

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий
(наименование института полностью)

Кафедра «Прикладная математика и информатика»
(наименование)

02.03.03 Математическое обеспечение и администрирование информационных систем
(код и наименование направления подготовки / специальности)

Мобильные и сетевые технологии
(направленность (профиль) / специализация)

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (БАКАЛАВРСКАЯ РАБОТА)

на тему «Разработка API сервера для государственных организаций»

Обучающийся

В.А. Круглов

(Инициалы Фамилия)

(личная подпись)

Руководитель

С.В. Митин

(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Консультант (ы)

А.В. Москалюк

(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Аннотация

Тема бакалаврской работы: «Разработка ПО сбора новостной информации и формирование рейтинга наиболее активных тенденций».

Бакалаврская работа посвящена разработке ПО сбора новостной информации и формирование рейтинга наиболее активных тенденций.

В ходе выполнения исследований по бакалаврской работе был проведен сравнительный анализ различных методов сбора и анализа данных, описание методов сбора и анализа новостной информации, функционально-ориентированное проектирование системы, а также разработка и тестирование ПО.

Во введении прописывается актуальность темы, написаны цель и задачи.

В первом разделе рассматривается предметная область исследования и проводится анализ аналогов.

Во втором разделе произведен обзор алгоритмов формирования рейтинга, способы хранения информации, описали и создали архитектуру системы, спроектировали систему.

В третьем разделе разработали ПО и протестировали её по двум критериям.

В заключении представлены результаты выполнения выпускной квалификационной работы.

Бакалаврская работа состоит из введения, трёх разделов, заключения и списка использованной литературы.

Бакалаврская работа состоит из 45 страниц, 14 рисунков, 2 таблиц, 31 источника и 1 листинга.

Abstract

The title of the bachelor's thesis is "Development of software for collecting news information and forming a rating of the most active trends".

The research is devoted to development of software for collecting news information and forming a rating of the most active trends.

When doing a research on the bachelor's work, a comparative analysis of various methods of data collection and analysis, a description of methods for collecting and analyzing news information, functionally oriented system design, as well as software development and testing were carried out.

The introduction reveals the relevance of the research and gives a brief description of the work done.

The first section the subject area of the study is considered and the analysis of analogues is carried out.

The second section the review of algorithms for rating formation, methods of information storage, described and created the architecture of the system, designed the system.

The third section we developed the software and tested it according to two criteria.

In conclusion, the conclusions of the entire work are drawn.

The bachelor's thesis consists of an introduction, three sections, a conclusion and list of used literature.

The volume of the bachelor's thesis is 45 pages, it also contains 14 figures, 2 tables, a list of 31 references and 1 listing.

Оглавление

Введение.....	5
1. Описание сбора новостной информации.....	7
1.1 Постановка задачи и описание существующих методов и средств сбора новостной информации.....	7
1.2 Методы поиска информации в Интернете	8
1.3 Анализ существующих подходов к сбору и анализу новостной информации	9
1.4 Сравнительный анализ различных методов сбора и анализа данных	13
2. Описание хранения новостной информации и формирования рейтинга	17
2.1 Обзор алгоритмов формирования рейтинга тенденций.....	17
2.2 Описание методов сбора и анализа новостной информации ..	20
2.3 Способы хранения информации – выбор БД	21
2.4 Описание архитектуры системы	23
2.5 Функционально-ориентированное проектирование системы ..	24
3 Реализация и тестирование ПО	33
3.1 Реализация ПО.....	33
3.2 Тестирование программного обеспечения.....	36
Заключение	44
Список используемой литературы	45

Введение

В современном динамичном мире быть в курсе последних новостей и тенденций является важным аспектом информированности и принятия обоснованных решений. С появлением технологий то, как люди потребляют новости, резко изменилось. Прошли те дни, когда вы ждали утреннюю газету или смотрели вечерние новости. Теперь у людей есть доступ к обновлениям новостей одним нажатием кнопки благодаря Интернету и социальным сетям [5].

В результате этого изменения в привычках потребления новостей резко увеличилось количество источников новостей, доступных в Интернете [16]. Эти источники различаются по качеству, своевременности и актуальности. В таком переполненном пространстве пользователям может быть сложно найти самые актуальные и надежные источники новостей.

Эта тема является актуальной, поскольку затрагивает необходимость оперативного и результативного сбора и анализа новостей в цифровую эпоху. Благодаря огромному объему информации, доступной в Интернете, автоматизированное программное обеспечение может помочь журналистам и новостным организациям отслеживать последние тенденции и новости. Рейтинговая система также может помочь определить, какие истории являются наиболее популярными и влиятельными, предоставляя ценную информацию журналистам и специалистам в области СМИ.

Альтернативой программному обеспечению для сбора новостной информации и формирования рейтинга наиболее активных трендов является использование ручных методов сбора и анализа новостей. Этот подход отнимает много времени и подвержен ошибкам, так как требует большого количества человеческого ввода и интерпретации.

Более того, использование ручных методов может привести к необъективному отбору новостей, поскольку отбор зависит от личных предпочтений и мнений. Использование программного обеспечения для

сбора новостной информации и формирования рейтинга наиболее активных тенденций позволит устранить эти ошибки и предубеждения, обеспечив более точное и непредвзятое представление о новостях.

Объектом исследования бакалаврской работы является разработка программного обеспечения для сбора новостной информации.

Предметом исследования – критерий оценки, алгоритмы, программные решения, библиотеки, которые будут использованы для реализации ПО.

Цель работы разработка программного обеспечения для сбора новостной информации и формирования рейтинга наиболее активного тренда и оценка его эффективности в предоставлении пользователям точной и актуальной новостной информации.

Задачи диплома включают в себя:

- обзор литературы по существующим методам и средствам сбора новостной информации и алгоритмам формирования рейтинга тенденций:
- разработку методологии исследования, включая описание методов сбора и анализа новостной информации, алгоритма формирования рейтинга тенденций, используемых технологий и инструментов;
- разработку программного обеспечения для сбора новостной информации, описание архитектуры разработанной системы, основных модулей системы и интерфейса пользователя;
- тестирование и оценку качества разработанного ПО, включая описание методов тестирования, результаты тестирования и сравнение результатов с другими программными продуктами;
- формулирование выводов по исследованию, рекомендаций по использованию разработанного ПО и перспектив дальнейшего развития темы.

1. Описание сбора новостной информации

1.1 Постановка задачи и описание существующих методов и средств сбора новостной информации

Задачей данной работы является разработка программного приложения, способного эффективно собирать новостную информацию из различных источников и использовать ее для создания системы рейтинга наиболее активных тенденций. Программное обеспечение будет предназначено для автоматического сбора и анализа данных из новостных статей, платформ социальных сетей и других источников, чтобы выявлять новые тенденции и ранжировать их в зависимости от уровня активности.

Цель этой работы – создать надежный и удобный инструмент, который поможет компаниям, СМИ и частным лицам оставаться в курсе последних новостей и тенденций.

В зависимости от целей и задач сбора новостной информации могут быть использованы различные методы и средства. Например, для сбора информации о новостных событиях в реальном времени могут использоваться RSS-каналы [3],[28], которые предоставляют обновления новостей в автоматическом режиме. Для более глубокого анализа новостной информации могут использоваться специализированные сервисы, такие как GDELT и NewsAPI, которые позволяют получать информацию из различных источников и анализировать ее для выявления тенденций и трендов.

Для сбора информации о новостных событиях в социальных сетях могут использоваться API соответствующих сервисов. Эти сервисы позволяют получать данные о постах, хэштегах и другой информации, связанной с новостными событиями. Также для сбора информации о новостных событиях могут использоваться специальные инструменты для мониторинга социальных сетей, такие как Hootsuite и Sprout Social [10],[14],[31].

1. Цифровые методы сбора новостной информации. В этом разделе будут рассмотрены цифровые методы сбора новостной информации. К ним относятся мониторинг социальных сетей, RSS-каналы и Web-скрапинг.

2. Проблемы, с которыми сталкиваются при сборе новостной информации. В этом разделе будут рассмотрены проблемы, с которыми сталкиваются новостные организации при сборе новостной информации. Эти проблемы включают точность информации, скорость сбора информации и надежность источников.

3. Обзор существующих систем рейтинга новостей. В этом разделе представлен обзор существующих систем рейтинга новостей. К ним относятся алгоритм ранжирования Google News, система ранжирования Alexa и система NewsWhip Spike.

Тема работы будет опираться на информацию, изложенную в этой главе. Программное обеспечение будет использовать как традиционные, так и цифровые методы сбора новостной информации для создания всеобъемлющей базы данных новостей. Затем программное обеспечение будет использовать эту базу данных для создания рейтинга наиболее активных тенденций в новостях. Рейтинговая система будет учитывать такие факторы, как количество упоминаний конкретной истории, настроение освещения и источники освещения [9]. Программное обеспечение также будет легко настраиваться, что позволит новостным организациям адаптировать систему ранжирования к своим конкретным потребностям. В целом, программное обеспечение предоставит новостным организациям мощный инструмент для сбора и анализа новостной информации.

