заслышав из уст матери медицинские термины, непременно высказывает дурацкие, невежественные соображения. Кровоизлияние в мозг. «Подключили бы какой-нибудь аппарат». А мать и бровью не повела. Как всегда. Эндрю жевал хлопья «Витабикс» и закипал ненавистью»* [Роулинг, 2013].

По результатам опроса, в данном фрагменте содержится эмоция «Гнев / Ярость». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Гнев / Ярость»: «*возмущался*», «*злорадствовал*», «*закипал ненавистью*».

**Фрагмент №6:** *«Измученная Кей Боден лежала без сна в кровати своего дома на Хоуп стрит, отдыхая в тишине пэзфордского рассвета и наблюдая за Гайей, которую уложила рядом с собой: в первых лучах солнца дочка выглядела бледной и дистрофичной. На полу стояло ведро, которое Кей принесла сюда поздно ночью, когда чуть ли не на руках притащила дочь в комнату из ванной, где до этого битый час придержививала ей волосы, чтобы они не падали в унитаз. Кей смотрела на лицо спящей дочери и вспоминала, как шестнадцать лет назад эта маленькая красавица так же спала вместе с ней. Она помнила, как Гайя лила слёзы, узнав о её*

*расставании с бойфрендом Стивом, с которым они прожили восемь лет. Стив ходил к Гайе в школу на родительские собрания, научил её кататься на велосипеде. Потом Кей мечтала, что они с Гэвином создадут семью и у Гайи наконец будет постоянный отчим и красивый дом за городом (сейчас эта фантазия казалась ей такой же глупой, как желание четырёхлетней Гайи иметь единорога). Как страстно она желала, чтобы у их истории был счастливый конец, чтобы Гайя жила в тех условиях, к которым привыкла; отъезд дочери надвигался на Кей неотвратно, как метеорит, и она предвидела, что расставание будет равносильно катастрофе, которая пошатнёт весь её мир» [Роулинг, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Гнев / Ярость». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Гнев / Ярость»: «лила слёзы», «неотвратно».

**Фрагмент №7:** *«В избирательных кабинках копались две старушки. В ожидании Саманта разглядывала их спины и серо стальную химическую завивку, толстые пальто и толстые лодыжки. Вот и она когда нибудь станет такой же. Заметив Майлза, более сгорбленная из двоих просияла и сообщила:*

*– Мы только что проголосовали за вас.*

*– О, большое вам спасибо. – Майлз был в восторге.*

*Войдя в кабинку, Саманта взяла подвешенный на верёвочке карандаш и стала изучать свой бюллетень с двумя именами: Майлз Моллисон и Колин Уолл. Потом написала по диагонали: «Ненавижу чёртов Пэгфорд», сложила бюллетень пополам и с серьёзным видом бросила в избирательную урну» [Роулинг, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Гнев / Ярость». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Гнев / Ярость»: «ненавижу», «с серьёзным видом».

**Фрагмент №8:** *«Когда я наконец вернулся домой, моя сестра сидела на стуле. Мама и папа стояли перед ней. И я понял, что Билл позвонил домой и всё им рассказал. Я почувствовал себя отвратительно. Это была моя вина.*

*Моя сестра плакала. Мама была очень, очень тихой. А отец говорил. Он сказал сестре, что запрещает ей видеться с тем парнем, который ударил её. Более того, он собирался поговорить с его родителями этим же вечером. Сестра сказала, что она сама виновата, что спровоцировала его, но отец сказал, что это не оправдание.*

*– Но я люблю его! – я никогда не видел, чтобы моя сестра так плакала.*

*– Неправда.*

*– Я ненавижу тебя!*

*– Неправда, – папа порой бывает очень спокойным.*

*– Он – вся моя жизнь.*

*– Никогда не говори так, о ком бы то ни было. Даже обо мне, – это сказала мама» [Чбоски, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Гнев / Ярость». Ключевые слова и словосочетания, которые, возможно, указывают на эмоцию «Гнев / Ярость»: «*отвратительно*», «*я ненавижу тебя!*».

Мы произвели анализ эмоций текстовых фрагментов «Страх / Ужас» и «Злость / Гнев». Далее, используя лингвистический инструмент основанный, на базе корпусов «Sketch Engine», мы выявили языковые маркеры для каждой проанализированной эмоции. Sketch Engine позволяет создавать собственный языковой корпус и используя имеющиеся функции, найти необходимые языковые особенности. Для поиска языковых маркеров мы применили такие функции для анализа как word list (генерируется списки часто употребляемых слов, которые могут быть отфильтрованы с помощью сложных критериев) и keyword / terms (автоматическое извлечение как одноязычных, так и двуязычных терминов из текстов (основанные на частоте и лингвистических критериях). Например, при создании собственного корпуса, был создан

критерий сравнения с уже имеющимся языковым корпусом русского языка. В дальнейшем поможет программному обеспечению быстрее произвести анализ тональности текста.

Таблица 1. Языковые маркеры: «Страх / Ужас» и «Гнев / Ярость»

Страх / Ужас	Гнев / Ярость
Страшно	Неправда
Пугался	Плакала
Вечером	Ненавижу
Страх	Игнорировала
Полночь	Гневным
Вскрикнула	Обидными
Полночью	Спровоцировала
Паника	Отвратительно
Ужасная	Катастрофе
Разрушающее	Приступ рвоты
Страха	Очередной приступ
Убить	Злорадствовал
Жаль	Закипал ненавистью
Ужасная паника	С серьезным видом
Ближайший крупнейший город	Отвратительно
Крупнейший город	Я ненавижу тебя!
Тишайший днем	
Испуганным	
Озабоченным	
Напугало	
Замертво	
Ему конец	
Ужасающей	
Навсегда потерять	

### 2.3. Эмоциональный класс текстовых данных «Стыд / Унижение (Shame / Humiliation), Грусть / Тоска (Distress / Anguish)»

Третьим эмоциональным классом для анализа стала группа «Стыд / Унижение». Далее мы рассмотрим несколько текстовых фрагментов и их дальнейший анализ.

**Фрагмент №9:** *«На неё поочередно накатывали тошнота и унижение. Как она мечтала забыть, что вчера перебрала, но у неё перед глазами всё ещё стояло лицо того мальчика, которого она лапала... Она помнила все выпуклости его тела, такого стройного, такого юного...»*

*Будь на его месте Викрам Джаванда, это бы ещё куда ни шло... Ей страшно захотелось кофе. Не сидеть же всю жизнь в ванной. Но перед тем, как открыть дверь, она увидела своё отражение в зеркале и сразу подрастеряла кураж. Отёчное лицо, заплавленные глаза, резкие морщины, подчеркнутые стрессом и обезвоживанием.*

*«Господи, что он обо мне подумал...»*

*В кухне Саманта застала Майлза. Даже не посмотрев в его сторону, она подошла прямо к шкафчику, где они держали кофе» [Роулинг, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Стыд / Унижение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Стыд / Унижение»: «унижение», «подрастеряла кураж».

**Фрагмент №10:** *«Те же немногие, что обращались, делали это с такой нарочитой учтивостью, что Эйлиш становилось неловко и даже стыдно. После того как на прилавке появились «Кофе» и «Сепия», ей пришлось показывать покупательницам эти чулки, рассказывать о преимуществах этих моделей, однако цветные покупательницы не уделяли ее объяснениям никакого внимания. К концу дня Эйлиш изматывалась и вскоре обнаружила, что едва ли не отдыхает на вечерних лекциях, испытывает облегчение, поскольку они отвлекают ее от мыслей о царившей*

*в магазине напряженности, которая в особенности сгущалась вокруг ее прилавка. Она жалела, что ее отобрали для этой работы, и гадала, не удастся ли ей со временем перебраться в другой отдел» [Тойбин, 2017].*

По результатам опроса, в данном фрагменте содержится эмоция «Стыд / Унижение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Стыд / Унижение»: «стыдно».

