

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Удмуртский государственный университет»
Институт математики, информационных технологий и физики
Кафедра теоретических основ информатики

М.М. Аббаси, А.П. Бельтюков

ЛИНГВИСТИЧЕСКИЙ ИНТЕРФЕЙС
С ЭМОЦИОНАЛЬНЫМ АНАЛИЗОМ ТЕКСТА
В КОМПЬЮТЕРНЫХ СИСТЕМАХ

Монография



Ижевск
2024

УДК 004.2
ББК 32.971.3
А137

Рекомендовано к изданию редакционно-издательским советом УдГУ

Рецензенты: д-р техн. наук, профессор, зав. каф. автоматизированных систем управления Уфимского университета науки и технологий **В.В. Антонов**,
канд. тех. наук, доцент, зав. каф. автоматизированных систем обработки информации и управления Ижевского государственного технического университета имени М.Т. Калашникова **М.Н. Мокроусов**.

Аббаси М.М., Бельтюков А.П.

А137 Лингвистический интерфейс с эмоциональным анализом текста в компьютерных системах : монография. – Ижевск : Удмуртский университет, 2024. – 114 с.

ISBN 978-5-4312-1192-8

DOI: 10.35634/978-5-4312-1192-8-2024-1-114

Монография предназначена для студентов, обучающихся по направлению «Прикладная информатика», и студентов других смежных направлений, изучающих принципы управления проектами и разработку нового компьютерного алгоритма. Монография также может быть использована молодыми исследователями и учёными, которые занимаются разработкой моделей для искусственного интеллекта и обработкой естественного языка. Её основная цель – расширить исследования в области автоматической обработки текстов, написанных на русском языке.

УДК 004.2
ББК 32.971.3

ISBN 978-5-4312-1192-8
DOI 10.35634/978-5-4312-1192-8-2024-1-114

© Аббаси М.М., Бельтюков А.П., 2024
© ФГБОУ ВО «Удмуртский
государственный университет», 2024

ВВЕДЕНИЕ

Актуальность темы исследования. Человеко-машинное взаимодействие приобрело популярность за последнее десятилетие и находит применение в различных областях техники. Для обеспечения простоты взаимодействия и облегчения данного процесса, необходим интерфейс между человеком и машинной. Интерфейс принимает информацию от человека, пересылает её машине и аналогичным образом передаёт ответ машины человеку. Информация для взаимодействия человека и машины представлена в различных формах, таких как текст, аудио и видеосигналы, биомедицинские сигналы и т. д.

В данной работе основное внимание уделяется взаимодействию человека и машины с помощью текста. Для такого взаимодействия требуется лингвистический интерфейс, который может облегчить языковое общение между человеком и машиной. Текст имеет различные характеристики: синтаксические, семантические, структурные и эмоциональные. Машина, получив текст через лингвистический интерфейс, анализирует его эмоциональную ориентацию и отвечает оппоненту, основываясь на этой характеристике. В качестве основного источника текста для анализа его эмоциональной составляющей выбран Интернет.

Основным источником информации в Интернете являются приложения для социальных сетей, где люди делятся своими повседневными событиями, фотографиями, эмоциями и т. д.

Для анализа данных в виде текста из Интернета были предложены различные методологии. Создан ряд систем для выявления и анализа эмоций из текста. Каждая система имеет определённый механизм, который анализирует конкретный аспект эмоций и, как правило, зависит от языка или платформы, на которой эта система расположена.

Сегодня основные системы лингвистического интерфейса, доступные онлайн, созданы для коммерческих целей, для анализа текста и определения мнений клиентов и отзывов о продукте или услуге. Такие системы широко используются в интернет-магазинах, таких

как: AliExpress, Amazon, Skyscanner, Booking и т. д. Самый лёгкий способ оставить отзыв по какому-либо поводу – просто ответить на заданные вопросы. Такая система может представлять мнение людей в широком масштабе, но для более подробного анализа этого недостаточно. Компаниям интересно выяснить как положительные, так и отрицательные характеристики своих продуктов и услуг.

Таким образом, выявилась потребность в создании гибкой платформы, которая может изучать эмоции, извлечённые из текста, и повышать точность результатов. Данная исследовательская работа сосредоточена на разработке системы, которая в состоянии подробно анализировать характеристики эмоций, извлеченных из текста.

Кроме этого, система имеет механизм, улучшающий взаимодействие человека и машины путём определения психолингвистических характеристик текста, которые представляют поведение человека.

Для проведения анализа и экспериментов выбраны эмоции, которые были выделены известными философами и исследователями в их работах, таких как Теория эмоций Плутчика и Теория эмоций Паррота. Работы К. Изарда и А.Н. Леонтьева, П. Зарубина также изучались и использовались в экспериментах с русским текстом.

Лингвистический интерфейс, созданный в ходе исследования, способен удовлетворить будущие потребности искусственного интеллекта в создании полностью независимой системы, способной общаться с людьми через текст. Предлагаемый интерфейс ориентирован на текст, написанный на русском языке. Выбор русского языка сделан с учётом исследований и достижений российских учёных в области искусственного интеллекта с обработкой естественного языка. Основным продуктом исследования является алгоритм Emotstemmer, первый в своем роде, который может обрабатывать и определять эмоциональную ориентацию текста с помощью небольшого словаря за очень короткое время по сравнению с системами, которые уже используются для русскоязычного текста. Также разработанная система использовалась на экспериментальной основе для изучения

психолингвистических характеристик текста наряду с его эмоциональной ориентацией для подготовки ответа на текст или ответа человеку при работе человеко-машинного интерфейса. Такие алгоритмы, как машина опорных векторов и алгоритм кластеризации, использовались для изучения эмоциональных свойств текста.

Взаимосвязь текста с его темой и зависимости, существующие между различными эмоциями и их влияние на общую полярность текста, являются частью этого исследования. Подробно исследование представлено в следующих главах.

ГЛАВА 1. ЗНАЧЕНИЕ И ВАЖНОСТЬ ЭМОЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК ТЕКСТА

Эмоции – это физические и психические состояния, вызванные нейрофизиологическими изменениями, по-разному связанными с мыслями, чувствами, поведенческими реакциями и степенью удовольствия или неудовольствия.

За последние два десятилетия исследования эмоций расширились, в них приняли участие многие области, включая психологию, медицину, историю, социологию эмоций и информатику. Текст стал основным средством передачи эмоций. Эмоции выражаются в текстах с помощью слов, предложений и абзацев, которые передают чувства и установки. Письменные тексты отражают эмоциональное состояние людей, которые их создают. Эмоции можно обнаружить и измерить с помощью статистического анализа и инструментов анализа настроений.

Эмоциональные характеристики текста – это те слова или предложения, которые отражают эмоции автора текста и его эмоциональное состояние во время написания текста. Эти характеристики также описывают эмоции, мышление или чувства человека, связанные с определенным событием, по отношению к людям, вещам и его идеям или взглядам на жизнь, политику и т. д. Существуют различные методы, используемые для анализа эмоциональных характеристик текста. Наиболее известной из этих моделей является модель валентно-энергетического пространства.

1.1. Валентно-энергетическое пространство в анализе эмоций, извлечённых из текста

Валентно-энергетическое пространство часто используется для анализа эмоций из текста, фокусируясь на уровнях позитивности/негативности и энергии написания. Тенденции в эмоциональных характеристиках текстов можно определить, проанализировав большие массивы исторических текстов, а также современных сообщений.

Различия в эмоциональных характеристиках также могут наблюдаться у разных групп людей, таких как пол, возраст и профессия. Математическое моделирование может быть использовано для интерпретации результатов эмоционального анализа текстов.

1.2. Важность анализа эмоций, извлечённых из текста

Анализ эмоций, извлечённых из текста, жизненно важен в различных аспектах бизнеса и личного выбора для принятия решений. Например:

1.2.1. В бизнес-аналитике

В деловом мире понимание настроений клиентов имеет решающее значение. Анализируя отзывы клиентов, обзоры и упоминания в социальных сетях, компании могут получить ценную информацию об удовлетворённости клиентов и определить области для изменения в лучшую сторону. Эти данные могут использоваться при разработке продукта, маркетинговых стратегиях и улучшении обслуживания клиентов.

1.2.2. В управлении репутацией бренда

Положительная репутация в Интернете бесценна. Анализ эмоций позволяет брендам отслеживать присутствие в Интернете и упоминания о них, и оперативно реагировать на негативные настроения. Поступая таким образом, компании могут защитить имидж своего бренда и сохранить доверие клиентов.

1.2.3. В маркетинговых исследованиях

Анализ эмоций позволяет компаниям оценить общественное мнение о своих продуктах или услугах и об услугах своих конкурентов. Эти данные помогают принимать обоснованные решения, запускать новые продукты и опережать тенденции рынка.

1.2.4. В политическом анализе

На политической арене анализ настроений используется для оценки общественного мнения, особенно во время избирательных кампаний. Это помогает политикам понять проблемы своих избирателей и соответствующим образом адаптировать свои сообщения.

1.2.5. В анализе отзывов клиентов

Анализируя отзывы клиентов, компании могут выявлять повторяющиеся проблемы и оперативно их устранять. Это повышает удовлетворённость и лояльность клиентов.

1.2.6. В прогнозах фондового рынка

Анализ эмоций все чаще используется в финансовом секторе для прогнозирования тенденций фондового рынка. Он анализирует новостные статьи, сообщения в социальных сетях и финансовые отчёты для оценки рынка, помогая инвесторам принимать более обоснованные решения.

1.2.7. В антикризисном управлении

Во время кризиса понимание общественных эмоций имеет решающее значение для эффективного антикризисного управления. Аналитика эмоций может помочь организациям оценить серьёзность общественных эмоций и отреагировать соответствующим образом.

1.2.8. В персонализации контента

Аналитика эмоций используется контент-провайдерами для персонализации рекомендаций по контенту для пользователей. Понимая эмоции и настрои пользователей, платформы могут предлагать контент, соответствующий их предпочтениям.

1.3. Что такое анализ эмоций в искусственном интеллекте?

Анализ эмоций – это процесс идентификации и извлечения человеческих эмоций из огромного количества текстовых, визуальных или слуховых данных. Это что-то вроде анализа настроений, в том смысле, что они оба используют схожие источники данных. Однако анализ эмоций даёт гораздо более целостное представление о чувствах и эмоциях клиентов по поводу продукта или услуги. В настоящее время технологические гиганты, такие как IBM, интегрируют искусственный интеллект и машинное обучение для измерения настроений и эмоций клиентов. Например, IBM Watson использует модели глубокого обучения для определения таких эмоций, как гнев, отвращение, страх, радость или печаль, на основе неструктурированных текстовых данных своих клиентов.

Методы и техника искусственного интеллекта помогают машинам понимать, воспроизводить человеческие эмоции и реагировать на них. Он сочетает в себе информатику и психологию для облегчения эмпатического взаимодействия между людьми и компьютерами и также известен как аффективные вычисления.

1.4. Схема системы

Системная схема используется для представления фактического физического состояния системы и является первым уровнем ее абстракции. Системная схема лингвистического человеко-машинного интерфейса начинается с определения всех фаз, которые влияют на работу системы. Каждая фаза лингвистического человеко-машинного интерфейса представлена в форме прямоугольника; кружок на схеме представляет элементы ввода и вывода данной системы. Скруглённые прямоугольники в схеме используются для представления функций, алгоритмов, которые применяются на каждом конкретном этапе системы. Она включает в себя компоненты, которые отфильтровываются из системы во время определённой фазы. Стрелка в схеме представляет направление потока управления процессами лингвистического человеко-машинного интерфейса.

Схема ниже представляет полный вид лингвистического человеко-машинного интерфейса. Она демонстрирует процессы, которые выполняются для выявления, анализа и обобщения эмоциональных компонентов письменного текста. Схема системы описана на ниже (рисунок 1).

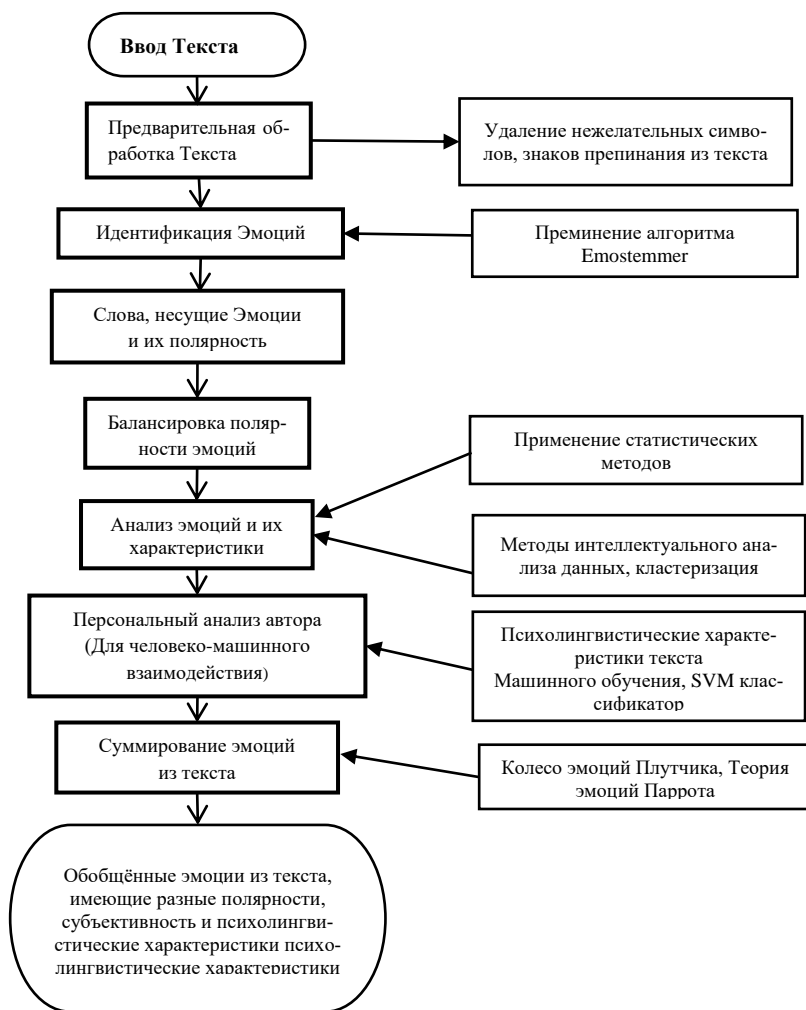


Рисунок 1 – Системная диаграмма для лингвистического человеко-машинного интерфейса

Схема представляет различные фазы лингвистического человеко-машинного интерфейса. На каждом этапе интерфейса используются определённые методы и алгоритмы. Детали этих методов и алгоритмов обсуждаются в следующих главах диссертации.

Лингвистический человеко-машинный интерфейс начинает работать с текстом в качестве входных данных. Обычно текст представляет собой блог из дискуссионной группы в Интернете или сайта социальной сети. Текст из интернета в основном содержит ненужные символы, пробелы, знаки препинания или орфографические ошибки, что усложняет анализ текста. Чтобы избежать этих осложнений, предварительная обработка выполняется для текста с использованием алгоритма, созданного во время повторного поиска. Алгоритм идентифицирует ненужные компоненты текста и исправляет орфографические ошибки, чтобы восстановить текст в его фактической форме.

На следующем этапе эмоциональные характеристики текста отслеживаются путём выявления эмоций, представленных в нем. Чтобы идентифицировать эмоции из текста, был создан алгоритм с именем Emostemmer. Emostemmer является первой в своем роде программой, которую используют для идентификации эмоций.

После этого полярность слов, представляющих эмоции, обрабатывается и уравнивается. Замечено, что отрицания, такие как «нет», когда эмоции в тексте изменяются, изменяют фактическую полярность эмоции. Полярность эмоции отражает ее природу, будь то положительная или отрицательная эмоция.

После выявления эмоций и уравнивания их полярности анализируются характеристики эмоций в тексте, которые включают в себя анализ синтаксических, семантических и логических характеристик. Для этого используются различные статистические показатели и методы интеллектуального анализа данных. Чтобы сделать систему способной эффективно взаимодействовать с человеком и машиной, используются проанализированные эмоциональные характеристики текста, а также база данных психологических слов для представления конкретного поведения человека. Объединение обеих

характеристик дает системе возможность анализировать поведение автора текста и соответственно общаться с ним. Здесь алгоритмы, такие как машина опорных векторов (SVM), оценка максимального правдоподобия (MLE), используются для классификации поведения автора в определенной категории.

На заключительном этапе лингвистического человеко-машинного интерфейса общие эмоции, выделенные из текста, в результате их взаимодействия, суммируются с использованием теорий эмоций. Обобщенная эмоциональная ориентация текста является результатом лингвистического человеко-машинного интерфейса.

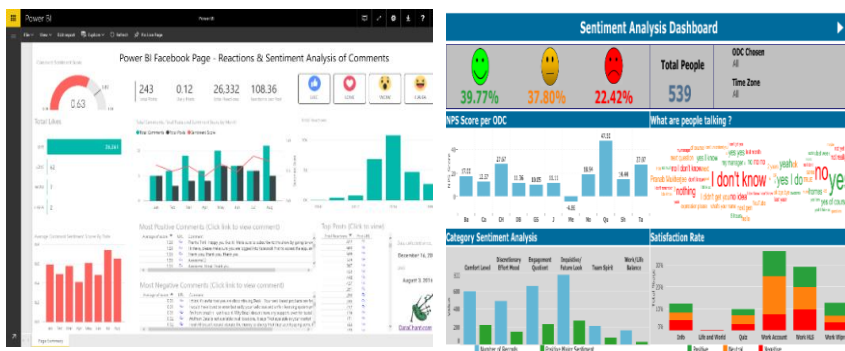
Лингвистический человеко-машинный интерфейс представляет собой полный интерфейс, который идентифицирует и анализирует те эмоциональные характеристики текста, которые не привлекли большого внимания исследователей. Перед созданием лингвистического интерфейса было проведено детальное изучение доступных систем и моделей, используемых для анализа эмоциональных характеристик текста. Эти модели обсуждаются в последующих главах.

Выводы по главе 1

В первой главе изучались эмоциональные характеристики текста и его значение в различных сферах жизни. Рассматривалась разработанная лингвистическая система для обработки текста, понимания его различных компонентов, выявления эмоций в тексте, изучения взаимосвязей между ними и последующего обобщения всех эмоций в тексте.

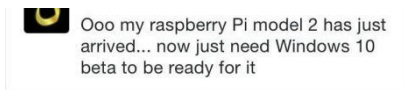
ГЛАВА 2. ИЗВЕСТНЫЕ СИСТЕМЫ, ИСПОЛЬЗУЕМЫЕ ДЛЯ АНАЛИЗА ЭМОЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК ТЕКСТА

Существуют системы, которые доступны для общего пользования и используются для анализа текста и идентификации эмоций, выделенных из текста. Основные существующие системы представлены ниже (рисунок 2).



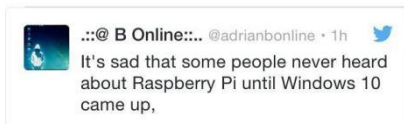
Power BI

Tableau



Mood: Anticipation

needed, impatient, anxious, anticipation, restless, unsure



Mood: Disgust

Q emotion

MoodPatrol API

Рисунок 2 – Доступные системы, используемые для анализа эмоций из текста

На рисунке 2 показаны несколько самых популярных систем, используемых для анализа текста. Они включают Power BI, Tableau, Qemotion и MoodPatrol API. Прежде чем разрабатывать новую систему или алгоритмы, необходимо изучить уже существующие алгоритмы и их преимущества и недостатки. Это поможет понять, чем новая предлагаемая система будет отличаться в лучшую сторону от существующих. Детали существующих систем и их использовании для анализа эмоциональных характеристик текста рассказывается ниже.

2.1. Программа Power BI

Система работает по принципу Facebook для анализа эмоций людей с помощью смайлика. В этом приложении одни люди добавляли интересующие их темы в виде вопроса, а другие отображали там ответ в виде улыбки. Система подсчитывает различные типы эмоций по каждой теме и суммирует результаты эмоций с помощью графиков. Это также даёт людям возможность добавить дополнительный комментарий, если кто-то хочет ещё больше выразить свои эмоции. Результаты анализа текста продолжают меняться, поскольку люди все больше и больше откликаются на эту тему.

Плюсы системы

Анализ очень прост и понятен. Результаты обновляются очень быстро. Система удобна для пользователя.

Минусы системы

Основной механизм выражения эмоций очень ограничен. Во-первых, количество и тип смайлов ограничены. Во-вторых, улыбка может представлять только абстрактную эмоцию, и одновременного выбора только одной недостаточно, чтобы выразить разные эмоции, которые могут возникать одновременно.

2.2. Программа Tableau

Приложение работает как в онлайн, так и в офлайн-режиме. Его можно загрузить на персональный компьютер, и при подключении к Интернету, приложение станет активным. Это приложение, которое в основном используется блогерами, чтобы писать об актуальной теме или событии, и люди могут высказывать свои эмоции по этому поводу. Основные темы или блоги в этом приложении посвящены политике, спорту и журналистике.

Приложение во время анализа выделяет ключевые слова из блога и использует их для анализа положительного или отрицательного мнения людей о блоге. Оно анализирует уровень удовлетворенности людей темой и их готовность выражать свои эмоции.

Плюсы системы

Анализ, который проводит система, более детален и анализирует эмоции до подуровня. Механизм выражения результатов анализа эмоций в графической и текстовой формах более интересен.

Минусы системы

Анализ настроений, выраженных в тексте – лишь малая часть системы. Система в основном предназначена для обслуживания потребителей. В данной системе отсутствуют механизм изучения взаимосвязи между разными эмоциями, анализ причин или мотивов выражения определённых эмоций. Это ограничивает область ее применения.

2.3. Программа Qemotion

Для взаимодействия с этим приложением используется его API, обеспечивающий очень интересный механизм анализа эмоций людей по шкале температуры. Приложение позволяет людям выбирать температуру по определённой теме. Температура в диапазоне от -20°C до -15°C отражает эмоции отвращения к этой теме. Диапазон температур от -14°C до -9°C представляет гнев, тогда как диапазон от -8°C до -4°C представляет печаль. Температура в диапазоне от -3°C

до + 2 °С представляет страх, а от +2 °С до +7 °С – эмоцию спокойствия. Положительная эмоция удивления представлена температурой в диапазоне от +8 °С до +12 °С, а счастье – от +13 °С до +17 °С.

Плюсы системы

Система определяет наиболее распространённые эмоции, понятные публике, связывает эмоции с температурой на градуснике, создавая интересный механизм для анализа эмоций. Семантический анализ выполняется автоматически.

Минусы системы

Система использует 5 чисел для обозначения эмоции и её интенсивности. Однако в некоторых случаях эмоции нуждаются в дополнительной шкале, чтобы классифицировать интенсивность. Эмоция страха наполовину отрицательна, наполовину положительна. Это затрудняет принятие решения об эмоциях людей по отношению к теме. Количество эмоций намного больше, чем представлено в приложении.

2.4. Программа MoodPatrol API

Приложение используется для анализа текста и его общих характеристик: семантических, синтаксических и эмоциональных. Эмоциональные характеристики – небольшая часть программы. Программа анализирует эмоции из текста и делит их на восемь основных эмоций. Эти восемь эмоций включают гнев, страх, печаль, отвращение, ожидание, удивление, доверие и радость.

Плюсы системы

Система проста и удобна в использовании.

Минусы системы

Определяются эмоции в 8-ми основных категориях. Связь между эмоциями и текстом отсутствует. Данной системы нет в свободном доступе.