1.2 Методы поиска информации в Интернете

Можно отметить следующие ключевые способы поиска данных в Интернете, которые, в связи с целью и задачами ищущего, применяются по отдельности либо в композиции друг с другом:

- использование поисковых машин;
- поиск с применением браузера;
- анализ новых ресурсов.

Так как все сайты в пространстве WWW по сути становятся сопряженными между собой, отбор данных можно произвести путем поочередного просмотра объединенных страниц с поддержкой браузера. Несмотря на то, что данный метод является абсолютно ручным способом поиска, представляющий абсолютным анахронизмом в Сети, включающий сотни миллионов узлов, "ручной" просмотр Web-страниц зачастую становится исключительно допустимым в завершающих стадиях информационного поиска, когда механическое "копание" уступает место наиболее углубленному рассмотрению [22]. Применение каталогов, систематизированных и предметных списков и различных маленьких справочников также принадлежит к данному типу поиска.

1.3 Анализ существующих подходов к сбору и анализу новостной информации

Сбор и анализ новостной информации имеет решающее значение в современном мире, где средства массовой информации постоянно развиваются и играют значительную роль в формировании общественного мнения. Существуют различные подходы к сбору и анализу новостной информации, которые можно разделить на две категории: ручные и автоматические [26].

Ручные методы включают в себя усилия человека по сбору новостной информации с помощью различных средств, таких как чтение газет, просмотр телевизионных новостей, проведение интервью и использование платформ социальных сетей [30]. Эти методы требуют много времени, а точность собранной информации зависит от навыков и опыта человека. Ручные методы также ограничены в возможности сбора информации из

широкого круга источников и могут упускать важные новости.

С другой стороны, автоматизированные методы используют программное обеспечение для сбора новостной информации из различных источников, таких как новостные Web-сайты, платформы социальных сетей и RSS-каналы [19]. Автоматизированные методы являются быстрыми, эффективными и могут собирать огромное количество информации из разных источников. Преимущество этих методов также в том, что они позволяют анализировать собранные данные и выявлять тенденции и закономерности, которые могут быть незаметны при использовании ручных методов.

Одним из подходов к анализу новостной информации является анализ настроений – важнейший аспект разработки программного обеспечения для сбора новостной информации и формирования рейтинга наиболее активного тренда. Этот процесс включает в себя анализ языка, используемого в новостных статьях, и определение настроения или отношения, выраженного по отношению к определенной теме или объекту. С ростом доступности цифровых источников новостей анализ настроений стал важным инструментом для предприятий, правительств и частных лиц, стремящихся быть в курсе и опережать новостной цикл.

В контексте сбора новостей и анализа тенденций анализ настроений может помочь определить новые темы и тенденции, анализируя тон и настроение публикуемых новостных статей. Отслеживая настроение новостных статей, можно определить области, вызывающие озабоченность или интерес, которые могут появиться в новостном цикле. Эта информация может быть использована для принятия решений, маркетинга и других бизнес-стратегий.

Существует несколько инструментов и алгоритмов, которые можно использовать для анализа настроений, включая алгоритмы обработки естественного языка (NLP) и методы машинного обучения. Эти инструменты могут помочь определить шаблоны в языке, используемом в новостных

статьях, и определить общее настроение статьи, независимо от того, является ли оно положительным, отрицательным или нейтральным.

Одним из основных преимуществ использования анализа настроений при разработке программного обеспечения для сбора новостей и анализа тенденций является возможность автоматизации процесса выявления возникающих тенденций. Используя алгоритмы машинного обучения, можно научить программное обеспечение распознавать шаблоны языка, используемого в новостных статьях, и автоматически определять темы и объекты, вызывающие высокий уровень интереса или беспокойства.

В целом, анализ настроений – незаменимый инструмент для разработки программного обеспечения для сбора новостной информации и формирования рейтинга наиболее активного тренда. Анализируя язык, используемый в новостных статьях, можно определить новые тенденции и темы и опережать новостной цикл. По мере того, как популярность и значение цифровых источников новостей продолжают расти, анализ настроений становится все более ценным инструментом для предприятий, правительств и частных лиц, стремящихся быть в курсе и опережать конкурентов.

Другим подходом является тематическое моделирование – популярный прием, используемый при разработке программного обеспечения для сбора новостной информации и формирования рейтинга наиболее активных тенденций. Этот метод включает в себя определение и анализ основных тем и тем, присутствующих в большом массиве данных, таких как новостные статьи, сообщения в социальных сетях и другой online-контент. Определяя эти темы и темы, программное обеспечение может быстро и точно классифицировать и ранжировать самые важные и актуальные новости.

Одним из ключевых преимуществ использования тематического моделирования является то, что оно позволяет получить более тонкое и детальное представление о новостях и тенденциях. Например, вместо того, чтобы просто смотреть на количество упоминаний определенной темы или

ключевого слова, программное обеспечение может определить конкретные подтемы и темы, которые являются наиболее актуальными и важными. Это может помочь пользователям лучше понять контекст и значение новостей, а также принимать более обоснованные решения на основе этой информации.

Кроме того, тематическое моделирование также можно использовать для выявления закономерностей и тенденций во времени. Анализируя темы и темы, наиболее распространенные в новостях, программное обеспечение может выявлять новые тенденции и сдвиги в общественном мнении [7],[24]. Это может быть особенно полезно для предприятий и других организаций, которым необходимо быть в курсе последних событий в своей отрасли или на рынке.

В целом разработка программного обеспечения для сбора новостной информации и формирования рейтинга наиболее активных тенденций является важным направлением исследований и разработок. Используя такие методы, как тематическое моделирование, разработчики могут создавать мощные и эффективные инструменты, которые позволяют пользователям оставаться в курсе, принимать взвешенные решения и оставаться на шаг впереди в постоянно меняющемся мире [12]. В общем случае технологический процесс обработки информации включает процедуры, представленные на рисунок 1.



Рисунок 1 – Процедура обработки информации

При разработке программного обеспечения для сбора новостной

информации и формирования рейтинга наиболее активных тенденций можно было бы использовать автоматизированные методы, такие как анализ настроений и тематическое моделирование, для сбора и анализа данных из различных источников. Программное обеспечение может быть разработано для сбора данных с новостных Web-сайтов, платформ социальных сетей и RSS-каналов, а затем использовать алгоритмы машинного обучения для анализа данных и выявления новых тенденций и закономерностей. Программное обеспечение также может быть разработано для оценки наиболее активных тенденций на основе собранных и проанализированных данных.

Сбор и анализ новостной информации имеют решающее значение в современном мире, и для этого существуют различные подходы [2]. Автоматизированные методы быстрые, эффективные и могут анализировать большие объемы данных, что делает их идеальными для сбора и анализа новостной информации. Использование программного обеспечения для сбора новостной информации и формирования рейтинга наиболее активных тенденций могло бы принести большую пользу журналистам, маркетологам и политикам в выявлении возникающих тенденций и разработке эффективных стратегий.

1.4 Сравнительный анализ различных методов сбора и анализа данных

Сравнительный анализ различных методов сбора и анализа данных для разработки ПО сбора новостной информации и формирования рейтинга наиболее активных тенденций является важной частью дипломной работы. При этом необходимо учитывать несколько аспектов [18].

Методы сбора данных могут варьироваться от использования Web-скрейпинга до сбора данных через открытые API. При сравнительном анализе методов необходимо учитывать скорость сбора данных, их точность,

алгоритмы сбора данных и доступность источников.

Предварительная обработка данных включает лемматизацию и стемминг. При сравнительном анализе методов необходимо учитывать обработку и подготовку данных перед анализом. Точность обработки может влиять на точность результатов анализа данных.

Анализ данных включает частотный анализ и анализ тональности. Частотный анализ позволяет определить частотность употребления определенных слов и терминов в тексте. Анализ тональности позволяет определять тональность текста. При выборе методов анализа необходимо учитывать возможность визуализации результатов, а также точность и полноту анализа.

При сравнительном анализе методов визуализации результатов необходимо учитывать, насколько понятно и наглядно они отображают результаты сбора и анализа данных, а также трудоемкость создания и поддержки визуализации.

При сравнительном анализе методов необходимо учитывать экономическую эффективность различных методов, а также жизненный цикл системы (например, затраты на разработку, поддержку, обновление и модернизацию).

После проведения сравнительного анализа необходимо выбрать наиболее оптимальные методы для конкретной задачи сбора новостной информации и формирования рейтинга наиболее активных тенденций. При выборе методов необходимо учитывать критерии, такие как доступность, точность, скорость сбора и анализа данных, возможность визуализации результатов и экономическую эффективность.