**Фрагмент №11:** *«Она только что совершила жуткий каждодневный ритуал: открыла свою страницу в «Фейсбуке» и удалила очередное сообщение от неизвестного отправителя. Как только она заносила его имя в чёрный список, профиль менялся – и мерзости продолжались. Они могли появиться в любое время. Сегодня ей прислали чёрно белую цирковую афишу девятнадцатого века: «Подлинная бородатая женщина, мисс Анна Джонс Эллиот». На ней была изображена одетая в кружевное платье темноволосая девушка с окладистой бородой и пышными усами. Сухвиндер считала, что за этим стоит Пупс Уолл, хотя, возможно, и ктонибудь другой. Тот же Дейн Талли с друзьями, которые всякий раз, когда она отвечала на уроках английского, издавали сдавленное обезьянье уханье. Они бы точно так же издевались и над другими ребятами с тёмным цветом кожи, просто в школе «Уинтердаун» таких больше не нашлось. Сухвиндер мучилась от позора и унижения, тем более что мистер Гэрри ни разу их не одёрнул. Как будто оглох. Наверное, он и сам считал, что Сухвиндер Каур Джаванда – обезьяна, волосатая обезьяна» [Роулинг, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Стыд / Унижение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Стыд / Унижение»: «позора», «унижения», «мерзости».

**Фрагмент №12:** *«Этот парень, Дэйв, сейчас выпускник. Он играет в футбольной команде. Он принимающий. Я видел, как в конце игры Дэйв поймал тачдаун, брошенный Брэдом. Наша школа победила. Болельщики на*

*стадионе с ума походили. Но всё, о чём я мог думать, была та вечеринка. Я долго думал об этом, потом посмотрел на Сэм.*

*– Он изнасиловал её, да?*

*Она только кивнула.*

*– Наверное, нам следует кому-нибудь рассказать?*

*Сэм покачала головой. Она объяснила, через что нужно пройти, чтобы это доказать, особенно если эти старшеклассники пользуются популярностью и всё ещё встречаются» [Чбоски, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Стыд / Унижение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Стыд / Унижение»: *в данном фрагменте отсутствуют типичные слова, связанные со стыдом или унижением, но из контекста ясно, что героиня испытывает стыд в связи с изнасилованием. Для чистоты исследования, данный текстовый фрагмент будет использован в Sketch Engine вместе с остальными.*

Четвертым эмоциональным классом для анализа стала группа «Грусть / Тоска». Далее мы рассмотрим несколько текстовых фрагментов и их дальнейший анализ.

**Фрагмент №13:** *«Пупсу было тоскливо, хотя одноклассников он веселил, как прежде. Он ведь не просто так стремился сбросить оковы морали, а лишь для того, чтобы восстановить всё то, что в нём подавили, всё, что он растерял, когда вышел из детства. Пупс хотел восстановить своего рода невинность и проторил к ней путь сквозь те качества, которые считались вредными, но, как ни парадоксально, виделась Пупсу единственно надёжными вехами на подступах к аутентичности, к некой чистоте. Любопытно, в самом деле, как часто все истины оказывались вывернутыми наизнанку и противоречили тому, что ему вдалбливалось; Пупс даже стал думать, что истина – это и есть житейская мудрость, поставленная с ног на голову. Его влекли тёмные лабиринты и неслыханные препоны, что*



*маячили внутри; он хотел расколоть благонравие и разоблачить ханжество; хотел смести запреты, чтобы выжать знание из их кровавого нутра; хотел обрести дар аморальности, пройти крещение неискушённой и простотой» [Роулинг, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Грусть / Тоска». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Грусть / Тоска»: «*тоскливо*».

**Фрагмент №14:** *«Мне стало так грустно. Я не понимал, что происходит. Мама старалась быть мягкой, потому что, когда я становлюсь таким, она обычно пытается всё уладить.*

*– Извини, мам.*

*– Нет-нет, не извиняйся. Ты просто хочешь купить папе отличный подарок. Это хорошо.*

*– Мам! – сестра и вправду начала злиться.*

*Мама даже не взглянула на неё.*

*– Чарли, можешь купить папе, что захочешь. Я знаю, ему понравится. Успокойся. Всё хорошо.*

*Мы с мамой зашли в четыре разных отдела. В каждом сестра садилась на первый же стул и стонала. Наконец я нашёл прекрасный магазин. Там продавались фильмы. Я нашёл видеокассету с последним эпизодом сериала «М\*А\*S\*Н» без рекламы. И я почувствовал себя гораздо лучше. Потом я начал рассказывать маме, как мы все вместе смотрели его» [Чбоски, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Грусть / Тоска». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Грусть / Тоска»: «*грустно*», «*извини*».

**Фрагмент №15:** *«И Хелен тоже. Однажды... я пошёл на фабрику... там для меня не было никакой работы... никакой... Я пришёл домой в два часа утра... был ужасно зол... ваша бабушка показала мне их табели... с*

*плохими оценками... а ведь они были умными девочками. Так что я пошёл в их комнату и попытался выбить из них всю дурь... когда я закончил, они плакали, а я взял их таблицы и произнёс: «Этого больше никогда не повторится». Она до сих пор говорит об этом... ваша мать... но знаешь что... это больше никогда не повторилось... они поступили в колледж... обе. Я хотел их отправить... всегда хотел отправить их... Жаль, что Хелен этого не понимала. Думаю, ваша мама поняла... в глубине души... она хорошая женщина... вы должны ею гордиться.*

*Когда я рассказал об этом маме, ей стало грустно, потому что дед никогда не говорил ей этого. Никогда. Даже когда вёл её к алтарю.*

*Но этот День благодарения был другим. Мы смотрели футбольную игру моего брата, которую специально для родственников принесли на видеокассете. Вся семья собралась перед телевизором, даже мои двоюродные бабушки, которые вообще не смотрят футбол» [Чбоски, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Грусть / Тоска». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Грусть / Тоска»: «грустно»

**Фрагмент №16:** *«Это был последний раз, когда я видел её. Мне нравится думать, что теперь у неё есть хорошая работа, для которой она училась. Мне нравится думать, что она встретила хорошего мужчину. Мне нравится думать, что она сбросила вес, который всегда мечтала сбросить без диеты.*

*Несмотря на все заверения моих родителей и доктора о том, что в случившемся не было моей вины, я не могу перестать думать, что это так. И я знаю, что тётя Хелен была бы жива сегодня, если бы просто купила мне один подарок, как и все остальные. Она была бы жива, если бы я родился не в такой заснеженный день. Я бы хотел как-то отогнать это чувство. Я ужасно скучаю по ней. Я должен перестать писать сейчас, потому что мне слишком грустно.*

С любовью, Чарли» [Чбоски, 2013].

По результатам опроса, в данном фрагменте содержится эмоция «Грусть / Тоска». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Грусть / Тоска»: *«слишком грустно», «скучаю»*.

Мы произвели анализ эмоций текстовых фрагментов «Стыд / Унижение» и «Грусть / Тоска». Далее, используя лингвистический инструмент, основанный на базе корпусов «Sketch Engine», мы выявили языковые маркеры для каждой проанализированной эмоции. Для поиска языковых маркеров мы применили такие функции для анализа как word list и keyword / terms.