Для сравнения вышеуказанных систем использовалась известная модель Delone Mclean Success Model (DMSM), предложенная

в 1992 году. Первоначально было измерено шесть элементов, чтобы определить эффективность системы. Эти элементы включают качество системы, время её отклика, простоту использования, доступность, количество пользователей, достоверность информации. Позже учёные добавили в эту модель больше компонентов. Было определено, что ни одна из данных систем не выполняет требования DMSM при экспериментировании с нашими текстовыми данными.

Выводы по главе 2

В данной главе изучены наиболее распространённые и известные системы, используемые для анализа текста и эмоций, выделенных из него. Данные системы в основном используют эмоциональный анализ как часть других услуг, которые они предоставляют.

Системы обладают способностью идентифицировать и классифицировать только определённые типы эмоций. В случае большого текста с разными типами эмоций системы не дают достоверных результатов. Кроме того, существующие системы не выполняют коррекцию слов и предварительную обработку текста. В результате большое количество ненужных компонентов является частью данных, полученных в результате анализа текста, и в нем упущены многие значимые эмоциональные компоненты текста, поскольку отсутствует коррекция орфографии.

ГЛАВА 3. МЕХАНИЗМЫ ДЛЯ ПОДГОТОВКИ ТЕКСТА К ЕГО АНАЛИЗУ ДЛЯ СОЗДАНИЯ ЛИНГВИСТИЧЕСКОГО ЧЕЛОВЕКО-МАШИННОГО ИНТЕРФЕЙСА

Для извлечения из текста ненужных символов, пробелов, знаков препинания и т. д. производится процесс фильтрации или предварительной обработки текста. Затем в предварительно обработанном тексте идентифицируются отрицания, чтобы определить действительную полярность эмоций, извлечённых из текста, и дальнейший анализ производится с уже подготовленным текстом. Процесс предварительной обработки текста перед его анализом отсутствует в большинстве доступных систем, что ухудшает качество итогового результата.

На тестовом наборе эксперимент показывает повышение эффективности для данных, генерируемых после предварительной обработки, по сравнению с данными, которые генерируются без нее.

Для текста, взятого из социальных сетей или блогов, не существует надлежащего формата для его написания. Пользователи обычно пишут на неформальном языке с ошибками в орфографии, пунктуации, сленгами, URL-адресами, тегами HTML и многими другими особенностями. Здесь процесс предварительной обработки данных сложнее, чем для текста, написанного в книгах или официальных документах.

3.1. Лингвистический человеко-машинный интерфейс для предварительной обработки текста

Исследование было применено к тексту, полученному в блоге онлайн. Чтобы извлечь из него ненужные символы, пробелы, знаки препинания и т. д. был построен лингвистический интерфейс. Первый этап интерфейса, заключающийся в предварительной обработке текста, до анализа эмоций, показан ниже (рисунок 3). При разработке лингвистического человеко-машинного интерфейса используются знания, полученные при обзоре исследований.

Ниже (рисунок 3) показан интерфейс перед загрузкой в него текста. Текст можно загрузить с помощью кнопки «Обзор файла», и загруженный текст также появляется в поле «Редактирование текста», как показано ниже на картинке.

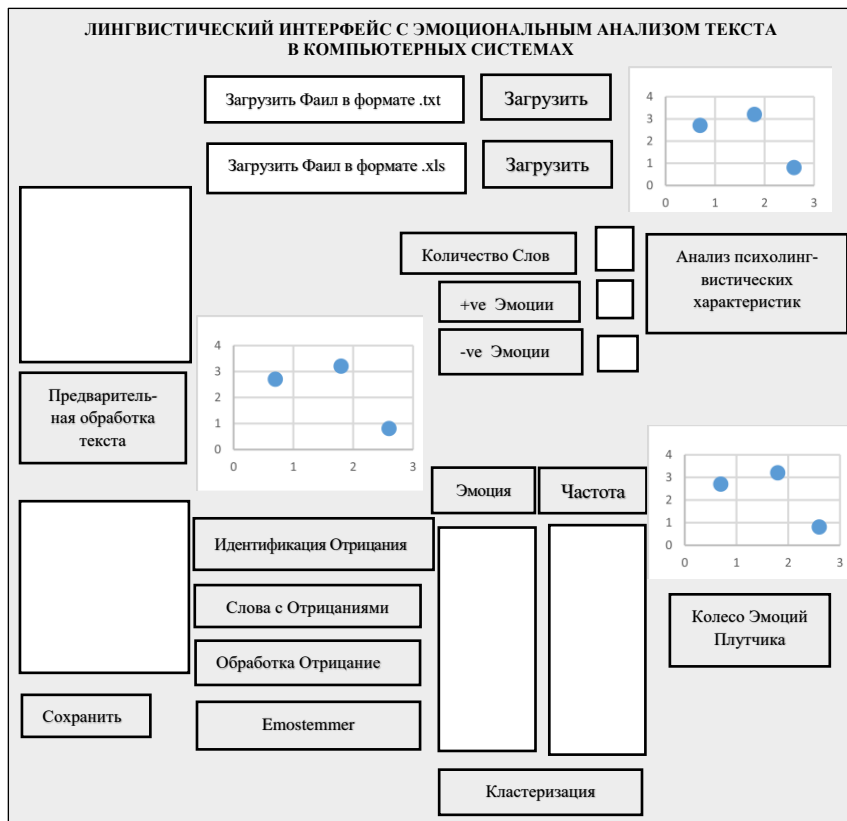


Рисунок 3 – Лингвистический человекo-машинный интерфейс

Текст может содержать ненужные символы, которые не требуются или могут нарушить процесс анализа текста. Различные методы и алгоритмы используются для очистки данных и исправления орфографических ошибок.

Эти алгоритмы при нажатии кнопки «Предварительная обработка текста» начинают предварительную обработку и отфильтровывают необязательные символы из текстового файла. Затем интерфейс создаёт новый файл, который можно сохранить и использовать для анализа текста, как показано ниже (рисунок 4).

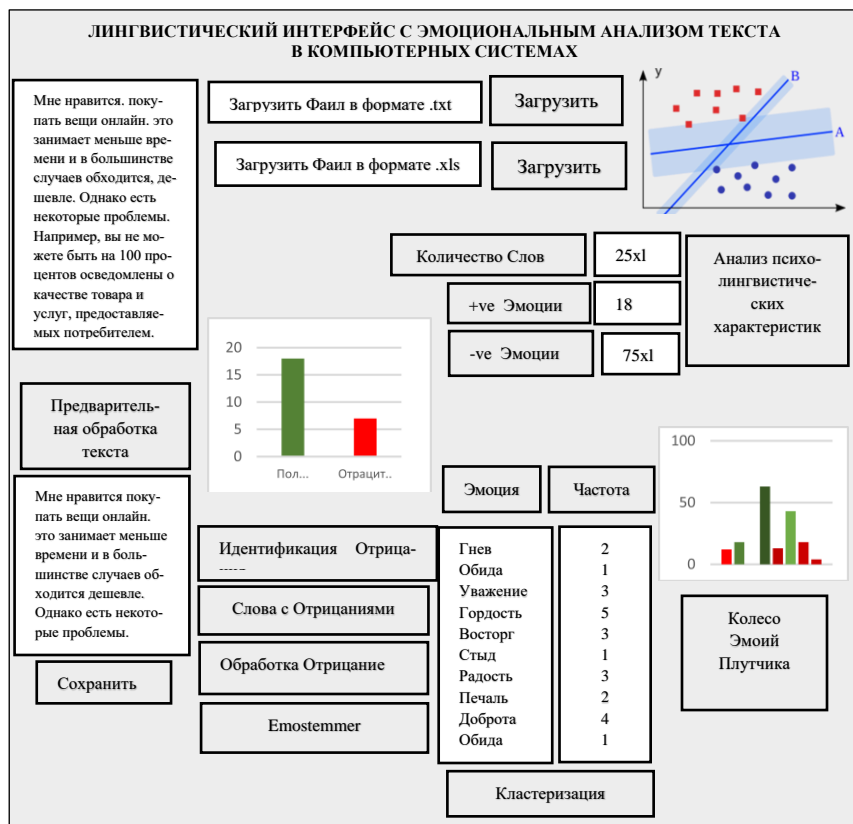
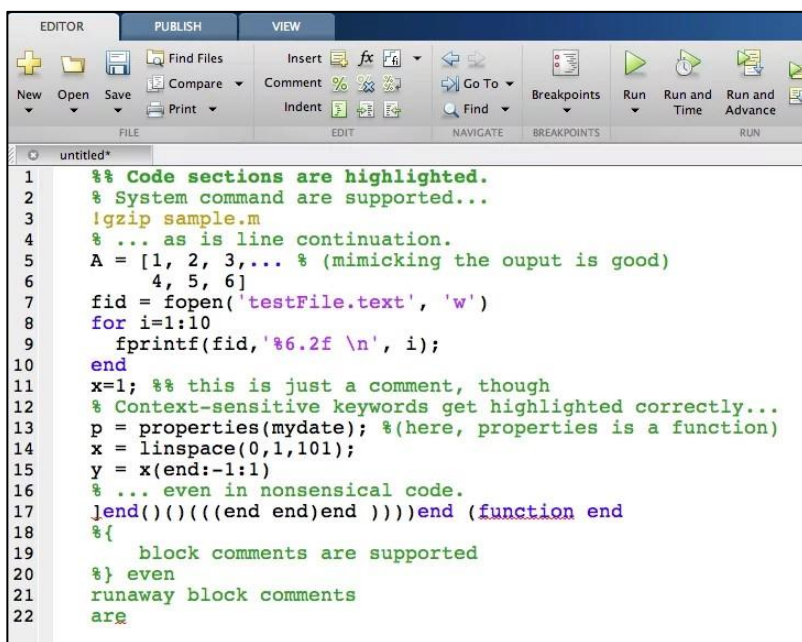


Рисунок 4 – Вид интерфейса после применения предварительной обработки текста

На рисунке 4 показан начальный вид лингвистического человеко-машинного интерфейса. Он имеет кнопки для применения

различных функций к тексту. Например, для применения предварительной обработки текста, для сохранения нового предварительно обработанного текста, для идентификации отрицаний текста, идентификации эмоций, извлечённых из текста, частоты каждой эмоции в тексте, анализа каждого предложения и его эмоциональных характеристик и т. д.

Небольшая часть программы или алгоритм для создания лингвистического человеко-машинного интерфейса показана ниже (рисунок 5).



```
1 %% Code sections are highlighted.
2 % System command are supported...
3 !gzip sample.m
4 % ... as is line continuation.
5 A = [1, 2, 3,... % (mimicking the ouput is good)
6      4, 5, 6]
7 fid = fopen('testFile.text', 'w')
8 for i=1:10
9     fprintf(fid,'%6.2f \n', i);
10 end
11 x=1; %% this is just a comment, though
12 % Context-sensitive keywords get highlighted correctly...
13 p = properties(mydate); %(here, properties is a function)
14 x = linspace(0,1,101);
15 y = x(end:-1:1)
16 % ... even in nonsensical code.
17 jend()(((end end)end )))end (function end
18 %{
19     block comments are supported
20 %} even
21 runaway block comments
22 arg
```

Рисунок 5 – Программный компонент лингвистического человеко-машинного интерфейса для предварительной обработки текста

На рисунке выше (рисунок 5) показан код, написанный с использованием программного обеспечения Matlab. Код включает в себя функции и переменные, которые используются для предварительной

обработки текста, для сохранения нового файла текста и применения различных механизмов.

3.2. Влияние алгоритма предварительной обработки на текст

Для понимания влияния алгоритма на текст был проведён эксперимент. Блоги для эксперимента были взяты из интернета по следующим темам: «Стоит ли покупать на Aliexpress?» (<https://chinas-lafa.ru/aliexpress-spravochnik/stoit-li-pokupat-na-aliekspress/>) и «Хороший ли магазин Aliexpress?» (<https://otvet.mail.ru/question/191447142>). «Удмуртский государственный университет (УдГУ), Ижевск – отзывы» <https://studika.ru/izhevsk/udgu/otzyvi>. Результаты приведены ниже (таблица 1).

Таблица 1

Влияние алгоритма предварительной обработки на различные компоненты текста

Текст и его компоненты	Количество компонентов до применения алгоритма	Количество компонентов после применения алгоритма	Изменение компонентов (%)
Количество символов в тексте	55	0	100,00 %
Количество слов в тексте	5625	5761	2,64 %
Количество предложений	804	804	0,00 %
Количество знаков пунктуации	55	0	100,00 %
Количество пробелов	1110	0	100,00 %

Из приведённой выше таблицы видно, что после применения предварительной обработки в тексте происходят значительные изменения. Это подразумевает, что анализ текста без его предварительной обработки может привести к потере некоторых наиболее важных его компонентов.

Алгоритмы работают как механизм, который идентифицирует и удаляет ненужные компоненты из текста с последующей коррекцией орфографии. При исправлении орфографии слова из текста сопоставляются со словарём RuWordNet. В случае, если он идентифицирует слово, написание которого не соответствует базе данных словаря русского языка, тогда используется наивный подход, чтобы сегментировать эту строку слов на подстроки, а затем сравнивать их с языковым пакетом словаря.

Это означает, что каждая строка длины n сегментируется на $2^n - 1$ подстроки и сравнивается для определения фактических слов из строки слов. Часть кода, используемого для сегментации строки слов, показана ниже.

```
public void CompositionGeneration(string input, string composition="")
{
    for (int i=1;i<=input.Length;i++)
    {
        string part1 = input.Substring(0, i);
        //recursion with the remainder of the string
        if (part1.Length < input.Length)
            CompositionGeneration(input.Substring(i),composition+part1+"");
        //display composition
        else Console.WriteLine(composition+part1);
    }
}
CompositionGeneration("Wordstring");
```

С помощью этого механизма исправляется орфография и удаляются ненужные компоненты из текста.

Для понимания взаимосвязи между группами текста с предварительной обработкой и без неё, к ним был применён Т-тест.

Т-тест – это статистический тест, в котором различные математические показатели, такие как среднее значение, стандартное отклонение, Р-значение, корреляция и т. д., измеряются в двух группах текста.

3.3. Применение Т-теста к тексту с предварительной обработкой и без нее

Основываясь на изменении различных компонентов текста до и после применения алгоритма предварительной обработки, можно сделать вывод, что без предварительной обработки текста результаты анализа не могут быть точными. Для анализа взаимосвязи между двумя группами текста, мы применили к ним Т-тест онлайн. <https://www.socscistatistics.com/tests/studentttest/default.aspx>. Результаты показаны ниже (таблица 2).

Таблица 2

Результаты Т-теста в двух группах

Статистические показатели	Группа 1	Группа 2
Mean	1529,10	1313,00
Variance	5455937,43	603923,11
Observations	5,00	5,00
Pearson Correlation	1,00	
Hypothesized Mean Difference	0,00	
Df	9,00	
t-Stat	-1,15	
P(T<=t) one-tail	0,01	
t Critical one-tail	1,83	
P(T<=t) two-tail	0,03	
t Critical two-tail	2,26	

Целью Т-теста является выявление общих отношений между двумя группами. Его расчёт требует трёх ключевых значений данных. Он включает в себя средние значения каждой группы данных, дисперсию каждой группы и количество значений данных в каждой группе.

Наряду со значениями P , которые вычисляют разницу между двумя группами, T -тест вычисляет несколько других параметров двух групп, таких как корреляция Пирсона, гипотетическая средняя разница, степень свободы (Df), значения t Stat и t Critical. Значение корреляции Пирсона лежит между -1 и 1 , что указывает на диапазон связи двух групп. Значение корреляции Пирсона -1 означает, что две группы связаны отрицательно линейно. Корреляция Пирсона 0 означает, что переменные не имеют никакого отношения друг к другу. Необходимо принять во внимание, что корреляция Пирсона 1 представляет, что две переменные связаны положительно линейно.

Предполагаемая средняя разница – это значение гипотезы, которое можно установить для эксперимента и применения T -теста для группы наблюдений. По умолчанию значение гипотезы равно 0 , это означает, что между группами нет никакой разницы.

Df – это степень свободы, представляющая собой объем информации, которую можно использовать для анализа изменчивости разницы в группах. Формула для расчёта Df – это $n-1$, где n – количество наблюдений. Значение t Stat – это отношение отличия расчётного значения параметра от его предполагаемого значения к его стандартной ошибке.

T -тест применяется для сравнения различных типов наборов данных, таких как текстовые, аудио и видеосигналы. Для того, чтобы понять линейные отношения между ними и составить эскиз их результатов, T -тест вычисляет значение P для одного тэйлд (Tailed) (распределение данных в одном направлении, от 0 до $+$ или от 0 до $-$), а также для двух тэйлд (Tailed) (распределение данных в обоих направлениях от 0 до $+$ и от 0 до $-$).

P -значение имеет пороговое значение $0,05$, которое представляет существующую разницу между двумя группами. Из приведенной выше таблицы значение P равно $0,03$, это означает, что после предварительной обработки в тексте существует большая разница, и анализ эмоций в обоих текстах может привести к большой разнице в их результатах.

Выводы по главе 3

Предварительная обработка текста является очень важным этапом во всех приложениях для интеллектуального анализа данных и, в частности, при анализе эмоций. Тем не менее, немногие работы были специально посвящены пониманию роли каждого из основных методов предварительной обработки, которые часто применяются к текстовым данным.

Фаза предварительной обработки зависит от характера анализируемого текста. Если текст взят из социальных сетей, он часто не структурирован; он содержит орфографические ошибки и ошибочные символы. В отличие от этого, в случае букв, книг, официальной информации текст правильно структурирован и не требует глубокой предварительной обработки.

Использование проверки орфографии может помочь в эффективной предварительной обработке данных. Важную информацию для токенизации данных можно извлечь после удаления некоторых символов, таких как скобки, знаки препинания. Это включает в себя разделение предложений на слова. Частоту слов можно использовать для категоризации слов по их значению.

Было отмечено, что работа по предварительной обработке не очень велика. Это зависит от выбора данных или текста исследователя и цели исследования. В большинстве случаев основной целью исследования было выявление и классификация чувств и проведение экспериментов, выбор определённого типа текста.

ГЛАВА 4. СОЗДАНИЕ УНИКАЛЬНОГО АЛГОРИТМА И ПРОГРАММЫ ДЛЯ ИДЕНТИФИКАЦИИ ЭМОЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК РУССКОЯЗЫЧНОГО ТЕКСТА

Определение эмоциональных характеристик текста с помощью искусственного интеллекта – это тема, требующая глубокого понимания алгоритмов машинного обучения, а также знания синтаксических и семантических особенностей языка, на котором написан текст. Созданная программа Emostemmer для идентификации эмоций, извлечённых из русскоязычного текста – это революционный алгоритм, который предлагается в данной работе с целью идентификации эмоций, извлечённых из текста, с использованием минимальных ресурсов, таких как время, память и дающий результаты с высокой точностью.

Перспектива создания и использования стеммера для идентификации эмоций из текста не привлекла большого внимания русскоязычных исследователей. Вместо этого они используют словарные методы, которые содержат все слова, выражающие эмоции, и их возможные вариации с использованием RuWordNet, RuWordNet – онлайн-словарь, содержащий 111 500 слов и фраз и используемый для анализа текста на русском языке. Из них более 12000 слов представляют эмоции в тексте на русском языке. Использование большого ресурса для идентификации эмоций путём сравнения каждого слова в тексте с каждым словом в словаре делает процесс идентификации эмоций более сложным и требует больше места и времени для их идентификации. Точно так же существует большая вероятность ошибок в случае, если текст является длинным и содержит некоторые изменения, которые отсутствуют в словаре.

Во многих языках романо-германской группы идентификация слов, представляющих эмоции, намного проще, чем в русском, поскольку одно слово представляет эмоцию независимо от грамматических форм и родов. В то время как для такого языка, как русский, где окончание слова, несущего эмоции, меняется в зависимости от рода, вида и др., анализ становится более сложным, как показано ниже (таблица 3).

Слова, которые представляют эмоции и их склонения

Слова, которые представляют эмоции	Однокоренные слова
Радость	Рада, Радостный, Нарадоваться, Радости, Обрадовать, Порадовать, На радостях, Обрадованный
Любовь	Любить, Любовь, Влюбляться, Влюблен, Себялюбие, Налюбоваться, Любимый, Любя, Любовный, Любитель
Сила	Сильный, Силач, Усиленно, Сильно, Осилить, Усилить, Пересилить
Удовлетворение	Удовлетворять, Удовлетворительно, Удовлетворенный, Удовлетворяемый, Удовлетворяет
Классно	Классный, Класс, Классика, Классический
Злость	Злой, Зло, Злоба, Озлобленный, Обозлиться, Зловеще
Довольно	Довольствоваться, Удовольствие, Недовольно
Грусть	Грустный, Грустить, Грустно, Взгрустнуть

Из приведённой выше таблицы видно, что слово с одинаковым значением может существовать в разных формах в тексте. В составе слова могут встречаться буквы, находящиеся до или после фактического слова, означающего эмоции. Наиболее распространенным способом является использование словарного метода, когда словарь состоит из сотен, а в некоторых случаях тысяч слов, и его необходимо поддерживать, чтобы идентифицировать все эмоции, выделяемые из текста. Это делает анализ сложным и трудоёмким. Как известно, качество алгоритма измеряется временем и объёмом памяти, которые требуются для его работы. В данном исследовании рассматриваются эти качественные показатели алгоритмов, а также точность определения эмоций, выделенных из текста с помощью программы Emostemmer.

4.1. Эмоциональные конструкты

Изначально математические модели выражения эмоций в словах применяют в эмоциональном анализе текстов только для рассмотрения слов как последовательностей букв. Не учитывается рукописная, устная речь, шрифтовое и иное графическое оформление текста.

Основная цель данного рассмотрения – выявление возможностей определения эмоционального выражения слов способами более компактными, чем составление всеобъемлющих словарей. Это актуально в силу большой изменчивости языка и словотворчества авторов сообщений. Кроме того, зачастую, тексты изобилуют опечатками и ненамеренными искажениями слов, затрудняющими использование словарей, но при этом не искажающими эмоциональное выражение этих слов.

Предполагается, что моделью комплекса эмоций, выраженных в словах, является элемент некоторого множества E_x – множества *эмоциональных выражений* и далее считаем, что зафиксировано некоторое такое множество E_x . На практике каждый его элемент отражает характер и силу различных эмоций, выраженных в слове. Кроме этого, анализируются эмоции, выраженные не только в словах, но и в более крупных единицах текста.

4.2. Функции эмоциональных выражений

Для того, чтобы промоделировать комплекс эмоций, который выражает слово (возможно, даже – зависимость этого комплекса от некоторых характеристик контекста, в котором это слово появилось), предполагаем, что имеется некоторая *функция эмоционального выражения* F_{ex} , которая отображает множество слов рассматриваемого языка L во множество эмоциональных выражений E_x :

$$F_{ex}: L \rightarrow E_x.$$

При этом лексика языка, понимается как потенциальная – расширяющаяся с развитием языка. Интерес вызывает зависимость

выражения этих эмоций от характеристик контекста. Это находит отражение в соответствующем усложнении структур объектов множества эмоциональных выражений Ex .

Алфавит языка обозначим буквой A . Так что здесь L – множество некоторых слов в алфавите A . Для того, чтобы не решать слишком сложную задачу построения таблиц функции Fex для быстро расширяющегося и меняющегося множества L , рассмотрим способ, которым можно определять, какую роль играют те или иные части слов в определении значений функции эмоционального выражения этих слов.

4.3. Эмоциональное наполнение префиксов слов

Поскольку слова во время чтения воспринимаются человеком (субъектом, получающим эмоции) в порядке следования их элементов (букв), то наиболее естественным подходом к эмоциональному анализу слов является автоматный подход, сущность которого заключается в попытке разделения эмоционального анализа по частям слова в порядке его чтения. Здесь рассматриваются линейные текстовые языки с относительно небольшим числом букв. Например, таким языком является традиционная текстовая форма выражения русского языка. Тогда как, например, традиционная текстовая форма выражения китайского языка не рассматривается.