Существует несколько методов сбора и анализа данных в рамках темы разработки программного обеспечения для сбора новостной информации и формирования рейтинга наиболее активного тренда. Наиболее распространенные методы включают Web-скраппинг, интеллектуальный анализ данных и обработку естественного языка. Проведем сравнительный

анализ этих методов:

– Web-скраппинг — это метод извлечения данных с Web-сайтов с помощью скриптов или программ. В контексте сбора новостной информации Web-скраппинг можно использовать для извлечения новостных статей из различных источников. Преимущество парсинга Web-страниц заключается в том, что его относительно легко реализовать, и он быстро предоставляет большой объем данных. Однако на Web-скраппинг могут влиять изменения Web-сайта, и он не всегда может собирать всю необходимую информацию;

– интеллектуальный анализ данных относится к процессу обнаружения закономерностей и тенденций в больших наборах данных. В контексте сбора новостной информации можно использовать интеллектуальный анализ данных для выявления наиболее активных тенденций путем анализа частотности ключевых слов в новостных статьях. Преимущество интеллектуального анализа данных заключается в том, что он может выявить информацию, которая не очевидна сразу. Однако для эффективной работы интеллектуального анализа данных требуется значительное количество вычислительных ресурсов;

– обработка естественного языка (NLP) – это метод анализа текстов для извлечения из них смысла. В контексте сбора новостной информации NLP можно использовать для определения настроения новостных статей и их классификации в соответствии с их темой. Преимущество NLP в том, что оно может обеспечить более тонкое понимание новостей, чем простой анализ ключевых слов. Однако для эффективности NLP требуется большой объем обучающих данных.

Каждый из этих методов имеет свои преимущества и недостатки, и выбор того, какой метод использовать, зависит от конкретных потребностей проекта. Web-скраппинг – хороший выбор, когда важны скорость и простота реализации, а интеллектуальный анализ данных и NLP лучше подходят для проектов, требующих более глубокого анализа [13]. В конечном счете, комбинация этих методов может быть лучшим подходом к эффективному и

точному сбору и анализу новостной информации. В таблице 1 представлены методы сбора и анализа данных.

Таблица 1 – Сравнительный анализ различных методов сбора и анализа данных

Метод	Описание	Преимущества	Недостатки
Web-скрейпинг	Извлечение данных с Web-сайтов с помощью программных средств	Предоставляет доступ к большому количеству данных и новостного контента	Правовые проблемы, связанные с интеллектуальной собственностью и нарушением авторских прав
Мониторинг социальных сетей	Отслеживание и анализ разговоров и тенденций в социальных сетях	Предоставляет информацию и анализ настроений в режиме реального времени	Ограничено общедоступными данными
Обзорное исследование	Сбор данных с помощью опросов и анкет	Позволяет собирать конкретную информацию о предпочтениях и поведении пользователей	Ограничено размером выборки и возможностью необъективных ответов
Интеллектуальный анализ данных	Использование статистических методов для анализа больших наборов данных и выявления закономерностей и тенденций	Обеспечивает целостное представление данных и раскрывает скрытые идеи	Требуются специальные навыки и инструменты
Машинное обучение	Использование алгоритмов для обучения на основе данных и создания прогнозов или классификаций	Более быстрый и точный анализ данных	Требуются большие объемы данных для обучения алгоритма

Выводы по разделу:

В первом разделе была поставлена задача на исследование, описаны методы поиска информации в интернете, произведен анализ существующих подходов к сбору и анализу новостной информации и произведен сравнительный анализ аналогов, где были показаны преимущества и недостатки каждого аналога.

2. Описание хранения новостной информации и формирования рейтинга

2.1 Обзор алгоритмов формирования рейтинга тенденций

В мире информации тенденции постоянно меняются и развиваются. Важно следить за тенденциями, чтобы быть в курсе последних событий. Однако идти в ногу с тенденциями – сложная задача, учитывая огромное количество информации, доступной в Интернете. В этой главе мы обсудим алгоритмы, используемые для создания рейтинга тенденций.

Тенденция относится к общему направлению, в котором что-то развивается или изменяется. В контексте информации тенденция относится к теме, которая в настоящее время популярна и широко обсуждается.

Рейтинг трендов – это список самых популярных тем, обсуждаемых в определенный момент времени. Этот список формируется на основе различных факторов, таких как частота упоминаний, количество написанных статей и источник информации.

Существует несколько алгоритмов, используемых для генерации рейтинга трендов:

- Скрытое распределение Дирихле (LDA) – это вероятностный алгоритм, который используется для создания рейтинга тенденций. LDA используется для идентификации тем в корпусе и слов, связанных с каждой темой. Алгоритм присваивает вероятность каждому слову в корпусе, указывающую, насколько вероятно, что оно связано с определенной темой. Темы с наибольшей вероятностью считаются наиболее важными и используются для формирования рейтинга тенденций.

- Алгоритм наивного Байеса – это вероятностный алгоритм, который используется для классификации текста. Алгоритм присваивает вероятность каждому слову в корпусе, указывающую, насколько вероятно, что оно будет связано с определенным классом. В контексте формирования

рейтинга тенденций каждая тема считается классом. Темы с наибольшей вероятностью считаются наиболее важными и используются для формирования рейтинга тенденций.

Существуют различные алгоритмы, которые можно использовать для создания рейтингов тенденций и типов, в зависимости от конкретного контекста и доступных данных [4],[25]. Некоторые распространенные подходы включают в себя:

– частотный анализ: этот алгоритм подсчитывает количество случаев каждого тренда или типа и генерирует рейтинг на основе их частоты.

Частотный анализ – это алгоритм, который анализирует частоту появления каждого элемента или события в наборе данных. Это помогает определить наиболее распространенные или частые элементы и закономерности в наборе данных, что может дать представление о тенденциях и закономерностях с течением времени. Это простой и понятный алгоритм, который хорошо работает для наборов данных с большим количеством переменных и наблюдений. Однако он может не подходить для наборов данных со сложными или нелинейными отношениями между переменными, поскольку не учитывает зависимости между переменными.

– Анализ настроений: этот алгоритм использует методы обработки естественного языка (NLP) для определения настроений, связанных с каждой тенденцией или типом, и генерирует рейтинг на основе их положительности или отрицательности.

Алгоритм анализа тональности – это метод обработки естественного языка, целью которого является классификация субъективных текстовых данных на положительные, отрицательные или нейтральные настроения. Он анализирует слова и фразы в тексте, чтобы определить основное настроение, и присваивает балл каждой части данных. Алгоритм использует различные методы, такие как машинное обучение, байесовский анализ и подходы на основе лексики для выполнения анализа настроений. Характеристики этого алгоритма включают его способность точно классифицировать настроения в

режиме реального времени, его масштабируемость и способность обрабатывать большие объемы данных из различных источников, таких как социальные сети, новости и отзывы клиентов.

– Машинное обучение: этот метод использует обученную модель для прогнозирования рейтинга каждой тенденции или типа на основе исторических данных и других соответствующих факторов.

Алгоритмы ранжирования тенденций – это алгоритмы машинного обучения, которые анализируют большие наборы данных для выявления закономерностей и ранжирования тенденций. Обычно они используют методы статистического анализа и обработки естественного языка для извлечения соответствующей информации из неструктурированных данных, таких как текст, изображения и видео.

Алгоритм работает, обрабатывая данные из нескольких источников и определяя наиболее часто встречающиеся ключевые слова и темы. Затем он ранжирует эти тенденции на основе их уровня активности и популярности с течением времени.

Некоторые из характеристик алгоритмов ранжирования тенденций включают их способность выявлять возникающие тенденции в режиме реального времени, их масштабируемость при обработке больших объемов данных и их способность адаптироваться к меняющимся тенденциям и предпочтениям пользователей. Их можно использовать в различных приложениях, включая маркетинг, анализ социальных сетей и финансовое прогнозирование.

Описание тенденций и типов может быть создано с использованием методов НЛП, таких как тематическое моделирование, которое определяет ключевые темы и темы в наборе данных, или распознавание образов, которое идентифицирует повторяющиеся образцы и структуры [8].

2.2 Описание методов сбора и анализа новостной информации

Существует несколько методов сбора новостной информации, в том числе:

- интервью. Журналисты могут брать интервью у соответствующих источников, экспертов или очевидцев, чтобы собрать информацию об истории;
- исследование: это включает в себя изучение предыдущих статей, новостей и других публикаций на эту тему для сбора информации;
- пресс-релизы. Пресс-релизы рассылаются организациями средствами массовой информации и журналистам для предоставления информации об их новостях или событиях;
- социальные сети. Платформы социальных сетей можно использовать для сбора информации, особенно рассказов очевидцев и обновлений событий в режиме реального времени;
- пресс-конференции. Организации проводят пресс-конференции для предоставления информации средствам массовой информации и ответов на вопросы;
- наблюдения. Журналисты могут наблюдать за событиями или ситуациями, чтобы собрать информацию об истории;
- анализ данных. Сбор и анализ данных может помочь журналистам выявить тенденции или закономерности, которые делают статьи достойными освещения в печати;
- пресс-агентства. Журналисты могут подписаться на информационные агентства, которые собирают и распространяют новости со всего мира.

Описание методов анализа данных:

После сбора данных их необходимо проанализировать, чтобы выявить наиболее активные тенденции.