Таблица 2. Языковые маркеры: «Стыд / Унижение» и «Грусть / Тоска»

Стыд / Унижение	Грусть / Тоска
Стыдно	Грустно
Жалела	Никогда
Напряженности	Сумрачным
Захлестнул	Чувствах
Напряженности	Скучаю
Оцепенение	Тосковать
Полное оцепенение	Темнота
Жутчайший	Страшная тоска
Ужаснейший стыд	Грустно
Приступ	Извини
Позора	
Унижения	
Мерзости	

## 2.4. Эмоциональный класс текстовых данных «Радость / Счастье (Enjoyment / Joy), Интерес / Возбуждение (Interest / Excitement)»

Пятым эмоциональным классом для анализа стала группа «Радость / Счастье». Далее мы рассмотрим несколько текстовых фрагментов и их дальнейший анализ.

**Фрагмент №17:** *«Джордж опустил стакан на стойку и, повернувшись к Нэнси, пригласил ее на танец, а затем взглянул на Джима, словно призывая его сделать то же самое. Нэнси улыбнулась Джорджу, а за ним – Эйлиш и Джиму, поставила свой стакан и вышла с Джорджем на середину зала. У нее явно полегчало на душе, выглядела она счастливой. Эйлиш, обводя взглядом зал, сознавала, что она и Джим остались у бара одни и что все места у стены, в которой прорезана ведущая в дамскую уборную дверь, заняты. Она словно в капкан попала и уйти может только в уборную или домой. На секунду ей показалось, что Джим Фаррелл сейчас шагнет к ней и пригласит потанцевать, и она готова была принять это приглашение, выбора у нее не осталось – не грубить же приятелю Джорджа» [Тойбин, 2017].*

По результатам опроса, в данном фрагменте содержится эмоция «Радость / Счастье». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Радость / Счастье»: «улыбнулась», «полегчало», «счастливой».

**Фрагмент №18:** *«Эйлиш всегда нравилось закрывать дверь своей комнаты, задергивать шторы. Сейчас ей хотелось лишь одного – спать, несмотря на то что прошлой ночью она хорошо поспала в отеле Рослэр-Харбора. Из Кова она послала Тони открытку с сообщением, что благополучно добралась до суши, а из Рослэра – письмо о плавании. И теперь порадовалась, что ей не придется писать, сидя в своей спальне, которая казалась Эйлиш лишенной признаков жизни, – она едва ли не испугалась,*

*осознав, сколь мало значит для нее эта комната. Эйлиш не представляла себе заранее, каким станет ее возвращение домой, ожидала лишь, что оно окажется легким; она так сильно скучала по привычным комнатам, что ожидала испытать, едва переступив порог, счастье и облегчение, а взамен того могла лишь считать в это первое утро оставшиеся до отъезда дни. Странное, виноватое чувство охватило ее; она свернулась в постели калачиком и закрыла глаза» [Тойбин, 2017].*

По результатам опроса, в данном фрагменте содержится эмоция «Радость / Счастье». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Радость / Счастье»: «благополучно», «счастье», «облегчение».

**Фрагмент №19:** *«Это напоминает о том, что порой, когда ты очень вдохновлён какой-то девушкой, и видишь людей, держащихся за руки, их сплетённые пальцы, то чувствуешь огромную радость за них. А в другое время ты видишь ту же пару, но они не вызывают ничего, кроме раздражения. И всё, чего тебе хочется – это всегда чувствовать за них радость, потому что это означает, что ты тоже счастлив.*

*Я только что вспомнил, что именно заставило меня размышлять обо всём этом. Я напишу причину, и тогда, возможно, получится меньше об этом думать. И я перестану расстраиваться. Дело в том, что я слышу, как Сэм и Крейг занимаются сексом, и впервые в жизни я понял конец того стихотворения. Хотя я никогда не хотел понимать этого. Поверь мне» [Чбоски, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Радость / Счастье». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Радость / Счастье»: «радость», «счастлив».

**Фрагмент №20:** *«Я хочу рассказать тебе, как мы бежали. Был очень красивый закат. И этот холм. Зелёный холм с восемнадцатой лункой, где мы с Патриком плевались от смеха вином. Всего за несколько часов до этого*

*Сэм, Патрик и все, кого я люблю и знаю, закончили свой последний день старшей школы. И я был счастлив, потому что они радовались. Сестра даже позволила мне обнять себя в коридоре. «Поздравляю» было словом дня. И мы с Сэм и Патриком отправились в «Большой парень» и курили. Потом гуляли, дожидаясь, когда наступит время идти на «Шоу ужасов Рокки Хоррора». И болтали о том, что в тот момент казалось важным. И смотрели на этот холм. И потом Патрик побежал вслед за уходящим солнцем. Сэм немедленно последовала за ним. Я видел их силуэты. Бежавших вслед за солнцем. Тогда и я побежал. И всё было так хорошо, как только могло быть» [Чбоски, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Радость / Счастье». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Радость / Счастье»: «смеха», «счастлив», «радовались», «хорошо».

Шестым эмоциональным классом для анализа стала группа «Интерес / Возбуждение». Далее мы рассмотрим несколько текстовых фрагментов и их дальнейший анализ.

**Фрагмент №21:** *«Её лёгкие, быстрые шаги приближались; вот она уже здесь; вот она села рядом. Слегка задев его стул, она задела и бок Эндрю. Ноздри его уловили аромат духов. Вся левая половина туловища горела от её соседства, и он порадовался, что у него на левой щеке (на той, что ближе к ней) почти нет прыщей. Никогда ещё он не находился от неё так близко и теперь не знал, решится ли на неё посмотреть, дать знак, что отметил её присутствие; впрочем, он тут же сказал себе, что слишком долго думал и уже не сумеет сделать это естественно» [Роулинг, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Интерес / Возбуждение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Интерес / Возбуждение»: «порадовался».

**Фрагмент №22:** *«Один парень на вечеринке, Марк, который дал мне это попробовать, появился из ниоткуда, и посмотрел на небо, и сказал мне взглянуть на звёзды. Тогда я поднял глаза, и мы были под огромным куполом, словно в стеклянном снежном шаре, и Марк сказал, что все эти удивительные белые звёзды на самом деле просто дырки в чёрном стекле купола. Он рассказал, что в момент, когда мы отправимся на небеса, стекло разобьётся, и не останется ничего, кроме белого звёздного пространства, которое светит ярче всего на свете, не причиняя боли глазам. Было так просторно и тихо, и я чувствовал себя таким маленьким.*

*Иногда я смотрю на улицу и думаю, как много других людей уже видели этот снег раньше. Или думаю, как много людей уже читали эти книги. Или слушали эти песни. Мне интересно, что они чувствуют этим вечером. Я правда не знаю, о чём говорю. Я, наверное, не должен писать всё это, потому что вещи вокруг по-прежнему движутся. Я хочу, чтобы они перестали, но это будет продолжаться ещё несколько часов. Так сказал Боб, перед тем как уйти в свою спальню с Джилл, девушкой, которую я не знаю. Думаю, всё, что я говорю сейчас, кажется знакомым. Это не кажется знакомым мне. Я имею в виду, что кто-то другой чувствовал подобное. Когда на улице всё спокойно, и вещи вокруг движутся, и ты хочешь, чтобы это прекратилось, а все спят. И все прочитанные тобой книги уже были прочитаны другими людьми. И все твои любимые песни уже слушали другие люди. И симпатичная тебе девушка симпатична и другим людям. И ты знаешь, что если вспомнить те моменты, когда ты был счастлив, ты почувствуешь себя превосходно потому, что был частью единого» [Чбоски, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Интерес / Возбуждение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Интерес / Возбуждение»: «попробовать», «интересно».