Для начала рассматриваем всевозможные разбиения слова w языка в порядке его написания и чтения на две части:

$$w = p s,$$

где p условно называем *префиксом*, а s – *суффиксом* этого разбиения.

Рассмотрим функцию *эмоционального выражения префикса* слов $Frex$ следующим образом. Пусть потенциальный префикс p – слово в алфавите A . Тогда значение $Frex(p)$ – это отображение, которое каждому слову s в алфавите A такому, что ps – слово рассматриваемого языка (из L), ставит в соответствие значение $Fex(ps)$.

4.4. Эмоциональное наполнение инфиксов

Для того, чтобы отражать вклад каждой части i слов в эмоциональные выражения всех слов, введём специальную функцию $Fiex$. Значение $Fiex(i)$ – это отображение, которое ставит в соответствие паре слов p и s в алфавите A , таких, что pis – слово рассматриваемого языка (из L), значение $Fex(pis)$. Слово i далее обобщённо называем *инфиксом*. Инфикс может быть основой слова (стемом), корнем или какой-либо другой частью, которую имеет смысл рассматривать при анализе эмоционального содержания слов.

Эмоциональный вклад инфикса i , кроме того, при автоматическом подходе может быть выражен как функция, выдающая по потенциальному префиксу p (слову в алфавите A) эмоциональный вклад его продолжения pi тоже как префикса – $Fpex(pi)$. Это выражается специальным оператором $Ftex$. Оператор $Ftex$ выдаёт к слову i такую функцию $f=Ftex(i)$: что

$$f(p) = Fpex(pi).$$

Связь между функцией $Fiex$ и оператором $Ftex$ выражается следующей теоремой.

Теорема 1. Каковы бы ни были слова i , p , и s в алфавите A , если слово pis принадлежит множеству L , то

$$Ftex(i)(p)(s) = Fiex(i)(p,s).$$

Доказательство теоремы нетрудно получить из приведённых выше определений функций $Ftex$, $Fiex$ и $Fpex$.

Таким образом, основная задача сокращения объёма работ при эмоциональном анализе текстов на развивающемся и расширяющемся языке состоит в поиске подходящих инфиксов (стемов), для которых алгоритм вычисления $Ftex$ выражается с приемлемой вычислительной (временной и зонной – *Time, Space*) и описательной (объём программы) сложностью.

4.5. Разрабатываемая методология

Программа Emostemmer работает, стремясь использовать минимальные время и объем памяти с максимальной точностью. В основе программы имеется словарь не менее чем на 164 слова, представляющих простые формы или части слов, выражающих эмоции на русском языке. Алгоритм работает на специальных N-граммах, чтобы идентифицировать слова в разных формах; N-грамма для выявления эмоциональной переносимости будет называться Эмоциогаммой. Эмоциогамма – это список необходимых частей слов, передающих эмоции в тексте.

Эмоциогамма из n элементов может быть представлена в виде:

$$D_{dict} = [DW_1, DW_2, DW_3 \dots \dots \dots DW_n],$$

где DW_n – n^{th} элемент эмоциогаммы.

Например, элементы эмоциогамм могут соотноситься со словами в соответствии со следующей таблицей.

Таблица 4

Эмоциогаммы для выявления соответствующих эмоций из текста

Эмоциогамма	Эмоции
Везуч	везучий, везучая, везучее, везучие, везучего, везучей, везучего, везучих, везучему везучим, везучую, везучими, везуч, везуча, везуче, везучи
Депресси	депрессия, депрессии, депрессий, депрессиям, депрессию, депрессией, депрессиях, депрессиями
Гнев	гнев, гневный, гнева, гневов, гневу, гневам, гневом, гневами, в гневе

Красів	красивый, красивая, красивое, красивые, красивого, красивой, красивых, красивому, красивому, красивым, красивую, красивое, красивого, красивых, красивыми, красивом, красив, красива, красиво, красивы, красивее, по-красивее, красивой, по-красивей
--------	--

Элементы эмоциограмм не всегда находятся в начале соответствующего им слова. Они могут размещаться в середине или даже в конце слова.

4.6. Описание этапов работы алгоритма Emostemmer

Для эксперимента мы использовали загруженный текстовый файл с разных сайтов и интернет-блогов на разные темы и применили к нему следующие операции:

$$T_{Text} = [Text]$$

(Получение начального текста),

$$Z_{extra} = \{.,;: \% @\&! \$\}$$

(Задание множества пропускаемых литер, символы),

$$T_{TextL} = f_{low}(T_{Text})$$

(Функция f_{low} осуществляет перевод текста в строчные буквы),

$$D_{dict} = \{DW_1 \dots \dots DW_{164}\}$$

(Задание начального списка эмоциограмм слов, выражающих эмоции)

$$T_c = f(Z_{extra})[T_{TextL}]$$

(Оператор $f(Z_{extra}) [T_{TextL}]$ удаляет символы Z_{extra} из текста, такие как русские «шумовые» слова),

$$W_s = f_{split}(T_c)$$

(Функция f_{split} разбивает текст на слова выполняя токенизацию текста),

$$L_S = f'_{split}(W_S).$$

(Функция f'_{split} представляет каждое слово как цепочку букв),

$$N_S = f_{\text{эмоциогр}}(L_S).$$

(Функция $f_{\text{эмоциогр}}$ находит N-граммы, входящие в каждое слово, $N = 3, 4, 5$),

$$E_T = f_{comp}(N_S, D_{dict}).$$

(Функция f_{comp} ищет Эмоциогранны в словаре Emostemmer, E_T – цепочка чисел вхождений эмоциогранмы каждого слова в словарь)

$$F = \sum_{i=1}^n E_T [i], n = length(E_T).$$

(Определение общего числа вхождений эмоциогранмы слов текста в словарь).

По результатам использования вышеприведённого алгоритма можно сделать вывод, что идентификация эмоций из текста им более эффективна по сравнению с существующими методами словаря, такими как RuSentiLex.

4.7. Эксперимент и сравнение производительности методов

Для проведения эксперимента с использованием Emostemmer для русскоязычного текста был загружен текстовый файл с разных сайтов и интернет-блогов на разные темы, например: образовательные организации. Одной из таких организаций был Удмуртский государственный университет Ижевск, и отзывы собирались из разных блогов, таких как (<https://udmurt.media/articles/obshchestvo/67638/>,

https://edunews.ru/universities-base/pfo_izhevsk/udsu.html, https://otzovik.com/review_8269223.html, <https://studika.ru/izhevsk/udgu/otzyvi> и др.) и в виде анкет студентов. Полученный текстовый файл содержит текст вместе с эмоциями на русском языке.

Для идентификации эмоций из текстового файла и классификации их, использовались два разных источника русских лексиконов эмоции: Эмоциограммы и RuSentiLex. Русский лексикон эмоции RuSentiLex, созданный Н.В. Лукашевич и А.В. Левчиком, представляет собой сборник слов и фраз, отражающих эмоции в русском языке. Он состоит из 12 тысяч слов, которые используются для обозначения эмоций. Сборник классифицирует слова-эмоции как положительные или отрицательные. Источник RuSentiLex был получен с сайта (<http://www.labinform.ru/pub/rusentilex>). Он был загружен и использован для эксперимента, чтобы идентифицировать эмоции из текстового документа, параллельно с помощью программы Emostemmer.

Для Emostemmer список самых распространенных эмоций в русском языке был загружен с сайта психолога Петра Зарубина (<https://peter-zarubin.ru/spisok-chuvstv-i-emotsij/>). Список содержит 164 эмоции на русском языке. Эксперты создали основы (Эмоциограммы) для этих 164 эмоций для Emostemmer. Было отмечено, что Эмоциограммы созданные для каждой эмоции, способны идентифицировать все возможные формы их выражения. Эксперимент был выполнен на Emostemmer, который состоит из 164 Эмоциограмм. Для сравнения – RuSentiLex состоит из 12000 слово, представляющих эмоции.

Изначально программа проходила испытания на 12000 слов, представляющих эмоции в RuSentiLex вместе с их категориями. Затем программа была протестирована с использованием текстового файла, чтобы идентифицировать и классифицировать эмоции из текстового файла в их соответствующей группе на основе обучения алгоритму. Результаты идентификации и классификации представлены ниже в матрице ошибок (Confusion Matrix). Так же программа обучалась с Эмоциограммами, с их положительными и отрицательными категориями. Текстовый файл был обработан с использованием алгоритма Emostemmer, чтобы идентифицировать эмоции

и классифицировать их в соответствующие положительные или отрицательные группы.

Матрица ошибок (Confusion Matrix) указывает на эффективность классификации алгоритма. При этом используются показатели, такие как Accuracy, Recall (полнота), Precision (точность) и F-measure (в общем случае), чтобы определить эффективность классификации алгоритма. Матрица ошибок для каждого алгоритма, генерируемая ниже:

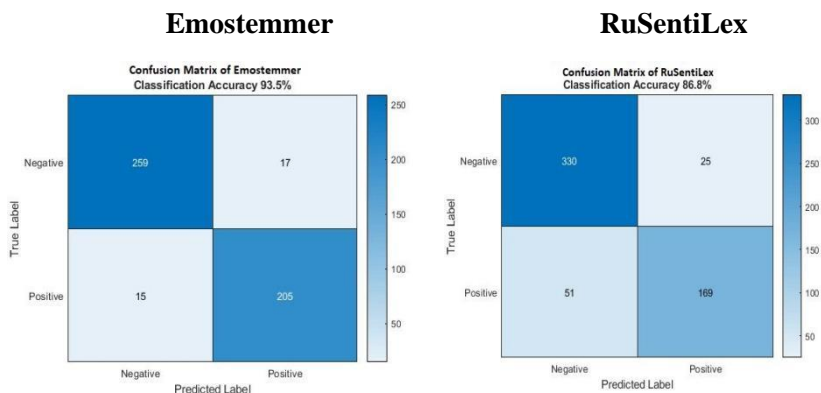


Рисунок 6 – Матрица ошибок для Emostemmer и RuSentiLex

TP (True Positive) и TN (True Negative) уточняют, что значения истины, определённые матрицей ошибок, являются актуальными, тогда как FN (False Negative) и FP (False Positive) детализируют ошибки или неправильную классификацию, сделанные матрицей ошибок во время выявления и классификации эмоций. Было замечено, что алгоритм RuSentiLex не смог распознать все эмоции в тексте и пропустил некоторые из них, которые не были упомянуты в его словаре.

Результаты матрицы ошибок были использованы для определения истинности значений, выдаваемых алгоритмами Emostemmer и RuSentiLex, и подробно описаны ниже в таблице 5.

Матрица ошибок для Emostemmer и RuSentiLex

Algorithm	Accuracy (Точность измерений)	Recall (Отзыв)	Precision (Точность результата измерений)	F-Measure (F-Мера)
RuSentiLex	86,78 %	86,61 %	92,96 %	89,67 %
Emostemmer	93,55 %	94,53 %	93,84 %	91,18 %

Показатель Accuracy был использован для определения правильности классификации алгоритма. Показатель Recall (Отзыв) – это отношение общего количества правильно классифицированных положительных примеров к общему количеству положительных примеров. Precision (точность результата измерений) показывает, что положительно помеченные примеры были правильно классифицированы. Показатель F-measure (F-Мера) использует гармонические средства для экстремальных значений. Он представляет значения Recall и Precision. Это равно или близко к наименьшему значению одного из них.

Из анализа матрицы ошибок (Confusion Matrix) видна более высокая точность результатов программы Emostemmer. Матрица показала более высокую производительность и эффективность алгоритма Emostemmer для идентификации и классификации эмоций из текста по сравнению с RuSentiLex.

Также было отмечено, что время выполнения RuSentiLex выше, чем Emostemmer. Словарь RuSentiLex состоит из 12000 слов и программа сравнивает все имеющиеся в его словаре слова с каждым словом из текста для идентификации и классификации эмоции из анализируемого текста до положительной или отрицательной категории. Для расчета времени выполнения каждого алгоритма в секундах, текст для эксперимента был разбит на разные файлы с разным количеством слов. Программы Emostemmer и RuSentiLex выполнялись

с использованием Software Matlab на компьютере Intel Core i5 7-го поколения. Полученные результаты представлены ниже в виде графика. Уменьшение времени выполнения задания Emostemmer по сравнению с RuSentiLex в среднем составляет 65 %. Полученные результаты представлены ниже в виде графика.

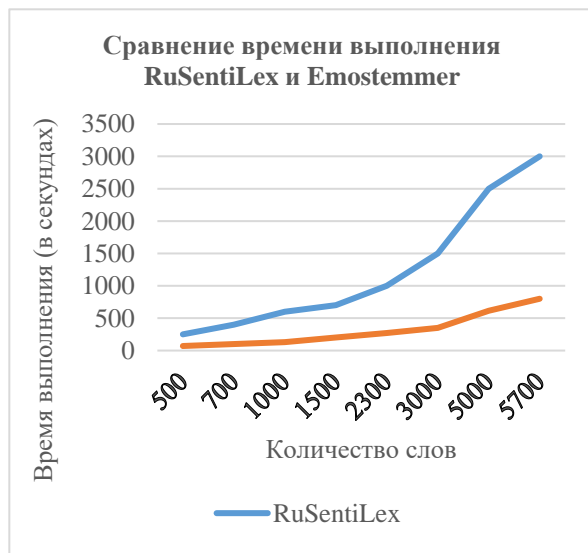


Рисунок – 7 Сравнение времени выполнения RuSentiLex и Emostemmer

Из приведённого графика видно, что время выполнения анализа увеличивается с увеличением размера текста. Заметная разница наблюдалась между временем выполнения программ Emostemmer и RuSentiLex с начала эксперимента с минимальным размером текстового файла. Это произошло из-за сложности RuSentiLex, во многих случаях размер RuSentiLex был больше размера самого текстового файла, используемого для анализа.

4.8. Отрицания и их влияние на эмоциональную полярность текста

Другими компонентами, влияющими на анализ текста, являются отрицания в нем. Они изменяют полярность эмоции и для определения реальной полярности слов в тексте, отрицаниям требуется обработка. Полярность эмоции описывает категорию эмоции, ее отношение к положительной или отрицательной группе эмоций. Отрицание меняет ориентацию эмоций, поэтому при выявлении всех эмоций, необходимо определить отрицания в тексте. Влияние отрицания на полярность эмоций следующее:

$$\text{NewPolscore}(B) = N(\text{Polscore}(B))$$

N представляет отрицание, которое, как правило, стоит перед словом и меняет его полярность. Для анализа взаимосвязи между отрицаниями и эмоциями использовались различные алгоритмы, такие как Машина Опорных Векторов, различные типы алгоритмов кластеризации и алгоритм Наивного Байеса. Отрицания лучше идентифицировать и кластеризовать с помощью программы кластеризации К-средних (K-Means кластеризация). Это алгоритм, который используется для разделения эмоций по группам таким образом, что эмоции в одних и тех же группах больше похожи друг на друга, чем в других группах или кластерах. Для кластеризации мы используем евклидово расстояние.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Где x и y – два вектора. Результаты программы «K-Means кластеризация» представлены на рисунке ниже. Перед применением кластеризации к тексту, эмоции, извлечённые из текста, рассредоточиваются в разных направлениях на основе их координаты x , то есть

номера индекса (положения) выражения эмоции в тексте, и координаты у, то есть её полярности, как видно на рисунке.

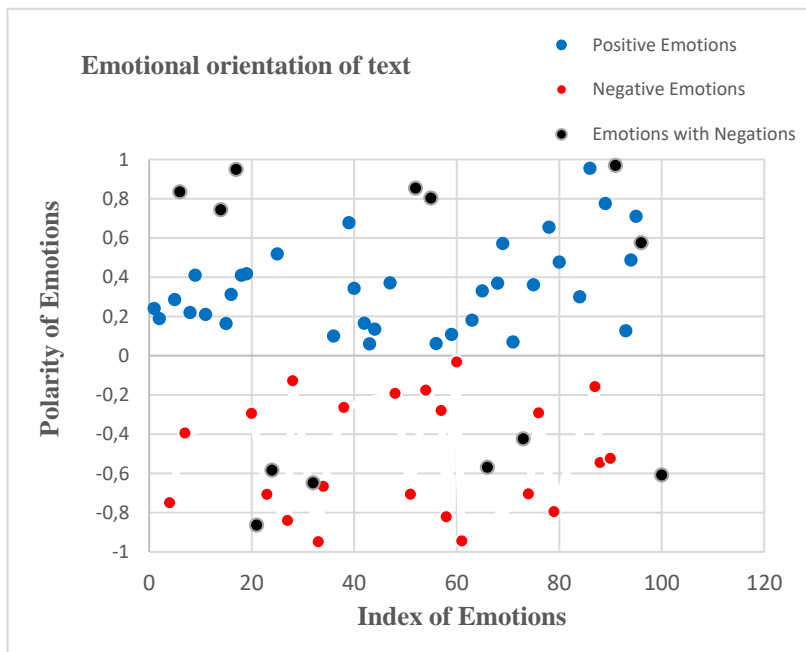


Рисунок 8 – Эмоциональная направленность текста

Рисунок отражает эмоциональную направленность текста. Ось X представляет собой место выражения эмоций в тексте, тогда как ось Y представляет полярность эмоций.

Полярность положительных эмоций принимает значения от 0 до +1, назначается экспертами на основе интенсивности и характеристики различных эмоций в тексте. Полярность отрицательных эмоций принимает значения от -1 до 0. Различные цвета на рисунке представляют характеристики эмоций. Красный цвет представляет отрицательные эмоции, синий цвет представляет положительные эмоции, а черный цвет – эмоции с отрицаниями.

При применении отрицания к эмоциям, эмоции меняют свою полярность, а также изменяются в графическом проявлении как показано ниже на следующем рисунке.

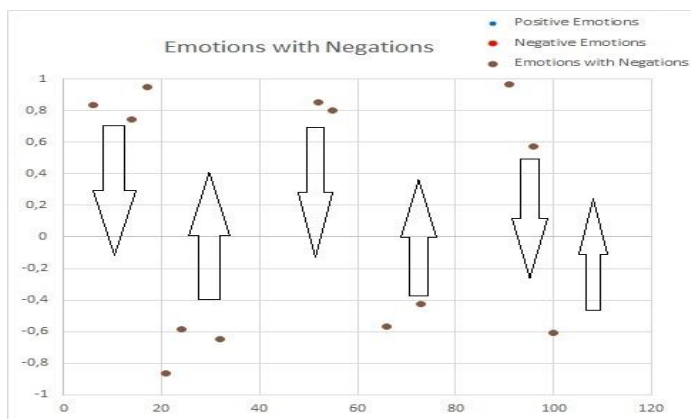


Рисунок 9 – Эмоции с ложной полярностью из-за отрицаний

Вышеприведённый рисунок представляет движение эмоций через граничное значение 0 при применении отрицаний, которые существуют перед ними. Новое расположение эмоций представлено на рисунке ниже.

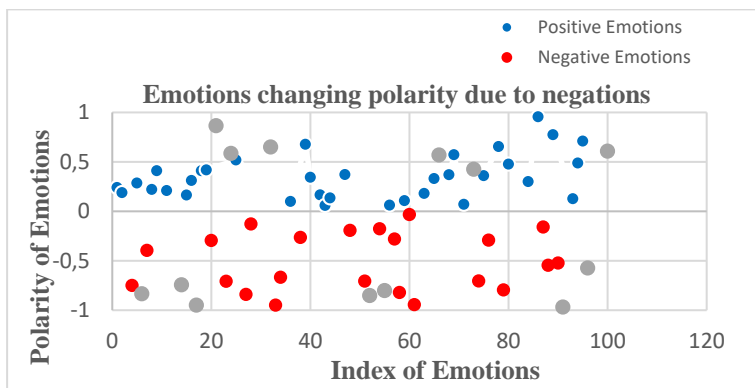


Рисунок 10 – Эмоции, меняющие свою полярность и графическое изображение при применении отрицаний

На рисунке выше представлено разделение эмоций на две группы: положительные и отрицательные. При применении к ним кластеризации можно наблюдать различные кластеры, которые можно увидеть на рисунке ниже.



Рисунок 11 – Кластеризация эмоций в тексте

Таким образом, в результате кластеризации было выявлено, что отрицания отражаются на полярности этих эмоций. Подсчёта частоты слов, выражающих положительные и отрицательные эмоции, не достаточно для определения полярности всех эмоций в тексте. В современном тексте люди часто используют отрицание с положительными эмоциями, а не с отрицательными, так как оно кажется более вежливым и приемлемым в обществе.

Выводы по главе 4

Изучены и проанализированы механизмы, используемые для анализа эмоций, выделенных из текста. При этом основное внимание было уделено стеммерам и их работе с русскоязычным текстом. Основная цель построения стеммера состоит в том, чтобы свести различные формы (словоформы) слова, такие как существительное, прилагательное, глагол, наречие и т. д. к его корневой форме.

В работе был использован Emostemmer для идентификации эмоций, которые несут слова из текста. До сих пор использованию стеммера для идентификации эмоций на русском языке не уделялось большого внимания. При сравнении производительности Emostemmer с RuSentiLex, выяснилось, что Emostemmer превзошел метод на основе RuSentiLex для идентификации эмоций из текста. Точность результатов Emostemmer также оказалась выше, чем точность RuSentiLex метода. Время выполнения анализа Emostemmer заметно меньше времени выполнения RuSentiLex. Emostemmer выполняет вычисления для идентификации эмоций из текста более просто и разумно.

ГЛАВА 5. АНАЛИЗ ЭМОЦИОНАЛЬНЫХ И ПСИХОЛИНГВИСТИЧЕСКИХ ХАРАКТЕРИСТИК ТЕКСТА, СОЗДАННОГО ПРИ ОБЩЕНИИ ЧЕЛОВЕКА С МАШИНОЙ

Проведён анализ характеристик эмоций, извлечённых из текста, идентифицированных программой Emostemmer в предыдущей главе. Выявленные в тексте эмоции лемматизируются в их простую или корневую форму. Рассмотрены синтаксические, семантические и логические характеристики эмоций и связь между ними. Также изучены психолингвистические характеристики текста. Эти характеристики используются для анализа текста и построения человеко-машинного интерфейса.

5.1. Логические характеристики эмоций в тексте

Для анализа логических характеристик эмоций в тексте использовались модальности суждений и модальная логика, а также было предложено несколько теорем.

Работа над модальной логикой началась во втором десятилетии 20-го века. Однако основная работа по модальной логике на уровне логики высказываний – процедуры доказательства, была введена К.И. Льюисом, основателем современной модальной логики.

Сол Крипке и А. Жояль в конце 1959 году создали реляционные и семантические решётки для неклассических логических систем. Сначала они были разработаны для модальной логики, а затем адаптированы к интуиционистской логике и другим неклассическим системам. Развитие семантики Крипке явилось прорывом в теории неклассических логик, поскольку теории моделей для таких логик до Крипке почти не существовало.

В этой работе использовали концепцию Крипке, чтобы предложить нашу модель логического анализа эмоциональных оценок. Целью данной работы является расширение области модальной логики до анализа логики эмоций, выраженной в текстах и определения

наиболее естественной, с нашей точки зрения, «логики эмоциональных оценок».