Обработка естественного языка (NLP). НЛП – это метод анализа текстовых данных. Он используется для выявления закономерностей в текстовых данных и извлечения значимой информации. НЛП будет использоваться для анализа собранных новостных статей и выявления наиболее часто встречающихся слов и фраз.

Лучшим алгоритмом среди обработки естественного языка (NLP), анализа настроений и машинного обучения для разработки программного обеспечения для сбора новостной информации и формирования рейтинга наиболее активных тенденций будет машинное обучение.

Алгоритмы машинного обучения могут анализировать большие объемы данных, изучать закономерности и делать прогнозы на основе закономерностей. Обучая модель машинного обучения на наборе данных новостных статей и данных об их читательской аудитории, модель может узнать, какие статьи вызывают наибольший интерес, и соответствующим образом ранжировать их.

Кроме того, методы машинного обучения могут включать НЛП и анализ настроений в процесс принятия решений. Это означает, что модель также может учитывать настроение новостных статей и язык, используемый для определения наиболее активного тренда. В целом машинное обучение обеспечивает гибкий и мощный подход к разработке системы ранжирования новостей.

2.3 Способы хранения информации – выбор БД

Существует несколько способов структурировать данные (несколько моделей баз данных):

- сетевая модель баз данных это модель позволяющая хранить сведения со структурой в виде свободного графа. Этот подход считается наиболее универсальным, разрешая держать сведения произвольной структуры. Но подобная многофункциональность не

дает возможность приобрести довольно результативные операции по обработке данных сведений, по этой причине эта модель значительного распространения не приобрела;

- иерархическая модель баз данных это модель позволяющая хранить данные со структурой в виде дерева – графа не содержащего циклы. Такое ограничение на структуру данных позволяет применить известные алгоритмы работы с деревьями. Такая модель часто применяется в различного рода каталогах, например, в современных файловых системах (в качестве каталога файлов);
- реляционная модель баз данных это модель позволяющая хранить данные в виде взаимосвязанных прямоугольных таблиц. Для таких структур данных существует проработанный математический аппарат, основанный на теории множеств. Этот аппарат хорошо подходит для решения многих практических задач, в связи с чем, реляционные базы данных в настоящее время широко распространены;
- объектно-ориентированная модель баз данных это модель позволяющая хранить данные в виде взаимосвязанных абстрактных объектов. Эта модель хорошо согласована с популярным методом проектирования и анализа информационных систем – объектно-ориентированным моделированием. Она в течение нескольких лет считается перспективной, однако к настоящему времени широко распространения не получила.

Таблица 2 – Сравнительный анализ реляционных СУБД

СУБД	Преимущества	Недостатки
Oracle	Функциональность; Надежность.	Стоимость; Потребление системой большого кол-ва ресурсов.

Продолжение таблицы 2

MySQL	Функциональность; Поддержка пользовательских интерфейсов; Работа с другими базами данных.	Отсутствие поддержки OLAP и XML; Платная поддержка.
Microsoft SQL сервер	Простота; Производительность; Работа на мобильном устройстве; Взаимодействие с другими продуктами Microsoft.	Потребление системой большого кол-ва ресурсов; Проблемы с импортом файлов.
PostgreSQL	Большой объем обработки; Доступность интерфейсов; Функциональность.	Неполная поддержка; Конфигурация; Скорость работы.

Сравнив СУБД (таблица 2) можно сделать вывод, что в дальнейшем будем использовать PostgreSQL, так как является хорошим выбором, поскольку он имеет открытый исходный код, хорошо масштабируется и предлагает мощную поддержку сложных запросов, текстового поиска и индексации данных. Кроме того, он имеет репутацию очень надежного и безопасного. Однако выбор в конечном итоге зависит от конкретных требований и ограничений проекта.

2.4 Описание архитектуры системы

Компонент сбора данных может использовать Web-скрапинг или API для сбора новостных статей из различных источников. Затем эти статьи сохраняются в базе данных для дальнейшей обработки [27].

Компонент обработки анализирует статьи, чтобы определить общие темы и темы, и присваивает каждой из них вес на основе таких факторов, как популярность и актуальность. Эта информация используется для

формирования рейтинга наиболее активных трендов. На рисунке 2 показана архитектура системы.



Рисунок 2 – Архитектура системы

Компонент визуализации может отображать рейтинг и соответствующие новостные статьи в удобном для пользователя интерфейсе, таком как информационная панель или Web-страница.

2.5 Функционально-ориентированное проектирование системы

Обзор CASE-средств для автоматизации проектирования и разработки информационных систем ранее применялась структурная методология, которая означала использование формализованных методов описания разрабатываемой системы и принимаемых решений. При этом использовались графические средства описания различных моделей информационных систем с помощью схем и диаграмм. При разработке таких информационных систем вручную эти графические модели разрабатывать и использовать весьма долго.

Отмеченные обстоятельства стали одной из причин появления программно-технологических средств, получивших название CASE-средств и реализующих CASE-технологии создания и сопровождения информационных систем. В ряде современных CASE-средств используется объектно-ориентированная методология проектирования.

Термин CASE дословно переводится как разработка программного обеспечения с помощью компьютера [20]. В настоящее время это понятие получило более широкое значение, означающее автоматизацию разработки информационных систем.

CASE-средства представляют собой программные средства, поддерживающие процессы сопровождения и создания таких информационных систем, как: анализ и формулировка требований, проектирование баз данных и приложений, генерация кода, тестирование, обеспечение качества.

CASE-технология определяется методологией проектирования информационных систем с инструментальными средствами, которые позволяют наглядно моделировать предметную область, анализировать ее модель на всех этапах разработки и сопровождения информационной системы и разрабатывать приложения для пользователей.

Основной целью CASE-систем и средств является отделение программного обеспечения от его кодирования и последующих этапов разработки (тестирование, документирование), а также автоматизация всего процесса создания программных систем или разработки [1],[29].

Чаще всего разработка программ начинается с некоторого предварительного варианта системы. В качестве такого варианта может быть разработанный специально для этого прототип или устаревшая и не удовлетворяющая новым требованиям система. В последнем случае для восстановления знаний о программной системе с целью последующего их использования применяют повторную разработку – реинжиниринг. На рисунке 3 представлена атрибутивная модель.

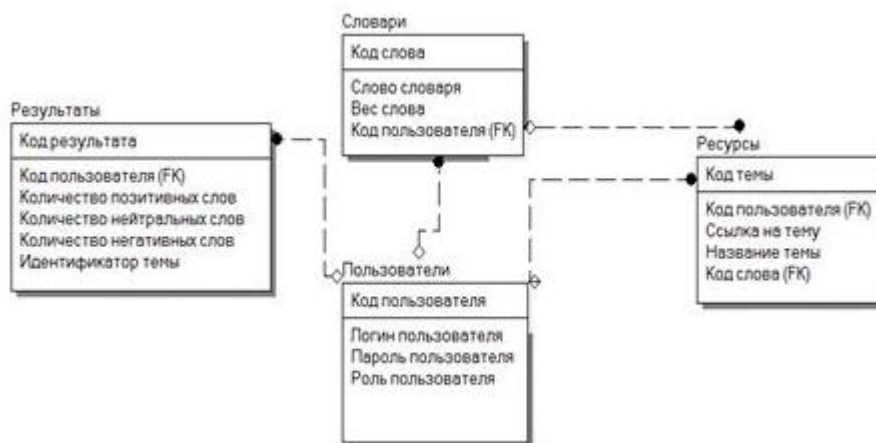


Рисунок 3 – Атрибутивная модель

Повторная разработка сводится к построению исходной модели программной системы путем исследования ее программных кодов. Имея модель, можно ее усовершенствовать, а затем перейти к разработке. В большинстве случаев так и делают при проектировании. Один из наиболее известных является принцип «возвратного проектирования» – Round Trip Engineering (RTE) [11].

Современные CASE-системы занимаются не только разработкой, а еще обеспечивают и повторную разработку. Это заметно ускоряет разработку приложений, а также повышается их качество.

Структурный подход к проектированию информационной системы заключается в рассмотрении ее с общих позиций с последующей детализацией и представлением в виде иерархической структуры [21]. На верхнем уровне иерархии обычно представляется функциональное описание системы. При проведении структурного анализа и проектирования для повышения наглядности используется графическое представление функций информационной системы и отношений между данными [17].

Наиболее распространенными моделями и диаграммами графического представления являются следующие:

- диаграммы сущность – связь или ER – диаграммы – модель данных, служащая для описания концептуальной схемы баз данных;
- диаграммы потоков данных (DFD) – моделирует систему как набор действий, соединенных друг с другом стрелками;
- метод структурного анализа и проектирования (SADT) – служит для построения функциональной модели объекта;
- схемы описания иерархии вход – обработка – выход служат для описания реализуемых программой функций и циркулирующих внутри нее потоков данных;
- диаграмма вариантов использования – графическое изображение возможных взаимодействий пользователя с системой (рисунок 4).