**Фрагмент №23:** *«Мать так противилась ее переезду в Англию, что эта мысль стала для Эйлиш потрясением. Интересно, пошли бы мама с сестрой на такой разговор, если бы она не начала работать в магазине, не рассказывала бы им об унижениях, которые еженедельно сносит от мисс Келли? И Эйлиш пожалела, что так много всего наболтала; но ведь она делала это главным образом потому, что ее рассказы смешили Роуз и маму, делали семейные трапезы веселыми, облегчали и украшали их жизнь лучше, чем все, что произошло после смерти отца и отъезда мальчиков. Ее только теперь осенило, что они вовсе не находили ее работу у мисс Келли забавной и именно поэтому ни словом не возразили отцу Флуду когда он закончил нахваливать свой приход и заверил всех, что сможет подыскать для Эйлиш подходящее место» [Тойбин, 2017].*

По результатам опроса, в данном фрагменте содержится эмоция «Интерес / Возбуждение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Интерес / Возбуждение»: «интересно», «пошли бы».

**Фрагмент №24:** *«Когда вновь заиграла музыка, многие их друзья ушли танцевать, но Джим Фаррелл остался у стойки. Джордж, выдав по стакану лимонада Нэнси и Эйлиш, официально представил их Джиму, и тот коротко кивнул, но руки для пожатия не протянул. Джордж, похоже, растерялся – стоял, потягивая лимонад. Потом сказал что-то Нэнси, та ответила. Джордж снова приложился к лимонаду. Интересно, что он собирается делать? – подумала Эйлиш; ясно же, что его другу ни Нэнси, ни Эйлиш не нравятся, разговаривать с ними он не желает; не стоило ей подходить к бару Она пила лимонад и смотрела в пол. А подняв взгляд, увидела, что Джим Фаррелл холодно изучает Нэнси, сообразив же, что Эйлиш наблюдает за ним, он переступил с ноги на ногу и повернул к ней лишенное всякого выражения лицо. На нем была дорогая спортивная куртка поверх рубашки с шейным платком» [Тойбин, 2017].*



По результатам опроса, в данном фрагменте содержится эмоция «Интерес / Возбуждение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Интерес / Возбуждение»: «интересно», «что он собирается делать?».

Мы произвели анализ эмоций текстовых фрагментов «Радость / Счастье» и «Интерес / Возбуждение». Далее, используя лингвистический инструмент, основанный на базе корпусов «Sketch Engine», мы выявили языковые маркеры для каждой проанализированной эмоции. Для поиска языковых маркеров мы применили такие функции для анализа как word list и keyword / terms.

Таблица 3. Языковые маркеры: «Радость / Счастье» и «Интерес / Возбуждение»

Радость / Счастье	Интерес / Возбуждение
<p>Радостно Хорошо Нравилось Огромная радость Благополучно Счастье Облегчение Радость Счастлив Радовались Хорошо Самый красивый закат Малейший намек Наибольший отрыв Наибольший парень</p>	<p>Интересно Обрадовало Удивительный Лучшая Заинтересовало Захватило Привлекательное Попробовать Что он собирается делать?</p>

## 2.5. Эмоциональный класс текстовых данных «Презрение / Отвращение (Contempt / Disgust)», «Удивление (Surprise)»

Седьмым эмоциональным классом для анализа стала группа «Презрение / Отвращение». Далее мы рассмотрим несколько текстовых фрагментов и их дальнейший анализ.

**Фрагмент №25:** *«Наверное, я кивнул, потому что она помогла мне одеться. Потом она надела блузку. И мне захотелось хорошенько пнуть самого себя за то, что веду себя как ребёнок. Потому что я любил Сэм. И мы были вместе. И я всё разрушил.*

*Просто разрушил. Ужасно. Я чувствовал себя отвратительно.*

*Она вышла со мной на улицу.*

*– Тебя отвезти? – спросила она. У меня была машина отца. Я не был пьян. Она выглядела очень обеспокоенной.*

*– Нет, спасибо.*

*– Чарли, я не позволю тебе вести машину в таком состоянии.*

*– Извини. Я тогда пойду пешком, – сказал я.*

*– Сейчас два часа утра. Я отвезу тебя домой» [Чбоски, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Презрение / Отвращение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Презрение / Отвращение»: «ужасно», «разрушил», «отвратительно».

**Фрагмент №26:** *«Они молча ели тосты, запивая их свежей порцией кофе. Жевательные движения умили раздражение Саманты. Она вспомнила, как с рассветом судорожно подскочила на кровати в тёмной спальне и до смешного быстро успокоилась, когда с благодарностью нащупала рядом с собой Майлза, такого большого, с брюшком, пахнущего камфорным маслом и застарелым потом. Вслед за тем она представила, как будет рассказывать покупательницам у себя в бутике, что прямо у неё*

на глазах человек упал замертво и что они с мужем помчались на «скорой» в больницу. Она до мелочей продумала, как опишет все этапы поездки, вплоть до самой драматичной сцены, когда появилась докторша» [Роулинг, 2013].

По результатам опроса, в данном фрагменте содержится эмоция «Презрение / Отвращение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Презрение / Отвращение»: «раздражение».

**Фрагмент №27:** *«Из потемневшего от времени зеркала на него уставилась собственная личность. Под глазами темнели круги, а жидкие и сухие светлые волосы торчали клочьями. Голая лампочка с жестокостью фоторобота обрисовывала слабовольную козлиную физиономию. «Тридцать четыре, – подумал он, – а выгляжу на все сорок». Он занёс бритву и приготовился осторожно срезать две толстые светлые волосины, которые упрямо росли по обеим сторонам выступающего кадыка. В дверь забарабанили кулаками. Рука у него дрогнула, и с тонкой шеи на свежую белую рубашку закапала кровь. – Твой бойфренд, кажется, уснул в ванной, – вскричал раздражённый девичий голос, – а я уже опаздываю! – Я закончил! – гаркнул он в ответ. Порез саднило; ну и что? Зато появился удобный предлог: Посмотри, что случилось из-за твоей дочери. Теперь мне придётся перед работой мчаться домой, чтобы сменить рубашку. Почти что с лёгким сердцем он схватил галстук и пиджак, висевшие на вбитом в дверь крючке, и открыл задвижку» [Роулинг, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Презрение / Отвращение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Презрение / Отвращение»: «раздражённый».

**Фрагмент №28:** *«Когда я наконец вернулся домой, моя сестра сидела на стуле. Мама и папа стояли перед ней. И я понял, что Билл позвонил домой и всё им рассказал. Я почувствовал себя отвратительно. Это была моя вина.*

*Моя сестра плакала. Мама была очень, очень тихой. А отец говорил. Он сказал сестре, что запрещает ей видеться с тем парнем, который ударил её. Более того, он собирался поговорить с его родителями этим же вечером. Сестра сказала, что она сама виновата, что спровоцировала его, но отец сказал, что это не оправдание» [Чбоски, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Презрение / Отвращение». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Презрение / Отвращение»: «отвратительно», «ударил».

Восьмым эмоциональным классом для анализа стала группа «Удивление». Далее мы рассмотрим несколько текстовых фрагментов и их дальнейший анализ.