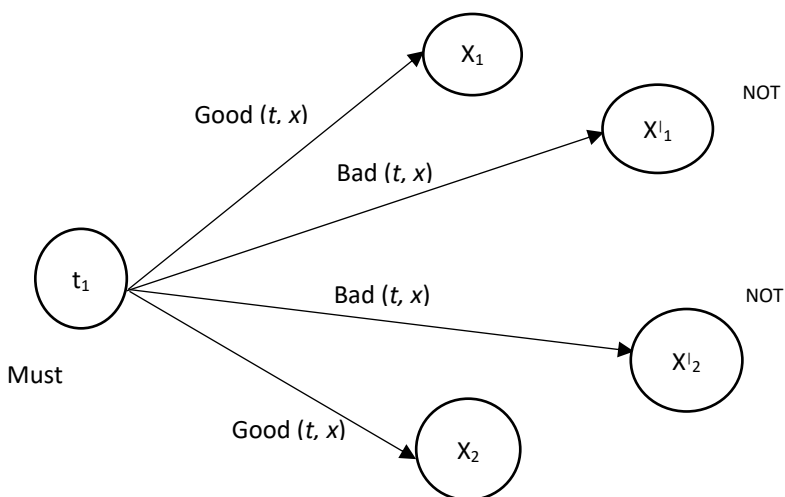
Гипотеза состоит в том, что набор субъектов с эмоциональными оценками, логика которых соответствует описанной ниже, достаточно велик, чтобы иметь смысл рассмотреть его и предложены определение модальности для эмоций.

5.1.1. Определение модальностей

I. Определение модальности $Must(Good)C$ – «желательно, чтобы было выполнено условие C »:

$$Force(t, Must(Good)C) \Leftrightarrow All(x)(Good(t, x) \Rightarrow Force(x, C)) \quad (1)$$

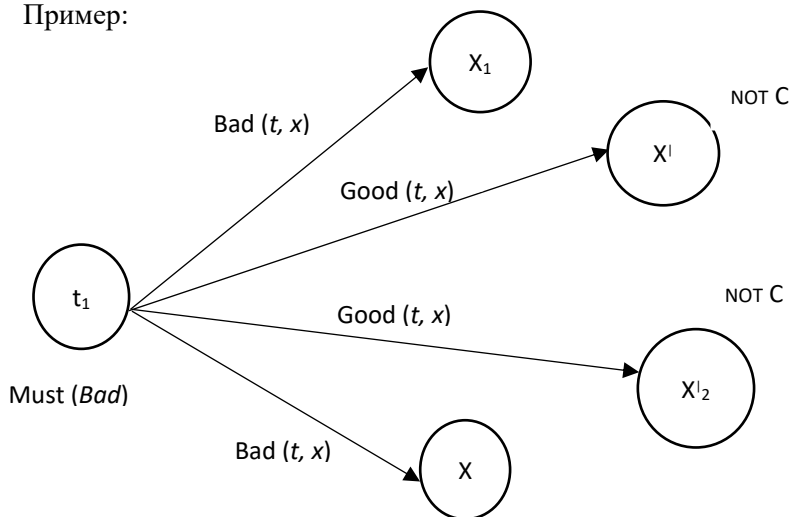
Пример:



II. Определение модальности $Must(Bad)C$ – «лучше бы не было выполнено условие C »:

$$Force(t, Must(Bad)C) \Leftrightarrow All(x)(Good(t, x) \Rightarrow Force(x, C)) \quad (2)$$

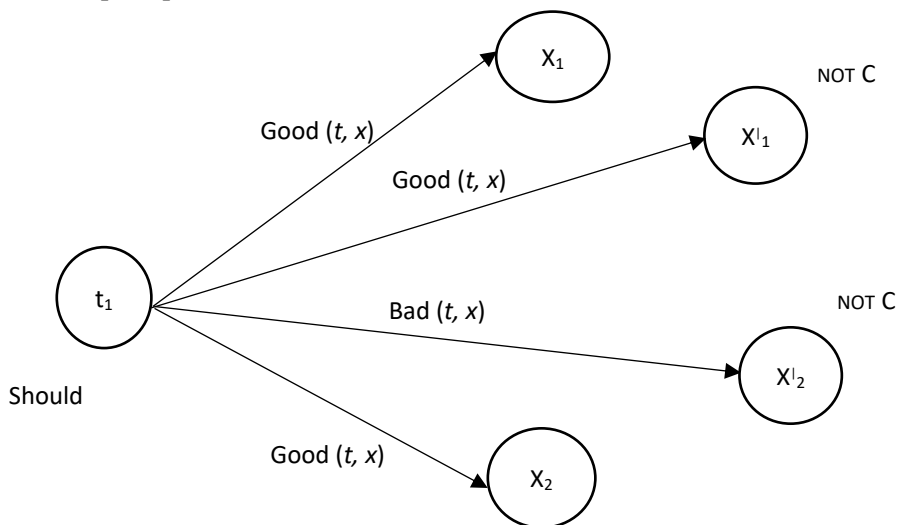
Пример:



III. Определение модальности $Should(Good)C$ – «хорошо, если выполнено условие C»:

$$Force(t, Should(Good)C) \Leftrightarrow All(x)(Force(x, C) \Rightarrow Good(t, x)) \quad (3)$$

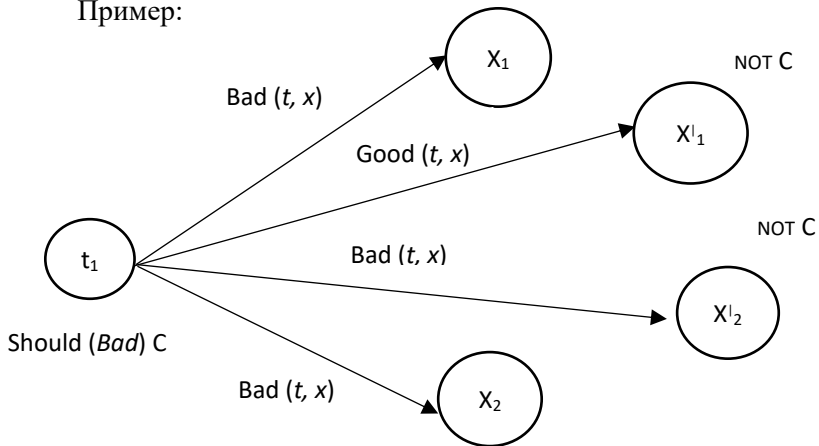
Пример:



IV. Определение модальности $\text{Should}(\text{Bad})C$ – «плохо, если выполнено условие C »:

$$\text{Force}(t, \text{Should}(\text{Bad})C) \Leftrightarrow \text{All}(x)(\text{Force}(x, C) \Rightarrow \text{Bad}(t, x)) \quad (4)$$

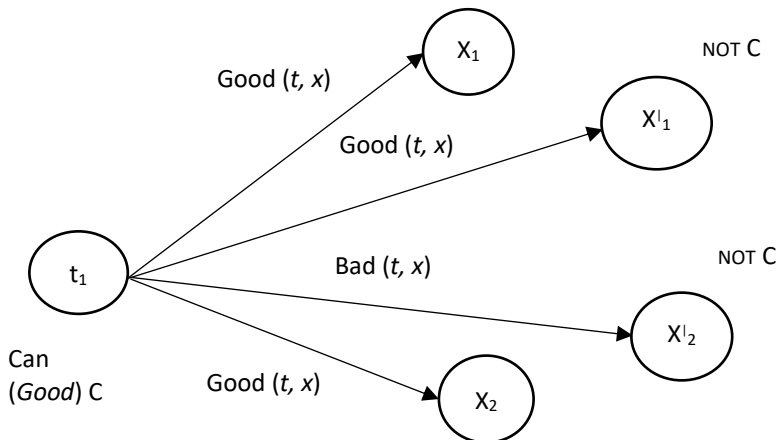
Пример:



VI. Определение модальности $\text{Can}(\text{Good})C$ – «условие C может быть хорошо»:

$$\text{Force}(t, \text{Can}(\text{Good})C) \Leftrightarrow \text{Exist}(x)(\text{Good}(t, x) \& \text{Force}(x, C)) \quad (5)$$

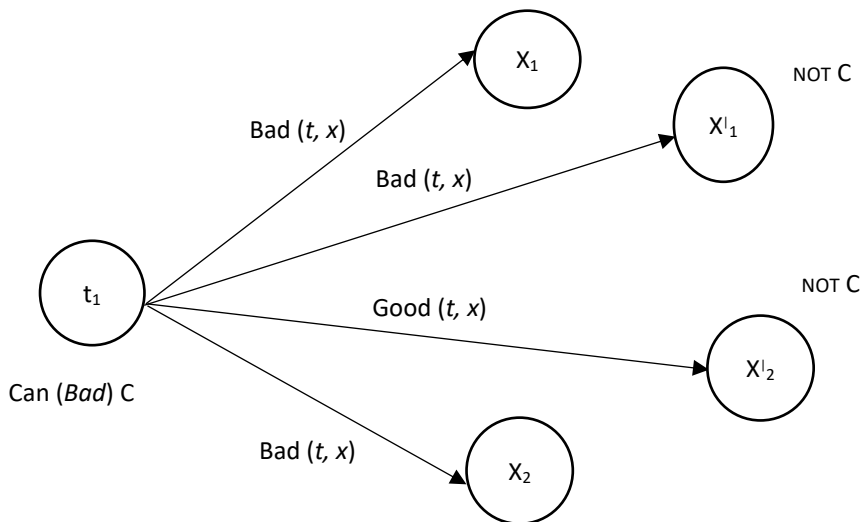
Пример:



VI. Определение модальности $\text{Can}(\text{Bad})C$ – «условие C может быть плохо»:

$$\text{Force}(t, \text{Can}(\text{Bad})C) \Leftrightarrow \text{Exist}(x)(\text{Bad}(t, x) \& \text{Force}(x, C)) \quad (6)$$

Пример:



Здесь были исследованы только так называемые экстенциональные оценки, зависящие от ситуаций, представляемых субъектом. Следует изучить также интенциональные оценки – оценки, которые вызывают сами формы утверждений. Зачастую они зависят не только от логической формы утверждения, но и от синтаксической формы на рассматриваемом естественном языке. С этим связано следующее замечание.

Фактически при изучении эмоционального анализа текстов касаемся следующих трёх видов смысла и трёх видов синтаксиса этих текстов:

Первый вид смысла – поверхностный «поисковый смысл», который, например, рассматривается при векторном анализе семантики

текстов поисковыми системами. Для извлечения такого поверхностного смысла можно не вникать глубоко в синтаксис предложений – достаточно анализировать, в каких сочетаниях встречаются различные слова.

Второй вид смысла – онтологический смысл. Он говорит о свойствах и отношениях элементов предметной области, к которой относится рассматриваемый текст. Для выявления онтологического смысла надо использовать традиционный синтаксический анализ текста и ещё более глубокую его обработку. В идеале полное выявление онтологического смысла должно приводить к точным формулам сложных логико-предметных языков. Это является очень трудной задачей.

Третий вид смысла – ещё мало изученный эмоциональный смысл текста, который говорит об эмоциях и эмоциональных оценках, связанных с текстом. По идее с эмоциональным смыслом должен быть связан свой анализ синтаксиса – анализ синтаксиса эмоций в тексте. Последнее является важной, ещё не решённой задачей.

Приведённые выше уравнения и формулы были использованы для анализа логических отношений, существующих между различными эмоциями в тексте, и их роли в определении будущего контекста.

5.2. Синтаксические характеристики эмоций в тексте

Для понимания характеристик эмоций в конкретном тексте, необходимо проанализировать синтаксис эмоций и его расположение в тексте. Данный анализ эмоций, выделенных из текста, не привлек внимание исследователей. Существование эмоции в определенной координате играет важную роль для определения эмоциональной поллярности текста.

Координата (x, y) эмоции в тексте – это ее местоположение в тексте. Координата x является положением эмоции в строке текста, тогда как координата y является положением эмоции в пространстве эмоций.

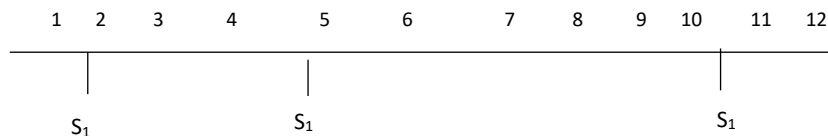
Для анализа семантических характеристик эмоций, извлечённых из текста, была использована математическая модель пространства слов (Word Space), в которой вхождения слов и расстояние между ними используются для определения силы и интенсивности эмоции в тексте.

Пространство слов – это понятие, производное от Метрического пространства, означающее набор элементов с определенным расстоянием между всеми его членами. В пространстве слов расстояния между словами и их вхождениями в текст являются значениями метрики. Более сильные эмоции несут слова с относительно высокой частотой появления в тексте и меньшим расстоянием между их вхождениями. Пространство слов используется для сочетания координат, соответствующих эмоциональным характеристикам и местоположению в тексте.

Чтобы определить силу выражения эмоций в тексте, необходимо обратить внимание на следующие факторы:

- расстояние между двумя вхождениями выражения одной и той же эмоции в тексте;
- идентификация изменения полярности эмоций, возникающего при появлениях слов, соответствующих одной и той же эмоции в разных контекстах.

Например, есть текст, в котором первые 12 слов символически изображены на рисунке ниже. В тексте присутствует эмоция $S(\mathcal{E}_1)$ с разными вхождениями $S_1(\mathcal{E}_1)$, $S_2(\mathcal{E}_1)$, и $S_3(\mathcal{E}_1)$ в координатах 2, 5 и 11.



Расстояние между первым и вторым появлением эмоции в тексте меньше расстояния между её вторым и третьим появлением. Значение функции расстояния α между двумя вхождением слов рассчитывается следующим образом:

$$\alpha(S(\Theta_1)) = \frac{1}{S_2(\Theta_1) - S_1(\Theta_1)} \times 100$$

По мере увеличения расстояния между вхождениями эмоций в текст, функция расстояния $\alpha(S(\Theta_1))$ уменьшается. Если есть эмоции противоположной полярности Θ_2 между двумя вхождениями одной и той же эмоции в текст, их значение будет добавлено в формулу для функции расстояния $\alpha(S(\Theta_1))$.

$$\alpha(S(\Theta_1)) = \frac{1}{S_2(\Theta_1) - S_1(\Theta_1) - (-S_1(\Theta_2))} \times 100$$

Среднее значение функции расстояния эмоции $S(\Theta)$

$$C(S(\Theta_1)) = \frac{1}{N} \sum_{i=1}^n \left(\frac{1}{S_{i+1}(\Theta_1) - S_i(\Theta_1)} \times 100 \right)$$

$$C(S(\Theta_1)) = \frac{1}{N} \sum_{i=1}^n \alpha(S_i(\Theta_1))$$

Сила определённой эмоции $I(\Theta_1)$ в тексте рассчитывается путём подсчёта количества появлений эмоции в тексте и добавления значения её функции расстояния α как

$$I(\Theta_1) = C(S(\Theta_1))$$

При использовании формул, приведённых выше, рассчитывается сила эмоций в тексте. Выявленная сила каждой эмоции затем используется для анализа взаимосвязи между эмоциями и темой текста. Выявленные эмоции и их силу можно увидеть в таблице ниже.

**Наиболее часто встречающиеся положительные
и отрицательные эмоции в тексте**

Эмоции	Количество в тексте	Сил	Эмоции	Количество в тексте	Сил
Радость	24	0.33	Сожаление	10	0.22
Восторг	11	0.11	Скука	3	0.004
Интерес	12	0.08	Приподнятос	11	0.15
Надежда	17	0.24	Увлечение	12	0.25
Благодарность	9	0.09	Забота	6	0.04
Гордость	3	0.001	Изумление	11	0.18
Восхищение	7	0.05	Оптимизм	13	0.22
Признание	11	0.21	Доверие	11	0.15
Дружелюбие	3	0.008	Спокойствие	18	0.31
Любовь	9	0.12	Искренность	8	0.21
Раздражение	3	0.005	Гнев	7	0.11
Негодование	4	0.002	Ужас	3	0.005
Досада	7	0.07	Тревога	11	0.27
Возмущение	9	0.11	Обида	6	0.010
Грусть	11	0.23	Отвращение	2	0.002
Разочарование	5	0.05	Счастье	17	0.26

Наряду с эмоциональными характеристиками текста, также изучаются его психолингвистические характеристики. На основе этих характеристик программа будет понимать особенности поведения человека и эффективно реагировать.

5.3. Анализ сложной природы эмоций и отношений между ними

Эмоции сложны по своей природе. Эмоция, в зависимости от её происхождения и характеристик, может представлять другую эмоцию в письменном тексте. Комплексное свойство эмоций аналогично характеристикам комплексных чисел, которые представлены

их действительной и мнимой частями. Различные алгоритмы используются для выявления и анализа эмоций в тексте.

Они обычно рассматривают каждую эмоцию как отдельный элемент со своими характеристиками. Связь между различными эмоциями и их влиянием на общую полярность текста не получила большого внимания. Идея состоит в том, чтобы проанализировать сложные характеристики эмоций, идентифицируя отношения и зависимости, которые существуют между ними. Для этого были использованы теории эмоций Плутчика и теория эмоций Паррота. Согласно этим теориям, эмоции делятся на разные группы и уровни в соответствии с их интенсивностью и природой происхождения.

Предложенная методология помогает проанализировать сложные характеристики эмоций и понять их реальную и мнимую части. Например, скажем, «Андрей любит Лену». Здесь единственная эмоция, явно представленная в предложении – это любовь. Однако это также означает, что он любит проводить с ней время. Он чувствует себя счастливым и проявляет к ней привязанность. Здесь эмоции «счастливый» и «привязанность» являются воображаемыми эмоциями действительной эмоции «любовь». Эти эмоции не представлены в тексте, но они существуют в значении. Они семантически и логически зависят от своей основной эмоциональной любви, что было рассмотрено Парротом в его работе. Точно так же: «Евгений боится экзамена». Здесь действительная эмоция, выраженная в тексте, «боится». Однако с этим связаны воображаемые эмоции, такие как паника, стресс, ужас, напряжение и т. д. Можно видеть, что эмоция выражает существование нескольких несуществующих эмоций в тексте, или эмоция может представлять множество других эмоций.

По словам Паррота, эмоции можно разделить на три разных уровня или группы в зависимости от их характера и интенсивности, такие как первичные, вторичные и третичные эмоции. Существуют разные отношения между эмоциями одного уровня и эмоциями другого уровня. Первичные эмоции – это самые сильные эмоции, а третичные – наименее интенсивные. Связь между эмоциями наблюдается ниже (таблица 7).

Теория эмоций Паррота 2001

Primary emotion	Secondary emotion	Tertiary emotions
Love	Affection	Adoration, affection, love, fondness, liking, attraction, caring, tenderness, compassion, sentimentality
	Lust	Arousal, desire, lust, passion, infatuation
	Longing	Longing
Joy	Cheerfulness	Amusement, bliss, cheerfulness, gaiety, glee, jolliness, joviality, joy, delight, enjoyment, gladness, happiness, jubilation, elation, satisfaction, ecstasy, euphoria
	Zest	Enthusiasm, zeal, zest, excitement, thrill, exhilaration
	Contentment	Contentment, pleasure
	Pride	Pride, triumph
	Optimism	Eagerness, hope, optimism
	Enthrallment	Enthrallment, rapture
	Relief	Relief
Surprise	Surprise	Amazement, surprise, astonishment
Anger	Irritation	Aggravation, irritation, agitation, annoyance, grouchiness, grumpiness
	Exasperation	Exasperation, frustration
	Rage	Anger, rage, outrage, fury, wrath, hostility, ferocity, bitterness, hate, loathing, scorn, spite, vengefulness, dislike, resentment
	Disgust	Disgust, revulsion, contempt
	Envy	Envy, jealousy
	Torment	Torment
Sadness	Suffering	Agony, suffering, hurt, anguish
	Sadness	Depression, despair, hopelessness, gloom, glumness, sadness, unhappiness, grief, sorrow, woe, misery, melancholy
	Disappointment	Dismay, disappointment, displeasure
	Shame	Guilt, shame, regret, remorse
	Neglect	Alienation, isolation, neglect, loneliness, rejection, homesickness, defeat, dejection, insecurity, embarrassment, humiliation, insult
	Sympathy	Pity, sympathy
Fear	Horror	Alarm, shock, fear, fright, horror, terror, panic, hysteria, mortification
	Nervousness	Anxiety, nervousness, tenseness, uneasiness, apprehension, worry, distress, dread

Из приведённой выше таблицы, различные эмоции, их уровень интенсивности и отношения между ними можно наблюдать в соответствии с теорией эмоций Паррота.

5.3.1. Методология для анализа эмоций

Согласно теории Паррота, эмоции делятся на три группы или уровня, называемые первичными, вторичными и третичными уровнями. Первичные эмоции – это независимые эмоции. Вторичные эмоции зависят от первичных эмоций, тогда как третичные эмоции напрямую зависят от вторичных и косвенно от первичных.

Это означает, что при распределении оценки полярности для каждой эмоции необходимо наблюдать за их связью с другими эмоциями и зависимостями, существующими между ними. Сложность состоит в том, что эмоция может проявлять свойства некоторых других эмоций или она может представлять другие эмоции в тексте.

Основываясь на этой гипотезе, оценки полярности присваиваются эмоциям с использованием приведённых ниже уравнений.

$$\begin{aligned} \text{polscore}(A) &= \text{polscore}(A) + \sum_{i=1}^N \text{polscore}(B(i)) \\ \text{polscore}(B) &= \text{polscore}(B) + \sum_{i=1}^N \text{polscore}(C(i)) \\ \text{polscore}(C) &= \text{polscore}(C) \end{aligned}$$

Из приведённых выше уравнений определяется значение полярности эмоций, принадлежащих основному уровню А, которое является суммой его собственного значения полярности и значений полярности всех от него зависимых эмоций уровней В и С. Аналогично, значение полярности эмоций, принадлежащих уровню В, является суммой значения его собственной полярности и полярностей всех от него зависимых эмоций на уровне С. В то время как эмоции на уровне С имеют только свои собственные значения полярности.

Другими словами, оценка полярности реальных эмоций – это их собственная полярность, дополненная оценкой полярности всех воображаемых эмоций. Результаты приведённых выше уравнений можно визуализировать, (рисунок 12).

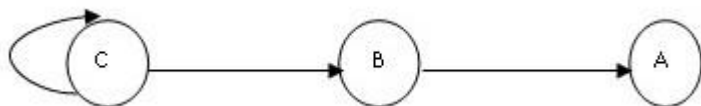


Рисунок 12 – Логические отношения между разными эмоциями

На Рисунке 12 позиция С определяет третичные эмоции, представленные в таблице 7. Эмоции на этом уровне обычно являются кратковременными и умеренными эмоциями. Если они существуют в течение более длительного периода, они переходят на более высокий и более экстремальный уровень эмоций в состоянии В. Эмоции в состоянии В являются вторичными эмоциями. У них есть свои особенности, плюс характеристики третичных эмоций, которые следуют за ними. Когда вторичные эмоции существуют на этом уровне в течение более длительного периода, они переходят на самый высокий и самый интенсивный уровень эмоций, называемый первичными эмоциями.

Первичные эмоции – это интенсивные и продолжительные эмоции. Это в основном необратимые эмоции, и для их отмены требуется более длительный период. Первичные эмоции содержат свои собственные свойства, а также свойства своих вторичных и третичных эмоций.

Приведённый выше рисунок детализирует отношения между эмоциями на основе категорий эмоций, предложенных Парротом в его теории. Можно наблюдать, что настоящие эмоции, как и в случае первичных эмоций, обладают свойствами своих воображаемых эмоций, таких как вторичные (уровень В) и третичные (уровень С) эмоции.

Для определения эмоций и их зависимости, каждое предложение из текста было проанализировано отдельно в эксперименте. Когда третичные эмоции явно существуют в тексте, они содержат только свою предполагаемую полярность. В то время как вторичные эмоции в тексте содержат все характеристики и значения полярности их третичных эмоций (мнимых эмоций), а также их собственные (реальные

эмоции) полярности. Точно так же первичные эмоции содержат свои собственные значения полярности наряду со значениями полярности их вторичных и третичных эмоций (мнимых эмоций).