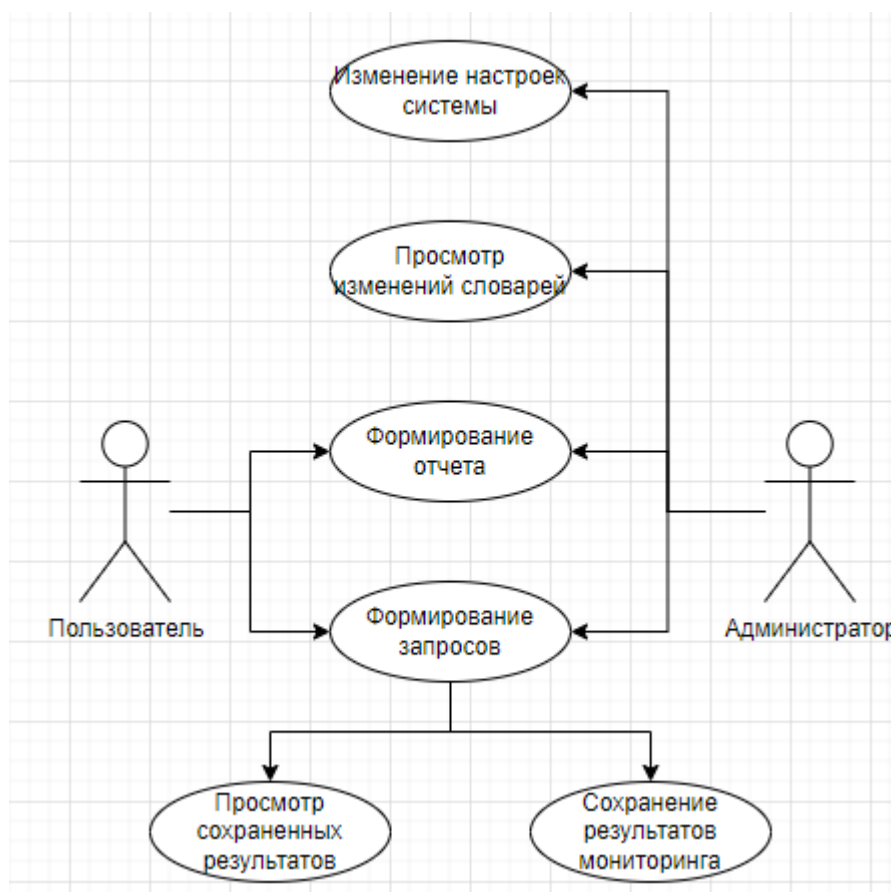


Рисунок 4 – Диаграмма вариантов использования

Построение контекстной диаграммы с помощью CASE-средства BPWin, рассмотрим методологию структурного анализа и проектирования IDEF0, основанную на методологии SADT. Результатом применения методологии SADT является модель, которая состоит из диаграмм, фрагментов текстов и глоссария, имеющих ссылки друг на друга. Диаграммы – главные компоненты модели, все функции ИС и интерфейсы на них представлены как блоки и стрелки. Место соединения стрелки с функциональным блоком определяет тип интерфейса. Стрелка управления входит в блок сверху и обозначает, кто отвечает за регулирование того, как и когда выполняется функциональный блок. Стрелка, входящая в блок слева показывает какая информация или источники взяты за входные данные.

Стрелка, которая выходит из функционального блока показывает готовый продукт и результаты [23]. Стрелка, входящая в функциональный блок снизу – это механизм (человек или автоматизированная система), то, что используется для выполнения процесса, но остается неизменным. Механизмы являются ресурсом, которые непосредственно исполняют моделируемые действия. Одной из наиболее важных особенностей методологии SADT является введение все больших уровней детализации по мере создания диаграмм, отображающих модель.

Контекстная диаграмма модели программной системы представлена на рисунке 5.

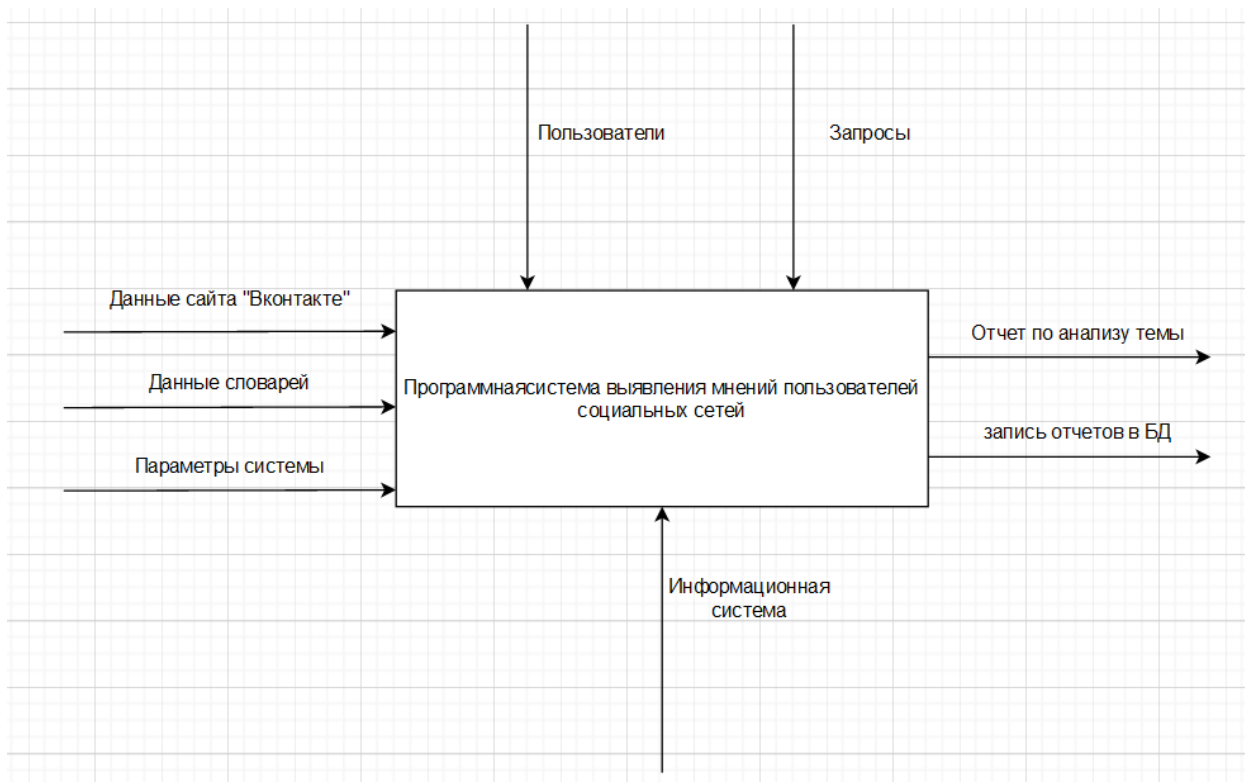


Рисунок 5 – Контекстная диаграмма модели программной системы

Вход представляет собой информацию, преобразуемую функциональным блоком. Для данной модели входной информацией являются данные сайта «Вконтакте». Управляющим механизмом является пользователь, запросы и словари. Исполняющим механизмами является сама система. В качестве выходной информации служит отчет по общественному мнению.

Построим диаграмму IDEF0 (рисунок 6), программная система состоит из трех этапов: «Ввод данных», «Обработка данных», «Формирование отчета»