**Фрагмент №29:** *«По правде говоря, он слегка удивился, когда Гэвин сказал ему, что удостоен такой чести. У Майлза теплилась надежда, что им с Самантой будет отведено особое место, окружённое ореолом тайны и значительности: не они ли находились рядом с Барри до последнего его вздоха? Мэри или её близкие могли бы из уважения попросить его, Майлза, прочесть, как положено, отрывок из Священного Писания или же сказать несколько слов у гроба. Но Саманта всем своим видом показывала, что выбор совершенно справедливо был сделан в пользу Гэвина.*

*– Вы ведь с Барри были очень дружны, правда, Гэв?*

*Гэвин кивнул. Его знобило и подташнивало. Ночь прошла беспокойно: проснулся он ни свет ни заря от кошмарного сна, в котором уронил гроб, да так, что Барри выкатился на каменный пол; затем решил ещё немного вздремнуть – и увидел другой сон, как проспал похороны и примчался в церковь Архангела Михаила и Всех Святых к шапочному разбору, а Мэри, бледная и злая, одиноко стоявшая на кладбище, стала кричать, что его убить мало.*

– Ума не приложу, где мне следует находиться, – сказал он, оглядываясь по сторонам. – Никогда не выступал в такой роли.

– Дело нехитрое, – отозвался Майлз. – Главное – ничего не уронить, хи хи хи.

Писклявый смех Майлза совершенно не вязался с его басовитым голосом. Ни Гэвин, ни Саманта не улыбнулись.

Из плотной толпы появился Колин Уолл. Большой и нескладный, с высоким шишковатым лбом, он всегда напоминал Саманте чудовище Франкенштейна.

– Гэвин, – проговорил он, – вот ты где. Думаю, нам нужно стоять на тротуаре, они с минуты на минуту придут.

– Есть, – сказал Гэвин, радуясь, что за него всё решили.

– Колин, – кивнул Майлз.

– Да, здравствуйте. – Колин засуетился и тут же исчез» [Роулинг, 2013].

По результатам опроса, в данном фрагменте содержится эмоция «Удивление». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Удивление»: «удивился», «беспокойно».

**Фрагмент №30:** «От усталости Тессе вдруг показалось, что он совсем чужой; она даже удивилась, когда он свернул к машине, и подумала, как ей не хочется вылезать на злющий, невообразимый ветер, чтобы только пропустить этого незнакомца на заднее сиденье. Но когда он, подойдя вплотную, одарил её кривой полуулыбкой, она тут же узнала в нём своего любимого – несмотря ни на что – мальчика и, пока он усаживался сзади, стоически перетерпела острые, как нож, порывы ветра; отец не шелохнулся» [Роулинг, 2013].

По результатам опроса, в данном фрагменте содержится эмоция «Удивление». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Удивление»: «удивилась», «подумала».

**Фрагмент №31:** *«Я очень сильно люблю Сэм. Чувства к ней причиняют мне ужасную боль.*

*Я решил спросить сестру о парне, с которым она была на танцах. Она отказалась делиться со мной своим секретом, если я не пообещаю никому о нём не рассказывать, даже Биллу. И я пообещал. Она сказала, что видится с этим парнем тайком, ведь отец запретил им встречаться. Когда его нет рядом, она думает лишь о нём. И, по её словам, они собираются пожениться, после того как оба закончат колледж, а он – юридический факультет.*

*Она попросила меня не волноваться по этому поводу, потому что больше он её не бил. И никогда не будет. Ничего другого она мне не сказала, только повторяла одно и то же.*

*Приятно было сидеть в тот вечер со своей сестрой, ведь обычно она не любит со мной разговаривать. Я удивился, что она так много мне рассказала, но мне кажется, что ей просто нужно было с кем-то поделиться.*

*Но после её рассказа я не перестал за неё волноваться. В конце концов, она моя сестра» [Чбоски, 2013].*

По результатам опроса, в данном фрагменте содержится эмоция «Удивление». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Удивление»: «удивился».

**Фрагмент №32:** *«Вообще-то он знал об абрикосах больше, чем мы: их прививку, этимологию, происхождение, распространение в Средиземноморье и его окрестностях. За завтраком тем утром мой отец рассказывал, что название фрукта пришло из арабского, так как слово – в итальянском «albicossa», «abricot» во французском, «aprikose» в немецком, как слова «алгебра», «алхимия» и «алкоголь» – образовалось от арабского существительного с приставкой «al-». Изначально «albicossa» было «al-biriqu». Мой отец, не умея сопротивляться желанию не только раскрыть*

какой-то факт, но и присовокупить какой-нибудь свежей информации, добавил, как на самом деле поразительно, что сейчас в Израиле и многих арабских странах этот фрукт зовется совершенно иначе: «*misbmish*».

Моя мать была в замешательстве. Все мы, включая двух младших кузин, посетивших нас в ту неделю, были готовы заплодировать.

Но по вопросу этимологии, однако, Оливер осмелился не согласиться. «*Ха?!*» – удивился отец» [Асиман, 2018].

По результатам опроса, в данном фрагменте содержится эмоция «Удивление». В ходе анализа, были выявлены следующие языковые маркеры эмоций «Удивление»: «в замешательстве», «*ха?!*», «поразительно» «удивился».

Мы произвели анализ эмоций текстовых фрагментов «Презрение / Отвращение» и «Удивление». Далее, при помощи корпусного менеджера Sketch Engine мы выявили языковые маркеры для каждой проанализированной эмоции, что в дальнейшем поможет программному обеспечению быстрее произвести анализ тональности текста.

Таблица 5. Языковые маркеры: «Презрение / Отвращение» и «Удивление»

Презрение / Отвращение	Удивление
Отвратительно	Интересно
Разрушил	Обрадовало
Виновата	Удивительный
Физиономия	Лучшая
Отвратительный	Заинтересовало
Мучительный	Захватило
Оправдание	Привлекательное
Злая ирония судьбы	Попробовать
Раздражение	

Проанализировав все эмоциональные группы, мы обнаружили некоторые особенности. В эмоциональных группах «Страх / Ужас» и «Радость / Счастье» присутствует большое количество языковых маркеров как прилагательные в простой и сложной превосходной степени. В эмоциональных группах «Гнев / Ярость» и «Удивление» присутствует большое количество таких языковых маркеров как восклицательные местоимения. В остальных же эмоциональных группах присутствует равное количество языковых маркеров в виде разных частей речи (существительные, глаголы, прилагательные, наречия).

В следующем параграфе, имея полученные результаты, мы составим базу данных «языковой маркер – имя эмоции – тег Last.fm» и рассмотрим процесс реализации программы для синхронизации саундтрека и эмоциональной тональности текста.

## 2.6. Реализация программы-приложения для синхронизации саундтрека и эмоциональной тональности текста: основные трудности и результаты

Одним из основных шагов для реализация программы-приложения для синхронизации саундтрека и эмоциональной тональности текста это составление базы данных «языковой маркер – имя эмоции – тег Last.fm».