Вышеупомянутая методология помогает понять логические отношения, которые существуют между различными эмоциями в тексте. Эти отношения между эмоциями могут использоваться для обновления полярности каждой эмоции в тексте и для определения общей силы эмоции в тексте.

5.3.2. Эксперимент и результаты

Для эксперимента был загружен текст из Интернета. Слова (эмоции) были проанализированы с использованием теории эмоций Паррота для классификации и обновления значений их полярности. Затем была применена предложенная методология, учитывающая различные уровни интенсивности эмоций, извлечённых из текста. Для простоты вычислений в тексте выставлена оценка полярности: 1 – для эмоций третичного уровня, 2 – для вторичных эмоций и 3 – для эмоций первичного уровня. Эксперимент и его результаты подробно изложены ниже (таблица 8).

Таблица 8

Интенсивность эмоций в тексте по теории эмоций Паррота

Эмоции, извлечённые из текста	Интенсивность эмоций в зависимости от местоположения в тексте $I_i(\text{Э})$	Полярность эмоций согласно теории Паррота $P_i(\text{Э}_L)$	Полярность зависимых эмоций $P_i(\text{Э}_d)$	Общая полярность эмоций согласно теории Паррота $P_i(\text{Э})$	Общая интенсивность эмоций в тексте по теории Паррота $IE_i = P_i(\text{Э}) + I_i(\text{Э})$
Безопасность	85,6	1	0	1	86,6
Надежность	97,1	1	0	1	98,1
Хорошо	34,3	1	0	1	35,3

Доверие	86,3	1	0	1	87,3
Легкость	73,6	1	0	1	74,6
Сюрприз	47	3	5	8	55
Страх	52,7	3	9	12	64,7
Горе	44,2	1	0	1	45,2
Опасность	38,5	3	9	12	50,5
Небезопасный	29,5	1	0	1	30,5

Для вычисления общей интенсивности эмоций, извлечённых из текста, определяются такие характеристики эмоции, как количество вхождений в текст, положение вхождения, расстояние между различными вхождениями, а также их природа и зависимости друг от друга, как рассуждал Паррот в своей теории эмоций. Общая интенсивность эмоций, извлечённых из текста, рассчитывается с использованием приведённых ниже уравнений:

$$IE_i = P_i(\mathcal{E}) + I_i(\mathcal{E})$$

где:

$$P_i(\mathcal{E}) = P_i(\mathcal{E}_L) + P_i(\mathcal{E}_d)$$

$$I_i(\mathcal{E}) = \sum_{i=1}^n C_i + \alpha (S(\mathcal{E}))$$

Из приведённой выше таблицы 8 можно заметить, что положительных или отрицательных эмоций в тексте недостаточно для определения фактической эмоциональной полярности текста. Но такие факторы, как природа эмоций, в зависимости друг от друга, частота появления в тексте, способствуют вычислению реальной интенсивности и эмоциональной полярности текста. Точно так же появление определённых эмоций в тексте зависит от названия текста.

5.3.3. Определение взаимосвязи между темой текста и эмоциями в нем

Тема текста играет важную роль в подборе человеком определённых эмоций для данного текста. Чтобы понять это явление, было проведено исследование эмоций, представленных людьми в разных текстах. Сначала необходимо определить темы, которые в настоящее время обсуждаются в Интернете. Для эксперимента были взяты блоги с похожими темами, размещённые в разные периоды, и объединяющиеся в одну группу. Общая тема, которая используется в повторном поиске, была извлечена из обзоров и блогов людей о каких-либо продуктах, науке, будущей жизни, текущей политической ситуации, оружии, религии и т. д. Используя предложенную ранее методологию, изучаем изменение эмоций и их силу в разные периоды.

Шкала 1/10 (шкала Лайкерта) использовалась, чтобы минимизировать значения для упрощения вычислений. Согласно этой шкале значения интенсивности умножаются на 0.1, а результирующая интенсивность представлена ниже (таблица 9).

Важным фактором, повышающим эффективность анализа, является связь между темой текста и эмоциями в нем. Для эксперимента были использованы блоги на похожие темы, размещённые в разные периоды. Применяя созданную методологию, были произведены наблюдения за изменением эмоций и их силы за указанный период. В ходе эксперимента было замечено, что время появления текстов по заданной теме играет важную роль в проявлении эмоций людьми. Результаты двух тем подробно описаны ниже (таблица 9, таблица 10) и (рисунок 13, рисунок 14).

**Эмоции выраженные в теме:
«Интернет-магазин безопасен или небезопасен?»**

Интернет-магазин безопасен или небезопасен?			
Период	2005–10	2011–15	2016–19
Тип эмоции	Интенсивность		
Безопасность	5.0	6.5	6.09
Надёжность	4.0	4.0	8.04
Хорошо	3.5	3.2	4.03
Доверие	3.5	3.0	4.52
Лёгкость	2.6	2.7	3.01
Страх	3.04	3.5	3.51
Горе	3.0	3.2	3.01
Опасность	3.5	4.0	3.01
Небезопасный	2.7	3.7	1.005
Беспокойство	2.1	2.9	1.50

Первая тема была о совершении покупок онлайн. Был проанализирован текстовый блог из разных периодов, и наиболее распространённые эмоции в тексте подробно описаны в таблице выше вместе с их интенсивностью в разные периоды. В целом можно заметить, что доверие к интернет-магазинам, надежда вовремя получить их товар и качество услуг со временем возрастают. Оценку эмоций по первой теме можно наблюдать в обобщённом виде на рисунке ниже.

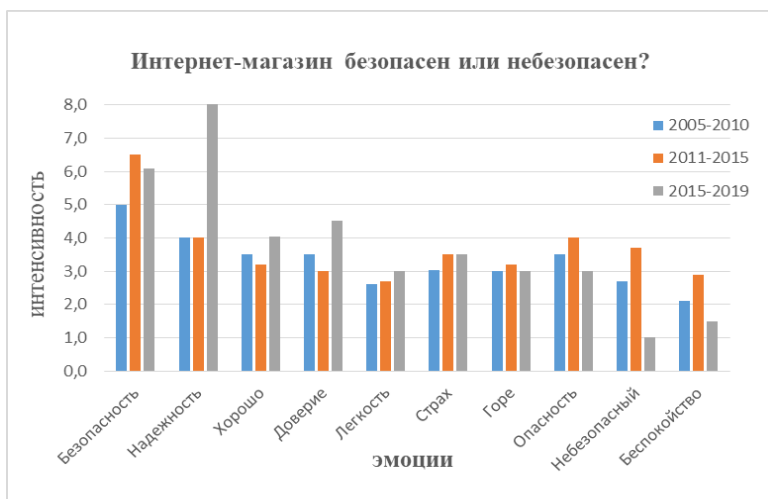


Рисунок 13 – Эмоции выраженные в теме:
«Интернет-магазин безопасен или небезопасен?»

Из приведённого выше рисунка можно детально наблюдать изменение эмоций людей по поводу различных компонентов онлайн-покупок за определённый период.

Вторая тема состояла в том, чтобы понять связь между темой и эмоциями людей, которая была посвящена религии, так как она была главной темой интернет-дискуссий в течение последних двух десятилетий, в частности из-за таких факторов, как война, экстремизм и т.д. Эмоции людей и слова, выражающие эмоции, также обобщены и представлены в Таблице ниже.

Таблица 10

Эмоции на тему: «Значение религии для общества»

Значение религии для общества?			
Год	2005–10	2011–15	2015–19
Тип слов, вызывающих эмоции	Интенсивность		
Насилие	2.55	4.5	3.9
Война	3.01	5.7	4.0
Погромы	2.01	4.0	4.2

Вера	4.52	4.0	3.7
Покорность	3.01	4.5	4.2
Отвлечение	3.51	5.5	5.4
Удовлетворение	8.03	7.1	5.0
Лояльность	3.01	4.02	3.5
Мораль	3.01	3.0	3.2
Мотивация	1.50	4.5	4.1
Опасность	3.51	5.5	4.9
Крайность	3.01	7.0	6.0
Мир	6.03	6.0	5.0
Обоснование	1.005	3.1	2.5
Человечество	8.04	5.5	4.5

В таблице можно заметить, что эмоции и слова, выражающие их, меняют свою интенсивность в зависимости от событий, которые происходят в течение определённого периода. Обобщённые результаты эксперимента представлены ниже.



Рисунок 14 – Эмоции на тему: «Значение религии для общества?»

Наблюдается заметное увеличение силы некоторых эмоций и такое же значительное снижение силы других эмоций за этот период, что означает изменение мышления людей с течением времени.

Для эксперимента были использованы блоги на похожие темы, размещённые в разные периоды. Применяя созданную методологию, были произведены наблюдения за изменением эмоций и их силы за указанный период.

Характеристики эмоций в тексте используются вместе с психолингвистическими характеристиками текста, чтобы сделать лингвистический человеко-машинный интерфейс способным проводить и улучшать взаимодействие между ними.

Основная цель работы – создать машины, способные анализировать поведение человека и его эмоции, извлечённые из текста, а затем разумно реагировать во время общения с людьми.

По мнению известных философов и психологов (И.Б. Майер, К.Г. Юнг), люди с разным психологическим характером имеют определённые типы поведения во время общения. Они предложили список слов, используемых разными группами людей во время общения.

Предлагаемая модель будет анализировать слова из текста, чтобы идентифицировать определённый тип человека, сравнивая их со списком слов, предложенным учёными.

5.4. Характеристики экстравертного типа личности

Люди экстравертного типа – это люди, которые общаются более открыто. Они часто выражают свои эмоции, чувства во время общения как показано в таблице ниже (таблица 11).

Таблица 11

Характеристики типа личности, общие фразы и их слова, используемые экстравертом при общении

Экстравертный тип	
Характеристики	Пример использования слов и фраз в общении
Доминирующий тип	Я, мой, хочу, нужно, решить, выбрать, управлять, вести, рекомендовать, мощный, выбор, большой, огромный, мой собственный, мне нравится, мое мнение, все, каждый, красивый, замечательный, исключительный, революционный, идея, сюрприз, неожиданный, вдохновляющий, экста-, супер-, мега-, гипер-, ультра-, создать, мечтать, странный, объяснить, проверить, сравнить, понять, технология, тайна, отношение, исследовать, я хотел бы знать, неправильно, искренне, не согласен, ошибка, падение, разрыв, замечание, уведомление, наблюдение, нечестность, жесткий, острый, критический, соперник, конкурентный, обострить, сожалеть, помочь, жалеть кого-то, вызов, трудный, смелый, мужество, риск, опасность, герой, защищать, дать, передать, подарок, помощь, благотворительность пожертвовать, уникальный, любимый, особенный, выборочный, выбранный, награда, продвижение по службе, разнообразить, акцент, другой.
Максималистский тип	
Вдохновляющий тип	
Тип Обнаружения	
Проверки тип	
Зачочки тип	
Эмпатический тип	
Смелый тип	
Щедрый тип	
Предпочтительный тип	

5.5. Характеристики интровертного типа личности

Во время общения люди с интровертным типом поведения более вежливы, чувствительны и осторожны в общении. Они покорны и ждут, пока другие проявят инициативу. Они следуют определенному набору слов, фраз во время общения, как описано ниже (таблица 12).

Таблица 12

Характеристики типа личности, общие фразы и их слова, используемые интровертом при общении

Интровертный тип	
Характеристики	Пример использования слов и фраз в общении
Подчиненный тип	Он, она, ваши, их, эксперты, мнения, статистика, другие, что бы вы мне посоветовали, предложение, просто, любое, минимум, неважно, меньше, средний, нормальный, стандартный, общий, типичный, предыдущий, следующий, последний, один за другим, по порядку, заметки, разделить, разложить, по времени, дате, крайнему сроку, тем же, доказанным, традицией, постоянная, терпимость, стабильная, конвенция, доверие, приказы, апелляция, все, присоединиться, объединить, классифицировать, мало, почти, легко, круг, примирить, уменьшить, соответствовать, я думаю об этом, вы не возражаете, воздержитесь, закончить, быстро, кратко, спешить, конкретно, точно, производительность, эффективно, сомневаться, осторожно, опасаться, предупредить, ограничить, угроза, безопасная охрана, слишком много, процент, ставка, специальные предложение, вес, уровень, оценка, сортировка, судья, закон, правило, погашение, последовательно, экономия, комиссия, энергосбережение, минимизация, выравнивание, оценка, справедливость, равномерно, правильно, дешево, продвижение по службе, переработка, выключение, рентабельно, энергоэффективно.
Минималистский тип	
Систематический тип	
Консервативный тип	
Не замечающий тип	
Гармоничный тип	
Фактический тип	
Защитный тип	
Экономный тип	
Тип Балансировки	

Характеристики различных типов людей и фразы, слова в таблицах рассматриваются и используются в качестве психолингвистического словаря для эксперимента. При получении текста, созданного пользователем, машина идентифицирует эмоциональное и психолингвистическое поведение, выраженное в тексте, и реагирует разумным образом. Это делает общение между человеком и машиной более эффективным.

5.6. Идентификация типа текста

Для идентификации и классификации текста, если он написан человеком с экстравертным или интровертным характером, система реализует различные этапы, как показано ниже (рисунок 15).

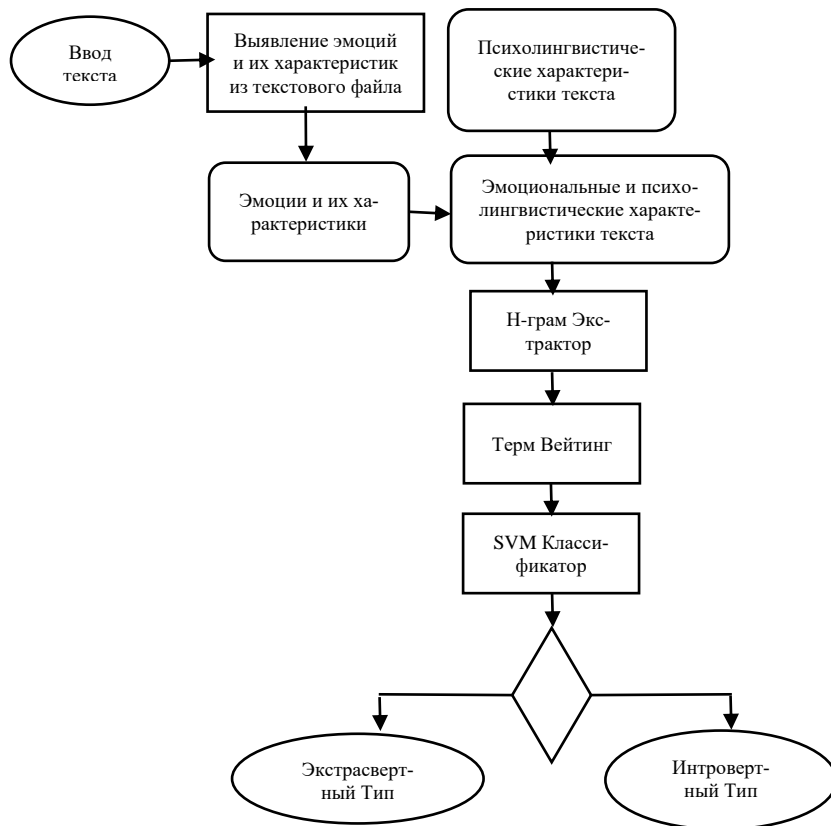


Рисунок 15 – Процесс классификации текста на экстравертный или интровертный тип личности автора

5.7. Методология и результаты

При получении текста, созданного пользователем, машина идентифицирует эмоциональное и психолингвистическое поведение, выраженное в тексте, и реагирует разумным образом. Это делает общение между человеком и машиной более эффективным. Список некоторых слов, показывающих психолингвистические характеристики людей, идентифицированных в ходе эксперимента:

Таблица 13

Характеристики экстраверта и интроверта и их слова, фразы, используемые в общении

Психологические характеристики текста	
Пример использования слов и фраз в тексте	Я, мой, идея, любое, по порядку, минимум, меньше, последний, собственный, вести, красивый, средний, мне, революционный, крайнему сроку, просто, каждый, управлять, типичный, меньше, неожиданный, выбрать, статистика, нравится, нормальный, тем же, рекомендовать, он, она, выбор, огромный, типичный.

Для подготовки соответствующего ответа автору текста характеристики, которые должен представлять ответ, взяты из различных теорий психологов и онлайн-блогов, таких как:

<https://peopleleaders.com.au/top-10-tips-for-communicating-with-extroverts/>
<https://peopleleaders.com.au/top-10-tips-for-communicating-with-extroverts/>

Один из простых ответов программы на анализируемый ею текст можно увидеть на рисунке ниже:

Текст представляет собой сочетание положительных и отрицательных эмоций разной интенсивности. Положительных эмоций больше и они имеют большую интенсивность. Психолингвистические характеристики выражают идею о том, что текст скорее написан экстравертом, чем интровертом. Таким образом, при общении необходимо учитывать следующие факторы, такие как сосредоточение внимание на теме, быстрое реагирование, высказывание мнения и ожидание больших ответов.

Рисунок 16 – Примерный ответ, генерируемый программой

Процесс начинается с выявления эмоций из текста и определения его психолингвистических характеристики. Затем программа сравнивает идентифицированные компоненты текста со словарем психолингвистических слов и фраз. После этого фразы извлекаются из текста с использованием N-грамм.

N-грамма – это известный механизм, в котором набор букв из каждого слова используется для идентификации конкретного слова и использует его появление для предсказания его будущего появления после вычисления его N-1 вхождений.

Простой механизм, используемый для предсказания слов и их появления в тексте, – это использование максимальной оценки правдоподобия (MLE). Например, чтобы вычислить вероятность появления биграммы слова y , следующего за x , нам нужно подсчитать количество вхождений их биграмм $C(xy)$ из текстового документа. Вероятность появления слова в тексте можно рассчитать с помощью приведенного ниже уравнения.

$$P(W_n | W_{n-1}) = \frac{C(W_{n-1}W_n)}{C(W_{n-1})}$$

где $C(W_{n-1})$ – используется для обозначения количества биграмм, начинающихся с W_{n-1} , т.к. число биграмм, начинающихся с W_{n-1} равно числу вхождений W_{n-1} . Уравнение для оценки вероятности для N-граммы с использованием оценки максимального правдоподобия имеет следующий вид:

$$P(W_n | W_{n-N+1}^{n-1}) = \frac{C(W_{n-N+1}^{n-1} W_n)}{C(W_{n-N+1}^{n-1})}$$

Каждой n-грамме слов и фраз, извлеченных из текстового документа, присваивается вес. Вес назначается в зависимости от частоты появления каждой n-граммы в тексте. Этот вес компонентов n-грамм используется для определения психолингвистического поведения автора текста.

На рисунках ниже показано существование интровертных и экстравертных слов, используемых людьми в анализируемом тексте. Алгоритм Support Vector Machine (SVM) используется для идентификации и классификации слов из текста на две группы.

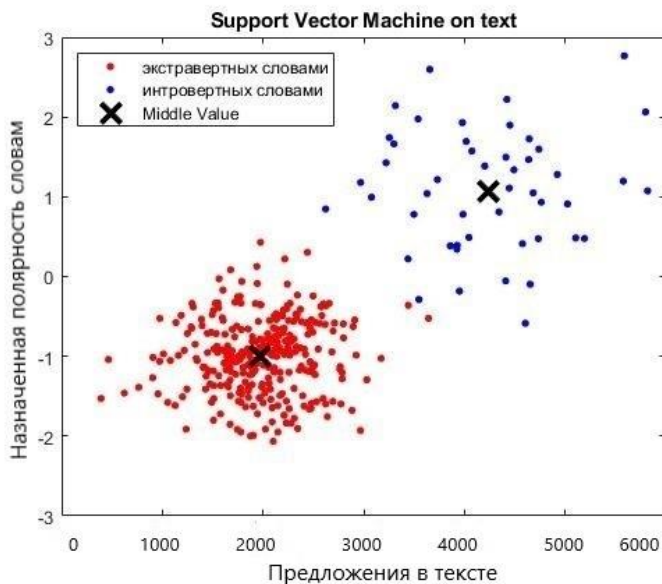


Рисунок 17– Применение (SVM) для идентификации групп интровертных и экстравертных слов в тексте

Из рисунка видно, что текст, использованный для эксперимента, больше относится к созданному экстравертной личностью (полярность от -3 до 0 для экстраверта и от 0 до $+3$ для интроверта по оси x и по оси y – ось для индексов предложений из текста). Используя SVM, можно заметить, что два типа характеристики личности человека разделены на две группы, как на картинке, красным цветом для экстраверта и синим для интроверта.

Предложенная модель была протестирована и опробована с использованием онлайн-текста, представляющего взаимодействие между человеком и виртуальным агентом. Модель используется машиной

для принятия решения об ответе человеку на основе его эмоциональных и психолингвистических характеристик.

Программа определяет характеристики эмоций в тексте для принятия решения об эмоциональной направленности текста. Она управляет процессом принятия решений, не только идентифицируя эмоции в тексте, но также вычисляя их силу, отношения между ними по их характеру и отношениям друг с другом, чтобы определить эмоциональную полярность текста.

Психолингвистические характеристики эмоций анализируются для определения психолингвистических характеристик текста, чтобы выбрать необходимый ответ автору текста и улучшить общение. Люди могут принимать решения относительно общественного мнения о продукте или политике правительства.

Необходимо принимать во внимание, что машина сфокусирована на теме и должна отвечать требуемым образом. Предлагаемая модель обучается фразам и их словам, которые отражают конкретное эмоциональное состояние и настроение человека, а также психолингвистические характеристики его текста. Когда программа получала текст от клиента-человека, она реагировала в зависимости от эмоционального состояния человека. Так что общение будет более эффективным и дружелюбным.

Выводы по главе 5

В этой главе были выявлены факторы, которые не привлекали внимания исследователей в процессе анализа текста и извлечения эмоций из него, и была предложена методология их анализа. Было отмечено, что эмоции и их начальная полярность, предложенные исследователями в RuWordnet и английском WordNet, недостаточны для глубокого анализа эмоций текста.

Для анализа текста и выявления связи между эмоциями требуется определить оба типа взаимосвязи, существующих между эмоциями. Внутренние отношения между эмоциями, которые являются зависимыми друг от друга, как объяснил Паррот в своей теории эмоций, и внешние отношения, которые являются отношениями между

эмоциями в определённом тексте. Логические модели облегчают процесс для понимания различной логической вероятности эмоций, извлечённых из текста.

Методология обеспечивает базовый интерфейс для взаимодействия между машиной и человеком-агентом с помощью системы динамического анализа текста в реальном времени. Исторически общение между оператором машины и человеком состоит из статического заранее определённого набора реакций, ответов на взаимодействие человек-машина. Предлагаемая методология обеспечивает механизм самообучения, которое происходит во время реального общения. Коммуникация должна совершенствоваться по мере того, как идёт дальнейшее обучение. Машина лучше изучает характер человека и взаимодействует более подходящим образом.

ГЛАВА 6. ОБОБЩЕНИЕ ЭМОЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК, ВЫДЕЛЕННЫХ ИЗ ТЕКСТА, СОЗДАННОГО ПРИ ОБЩЕНИИ ЧЕЛОВЕКА С МАШИНОЙ

В случае, если текст слишком велик, и в нем много эмоций, например, в книге, или в блоге из Интернета, посвященном теме актуальных событий, необходимо обобщить эмоции. На этом этапе эмоции, извлеченные из текста, обобщаются в минимальной форме для наблюдения за эмоциональной полярностью текста. Проведена проверка способа обобщения эмоций, в соответствующей форме на основе их конкретных характеристик, которые изучали такие ученые, как Плутчик и Паррот в своих работах. По их данным, существует определенное количество независимых эмоций. Эмоции, отличные от независимых, являются либо их продолжением, либо зависимыми от них. Сведения об отношениях между этими зависимыми и независимыми эмоциями были использованы для обобщения эмоций до их первоначальных или первичных эмоций к словам в соответствии с теорией Колеса эмоций или Теорией эмоций Плутчика, что показано на рисунке ниже.