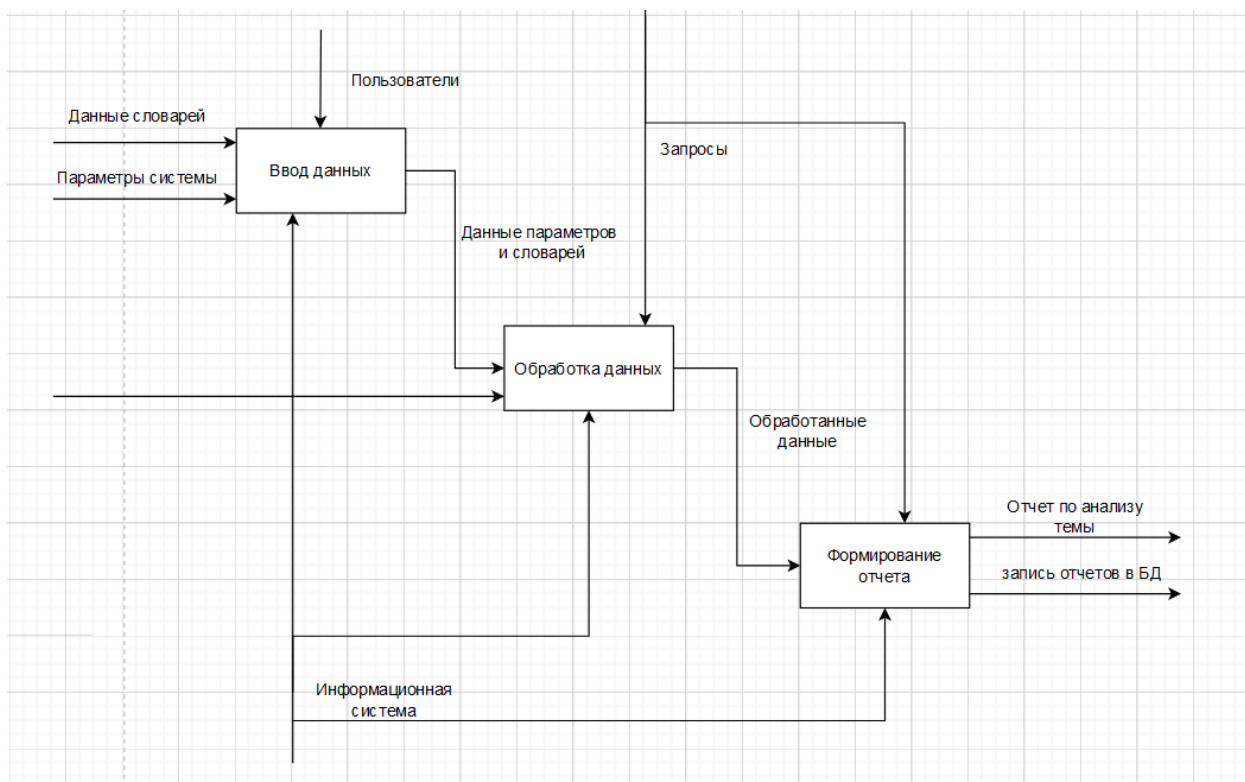


Рисунок 6 – Диаграмма декомпозиции 1-го уровня

Этап «Обработка данных» состоит из трех функциональных блоков: «Поиск и отбор тем», «Анализ содержимого темы» (рисунок 7), «Операции с результатами» (рисунок 8).

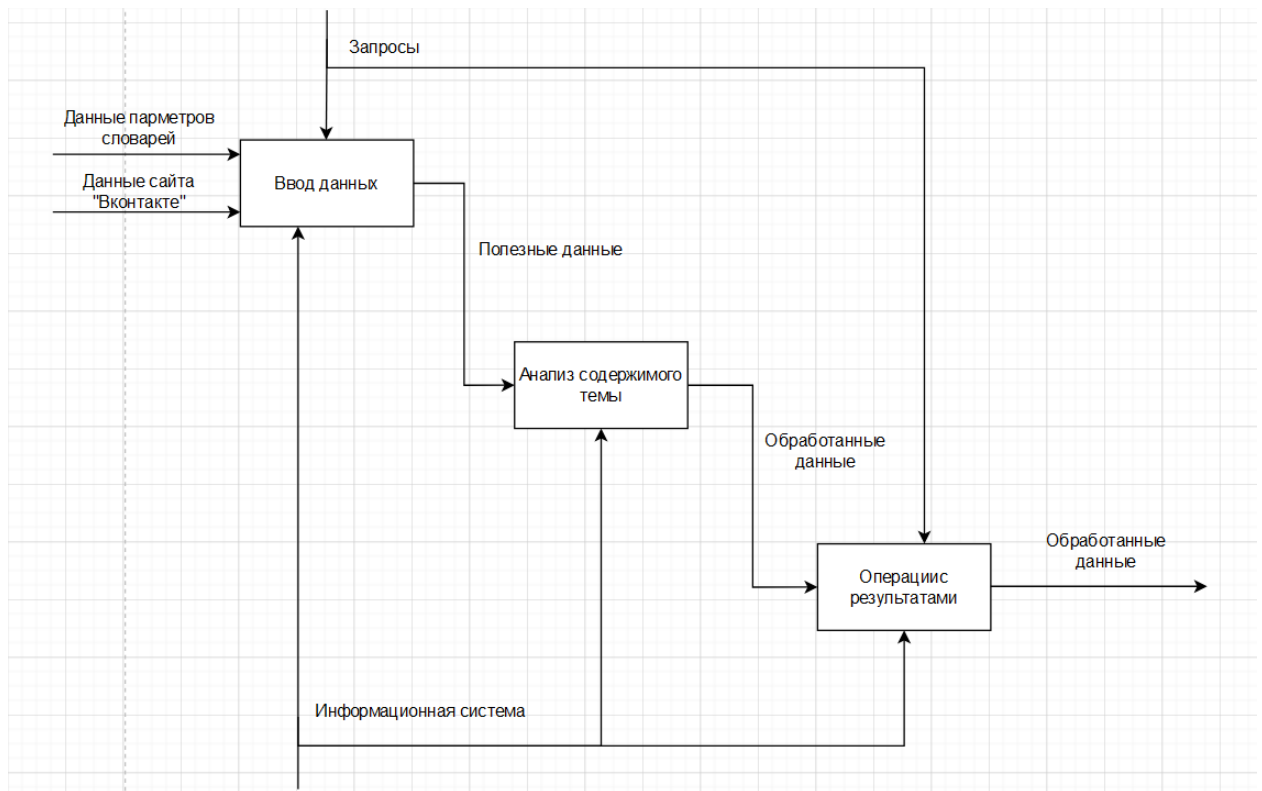


Рисунок 7 – Диаграмма декомпозиции 2-го уровня

Этап «Поиск и отбор темы» состоит из трех функциональных блоков: «Выбор темы», «Изучение текста темы», «Подготовка к анализу»

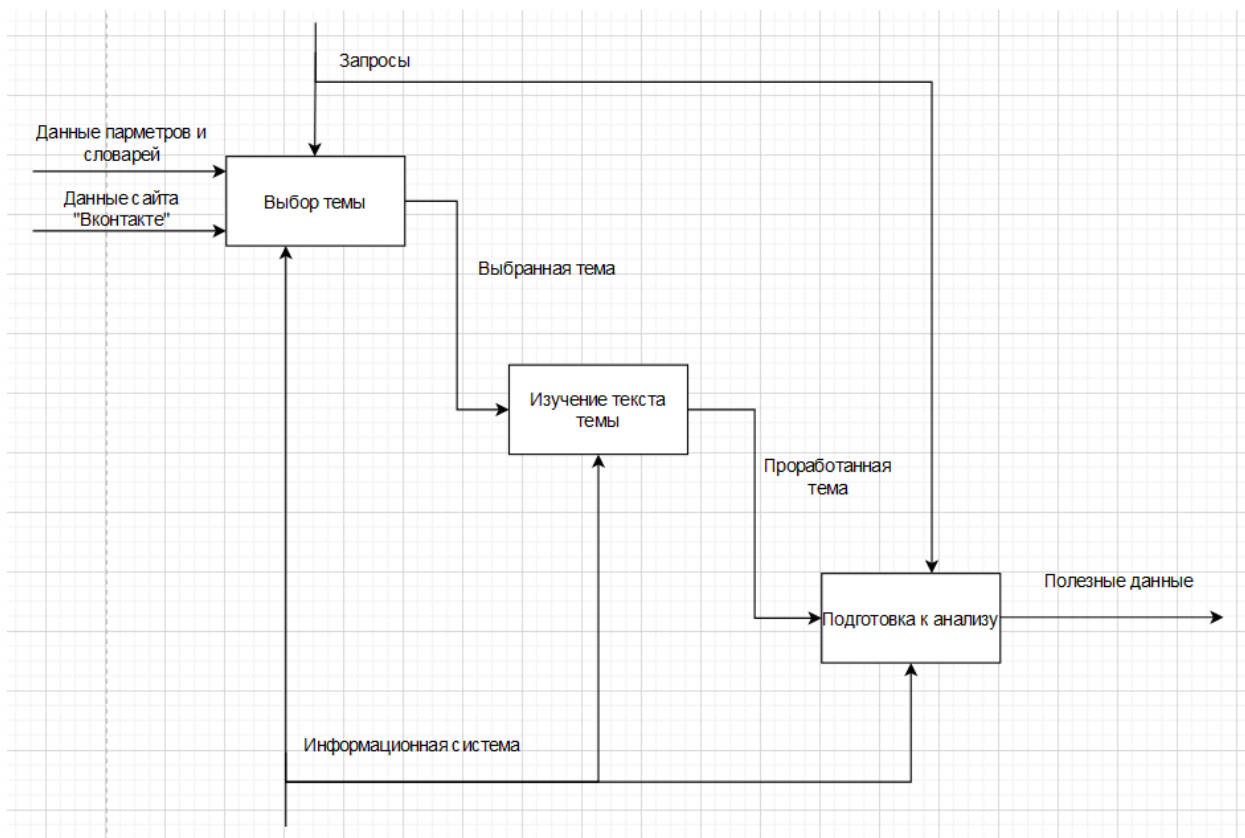


Рисунок 8 – Диаграмма декомпозиции 3-го уровня

Выводы по разделу 2

Второй раздел ВКР посвящен описанию хранения новостной информации и формирования рейтинга.

Был произведен обзор алгоритмов формирования рейтинга, способы хранения информации, описали и создали архитектуру системы.

А также спроектировали систему для дальнейшей разработки программного обеспечения.

3 Реализация и тестирование ПО

3.1 Реализация ПО

Существует несколько библиотек, которые можно использовать для сбора и оценки информации в рамках темы разработки ПО для сбора новостной информации и формирования рейтинга наиболее активного тренда.