Для сбора данной информации, мы использовали представленную в открытом доступе данные на веб-сайте Last.FM и составили следующую базу данных. В данной таблице указаны метаинформация, которые мы планируем использовать для создания программы-приложения:

Таблица 6. База данных «имя эмоции – тег Last.fm»

«Страх / Ужас»	«Гнев / Ярость»
#terror, #speedcore, #extratone, #witchcraft	#rage, #ocultism, #ragebrake, #progressivemetal, #ragerap



<b>«Стыд / Унижение»</b>	<b>«Грусть / Тоска»</b>
<i>#guilt, #bittersweet, #folk, #punk, #shame, #guiltypleasures, #humility</i>	<i>#sad, #pop, #sadcore, #indie, #melancholic, #ambient</i>
<b>«Радость / Счастье»</b>	<b>«Интерес / Возбуждение»</b>
<i>#joy, #fun, #catchy, #comedy, #dance, #happy, #uplifting, #happyhardcore</i>	<i>#experimental, #ambient, #catchy, #melodic, #exiting, #</i>
<b>«Презрение / Отвращение»</b>	<b>«Удивление»</b>
<i>#contemporaryfolk, #rent, #grandeur, #contempt</i>	<i>#surprising, #uplifting, #melodic, #indie, #alternative</i>

Чтобы получить подходящую метаинформацию, относящуюся к музыке, мы решили использовать службу LastFM. Причинами выбора стал большой набор данных для различных жанров музыкальных композиций, включая популярность и теги поиска, которые были настроены на различные композиции. На сайте действует система тегов. Пользователи могут пометить альбом, песню или исполнителя произвольным наименованием – тегом (как правило, это название жанра или страны исполнителя). Сайт позволяет пользователям прослушивать музыку, хранящуюся на сайте (при помощи программы – клиента или плеера, встроенного на страницы сайта), помеченную каким-нибудь тегом (или несколькими тегами). Это ускорило поиск поиску различных музыкальных треков с определенной эмоцией.

Для создания списка музыкальных треков мы сформировали ранее помеченные текстовые фрагменты в определенной последовательности, где ни один элемент не имел те же эмоции, что и его сосед (см. Рис. 2).

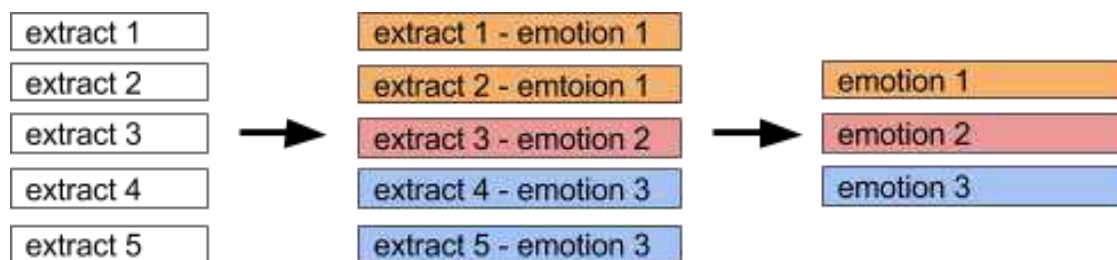


Рисунок 2. Схема формирования музыкальных списков

Для каждого элемента в этой последовательности мы сделали поисковый запрос для LastFM API, чтобы получить 100 треков с определенной меткой, которая соответствует ярлыку эмоций. После того, как каждый запрограммированный список тэгов был выбран из списка музыкальных треков, все песни для каждого элемента эмоции были перетасованы (см. Рис. 3).



Рисунок 3. Схема получения песни для подходящей эмоции

Используя созданную базу данных «языковой маркер – имя эмоции – тег Last.fm» и заранее сформированную обучающую выборку, будет предпринята попытка машинного обучения с учителем. Далее, будет произведена тестовая выборка, выборка, созданная по тем же принципам что и обучающая, но по которой будет оцениваться качество построенной программы для анализа. Для получения точной оценки созданной программы, появится необходимость провести третью, проверочную выборку,

что позволит нам в полной мере оценить качество анализа созданной программы.

В дальнейших исследованиях мы планируем реализовать программу-приложение и усовершенствовать скорость и результативность алгоритма. Также планируется разработать более детальный классификатор эмоций (несколько оттенков для каждой эмоции), что поможет алгоритму машинного обучения повысить точность определения эмоции.

## ВЫВОДЫ ПО ГЛАВЕ 2

Во второй главе настоящей магистерской диссертации были проанализированы 4 художественных произведения: Дж. К. Роулинг «Случайная Вакансия», К. Тойбин «Бруклин», С. Чбоски «Хорошо быть тихоней», А. Асиман «Зови меня своим именем». Методом сплошной выборки, мы отобрали 100 фрагментов, за каждым из которых была закреплена определенная эмоция.

После этого мы обратились к 35 экспертам для 35 оценки коллекции текстовых фрагментов, приписав каждому из них определенную эмоцию из восьми предложенных.

Далее, используя такой лингвистический инструмент как корпусный менеджер «Sketch Engine», мы выявили языковые маркеры для каждой проанализированной эмоции. Стоит упомянуть, что для поиска языковых маркеров мы применили такие функции для анализа как word list и keyword / terms. Например, при создании собственного корпуса, нами был создан критерий сравнения с уже имеющимся языковым корпусом русского языка.

После этого, мы подошли к созданию базы данных «языковой маркер – имя эмоции – тег Last.fm». Для сбора данной информации, мы использовали представленную в открытом доступе данные на веб-сайте Last.FM. Причинами выбора стал большой набор данных для различных жанров музыки, включая популярность и теги поиска. Это способствовала поиску различных музыкальных треков с определенной эмоцией.

Далее, мы подошли к реализации программы-приложения для синхронизации саундтрека и эмоциональной тональности текста: проанализировали имеющиеся трудности и результаты.

## ЗАКЛЮЧЕНИЕ

Целью данной научно-исследовательской работы явилась разработка компьютерной программы для синхронизации музыкального сопровождения и эмоциональной тональности текста на электронном носителе на основе использования принципов сентимент-анализа текста и классификации эмоций «Куб Лёвхейма».

Необходимость исследования была обусловлена тем, что на данный момент проектов по созданию специального классификатора, способного автоматически определять точную эмоцию, отражённую в тексте, пока немного, а для русскоязычных данных – нет. Также одной из целей стала разработка программы автоматического подбора саундтрека для русскоязычного текста.

Так, в первой главе настоящей научно-исследовательской работы были рассмотрены и изучены труды отечественных и зарубежных ученых в сферах, касающихся сентимент-анализа как направление прикладной лингвистики. Проанализированы принципы и методы формирования обучающей выборки для технологии машинного обучения.

Мы убедились, что точность сентимент-анализа определяется выбранным набором средств классификации обучающей выборки. Следует упомянуть, что художественный текст имеет воздействующую функцию на читателя. Также в художественном тексте реализованы языковые единицы всех уровней (от фонемы до предложения), следовательно, наличие языковых единиц всех уровней позволит нам найти подходящие языковые маркеры для каждой эмоциональной группы, что доказывает правильность выбора художественного текста как особого источника данных для формирования обучающей выборки.

Для повышения точности исследования, мы взяли за основу классификацию эмоций Г. Лёвхейма. Он установил, что, хотя сами по себе

эмоциональные состояния, являясь функцией от адаптивных систем человеческого организма, порождаются в лимбической системе и миндалевидном теле головного мозга, но дальнейший сигнал об эмоции активируется и распространяется на другие отделы головного мозга благодаря действию трех моноаминов: серотонина, допамина и норадреналина.

В практической части настоящего исследования мы сформировали принципы формирования обучающей выборки.

В качестве источника обучающей выборки, мы выбрали художественное произведение – Дж. К. Роулинг «Случайная Вакансия», К. Тойбин «Бруклин», С. Чбоски «Хорошо быть тихоней», А. Асиман «Зови меня своим именем». Выбор был обусловлен тем, что после просмотра рецензий на данные книги, мы выяснили, что текст вызывает сильный эмоциональный отклик у читателей, создавая максимально эмоциональное состояние.