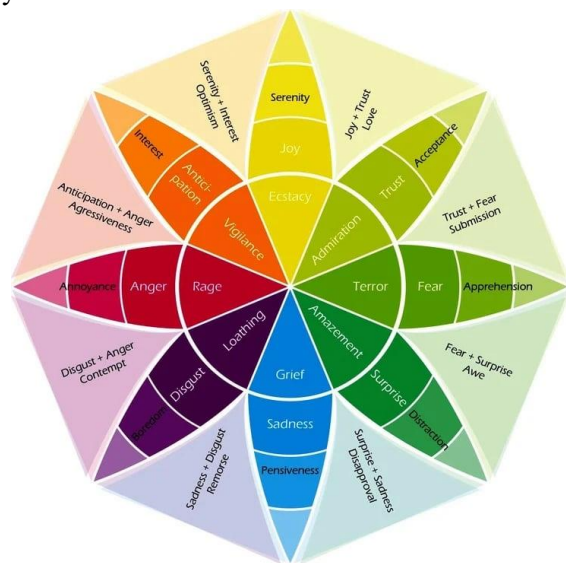


Рисунок 18 – Колесо эмоций или Теория эмоций Плутчика

На рисунке ниже показано Русскоязычное Колесо эмоций Плутчика.

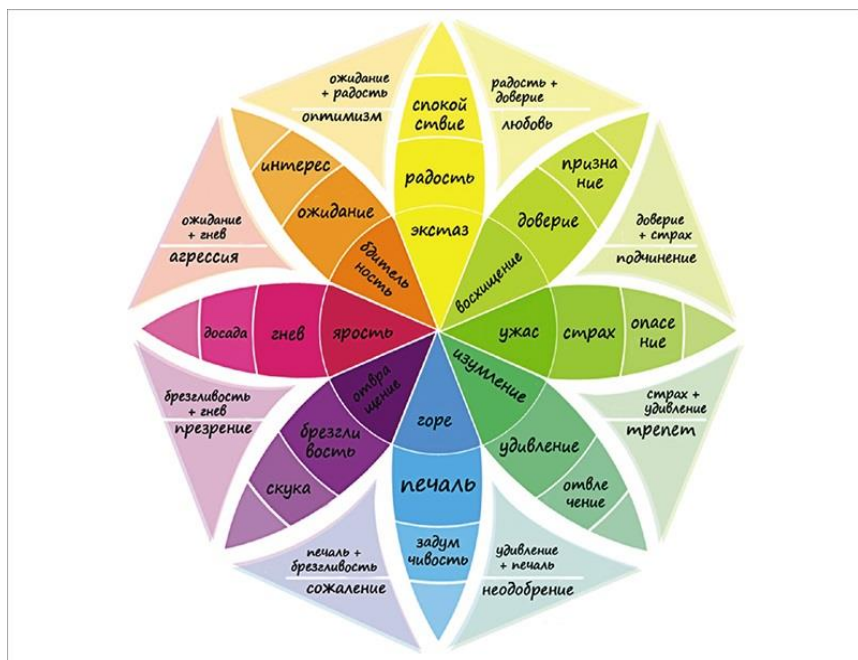


Рисунок 19 – Русскоязычное Колесо эмоций Плутчика

В соответствии с теорией Колеса эмоций Плутчика, интенсивность эмоций зависит от их места на колесе. Их максимум в центре колеса и он уменьшается по мере увеличения расстояния от центра колеса. Вес эмоций вблизи центра колеса равен 3, что также является максимальным уровнем интенсивности в колесе эмоций Плутчика. Для эмоций во втором слое колеса вес равен 2, и он постепенно уменьшается на 1 от слоя к слою колеса, когда эмоции находятся далеко от центра. Восемь независимых эмоций существуют вне колеса и представляют собой комбинацию двух эмоций уровня 2, имеющих вес 4, и их вес, как и вес эмоций последнего слоя, остается таким же, поскольку никакие эмоции от них не зависят. Обобщение эмоций, осуществляется с помощью алгоритма, представленного ниже.

$$\text{Количество эмоций } C_1 = \sum_{j=1}^n E_j$$

где количество эмоций C_1 равно сумме всех проявлений эмоций E_j).

$$\text{Интенсивность эмоций } IE_1 = C_1 \times W_{i1} + IE_d$$

где (IE_1 равно количеству эмоций C_1 , умноженных на вес эмоции в соответствии с колесом эмоций Плутчика W_{i1} плюс интенсивность зависимых эмоций IE_d).

Таким образом вычисляется интенсивность всех эмоций в тексте. Однако вес эмоции зависит от ее природы, является ли она зависимой эмоцией или независимой эмоцией. Согласно Плутчику, эмоции, которые лежат в центре, все единственные независимые эмоции, а все остальные являются зависимыми и являются продолжением эмоции, лежащей в слое перед ними и близко к центру. Таким образом, интенсивность эмоции равна количеству ее появлений, кратных ее весу, и интенсивности зависимых эмоций.

Процесс определения интенсивности эмоций начинается с самых крайних слов эмоций. Затем они добавляются к интенсивности эмоций внутреннего слоя, и это продолжается до тех пор, пока все эмоции не будут суммированы в основные эмоции, согласно колесу эмоций Плутчика. В случае таких языков, как русский, в которых существует более 32 эмоций, остальные эмоции будут иметь наименьший вес 1, а их интенсивность рассчитывается путем подсчета их появления, умноженного на их вес.

Пример результата после выявления и обобщения эмоций из текста с помощью теории эмоций Плутчика, подробно описан ниже в графиках.



Рисунок 20 – Обобщенные эмоции в соответствии с Колесом эмоций Плутчика

Некоторые эмоции во время обобщения исчезли из списка эмоций, поскольку они были зависимыми эмоциями, но они передали свою интенсивность эмоциям, от которых они зависели. По мере увеличения типа эмоций обобщение работает более эффективно.

6.1. Обобщенные исследования

В ходе описанного процесса лингвистический интерфейс при получении анализируемого текста сначала предварительно обрабатывает его, отфильтровывает ненужные или нерелевантные символы, слова, пробелы и т. д. Затем «отфильтрованный» текст анализируется с использованием алгоритма Emostemmer для выявления в нём эмоциональных компонентов.

Характеристики выявленных эмоций исследуются интерфейсом на основе координат их появления в тексте. Выявляется связь между эмоциями одинаковой полярности и эмоциями различной полярности, а также их связь с темой текста. Эти характеристики эмоций, наряду с психолингвистическими характеристиками, используются

для выявления особенностей поведения автора текста. Основываясь на выявленном поведении, интерфейс начинает работать соответствующим способом. Во время взаимодействия интерфейс предсказывает фразы, эмоции или слова, которые человек может использовать в общении. В конце взаимодействия все эмоции суммируются с помощью колеса эмоций Плутчика для использования в дальнейшем.

Система делает вывод об эмоциональной направленности текста. Он необходим для того, чтобы узнать эмоциональное состояние человека и подготовить ответ. Данные сведения используются для принятия решений об эмоциональной полярности текста, а также о психолингвистических характеристиках текста для подготовки ответа. Таким образом, используется мнение людей о принятии решения по поводу конкретной политики управления продуктом.

Обобщены результаты эксперимента в блогах с текстом отзывов об Удмуртском государственном университете.

Анализируемый текст отражает преимущественно положительные эмоции об Удмуртском государственном университете в соотношении 207 положительных к 79 отрицательных. Это означает, что около 25 % эмоций в адрес Удмуртского государственного университета отрицательные. В социальных блогах студенты были более открыты для представления своих идей, чем при использовании анкеты. На основе эксперимента были выявлены моменты, над которыми следует работать, чтобы улучшить мнение студентов (например, процесс выставления оценок студентам, поведение преподавателей по отношению к студентам, начисление стипендии и т. д.).

Математические модели и статистические инструменты, которые использовали для анализа текста, уже были объяснены выше. Для реализации лингвистического человеко-машинного интерфейса для исследования характеристик эмоций в тексте, было использовано контролируемое обучение вместе с основными методами кластеризации, такими как К-средние (K-Means), С-средние (Fuzzy C-Means), иерархическая кластеризация (Hierarchical clustering) и гауссовский метод (Gaussian).

6.2. Кластеризация как механизм классификации эмоций в группах

В лингвистическом интерфейсе кластеризация используется в качестве механизма для классификации эмоций в их соответствующих группах на основе их природы и категории, как это определено Плутчиком в его работе. Эмоции были разделены на основные группы с помощью колеса эмоций Плутчика, как уже обсуждалось. Для анализа работы текстового интерфейса и производительности алгоритмов часть текстов создавалась вручную.

Кластеризация также используется во второй главе этого исследования для классификации эмоций на две основные группы (положительные эмоции и отрицательные эмоции) после идентификации отрицаний в тексте.

Для классификации эмоций из текста, были проверены различные типы кластеризации для определения лучших из них. Для эксперимента был взят текст с эмоциями разных типов, где их количество было примерно одинаковым для определения производительности и возможности классификации каждого алгоритма. Работоспособность алгоритмов кластеризации определялась путём анализа их характеристик, таких как время выполнения, распределение эмоций в разных кластерах, стандартное отклонение кластеров. Стандартные отклонения их индивидуальных характеристик были дополнительно проанализированы.

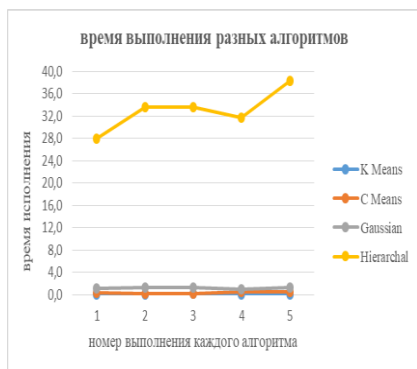
6.2.1. Время выполнения алгоритма кластеризации

Первое сравнение между различными алгоритмами кластеризации, используемое для классификации эмоций по различным группам, было проведено для измерения времени выполнения анализа текста и эмоций, как показано ниже (таблица 14).

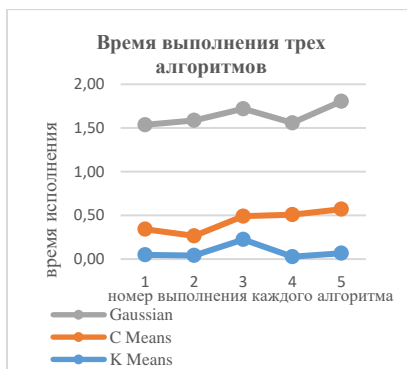
Время выполнения различных алгоритмов для анализа текста

Алгоритм для анализа эмоций	Количество кластеров	Итерация					Время в секундах
		1	2	3	4	5	
K-Means	5	0.0488	0.0409	0.224074	0.02770	0.0671	0.0817
Fuzzy C-Means		0.2927	0.2247	0.267274	0.48069	0.5033	0.3537
Gaussian		1.1948	1.321322	1.228191	1.0506	1.2354	1.2061
Hierarichal		27.908	33.6599	33.60246	31.6907	38.3762	33.0476

По таблице можно видеть эффективность алгоритмов кластеризации эмоций в разных кластерах. Было применено пять испытаний для объединения эмоций в восемь групп, а затем определялось среднее время выполнения каждого алгоритма кластеризации эмоций. Было отмечено, что иерархическая кластеризация требует максимального времени для кластеризации эмоций, тогда как алгоритму K-средние (K-Means) требуется минимальное количество времени. Результаты выполнения каждого алгоритма при кластеризации можно видеть на рисунках ниже.



Время выполнения четырёх алгоритмов анализа текста



Время выполнения трёх алгоритмов анализа текста

Рисунок 21 – Время выполнения алгоритмов анализа текста

График слева отображает время выполнения четырёх алгоритмов: К-средние (K-Means), С-средние (Fuzzy C-Means), Иерархическая кластеризация (Hierarchical clustering) и Гауссовский метод (Gaussian) в то время как график справа показывает сравнительное время выполнения трех алгоритмов: К-средние (K-Means), С-средние (Fuzzy C-Means) и Гауссовский метод (Gaussian).

Аналогичным образом наблюдается возможность кластеризации каждого из алгоритмов. В результате сравнения частот данных наблюдается, что кластеризация К-средние (K-Means), С-средние (Fuzzy C-Means) работает лучше всего, когда речь идет о распределении данных по различным кластерам. Видно, что они имеют высокую степень точности и ожидаемые результаты. Результаты можно видеть ниже (таблица 15).

Таблица 15

Процент данных в разных кластерах

Алгоритм для анализа	Процент данных в кластере							
	Кластер							
	1	2	3	4	5	6	7	8
K-Means	5.18	13.64	14.06	17.20	14.16	10.71	7.91	17.104
Fuzzy C-Means	14.96	9.98	18.24	12.33	7.60	14.09	17.58	5.183
Gaussian	18.62	11.74	23.60	7.63	6.59	9.46	10.71	11.61
Hierarchical	0.034	0.06	98.99	0.31	0.03	0.24	0.24	0.069

Из таблицы видно, что эмоции делятся на восемь кластеров по каждому алгоритму кластеризации. Возможность кластеризации K-Means и Fuzzy C-Means лучше, чем у других алгоритмов.

Аналогичным образом измеряется стандартное отклонение для каждого кластера каждого алгоритма. Это важная мера для выявления данных, показывающих необычное или ненормальное поведение из оставшихся данных. Данные с высоким значением стандартного отклонения считаются более ненормальными или необычными

по сравнению с данными с низкими значениями стандартного отклонения. Результаты приведены ниже (таблица 16).

Таблица 16

Стандартные отклонения в кластере

Алгоритм для анализа	Значения стандартного отклонения в кластере							
	Кластер							
	1	2	3	4	5	6	7	8
K-Means	0.53	0.25	0.21	0.20	0.22	0.30	0.35	0.19
Fuzzy C-Means	0.22	0.28	0.20	0.23	0.34	0.22	0.20	0.53
Gaussian	0.22	0.19	0.51	0.08	0.08	0.21	0.65	0.17
Hierarichal	0	0.03	1.52	0.11	0	0	0.03	0.02

Из приведённой таблицы, замечено, что K-Means Clustering – самый простой и эффективный способ кластеризации данных. Он так же хорошо группирует данные. Кластеризация нечётких C-Means всегда использует данные и представляет их как отдельный элемент, имеющий степень принадлежности к каждому кластеру. Это облегчает группирование данных в кластерах, показывающих их принадлежность к каждой группе. Кластеризация данных занимает немного больше времени, чем K-Means значение. Иерархический алгоритм кластеризации работает совершенно иначе, чем остальные алгоритмы кластеризации.

При проверке результатов было замечено, что кластеризация K-Means является лучшим выбором для разделения эмоций на разные кластеры. Он использовался в лингвистическом интерфейсе для обобщения и классификации эмоций в своих группах.

Выводы по главе 6

В данной главе был использован механизм обобщения эмоций, извлечённых из текста. Теория Плутчика использовалась для их

объединения в восемь основных групп. Для суммирования были проанализированы различные алгоритмы кластеризации и их производительность с использованием характеристик, таких как время выполнения, способность кластеризовать эмоции в разные группы и стандартное отклонение каждого кластера. Основываясь на результатах, K-Means был выбран в качестве алгоритма кластеризации для лингвистического человеко-машинного интерфейса.

ЗАКЛЮЧЕНИЕ

В ходе данного исследования была изучена существующая система выявления и анализа эмоций из текста и определены её недостатки. В связи с этим, был предложен новый алгоритм предварительной обработки текста перед его анализом, и результаты алгоритма сравнивались с существующими системами с использованием Т-теста. Алгоритм подготавливает текст, удаляя ненужные компоненты и исправив слова с ошибками.

Алгоритм Emostemmer был предложен для идентификации эмоций в тексте, написанном на русском языке. Точность алгоритма определения эмоций в тексте составляет 93,5 % по сравнению с существующим механизмом, таким как RuSentiLex, который имеет около 85,8 %. Точность рассчитывается с использованием Матрицы ошибок. Оба алгоритма выполняются в одной и той же системе, и работа Emostemmer для определения эмоций в тексте, занимает в среднем на 65 % меньше времени, чем RuSentilex. Также, он работает с гораздо меньшей базой эмоций по сравнению с RuSentiLex

В системе впервые на экспериментальной основе использовались психолингвистические характеристики текста, они изучались наряду с эмоциональной направленностью текста с помощью теории И.Б. Майера и К.Г. Юнга, которая делит людей на экстравертов и интровертов на основе слов, использовавшихся в письменном тексте. Это облегчает процесс ответа машины человеку в доступной форме во время их взаимодействия. Здесь машина опорных векторов применяется для определения различных психолингвистических характеристик текста, а оценка максимального правдоподобия используется для прогнозирования будущего текста на основе эмоционального и психолингвистического поведения конкретного человека.

Природа различных эмоций, а также зависимости между ними были изучены с использованием известных теорий, таких как колесо эмоций Плутчика и теория эмоций Паррота. Была рассчитана и обобщена полярность текста, основанная на эмоциях в нем.

В целом система обеспечивает хорошую платформу и направление для будущих исследований и создает машины, полностью способные независимо взаимодействовать с людьми, отвечать на их вопросы и решать их проблемы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Создан лингвистический интерфейс для взаимодействия человека с машиной. Данный интерфейс тестировался с русскоязычным текстом, полученные результаты были проанализированы.

2. Разработана модель использования эмоциональных и психолингвистических характеристик текста для улучшения взаимодействия между человеком и машиной. Характеристики были определены в тексте и протестированы с использованием предложенной модели.

3. Разработан механизм для предварительной обработки текста перед анализом для повышения эффективности конечного результата анализа. Механизм эффективно отфильтровывал ненужную составляющую текста, выполнял орфографическую коррекцию и выявлял настоящую полярность эмоций в тексте.

4. Создан алгоритм Emostemmer для идентификации эмоций, извлеченных из текста на русском языке. В сравнении со словарным методом Emostemmer эффективно идентифицирует эмоции с минимальным расходом технических ресурсов. Алгоритм имеет короткое время выполнения и более точно определяет эмоции в тексте.

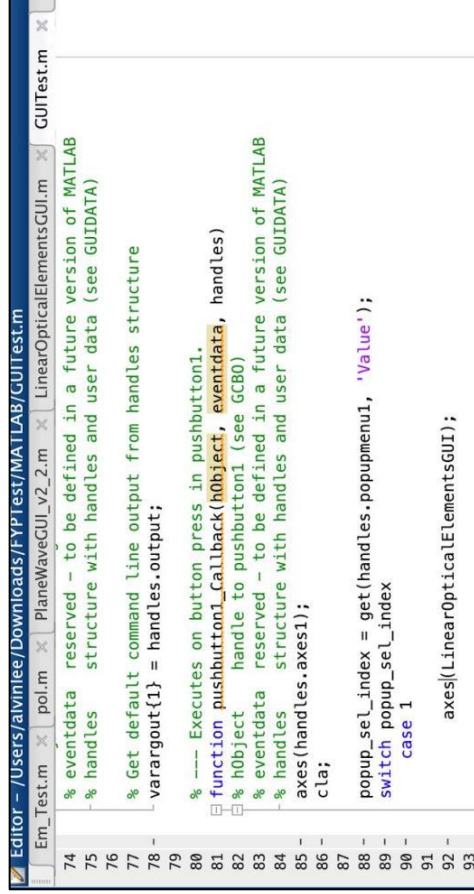
5. Разработан метод обобщения эмоций, выраженных в тексте, с помощью известных теорий эмоций: Колесо эмоций Плутчика и Теория эмоций Паррота. Метод, основанный на природе эмоций и взаимосвязи между ними, суммировал их в основную категорию эмоций.

ПРИЛОЖЕНИЯ

Приложение 1

Фрагмент кода информации о сборке

Лингвистический интерфейс подготовлен с использованием функций и компонентов языка программирования Matlab. Несколько рисунков лингвистического человеко-машинного интерфейса показаны ниже.



```
74 % eventdata reserved - to be defined in a future version of MATLAB
75 % handles structure with handles and user data (see GUIDATA)
76
77 % Get default command line output from handles structure
78 varargout{1} = handles.output;
79
80 % ---- Executes on button press in pushbutton1.
81 function pushbutton1_Callback(hObject, eventdata, handles)
82 % hObject handle to pushbutton1 (see GCBO)
83 % eventdata reserved - to be defined in a future version of MATLAB
84 % handles structure with handles and user data (see GUIDATA)
85 axes(handles.axes1);
86 cla;
87
88 popup_sel_index = get(handles.popupmenu1, 'Value');
89 switch popup_sel_index
90     case 1
91         axes(LInearOpticalElementsGUI);
92
93
```

Рисунок 22– Код GUI при запуске программы лингвистического человеко-машинного интерфейса

```
Editor - H:\Code\whitebalance_cuda.m
whitebalance.m x whitebalance_gpu.m x whitebalance_kernel.cu x whitebalance_cuda.m x
1 function adjustedImage = whitebalance_cuda(imageData)
2 % WHITEBALANCE forces the average image color to be gray
3
4 % Find the average values for each channel
5 - pageSize = size(imageData,1) * size(imageData,2);
6 - avg_rgb = mean( reshape(imageData, [pageSize,3]) );
7
8 % Find the average gray value and compute the scaling factor
9 - avg_all = mean(avg_rgb);
10 - factor = max(avg_all, 128) ./ avg_rgb;
11 - factor = reshape(factor,1,1,3);
12
13 % Load the kernel and set up threads
14 - kernel = parallel.gpu.CUDAKernel('whitebalanceKernel.ptxw64',...
15   'whitebalanceKernel.cu' );
16 - [nRows, nCols, ~] = size(imageData);
17 - blockSize = 256;
18 - kernel.ThreadBlockSize = [blockSize, 1, 3];
19 - kernel.GridSize = [ceil(nRows/blockSize), nCols];
20
```

Рисунок 23— Загрузка файла в текстовом формате или формате Excel в лингвистический интерфейс для предварительной обработки текста

```

1 % MATLAB test for XIMC library
2 % Tested R2014b 32-bit WinXP, R2014b 64-bit Win7
3
4 [~,maxArraySize]=computer;
5 is64bit = maxArraySize > 2^31;
6 if (is64bit)
7     disp('Using 64-bit version')
8     disp('NOTE! Copy ximc.h, libximc.dll, bindy.dll, xiwrapper.dll, wrappers/matlab/libximc_thunk_pcwin64
9
10     disp('Using 32-bit version')
11     disp('NOTE! Copy ximc.h, libximc.dll, bindy.dll and xiwrapper.dll to the current directory')
12     end
13
14 if not(libisloaded('libximc'))
15     disp('Loading library')
16     if (is64bit)
17         [notfound,warnings] = loadlibrary('libximc.dll', @ximcm);
18     else
19         [notfound, warnings] = loadlibrary('libximc.dll', 'ximcm.h', 'adheader', 'ximc.h');
20     end
21     end
22
23 device_names = ximc_enumerate_devices_wrap(0);
24 devices_count = size(device_names,2);