Одной из таких библиотек является Beautiful Soup, представляющий собой пакет Python, который используется для очистки Web-страниц. Он позволяет анализировать документы HTML и XML, упрощая извлечение данных с Web-сайтов. Beautiful Soup также предоставляет полезные инструменты для навигации по дереву синтаксического анализа, а также для поиска и изменения дерева синтаксического анализа.

Еще одна полезная библиотека – Pandas, которая представляет собой библиотеку обработки данных для Python. Его можно использовать для сбора и анализа данных из различных источников, включая Web-сайты и API. Он предоставляет функции для чтения данных в различных форматах (например, базы данных CSV, Excel, SQL), а также для очистки и преобразования данных [15].

Для задач обработки естественного языка (NLP) обычно используется набор инструментов Natural Language Toolkit (NLTK). NLTK можно использовать для выполнения таких задач, как токенизация, выделение корней и тегирование частей речи. Он также включает предварительно обученные модели для анализа настроений, распознавания именованных сущностей и других задач НЛП.

Наконец, для задач машинного обучения популярная библиотека scikit-learn предоставляет широкий набор алгоритмов для классификации, кластеризации и регрессии. Его можно использовать для построения моделей

для прогнозирования тенденций и выявления закономерностей в собранных данных.

В целом, эти библиотеки можно комбинировать и по-разному использовать для разработки программного обеспечения для сбора новостной информации и формирования рейтинга наиболее активного тренда. В листинге 1 представлен код программы.

Листинг 1 – Реализация ПО

```
// Import necessary libraries:
import requests
from bs4 import BeautifulSoup

// Define the function to collect news from the Web:
def get_news():
    url = ''
    result = requests.get(url)
    soup = BeautifulSoup(result.text, 'html.parser')
    headlines = soup.find_all('h3')
    news_list = []
    for headline in headlines:
        news_list.append(headline.text)
    return news_list

// Define the function to track trends:
def track_trends(news_list):
    trends = {}
    for news in news_list:
        if news in trends:
            trends[news] += 1
        else:
```

```

trends[news] = 1
sorted_trends = sorted(trends.items(), key=lambda x: x[1], reverse=True)
return sorted_trends

// Call the functions:
news_list = get_news()
sorted_trends = track_trends(news_list)

// Print the results:
print('Most active trends:')
for trend in sorted_trends:
print(trend[0], trend[1])
import requests
import json

# Set up the base URL of the news API
base_url = ''

# Get the latest news stories
news_stories = requests.get(base_url + '/stories')

# Parse the response into a JSON object
news_stories_json = json.loads(news_stories.text)

# Create a dictionary to hold the trending topics
trending_topics = {}

# Iterate over the stories in the response
for story in news_stories_json:
# Get the list of topics for this story

```

```

story_topics = story['topics']

# Iterate over the topics for this story
for topic in story_topics:
# Increment the count of the topic in our trending topics dictionary
if topic in trending_topics:
trending_topics[topic] += 1
else:
trending_topics[topic] = 1

# Sort the trending topics dictionary by the number of stories
sorted_trending_topics = sorted(trending_topics.items(), key=lambda x:
x[1], reverse=True)

# Print the top 10 trending topics
for i in range(10):
print(f'{sorted_trending_topics[i][0]}:          {sorted_trending_topics[i][1]}
stories')

```

3.2 Тестирование программного обеспечения

Для тестирования приложения использован метод функционального тестирования.

Функциональное тестирование – это метод тестирования программного обеспечения, который проверяет функциональность приложения или системы на соответствие заданным требованиям. Он включает в себя тестирование функций и функций программного обеспечения, чтобы убедиться, что оно работает должным образом и соответствует потребностям пользователя [6].

Цель функционального тестирования – убедиться, что программная система выполняет все функции, которые она должна выполнять, и не выполняет функции, для которых она не предназначена.

Пользовательский интерфейс (UI) — это визуальная и интерактивная часть программного приложения или Web-сайта, которая позволяет пользователям взаимодействовать с системой. Он включает макет, дизайн и функциональность кнопок, меню, форм и других элементов, с которыми пользователи взаимодействуют для выполнения задач и достижения целей. Хорошо продуманный пользовательский интерфейс интуитивно понятен, прост в навигации и визуально привлекателен, что может повысить вовлеченность и удовлетворенность пользователей.

В верхней части экрана находится панель навигации, на которой располагаются три кнопки «Ваша подборка», «Все новости» и «Профиль» (рисунок 9).

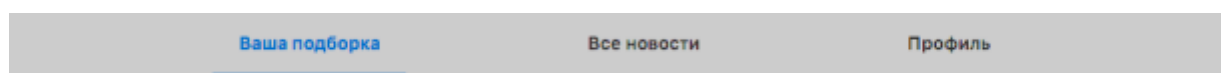


Рисунок 9 – Панель навигации

Первая вкладка, «Ваша подборка» (рисунок 10), отображает все персональные новости, на основе предпочтений пользователя. Если же у пользователей не вошел в профиль, то эта вкладка будет пустой.

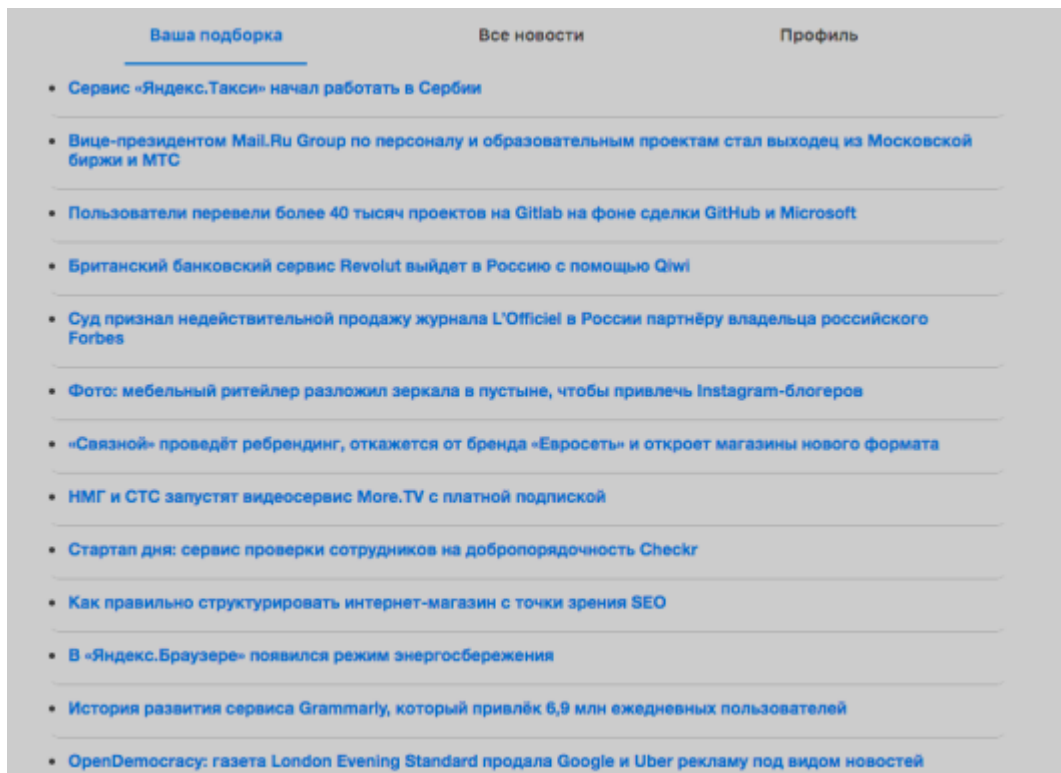


Рисунок 10 – Вкладка «Ваша подборка»

Вкладка «Все новости» доступна для всех пользователей, даже если они создали свой профиль. В ней отображаются все новости без разделения на тематики (рисунок 11).