Затем, основываясь на уже выбранной восьмеричной классификации эмоций, мы предложили 35 экспертам оценить коллекцию текстовых фрагментов, приписав каждому из них определенную эмоцию из восьми предложенных. После этого, нами были созданы текстовые корпуса для выявления необходимых языковых маркеров для каждой эмоции для последующей разработки теоретических принципов базы данных «языковой маркер – имя эмоции – тег Last.fm».

Для технической реализации сентимент-анализа будет использовано машинное обучение с учителем. Обучение с учителем – один из способов машинного обучения, в ходе которого испытываемая система принудительно обучается с помощью уже созданной *обучающей выборки*. Далее, будет произведена *тестовая выборка*, выборка, по которой оценивается качество построенной программы для анализа. Для получения точной оценки созданной программы, появится необходимость провести третью,

*проверочную выборку*, что позволит нам в полной мере оценить качество анализа созданной программы.

Говоря о будущих перспективах данного научного исследования, интересным продолжением работы может стать более детальная настройка и дальнейшая реализация программы для синхронизации саундтрека и тональности текста.

Подводя итоги вышесказанному необходимо отметить, что создание программы для анализа тональности текста является сложной задачей, но вполне посильной, если имеются данные для обучения и заранее определена тема. При использовании машинного обучения важно тестировать разные параметры, чтобы подобрать те, которые работают лучше на тестовых данных.

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Азимов Э.Г., Щукин А.Н. Новый словарь методических терминов и понятий (теория и практика обучения языкам). М.: ИКАР. 2009. 448 с.
2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983. 471 с.
3. Алифанова О.Г., Исаев, Павлова А.В., Рыхлова О.С., Турлова Е.В., Хрущева О.А., Щербакова М.В. Перспективные научные исследования в языкознании: колл. монография. М.: Флинта: Наука, 2012. 200 с.
4. Антонова А.С., Соловьев Ю.А. Использование метода условных случайных полей для обработки текстов на русском языке // Компьютерная лингвистика и интеллектуальные технологии: «Диалог-2013». М.: РГГУ, 2013. 44 с.
5. Арнольд И.В. Основы научных исследований в лингвистике. М.: Высшая школа, 2011. 140 с.
6. Бабенко Л.Г., Казарин Ю.В. Лингвистический анализ художественного текста: учебник; практикум. М.: Флинта, 2009. 496 с.
7. Баранов А.Н. Лингвистическая экспертиза текста: Теоретические основания и практика. М.: Флинта: Наука, 2007. 592 с.
8. Болотнова Н.С. Филологический анализ текста: Учебное пособие. М.: Флинта: Наука, 2009. 520 с.
9. Виноградов В.В. О языке художественной литературы. М.: Гослитиздат, 1959. 655 с.
10. Воронцов К.В. Математические методы обучения по прецедентам. М.: ВМК МГУ и МФТИ, 2011. С. 23–54.
11. Гвишиани Н.Б. Язык научного общения. Вопросы методологии. М.: ЛКИ, 2008. 280 с.



12. Гельгардт Р.Р. Исследование стиля художественной речи. М., 2011. 203 с.
13. Голованова Д.А., Кудинова Р.И., Михайлова Е.В. Русский язык и культура речи. Краткий курс. М.: Окей-книга, 2008. 144 с.
14. Голуб И.Б. Стилистика русского языка. М.: Рольф, 2001. 448 с.
15. Денисенко В.Н., Чеботарёва Е.Ю. Современные психолингвистические методы анализа речевой коммуникации. М.: РУДН, 2008. 258 с.
16. Ермаков А.Е., Киселев С.П. Лингвистическая модель для компьютерного анализа тональности публикаций СМИ. языке // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной международной конференции «Диалог». М.: Изд-во РГГУ, 2011. С. 282–285.
17. Земская Е.А. Русская разговорная речь: Лингвистический анализ, проблемы обучения. М.: Наука: Флинта, 2006. 240 с.
18. Клековкина М.В., Котельников Е.В., Метод автоматической классификации текстов по тональности, основанный на словаре эмоциональной лексики // RCDL-2012, Переславль-Залесский: Изв-во ИПС РАН, 2012. С. 135–141.
19. Кожин А.Н. Функциональные типы русской речи. М.: Высшая школа, 1982. 223 с.
20. Кожина М.Н. Стилистика русского языка. М.: Просвещение, 1987. 464 с.
21. Колмогорова А.В., Калинин А.А., Маликова А.В. Лингвистические принципы и методы компьютерной лингвистики для решения задач сентимент-анализа русскоязычных текстов // Актуальные проблемы филологии и педагогической лингвистики. Владикавказ: Изд-во СОГУ, 2018. 1(29) С. 139–148.

22. Лукашевич Н.В., Четверкин И.И. Извлечение и использование оценочных слов в задаче классификации отзывов на три класса, Вычислительные методы и программирование // Вычислительные методы и программирование. М.: НИВЦ МГУ, 2011. С. 73-81.
23. Меньшиков И.Л., Анализ тональности текста на русском языке при помощи графовых моделей. // Конференция УРФУ. Екатеринбург: Изд-во УРФУ, 2012. С. 145–151.
24. Мерков А.Б. Распознавание образов. Введение в методы статистического обучения. М.: Едиториал УРСС, 2011. 260 с.
25. Обучаем компьютер чувствам (sentiment analysis) [Электронный ресурс] // Хабрхабр <https://habrahabr.ru/post/149605/> (дата обращения: 25.05.17).
26. Пазельская А.Г., Соловьев А.Н. Метод определения эмоций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной международной конференции «Диалог». М.: Изд-во РГГУ, 2011. С. 574–586.
27. Розенталь Д.Э. Практическая стилистика русского языка. М.: Айрис, 1997. 381 с.
28. Рузавин Г.И. Методология научного познания: Учеб. пособие для вузов. М.: ЮНИТИ-ДАНА, 2012. 287 с.
29. Сентимент анализ текста [Электронный ресурс] // Хабрхабр <https://habrahabr.ru/company/palitrumlabor/blog/262595/> (дата обращения: 25.05.17).
30. Усталов Д.И. Извлечение терминов из русскоязычных текстов при помощи графовых моделей. Екатеринбург: УРФУ, 2012. 12 с.
31. Хроленко А.Т., Денисов А.В. Современные информационные технологии гуманитария: Практ. руководство. М.: Флинта: Наука, 2007. 128с.

32. Четвёркин И.Ч. Автоматизированное формирование базы знаний для задачи анализа мнений: автореф. дис. ... канд. физ-мат. наук: 05.12.13 Москва 2013, 147 с.
33. Analyze and predict sentiment with machine learning [Электронный ресурс] // MathWorks <https://goo.gl/9f6Cko> (дата обращения: 17.10.17).
34. Anstey M.L, Rogers S.M., Ott S.R., Burrows M., Simpson S.J. Serotonin mediates behavioral gregarization underlying swarm formation in desert locusts // Science Journal, 2009. P. 27–30.
35. Aue A., Gamon M. Customizing sentiment classifiers to new domains: A case study // Proceedings of recent advances in natural language processing. Varna: RANLP, 2005. P. 198–205.
36. Aue A., Gamon M. Customizing sentiment classifiers to new domains: a case study // Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP). 2005. P. 234–240.
37. Banea C., Mihalcea R., Wiebe J. Multilingual subjectivity: are more languages better? // C.-R. Huang, D. Jurafsky (Eds.) COLING 2010. 23 rd International Conference on Computational Linguistics. Proceedings of the Conference. 2010. Vol.2. P. 28–36.
38. Big Data And Power Of Sentiment [Электронный ресурс] // Forbes <https://goo.gl/tWSiqL> (дата обращения: 18.10.17).
39. Boiy E., Marie-Francine M. A machine learning approach to sentiment analysis in multilingual web texts // Inf Retrieval. 2009. № 12(5). P. 526–558.
40. Bollen J., Mao H., Zeng X. Twitter mood predicts the stock market // Journal of Computational Science, 2011. 1 (2), P. 1–8
41. Ceron L., Curini S., Iacus S., Porro G. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens political preferences with an application to Italy and France // New Media Soc. 2014. №16 (2). P. 340–358.