```

Рисунок 24 – Код Emostemmer для идентификации эмоций, извлеченных из текста

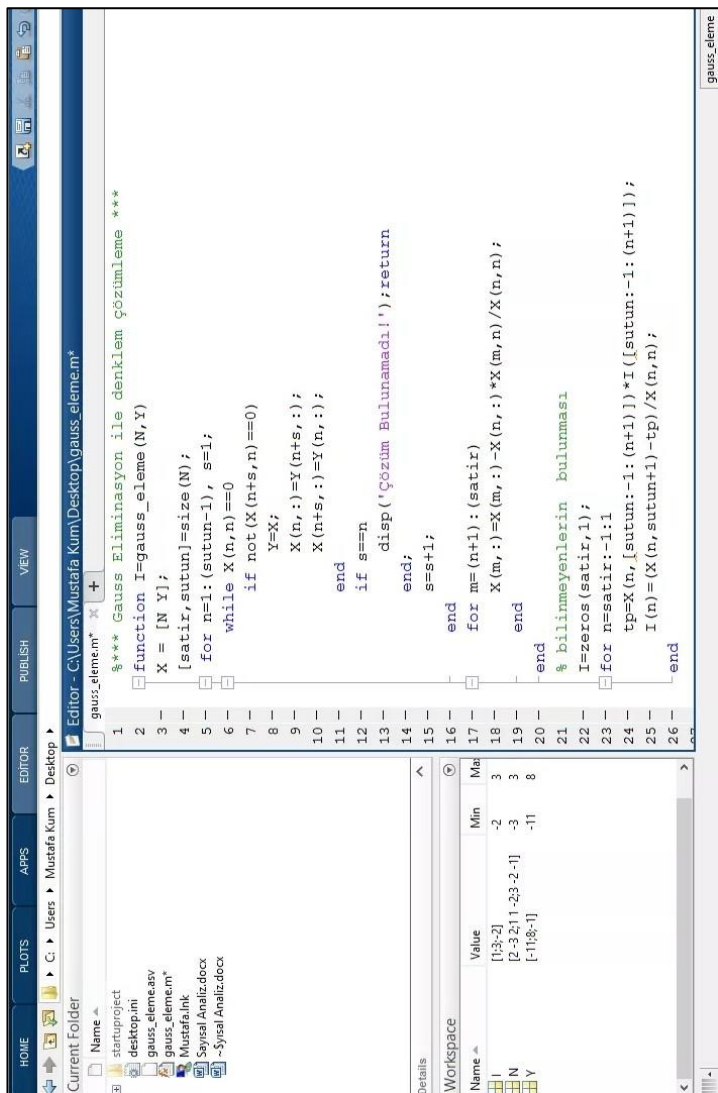


Рисунок 25 – Применение кластеризации для разделения эмоций на разные группы

```

1 % Hermite interpolation
2 % Find the approximate value of f(1.5) from
3 % (x,y,y') = (0,1,1), (1,e,e), (2,e^2,e^2) & (3,e^3,e^3).
4
5 n = input('Enter n for (n+1) nodes, n: ');
6 x = zeros(1,n+1);
7 q = zeros(2*n+2,2*n+2);
8
9 for i = 0:n
10     fprintf('Enter x(%d) and f(x(%d)), and f''(x(%d)) on separate lines:
11         x(i+1) = input(' ');
12         q(2*i+1,1) = input(' ');
13         q(2*i+2,2) = input(' ');
14     end
15
16     z = zeros(1,2*n+2);
17     for i = 0:n
18         z(2*i+1) = x(i+1);
19         z(2*i+2) = x(i+1);
20         q(2*i+2,1) = q(2*i+1,1);
21         if i ~= 0
22             q(2*i+1,2) = (q(2*i+1,1)-q(2*i,1))/(z(2*i+1)-z(2*i));
23         end
24     end
25
26     k = 2*n+1;
27     for i = 2:k

```

Рисунок 26 – Сравнения различных алгоритмов кластеризации и построения графика между ними


```
1 - clear all
2 - clc
3 - x = -1:0.005:-0.3;
4 - f = sin(x.^-2);
5 - g = sin(1.2*x.^-2);
6 - plot(x, f, 'm', 'LineWidth', 2)
7 - grid
8 - hold on
9 - plot(x, g, 'c', 'LineWidth', 2)
10 - xlabel('x')
11 - ylabel('f, g')
12 - title('ГРАФИКИ ФУНКЦИЙ f и g', 'FontName', 'Courier')
13 - legend('f(x)=sin(1/x^2)', 'g(x)=sin(1,2/x^2)', 0)
```

Рисунок 27 – Применение машинного обучения

Фрагмент программного кода

```

function varargout = start(varargin)
% START MATLAB code for start.fig
%   START, by itself, creates a new START or raises the existing
%   singleton*.
%
%   H = START returns the handle to a new START or the handle to
%   the existing singleton*.
%
%   START('CALLBACK',hObject,eventData,handles,...) calls the local
%   function named CALLBACK in START.M with the given input arguments.
%
%   START('Property','Value',...) creates a new START or raises the
%   existing singleton*. Starting from the left, property value pairs are
%   applied to the GUI before start_OpeningFcn gets called. An
%   unrecognized property name or invalid value makes property application
%   stop. All inputs are passed to start_OpeningFcn via varargin.
%
%   *See GUI Options on GUIDE's Tools menu. Choose "GUI allows only one
%   instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help start

% Last Modified by GUIDE v2.5 21-Jan-2019 20:15:56

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                  'gui_Singleton',  gui_Singleton, ...
                  'gui_OpeningFcn', @start_OpeningFcn, ...
                  'gui_OutputFcn',  @start_OutputFcn, ...
                  'gui_LayoutFcn',  [], ...
                  'gui_Callback',    []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
% End initialization code - DO NOT EDIT

% --- Executes just before start is made visible.
function start_OpeningFcn(hObject,eventdata,handles,varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
% varargin   command line arguments to start (see VARARGIN)

% Choose default command line output for start
handles.output = hObject;

% Update handles structure
guidata(hObject, handles);

% UIWAIT makes start wait for user response (see UIRESUME)
% uiwait(handles.figure1);

```

```

% --- Executes on button press in pushbutton2.
function pushbutton2_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton2 (see GCBO) for text file read
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

[fileName,pathName,filterIndex] = uigetfile('*.txt', 'Pick a text File')
if fileName ~= 0
    set(handles.edit3, 'String', strcat(pathName,fileName));

fid = fopen(fileName)
handles.fidd=fid
guidata(hObject,handles)

s=[];
    tline = fgetl(fid)
    mtext = tline

    Idx = regexp(mtext, '[^,!]')
    result= mtext(Idx)
    tline = result

    handles.fileNames=result
    guidata(hObject,handles)

    while ischar(mtext)
        s=[s;mtext]
        mtext = fgetl(fid)
    end

set(handles.edit2,'String',s);

end

% --- Executes on selection change in listbox2.
function listbox2_Callback(hObject, eventdata, handles)
% hObject    handle to listbox2 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: contents = cellstr(get(hObject,'String')) returns listbox2 contents as
cell array
%         contents{get(hObject,'Value')} returns selected item from listbox2

% --- Executes during object creation, after setting all properties.
function listbox2_CreateFcn(hObject, eventdata, handles)
% hObject    handle to listbox2 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: listbox controls usually have a white background on Windows.
%         See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUiControlBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
end

```

```

% --- Executes on button press in pushbutton4.
function pushbutton4_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton4 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

result=handles.filename5
tline=result

fid=handles.fidd

s={};
while ischar(tline)
    s=[s;tline]
    tline = fgetl(fid)
end

set(handles.edit4,'String',s);

% --- Executes on button press in pushbutton5.
function pushbutton5_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton5 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% fid= handles.fidd

fid = fopen('C:\Users\qmc\Documents\ProgrammingPart\newtext.txt');

str=textscan(fid,'%s')
text = str{:}
count=size(text,1)
set(handles.text1,'String',count);
% fclose(fid)

j=1
index={}

for k=1:numel(A)

    freq(k)=sum(ismember(text,A{k}))
    if (freq>0)
        index{j}=A{k}
        j=j+1
    end
end
index
set(handles.listbox1,'String',index);

    set(handles.listbox2,'String',num2str(freq));

%         set(handles.uitable1,'data',freq');
%         set(handles.uitable1,'data',str2double(index'));

```

```

% For Positive Emotions

handles.positivelist=PosEmo
guidata(hObject,handles)

% fid=fopen(fileName)
% str=textscan(fid,'%s')
% % fclose(fid)
% text=str(:)

for k=1:numel(PosEmo)
    Pfreq(k)=sum(ismember(text,PosEmo{k}))
end
ppofreq=sum(Pfreq) % For adding Number of Array
handles.ppofreq=ppofreq
guidata(hObject,handles)

PosFreq= nnz(Pfreq); % For non Zero elements of Array
set(handles.text4,'String',ppofreq)

% For Negative Emotions

handles.negativelist= NegEmo
guidata(hObject,handles)

% fid=fopen(fileName)
% str=textscan(fid,'%s')
% fclose(fid)
% text=str(:)

for k=1:numel(NegEmo)
    Nfreq(k)=sum(ismember(text,NegEmo{k}))
end
neggfreq=sum(Nfreq)
handles.negfre= neggfreq
guidata(hObject,handles)

NegFreq= nnz(Nfreq);
set(handles.text7,'String',neggfreq)

axes(handles.axes3);
% bar(ppofreq,neggfreq)
y=[ppofreq,neggfreq]

bar(1, ppofreq, 'g');
hold on
bar(2, neggfreq, 'r');
set(gca,'XTickLabel',{'Positive Emotions','Negative Emotions'});
ylabel('Number of Emotions');
% bar(y,'FaceColor',[0 .5 .5],'EdgeColor',[0 .9 .9],'LineWidth',1.5)
% bar(y,'FaceColor',[0 .5 .5],'EdgeColor',[0 .9 .9],'LineWidth',1.5)

set(gca,'XTickLabel',{'Positive Emotions','Negative Emotions'})
ylabel('Number of Emotions')

legend(legends);

```

```

matlabImage = imread('smile1.jpg');
image(matlabImage)
axis off
axis image

axes(handles.axes5)
matlabImage = imread('anger.jpg');
image(matlabImage)
axis off
axis image

axes(handles.axes6)
matlabImage = imread('textt.jpg');
image(matlabImage)
axis off
axis image

% --- Executes on button press in pushbutton6.
function pushbutton6_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton6 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

fid = fopen('C:\Users\qmc\Documents\ProgrammingPart\newtext.txt')

str=textscan(fid,'%s')
text = str{:}

B = {'he','het'}
% fid=fopen(fileName)
% str=textscan(fid,'%s')
% fclose(fid)
% text=str{:}

% text =result
j=1
index={}

for k=1:numel(B)

    freq(k)=sum(ismember(text,B{k}))

    if (freq>0)
        index[j]=B{k}
        j=j+1
    end
    index
end

c= ismember(text,'he')

idx1 = find(ismember(c, 1))
idx2=idx1+1

newStr = text(idx2)
handles.wordnegs=newStr
quidata(hObject, handles)

```

```

% --- Executes on button press in pushbutton7.
function pushbutton7_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton7 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

wordsnegations= handles.wordnegs

set(handles.listbox1,'String',wordsnegations);

%     set(handles.listbox2,'String',num2str(freq));

% --- Executes on button press in pushbutton8.
function pushbutton8_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton8 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

wordsnegations= handles.wordnegs
negadj=handles.negativeList
posadj=handles.positiveList

c= ismember(negadj,wordsnegations)
d= ismember(posadj,wordsnegations)

freqnn=handles.negfre
freqppo=handles.ppfreq

if(c==1)
freqnn=freqnn-1
freqppo=freqppo+1

else
freqppo=freqppo-1
freqnn=freqnn+1
end
set(handles.text4,'String',freqppo)
set(handles.text7,'String',freqnn)

% set(freqppo,'FaceColor','red');
% set(freqnn,'FaceColor','blue');
axes(handles.axes3);
% figure;

y=[freqppo,freqnn]

% bar(y,'FaceColor',[0 .5 .5],'EdgeColor',[0 .9 .9],'LineWidth',1.5)
% xlabel(handles.axes,'Radius')
% bar(y)
% bar(y[2],'FaceColor',r,'EdgeColor',[0 .9 .9],'LineWidth',1.5)
% axes(handles.axes3);
%
bar(1, freqppo, 'g');
hold on
bar(2, freqnn, 'r');
set(gca,'XTickLabel',{'Positive Emotions','Negative Emotions'});
ylabel('Number of Emotions');

```

```

tic
t=fileread('dictionary.txt')
dictt=string(t) % converting to string
dictt=splitlines(dictt) % split into different lines
dictt(1:5)
TF=(dictt== '')
dictt(TF)=[]
dictt(1:3)

p= [".", ",", "?", "!", ":", ";", ":", "&", "/"] % list of symbol to remove
dictt= replace(dictt,p, "") % replace them with empty place
dictt(1:3)

dictt=strip(dictt)
dictt(1:3)

dicWords=strings(0)
for i=1:length(dictt)
    dicWords=[dicWords;split(dictt(i))] %split into separate words
end

dicWords=lower(dicWords);

Tex=fileread('Opinionontext.txt')

texx=string(Tex) % converting to string
texx=splitlines(texx) % split into different lines
texx(1:5)
TF=(texx=="")
texx(TF)=[]
texx(1:3)

texxWords=strings(0)
for i=1:length(texx)
    texxWords=[texxWords;split(texx(i))] %split into separate words
end
texxWords(1:10)

texxWords=lower(texxWords);

k=1;
dicloc=""
loc=""
for i=1:length(texxWords)
    for j=1:length(dicWords)
        a=sum(isstrprop(dicWords(j), 'alpha'));
        tf=strncmpi(texxWords(i), dicWords(j), a);

        if (tf==1)
            loc(k)=texxWords(i);
            dicloc(k)=dicWords(j);

            k=k+1;
        else
            end
            j=j+1;
        end
        i=i+1;
    end
    loc'
    toc

```



```

textData = fileread('myopinion.txt');
textData = replace(textData, " ", "");
textData = split(textData, [newline newline])

startOfTextCharacter = compose("\x0002");
whitespaceCharacter = compose("\x00B7");
endOfTextCharacter = compose("\x2403");
newlineCharacter = compose("\x00B6");

textData = startOfTextCharacter + textData;
textData = replace(textData, [" " newline], [whitespaceCharacter
newlineCharacter]);
uniqueCharacters = unique([textData{:}]);
numUniqueCharacters = numel(uniqueCharacters);

numDocuments = numel(textData);

XTrain = cell(1,numDocuments);
YTrain = cell(1,numDocuments);

for i = 1:numel(textData)
    characters = textData{i};
    sequenceLength = numel(characters);

    % Get indices of characters.
    [~, idx] = ismember(characters, uniqueCharacters);

    % Convert characters to vectors.
    X = zeros(numUniqueCharacters, sequenceLength);
    for j = 1:sequenceLength
        X(idx(j), j) = 1;
    end

    % Create vector of categorical responses with end of text character.
    charactersShifted = [cellstr(characters(2:end))' endOfTextCharacter];
    Y = categorical(charactersShifted);
    XTrain{i} = X;
    YTrain{i} = Y
end
textData{1}
size(XTrain{1})
YTrain{1}

inputSize = size(XTrain{1},1);
numHiddenUnits = 200;
numClasses = numel(categories([YTrain{:}]));

layers = [
    sequenceInputLayer(inputSize)
    lstmLayer(numHiddenUnits, 'OutputMode', 'sequence')
    fullyConnectedLayer(numClasses)
    softmaxLayer
    classificationLayer]
options = trainingOptions('adam', ...
    'MaxEpochs', 500, ...
    'InitialLearnRate', 0.01, ...
    'GradientThreshold', 2, ...
    'MiniBatchSize', 77, ...
    'Shuffle', 'every-epoch', ...
    'Plots', 'training-progress', ...
    'Verbose', false);
net = trainNetwork(XTrain, YTrain, layers, options);

```

```

function generatedText =
generateText(net,uniqueCharacters,startOfTextCharacter,newlineCharacter,whitespaceCharacter,endOfTextCharacter)

numUniqueCharacters = numel(uniqueCharacters);

X = zeros(numUniqueCharacters,1);

idx = strfind(uniqueCharacters,startOfTextCharacter);
X(idx) = 1;

generatedText = "";
vocabulary = string(net.Layers(end).Classes);

maxLength = 500;

while strlength(generatedText) < maxLength
    % Predict the next character scores.
    [net,characterScores] =
predictAndUpdateState(net,X,'ExecutionEnvironment','cpu');

    % Sample the next character.
    newCharacter = datasample(vocabulary,1,'Weights',characterScores);

    % Stop predicting at the end of text.
    if newCharacter == endOfTextCharacter
        break
    end

    % Add the character to the generated text.
    generatedText = generatedText + newCharacter;

    % Create a new vector for the next input.
    X(:) = 0;
    idx = strfind(uniqueCharacters,newCharacter);
    X(idx) = 1;
End

generatedText = replace(generatedText,[newlineCharacter
whitespaceCharacter],[newline " "]);

end

```

```

B1= K;
B2= C;
B3= G;
B4= H;

for i=1:6
    D(i) = B1(i)/sum(B1);
end
M = mean(D)
V = var(D)

pdf_D= pdf('norm',-1:0.01:1,M,V)
plot(pdf_D,'r')

hold on;
for i=1:6
    D(i) = B2(i)/sum(B2);
end
M = mean(D)
V = var(D)

pdf_D= pdf('norm',-1:0.01:1,M,V)
plot(pdf_D)

hold on;
for i=1:6
    D(i) = B3(i)/sum(B3);
end
M = mean(D)
V = var(D)

pdf_D= pdf('norm',-1:0.01:1,M,V)
plot(pdf_D,'--c')

hold on;
for i=1:6
    D(i) = B4(i)/sum(B4);
End

M = mean(D)
V = var(D)

pdf_D= pdf('norm',-1:0.01:1,M,V)
plot(pdf_D,'-g')
hold off;

legend('k means','fuzzy c means','Gaussian','Hirerchal')

```

СПИСОК ЛИТЕРАТУРЫ

1. Abbasi M.M, Beltiukov A.P, Hussain Lal, Abbasi A.Q. Spam Detection In Short Text Messages (Sms) Using Word Embedding And Term Frequency – Inverse Document Frequency (Tf-Idf) // XXI century: Resumes of the Past and Challenges of the Present plus. – 2020. – V. 9. № 2 (50). Pp. 143–148.

2. Аббаси М.М., Бельтюков А.П. Изучение доступных систем для анализа эмоций, извлеченных из текста, и обеспечения механизма для улучшения взаимодействия человека с машиной // Интеллектуальные системы в производстве. – ИжГТУ. – Том 17, № 4 (2019) [Http://Dx.Doi.Org/10.22213/2410-9304-2019-4-53-62](http://Dx.Doi.Org/10.22213/2410-9304-2019-4-53-62).

3. Beltiukov A.P., Abbasi M.M. Logical analysis of Emotions in Text from Natural language // Vestnik Udmurtskogo Universiteta. Matematika. Mekhanika. Komp'yuternye Nauki. – Izhevsk, 2019. – No 1 (29). – С. 106–116.

4. Abbasi M.M, Abbasi A.Q, Beltiukov A.P, Hussain Lal. Conflict matrix as a mechanism of identifying the conflict in emotions of written text // International Journal of Engineering & Technology, UAE. – 2020. – No. 2 (9). – С. 541–545.

5. Аббаси М.М., Бельтюков А.П., Хусейн Лалл., Аббаси А.К. Анализ эмоций из текстов для управления обществом // Инфокоммуникационные технологии // Академия телекоммуникаций и информатики. – Самара. – 2019.- № 2 (17) – С. 246–254.

6. Аббаси М.М. Анализ влияния отрицаний на текст с использованием кластеризации и колеса эмоций Плутчика // Научно -технический вестник Поволжья. – Казань. – 2019. – № 12 (66) – С. 23–28.

7. Аббаси М.М., Бельтюков А.П. Анализ настроений: полидисциплинарная область исследований // Конвергенция в сфере научной деятельности: проблемы, возможности, перспективы: материалы Всерос. науч. конф., 11 дек. 2018 г. / М-во науки и высш. образования РФ, М-во образования и науки УР, УРНИД, Всерос. науч. конф.

(декабрь, 2018); отв. ред. А.М. Макаров. - Ижевск: Удмуртский университет, 2018. – Электрон. ресурс. Лицензион. договор № 28лб от 14.01.2019 (Интернет). – С. 377–338.

8. Abbasi M.M., Beltiukov A.P. Analyzing emotions from text corpus using word space CSIT`2018 // Proceedings of the 20th International Workshop on Computer Science and Information Technologies, Varna-Bulgaria, 2018. – С. 90–94 // Industry 4.0. – 2018. – 3 (4). – С. 161–164.

9. Abbasi M.M., Beltiukov A.P. Identifying the strength of emotions in relation with the topic of text using Word space // Proceedings of the 21th inter-national workshop on Computer Science and Information Technologies, Austria, Vienna, 2019. Journal of Atlantis Highlights in Computer Sciences, 3. – С. 1–5.

10. Abbasi M.M, Beltiukov A.P. Analyzing the Complex Nature of Emotions from Written Text Using Parrott Theory of Emotions // 8-ая международная научная конференция Информационные технологии и системы. г. Ханты-Мансийск, Россия, 17–21 марта 2020 г.

11. Аббаси М.М. Инструменты Эмоционального анализа текстов в процессе управления и моделирования в сложных системах // 10 International Conference on Complex Systems: Control and Modeling problems (XX Международная конференция Проблемы управления и моделирования в сложных системах (ПУМСС),) Самара, 2018. September 3–6.– С. 236–242.

12. Аббаси М.М., Бельтюков А.П. Механизм предварительной обработки текста перед анализом настроений // 6th All Russian Conference on In-formation technology for intelligent decision making support (ITIDS), Ufa, Russian Federation, May 28–31, 2018.

13. Аббаси М.М., Бельтюков А.П. Информационные технологии изменения стилистики текста для его эмоциональной модификации // Материалы 7-ой международной научной конференции Информационные технологии и системы. г. Ханты-Мансийск, Россия, 12–16 марта 2019 г. – С. 137–142.

14. Abbasi M.M., Beltiukov A.P. Summarizing emotions from text using Plutchik wheel of emotion // 7th All Russian Conference on

Information technology for intelligent decision making support (ITIDS), Ufa, Russian Federation, May 28–30, 2019. – С. 291–294.

15. Аббаси М.М., Бельтюков А.П. Изучение эмоций по текстам, написанным людьми в Удмуртской Республике // Итоговая студенческая научная конференция (46; Апрель, 2018). XLVI итоговая студенческая научная конференция Удмуртского государственного университета: материалы всерос. конф. (апрель 2018 г.) Итоговая студенческая научная конференция (46; Апрель, 2018), М-во образования и науки РФ, ФГБОУ ВО «Удмуртский государственный университет»; отв. ред. А.М. Макаров. – Ижевск: Удмуртский университет, 2018. – Лицензион. договор №787лб от 12.11.2018(Интернет). – С. 26–28.

16. Аббаси М.М., Бельтюков А.П. Взаимосвязь между информатикой и языком для анализа текста // Relationship between informatics and linguistic for text analysis // Итоговая студенческая научная конференция (45; Апрель, 2017) XLV итоговая студенческая научная конференция: материалы конф. (апр. 2017 г.). Итоговая студенческая научная конференция (45; Апрель, 2017), М-во образования и науки РФ, ФГБОУ ВО «Удмуртский государственный университет»; отв. ред. А.М. Макаров. – Ижевск: Удмуртский университет, 2017. – Лицензион. договор № 236лб от 20.04.2018 (Интернет). – С. 13–14.

17. Аббаси М.М., Бельтюков А.П. // Цифровое развитие экономики и социальной сферы Удмуртской Республики // Цифровая экономика: сфера, которая изменит будущее человечества: актуальные вопросы и роль высшей школы: материалы Всерос. научно-практ. конф. 15 ноября 2019 г. М-во науки и высш. образования РФ, ФГБОУ ВО «Удмуртский государственный университет», М-во информатизации и связи УР, М-во образования и науки УР, Нац. центр цифровой экономики МГУ им. М.В. Ломоносова, Цифровое развитие экономики и соц. сферы УР: актуальные вопросы и роль высш. шк., Всерос. науч-практ. конф. ; отв. ред. А. М. Макаров. – Ижевск: Удмуртский университет, 2019. – Лицензион. договор № 843лб от 09.12.2019 (Интернет). – С. 3–6.

18. Abbasi M.M., Beltiukov A.P. Analysis of sentiment and emotion from text written in Russian language // 5th All Russian Conference on Information technology for intelligent decision making support (ITIDS), Ufa, Russian Federation, May 16–19, 2017. 1 (1). – C. 42–47. <https://www.elibrary.ru/item.asp?id=30520199>.
19. Stone P.J., Dunphy D.C., Smith M.S. The General Inquirer: A Computer Approach to Content Analysis / MIT Press – Cambridge, 1966. – P. 519.
20. Wiebe, Janyce M. Identifying Subjectivity characters in Narrative // Proc. 13th International Conference on Computational Linguistics. Helsinki, 1990. – P. 401–408.
21. Vasileios H., Kathleen R. M. (1997). Predicting the Semantic Orientation of Adjectives // Proc. 8th Conference on European chapter of the Association for Computational Linguistics. Spain, pp 174-181.
22. Peter D.T. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews // Proc. of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, pp. 417–424.
23. Pang B., Lee L. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques // Proc. Conference on Empirical Methods in Natural Language Processing. Philadelphia, pp. 79–86.
24. Denecke K. (2009). Are SentiWordNet scores suited for multi-domain sentiment classification? // Proc. 4th International Conference on Digital Information Management. USA, pp. 33–38.
25. Gupta, Vishal, Lehal, Gurpreet. (2009). A Survey of Text Mining Techniques and Applications // Journal of Emerging Technologies in Web Intelligence, 1 (1) C. 60–76.
26. Boley D., Gini M., Gross R. (1999) Partition based clustering for web document categorization // Elsevier Journal for Decision Support Systems, 27 (3). C. 329–341.
27. Emiliano L., Dominique L., Eunat M. (2014). A logical analysis of responsibility attribution: emotions, individuals and collectives // Journal of Logic and Computer. 24 (1). C. 1313–1339.

28. Ermakov A. (2009) . Knowledge extraction from text and its processing: Current state and prospects // Proc. of the Computational Linguistics and Intellectual Technologies. 2009, pp. 50–55.

29. Pivovarova L., Yangarber R. (2013). Adapting the PULS event extraction framework to analyze Russian text // Proc. of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing. Sofia, Bulgaria, 8–9, pp. 100–109.

30. Zagibalov T., Belyatskaya K. et al. (2010). Comparable English-Russian Book Review Corpora for Sentiment Analysis. Russia // Proc. of the 1st Workshop on Computational Approaches to Subjective and Sentiment Analysis, Lisbon, Portugal pp. 67–72.

31. Steinberger J., Lenkova P., Kabadjov M. (2011). Multilingual Entity-Centered Sentiment Analysis Evaluated by Parallel Corpora // Proc. of Recent Advances in Natural Language Processing. Bulgaria, pp. 770–775.

32. Chetviorkin I., Braslavskiy P., Loukachevich N. (2012). Sentiment Analysis Track at ROMIP 2011 // Computational Linguistics and Intellectual Technologies. Proc. of the International Conference (Dialog). Bekasovo, pp. 1–14.

33. Sokolova M., Bobicev V. (2009). Classification of emotion words in Russian and Romanian languages // International Conference RANLP. Borovets, Bulgaria, pp. 416–420.

34. Marina Sokolova, Victoria Bobicev (2009). Classification of Emotion Words in Russian and Romanian Languages // International Conference RANLP 2009. – Borovets, Bulgaria, pp. 416–420.

35. Богданова Л.И. Оценочные смыслы в русской грамматике(на материале глаголов эмоционального отношения) // Russian Journal of Linguistics. – 2018. – № 22 (4). – С. 844–873.

36. Jaruskululchi C., Kruengkrai. (2003). Generic text summarization using local and global properties of sentences // IEEE/WIC International conference on web intelligence. pp. 13–16.

37. Россеева О.И., Загорулько Ю.А. Организация эффективного поиска на основе онтологий // Труды международного семинара

Диалог'2001 по компьютерной лингвистике и её приложениям. – Т.2. – Аксаково, 2001. с. 333–342.

38. Ахмадеева И.Р., Загорулько Ю.А., Саломатина Н.В., Серый А.С., Сидорова Е.А., Шестаков В.К. Подход к формированию тематических коллекций текстов на основе интернет-ресурсов // Вестник НГУ. Серия: Информационные технологии. – 2013. – Т. 11, вып. 4. С. 5–15.

39. Непейвода Н.Н. Семантика алгоритмических языков. Итоги науки и техн. Сер. Теор. вероятн. Мат. стат. Теор. кибернет., 20, ВИНТИ, М., 1983, 95–166; J. Soviet Math., 25:6 (1984), 1558–1606.

40. Н. Н. Непейвода. О некоторых возможностях локальных вычислений в теории систем и базах данных. Программные системы: теория и приложения, 7:4 (2016), 145–160.

41. Мокроусов М.Н., Чиркова Н.Н. Извлечение данных из коммерческих веб-форумов // Интеллектуальные системы в производстве. – Том 14, № 1. – С. 659–673.

42. Мокроусов М.Н., Автоматизация нечеткого поиска в задаче распознавания старопечатных кириллических текстов // V международная конференция и молодежная школа «Информационные технологии и нанотехнологии». Самара, 21–24 мая 2019 года.

43. Gil R. Sentiment Analysis in Power BI – Part 2. [Электронный ресурс]. – URL: <https://datachant.com/2016/08/09/sentiment-analysis-power-bi-part-2/> (дата обращения: 10.02.2019).

44. Tableau Desktop: Start your free 14-Day trial [Электронный ресурс]. – URL: <https://www.tableau.com/products/desktop/download#system-requirements> (дата обращения: 05.04.2019).

45. Q emotionapp login. [Электронный ресурс]. – URL: <https://twitter.com/qemotionapp> (дата обращения: 11.04.2019).

46. MoodPatrol login. [Электронный ресурс]. – URL: <https://twitter.com/hashtag/MoodPatrol?src=hash> (дата обращения: 19.04.2019).

47. DeLone W.H., McLean E.R. (1992). Information systems success: The quest for the dependent variable // Information Systems Research, 3(1) С. 60–95.

48. Rai A., Lang S. S., Welker R B. (2002). Assessing the Validity of IS Success Models: An Empirical Test and Theoretical Analysis // *Information Systems Research*. 13(1) C. 50–69.
49. Bryd T.R., Thrasher E.H., Lang T., Davidson N.W. (2006). A process-oriented perspective of IS success: Examining the impact of IS on operational cost. *Omega*, 34 (1). Pp. 448-460.
50. Wu J H., Wang Y M. (2006). Measuring KMS success: A re-specification of the DeLone and McLean’s model // *Information & Management*, 43(1). Pp. 728–739.
51. Lai J.Y., Yang C.C., Tang W.S. (2008). Exploring the Effects of Dependability on Enterprise Applications Success in e-Business // *SIG-MIS-CPR’06* . Claremont, California, USA. Pp. 244–252. DOI: 10.1145/1125170.1125229
52. Jan Rijsbergen, C.J & Robertson, S.E. (Stephen Edward), 1946 & Porter, M.F. (Martin F.) (1980). *New models in probabilistic information retrieval*. British Library Research and Development Dept.], [London] British Library. Research and Development Department. British Library research & development reports; no. 5587.
53. Porter M.F (1980). An algorithm for suffix stripping // *Emerald Publishing*, 14 (3) C. 130-137. DOI:10.1108/00330330610681286
54. Krovetz Robert (2000). Viewing morphology as an inference process // *Artificial Intelligence Journal*, Q1 SJR 1.01. 118(2) C. 277–294 [https://doi.org/10.1016/S0004-3702\(99\)00101-0](https://doi.org/10.1016/S0004-3702(99)00101-0)
55. Paice Chris D (1990). Another Stemmer // *ACM SIGIR Forum*, 24 (3) Pp. 56–61. DOI:10.1145/101306.101310
56. Kiabod M., Naderi M., Sharafi S.M. (2012). A novel method of significant words identification in text summarization // *Journal of Emerging Technologies in Web Intelligence*, 4 (3).
57. William B. Frakes , Christopher J. Fox (2003). Strength and similarity of affix removal stemming algorithms // *ACM SIGIR Forum*, 37(1) Pp. 26–30. DOI:10.1145/945546.945548.
58. Michela Bacchin , Nicola Ferro , Massimo Me lucci (2005). A probabilistic model for stemmer generation // *Information Processing and Management* , 41(1). doi: 10.1016/j.ipm.2004.04.006.

59. Wiebe, J., T. Wilson, C. Cardie (2005). Annotating expressions of opinions and emotions in language // *Language Resources and Evaluation*, 39 (2) C. 165–210. <https://link.springer.com/article/10.1007/s10579-005-7880-9>.

60. Peng F., Ahmed N., Li X., Lu Y (2007). Context sensitive stemming for web search // *In Proc. of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* SJR Conference, pp. 639–646, 2007. <https://dl.acm.org/citation.cfm?id=1277851>

61. Prasenjit M., Mandar M., Swapan K., Gobinda K., Pabitra M., Kalyankumar D. (2007). YASS: Yet another suffix stripper // *ACM Transactions on Information Systems*. 25 (4). Article No. 18. <https://dl.acm.org/citation.cfm?doid=1281485.1281489>

62. Giorgos A., Konstantinos A., Christos B., Vassilis P (2010). An efficient mechanism for stemming and tagging: the case of Greek language // *In Proc. of the 14th international conference on Knowledge-based and intelligent information and engineering systems: Part III* , pp. 389–397, September 8-10, 2010, UK <https://dl.acm.org/citation.cfm?id=1885501>

63. Feinerer I (2010). Analysis and Algorithms for Stemming Inversion. In: Cheng PJ., Kan MY., Lam W., Nakov P. (eds) *Information Retrieval Technology. AIRS 2010 // Lecture Notes in Computer Science*, vol. 6458. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-17187-1_28

64. Jiaul H. Paik , Mandar Mitra , Swapan K. Parui , Kalervo Järvelin (2011). GRAS: An effective and efficient stemming algorithm for information retrieval // *ACM Transactions on Information Systems (TOIS)* 29 (4) C. 1-24, December 2011. doi: 10.1145/2037661.2037664.

65. Fernández A., Díaz J., Gutiérrez Y., Muñoz R (2011). An Unsupervised Method to Improve Spanish Stemmer // In: Muñoz R., Montoyo A., Métails E. (eds) *Natural Language Processing and Information Systems. NLDB 2011. Lecture Notes in Computer Science*, vol. 6716. Springer, Berlin, Heidelberg https://doi.org/10.1007/978-3-642-22327-3_24

66. Cherif W., Madani A., Kissi M (2014) . Building a syntactic rules-based stemmer to improve search effectiveness for Arabic language // 9th International Conference on Intelligent Systems: Theories and Applications (SITA-14), pp. 1–6. doi: 10.1109/SITA.2014.6847295. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6847295&isnumber=6846554>

67. Danilova V., Alexandrov M., Blanco X (2014). A Survey of Multilingual Event Extraction from Text. In: Métails E., Roche M., Teisseire M. (eds) // Natural Language Processing and Information Systems. NLDB. Lecture Notes in Computer Science, vol. 8455. Springer, Cham. https://doi.org/10.1007/978-3-319-07983-7_13

68. Moral C., de Antonio A., Imbert R., Ramírez J (2014). A survey of stemming algorithms in information retrieval // Information Research ISSN – 1368–1613 19 (1) P. 605. [Available at <http://www.informationr.net/ir/19-1/paper605.html#.XezgY-gzbIU>

69. Loukachevitch, N.V., Chetviorkin I.I (2014). Open evaluation of sentiment-analysis systems based on the material of the Russian language // Scientific and Technical Information Processing, 41 (6) Pp. 370–76. <https://doi.org/10.3103/S0147688214060057>

70. Gadri S., Moussaoui A. (2015). Information retrieval: A new multilingual stemmer based on a statistical approach // 2015 3rd International Conference on Control, Engineering & Information Technology (CEIT), Tlemcen, Algeria, 2015, pp.1–6. doi: 10.1109/CEIT.2015.7233113. <https://ieeexplore.ieee.org/document/7233113>

71. Lewis, C. I (1912). Implication and the Algebra of Logic // Mind, Oxford University Press, 21 (1). Pp. 522–531. <https://doi.org/10.1093/mind/XXI.84.522>

72. Marcus, Barcan C (1946). A Functional Calculus of First Order Based on Strict Implication // Journal of Symbolic Logic 11 (1) Pp. 1–16. <https://b-ok.xyz/ireader/1324986>

73. Georgette L (1977). Specificity and the interpretation of quantifiers // Linguistics and Philosophy, 1 (2) Pp. 233–245. <https://doi.org/10.1007/BF00351105>.

74. McKinsey J C., Tarski A., (1948). Some theorems about the sentential calculi of Lewis and Heyting // *Journal of Symbolic Logic*, 13 (3) Pp. 1–15. <https://doi.org/10.2307/2267873>

75. Büchi J., Bjarni J., Alfred T (1953). Boolean algebras with operators // *American Journal of Mathematics*, vol. 73 (1951), pp. 891–939, and vol. 74 (1952), pp. 127–162. *Journal of Symbolic Logic*, 18 (1). Pp. 70–71. doi: 10.2307 / 2266339

76. Kripke S. A (1959). A Completeness Theorem in Modal Logic // *Journal of Symbolic Logic*, 24 (1), pp. 1–4. <https://projecteuclid.org/euclid.jsl/1183733464>

77. Kripke S.A. (1962). The undecidability of monadic modal quantification theory // *Mathematical Logic Quarterly*, 8(2). Pp. 113–116. <https://doi.org/10.1002/malq.19620080204>

78. Kaplan D., Kripke S A (1966). Semantical analysis of modal logic I. Normal modal propositional calculi // *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, vol. 9 (1963), pp. 67–96. *Journal of Symbolic Logic*, 31(1). Pp. 120–122. doi:10.2307/2270649

79. Makinson D., Kripke Saul A (2014). Semantical analysis of modal logic II. Non-normal modal propositional calculi // *The theory of models*, In Proc. of the 1963 International Symposium at Berkeley, edited by Addison J. W., Henkin Leon, and Tarski Alfred, *Studies in logic and the foundations of mathematics*, North-Holland Publishing Company, Amsterdam 1965, pp. 206–220. Routley R. and Montgomery H.. The inadequacy of Kripke's semantical analysis of D2 and D3. *The Journal of Symbolic Logic*, vol. 33 (1968), pp. 568. *The Journal of Symbolic Logic*. 35. 135. DOI: 10.1017/S0022481200092434

80. Jaakko H., Cogito E (1963). Sum as an Inference and a Performance // *The Philosophical Review*, 72(4). Pp. 487–496. Doi: 10.2307/2183033

81. Bull R., Segerberg K (1984). Basic Modal Logic // In: Gabbay D., Guentner F. (eds) *Handbook of Philosophical Logic*. Syntheses Library (Studies in Epistemology, Logic, Methodology, and Philosophy of Science), vol. 165. Springer. https://doi.org/10.1007/978-94-009-6259-0_1

82. Boolos G (1993). The Logic of Provability // *Journal of Symbolic Logic*. Cambridge: Cambridge University Press, New York, 60 (4) Pp. 85–91. <https://doi.org/10.2307/2275893>

83. Carnielli W A., Pizzi C. (2009). Modalities and Multimodalities // *Logic, Epistemology, and the Unity of Science*. Springer, Berlin. 12 (1) Pp. 312–318. <https://doi.org/10.1007/978-1-4020-2808-3>

84. Lellmann B., Pattinson D. (2012). Sequent Systems for Lewis' Conditional Logics // In: del Cerro L.F., Herzig A., Mengin J. (eds) *Logics in Artificial Intelligence. JELIA 2012. Lecture Notes in Computer Science*, vol. 7519. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33353-8_25

85. Tishkovsky D., Schmidt R.A., Khodadadi M. (2012). The Tableau Prover Gen-erator MetTeL2 // In: del Cerro L.F., Herzig A., Mengin J. (eds) *Logics in Artificial Intelligence. JELIA 2012 Volume 7519*, pp. 492–495. *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33353-8_41

86. Lyaletski A (2015). Evidence Algorithm and Search in First-Order Logics // *Journal of Automated Reasoning*. 55(3). Pp. 269–284. <https://doi.org/10.1007/s10817-015-9346-0>

87. Jesse A., Heskens T., Kühlwein D., Tsvitshivadze E., Urban J. (2014). Premise Selection for Mathematics by Corpus Analysis and Kernel Methods // *Journal of Automated Reasoning*, 52 (2). Pp. 191–213. <https://link.springer.com/article/10.1007/s10817-013-9286-5> Q1, SJR 0.65

88. Orlandelli E. (2014). Proof Analysis in Deontic Logics // In: Cariani F., Grossi D., Meheus J., Parent X. (eds) *Deontic Logic and Normative Systems* pp. 139–14,. *DEON 2014. Lecture Notes in Computer Science*, vol 8554. Springer, Cham. Doi https://doi.org/10.1007/978-3-319-08615-6_11

89. Lellmann B., Pimentel E. (2015). Proof Search in Nested Sequent Calculi // In: Davis M., Fehnker A., McIver A., Voronkov A. (eds) *Logic for Programming, Artificial Intelligence, and Reasoning. LPAR. Lecture Notes in Computer Science*, vol. 9450. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-48899-7_39

90. Plutchik R (1980). The Nature of Emotions. Archived from the original on July 16, 2020.

91. Parrott R (2001). Emotions in Social Psychology, Key Readings in Social Psychology. Philadelphia: Psychology Press. ISBN 978-0863776830.

92. Лукашевич Н.В., Левчик А.В. Создание лексикона оценочных слов русского языка РуСентилекс // Труды конференции OS-TIS. – 2016. С. 377–382.

93. Loukachevitch N., Levchik A. Creating a General Russian Sentiment Lexicon. In Proceedings of Language Resources and Evaluation Conference LREC-2016.

94. Статья «Список чувств и эмоций» // Блог психолога Петра Зарубина из г. Новосибирска. – URL: <https://peter-zarubin.ru/spisok-chuvstv-i-emotsij/>

Оглавление

ВВЕДЕНИЕ	3
ГЛАВА 1. ЗНАЧЕНИЕ И ВАЖНОСТЬ ЭМОЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК ТЕКСТА	6
1.1. Валентно-энергетическое пространство в анализе эмоций, извлечённых из текста	6
1.2. Важность анализа эмоций, извлечённых из текста.....	7
1.2.1. В бизнес-аналитике	7
1.2.2. В управлении репутацией бренда	7
1.2.3. В маркетинговых исследованиях	7
1.2.4. В политическом анализе	8
1.2.5. В анализе отзывов клиентов	8
1.2.6. В прогнозах фондового рынка.....	8
1.2.7. В антикризисном управлении.....	8
1.2.8. В персонализации контента	8
1.3. Что такое анализ эмоций в искусственном интеллекте?.....	9
1.4. Схема системы	9
Выводы по главе 1	12
ГЛАВА 2. ИЗВЕСТНЫЕ СИСТЕМЫ, ИСПОЛЬЗУЕМЫЕ ДЛЯ АНАЛИЗА ЭМОЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК ТЕКСТА.....	13
2.1. Программа Power BI	14
2.2. Программа Tableau.....	15
2.3. Программа Qemotion.....	15
2.4. Программа MoodPatrol API.....	16
Выводы по главе 2	17
ГЛАВА 3. МЕХАНИЗМЫ ДЛЯ ПОДГОТОВКИ ТЕКСТА К ЕГО АНАЛИЗУ ДЛЯ СОЗДАНИЯ ЛИНГВИСТИЧЕСКОГО ЧЕЛОВЕ- КО-МАШИННОГО ИНТЕРФЕЙСА	18
3.1. Лингвистический человеко-машинный интерфейс для предварительной обработки текста	18
3.2. Влияние алгоритма предварительной обработки на текст.....	22

3.3. Применение Т-теста к тексту с предварительной обработкой и без нее	24
Выводы по главе 3	26
ГЛАВА 4. СОЗДАНИЕ УНИКАЛЬНОГО АЛГОРИТМА И ПРОГРАММЫ ДЛЯ ИДЕНТИФИКАЦИИ ЭМОЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК РУССКОЯЗЫЧНОГО ТЕКСТА.....	27
4.1. Эмоциональные конструкты.....	29
4.2. Функции эмоциональных выражений.....	29
4.3. Эмоциональное наполнение префиксов слов.....	30
4.4. Эмоциональное наполнение инфиксов	31
4.5. Разрабатываемая методология.....	32
4.6. Описание этапов работы алгоритма Emostemmer.....	33
4.7. Эксперимент и сравнение производительности методов.....	34
4.8. Отрицания и их влияние на эмоциональную полярность текста.....	39
Выводы по главе 4	43
ГЛАВА 5. АНАЛИЗ ЭМОЦИОНАЛЬНЫХ И ПСИХОЛИНГВИСТИЧЕСКИХ ХАРАКТЕРИСТИК ТЕКСТА, СОЗДАННОГО ПРИ ОБЩЕНИИ ЧЕЛОВЕКА С МАШИНОЙ	44
5.1. Логические характеристики эмоций в тексте.....	44
5.1.1. Определение модальностей	45
5.2. Синтаксические характеристики эмоций в тексте.....	49
5.3. Анализ сложной природы эмоций и отношений между ними ...	52
5.3.1. Методология для анализа эмоций	55
5.3.2. Эксперимент и результаты.....	57
5.3.3. Определение взаимосвязи между темой текста и эмоциями в нем.....	59
5.4. Характеристики экстравертного типа личности	63
5.5. Характеристики интровертного типа личности	64
5.6. Идентификация типа текста.....	65
5.7. Методология и результаты.....	66
Выводы по главе 5	69

ГЛАВА 6. ОБОБЩЕНИЕ ЭМОЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК, ВЫДЕЛЕННЫХ ИЗ ТЕКСТА, СОЗДАННОГО ПРИ ОБЩЕНИИ ЧЕЛОВЕКА С МАШИНОЙ	71
6.1. Обобщенные исследования	74
6.2. Кластеризация как механизм классификации эмоций в группах	76
6.2.1. Время выполнения алгоритма кластеризации	76
Выводы по главе 6	79
ЗАКЛЮЧЕНИЕ	81
ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ	83
ПРИЛОЖЕНИЯ	84
Фрагмент кода информации о сборке	84
Фрагмент программного кода	90
СПИСОК ЛИТЕРАТУРЫ	100

Научное издание

Аббаси Мохсин Маншад
Бельтюков Анатолий Петрович

**ЛИНГВИСТИЧЕСКИЙ ИНТЕРФЕЙС С ЭМОЦИОНАЛЬНЫМ
АНАЛИЗОМ ТЕКСТА В КОМПЬЮТЕРНЫХ СИСТЕМАХ**

Монография

*Авторская редакция
Компьютерная верстка: Т.В. Опарина*

Подписано в печать 10.10.2024. Формат 60x84 1/16.

Усл. печ. л. 6,68. Уч. изд. л. 5,97.

Тираж 300 экз. Заказ № 1880.

Издательский центр «Удмуртский университет»
426034, г. Ижевск, ул. Ломоносова, 4Б, каб. 021
Тел. + 7 (3412) 916-364, E-mail: editorial@udsu.ru

Типография Издательского центра «Удмуртский университет»
426034, г. Ижевск, ул. Университетская, 1, корп. 2.
Тел. 68-57-18