42. Chatzakou D., Vakali A., Kafesios K. Detecting variation of emotions in online activities // *Expert Systems with Applications: An International Journal*. 2017. 89. P. 318–332.
43. Dennis R.L., Chen Z.Q., Cheng H.W. Serotonergic mediation of aggression in high and low aggressive chicken strains // *Poult Sci*, 2008. P. 12–20.
44. Hall M. A. Correlation-based feature selection for machine learning. Waikato: The University of Waikato, 1999. 179 p.
45. Introduction to Sentiment Analysis [Электронный ресурс] // *Algorithmia* <https://goo.gl/5EEvvu> (дата обращения: 15.10.17).
46. Jakob N., Gurevych I. Extracting opinion targets in a single-and cross-domain setting with conditional random fields // *Proceedings of Conference on Empirical Methods in Natural Language Processing*, 2010. P. 1035–1045.
47. Lövheim H. A new three-dimensional model for emotions and monoamine neurotransmitters // *Medical Hypotheses*. 2012. 78. P. 341–348.
48. Lu Y., Castellanos M., Dayal U. Automatic construction of a context-aware sentiment lexicon: an optimization approach // *Proceedings of the 20th international conference on World Wide Web, ACM*. 2011. P. 347–356.
49. Maas A., Daly R., Pham P. Learning word vectors for sentiment analysis // *Proceedings of the 49th annual meeting of the association for computational Linguistics*. Portland: HLT, 2011. P. 142–150.
50. Machine Learning For Begginers [Электронный ресурс] // *Kaggle* <https://goo.gl/YzAZqe> (дата обращения: 14.10.17).
51. Machine Learning For Software Engineers [Электронный ресурс] // *GitHub* <https://goo.gl/7YvKtc> (дата обращения: 14.10.17).
52. Pang B., Lee L. Thumbs up? Sentiment Classification using Machine Learning Techniques // *EMNLP*, 2002. P. 79–86.
53. Pang B., Lee L. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts // *ACL*, 2004. P. 271–278.

54. Pang B., Lee L. Opinion Mining and Sentiment Analysis // RETR, 2008. P. 1–135.
55. Rules of machine learning [Электронный ресурс] // martin.zinkviech.org <https://goo.gl/bPmm3B> (дата обращения: 16.10.17).
56. Sentiment Analysis: Concept, Analysis and Applications [Электронный ресурс] // Towards Data Science <https://goo.gl/8Xp6Dx> (дата обращения: 19.10.17).
57. Snyder B., Barzilay R. Multiple Aspect Ranking using the Good Grief Algorithm // New Orleans: HLT-NAACL, 2007. P. 300–307.
58. Twitter Sentiment Analysis using Python [Электронный ресурс] // GeeksForGeeks <https://goo.gl/CB4oJ7> (дата обращения: 19.10.17).
59. Wiebe J., Riloff E. Creating subjective and objective sentence classifiers from unannotated texts // Computational Linguistics and Intelligent Text Processing. Berlin: Springer, 2005. 486 p.
60. Witten I.H., Frank E. Data Mining: Practical Machine Learning Tools and Techniques (Second Edition) // Burlington: Morgan Kaufmann, 2005. P. 56–63.

Иллюстрации

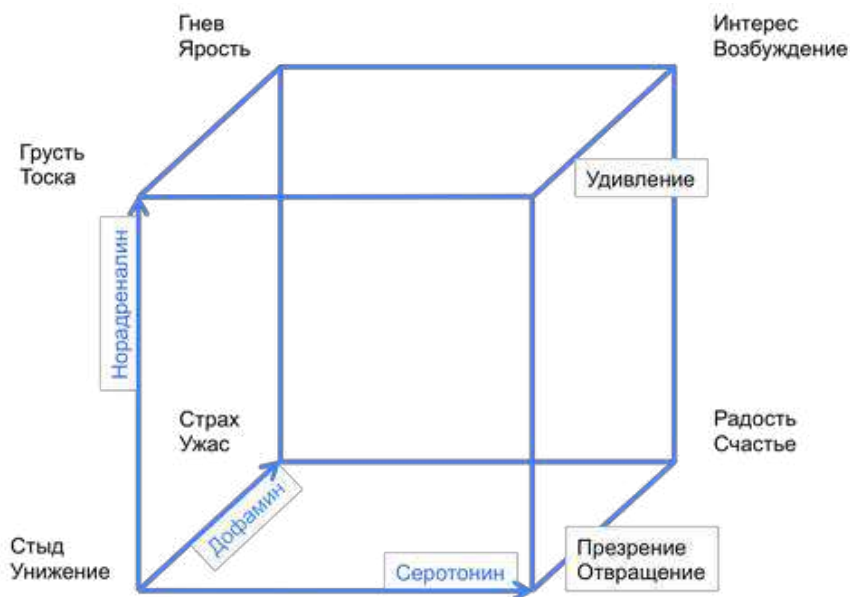


Рисунок 1. Куб Левхейма

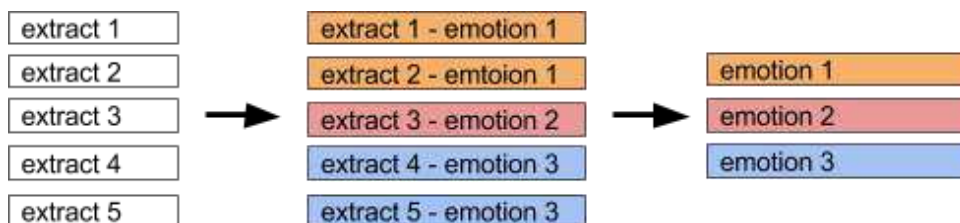


Рисунок 2. Схема формирования музыкальных списков



Рисунок 3. Схема получения песни для подходящей эмоции

Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
Институт филологии и языковой коммуникации  
Кафедра теории германских языков и межкультурной коммуникации

УТВЕРЖДАЮ

Заведующий кафедрой ТГЯиМКК

 /О.В. Магировская/

« 20 » июня 2018 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**ИСПОЛЬЗОВАНИЕ СЕНТИМЕНТ-АНАЛИЗА ДЛЯ  
СИНХРОНИЗАЦИИ САУНДТРЕКА И ЭМОЦИОНАЛЬНОЙ  
ТОНАЛЬНОСТИ ТЕКСТА**

45.04.02 Лингвистика

45.04.02.01 Межкультурная коммуникация и перевод

Магистрант



М.К. Кузлин

Научный руководитель



д-р. филол. наук,  
проф. каф. РЯиПЛ  
А.В. Колмогорова

Нормоконтролер



Э.А. Тарасенко

Красноярск 2018