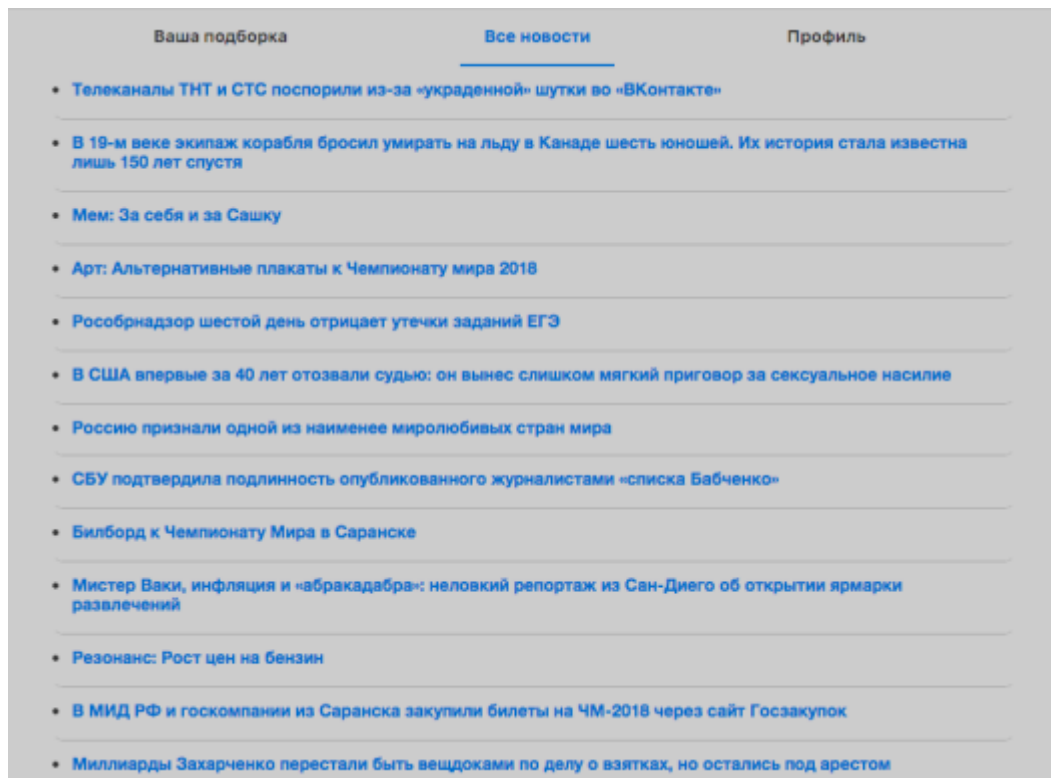


Рисунок 11 – Вкладка «Все новости»

Вкладка «Профиль» отвечает за регистрацию в системе и авторизацию, если учетная запись уже заведена. При первом посещении отображается форма авторизации и кнопка регистрации (рисунок 12).

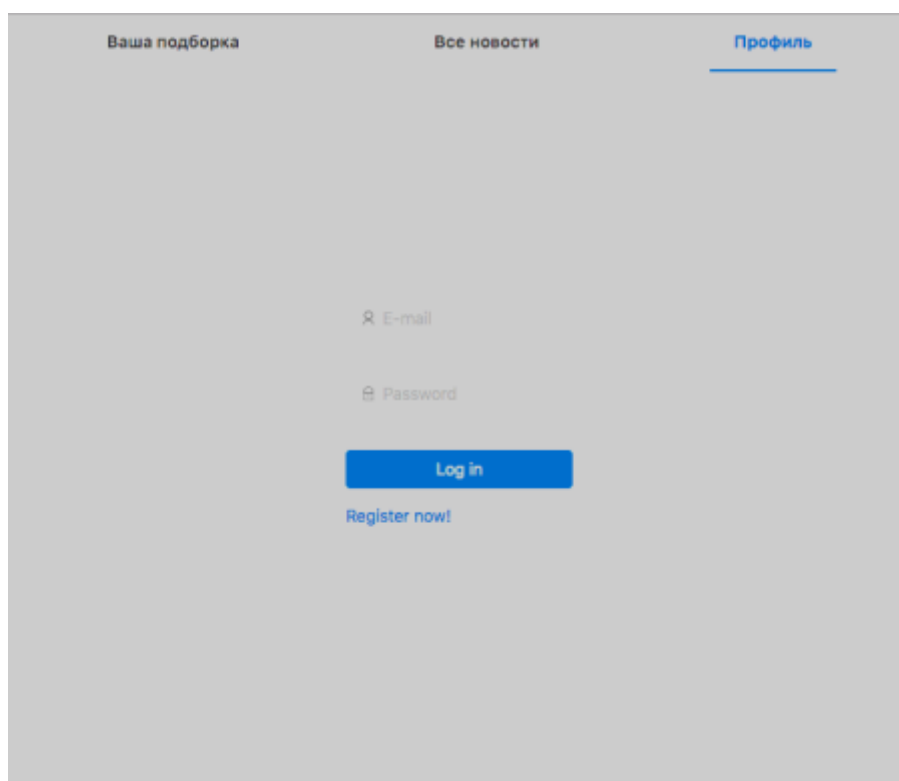


Рисунок 12 – Вкладка «Профиль»

Тестирование будет производиться два раза и ниже будут представлены результаты тестирования по:

- рейтингу цитируемости журналов;
- рейтингу цитируемости Интернет-ресурсов.

На рисунке 13 представлена таблица результата рейтинга цитируемости журналов.

2021			2022			2023		
Место в рейтинге	СМИ	ИЦ	Место в рейтинге	СМИ	ИЦ	Место в рейтинге	СМИ	ИЦ
1	Forbes	9 721,70	1	Forbes	14 527,10	1	Forbes	37595,4
2	The New Times	1 796,80	2	The New Times	1 456,60	2	The New Times	2422
3	Финанс	937,9	3	Сноб	757,4	3	Итоги	1300,6
4	Итоги	772	4	Итоги	657,9	4	Коммерсантъ-Власть	1220,6
5	Коммерсантъ-Власть	677	5	Коммерсантъ-Деньги	556,4	5	Эксперт	1095,8
6	Эксперт	634,8	6	GQ	517,6	6	СЕО	1019,4
7	Русский репортер	586,4	7	Огонек	509	7	Сноб	757,7
8	Профиль	517	8	Русский репортер	506,9	8	GQ	628,5
9	Огонек	493,7	9	Коммерсантъ-Власть	434,5	9	Профиль	587
10	Коммерсантъ-Деньги	358	10	Афиша	413,8	10	Русский пионер	555,2

Рисунок 13 – Рейтинг цитируемости журналов

Судя по таблице, два ведущих журнала за 2022 год, Forbes и The New Times, не изменились. Однако в рейтинг вошли три новых журнала, в том числе «Сноб», учредивший премию «Сделано в России», GQ, опубликовавший рейтинг самых невлиятельных мужчин, и «Афиша», взявшая интервью у Дмитрия Пескова. Наиболее существенное движение в рейтинге наблюдает журнал «Коммерсантъ-Деньги», опубликовавший рейтинг самых квалифицированных управляющих компаний.

В 2023 году Forbes и The New Times продолжают оставаться самыми цитируемыми журналами. Самое большое изменение в рейтинге произошло в журнале «Коммерсантъ-Власть», который поднялся на пять позиций после интервью с мэром Москвы Сергеем Собяниным. Новичками рейтинга стали «Эксперт», «СЕО», «Профиль» и «Русский пионер». После выборов мэра Москвы «Профиль» (9 место) опубликовал интервью с чиновниками Андреем Шароновым и Маратом Хуснуллиным. Кроме того, журнал «Эксперт» (5-е место) активно цитировался для интервью с советником

президента. «Русский пионер» (10 место) опубликовал интервью с Владиславом Сурковым после его ухода с поста заместителя главы Администрации Президента РФ. В течение года журнал СЕО (6-е место) цитировался в рейтингах предприятий и компаний.

Также проведено тестирование по рейтингу цитируемости Интернет-ресурсов, рисунок 14.

2021			2022			2023		
Место в рейтинге	СМИ	ИЦ	Место в рейтинге	СМИ	ИЦ	Место в рейтинге	СМИ	ИЦ
1	lenta.ru	12 814,50	1	Lenta.ru	16 493,10	1	Lenta.ru	47542
2	gazeta.ru	8 973,60	2	Gazeta.ru	13 282,70	2	Gazeta.ru	36899,4
3	lifenews.ru	5 177,40	3	Lifenews.ru	9 707,40	3	Vesti.ru	11763,7
4	vesti.ru	4 114,80	4	Vesti.ru	4 547,40	4	Fontanka.ru	10901,6
5	bfm.ru	3 719,10	5	Bfm.ru	3 550,50	5	Lifenews.ru	9345,6
6	inopressa.ru	2 810,70	6	Newsru.com	3 287,70	6	Newsru.com	8511,1
7	newsru.com	2 691,90	7	Fontanka.ru	2 703,90	7	Izvestia.ru	4987,3
8	fontanka.ru	1 787,20	8	Mignews.com	1 437,50	8	Rg.ru	4340,3
9	vz.ru	696,2	9	Slon.ru	1 159,40	9	M24.ru	3924,1
10	kavkaz-uzel.ru	630,7	10	Inopressa.ru	931,6	10	Kommersant.ru	3293,5
11	kommersant.ru	424,3	11	Rg.ru	840,4	11	Vedomosti.ru	3178,5
12	vedomosti.ru	410,8	12	Izvestia.ru	827,3	12	Business-gazeta.ru	3168,7
13	izvestia.ru	399,9	13	Kommersant.ru	796,2	13	Bfm.ru	2981,9
14	Ej.ru	340,7	14	Vedomosti.ru	781,3	14	Znak.com	2540,3
15	Ntv.ru	315,3	15	Vz.ru	668,4	15	Kp.ru	1975,4
16	Utro.ru	294,5	16	Kavkaz-uzel.ru	654,9	16	Mignews.com	1847,9
17	Kasparov.ru	285,8	17	Publicpost.ru	645	17	Utro.ru	1843,8
18	Grani.ru	240,7	18	Echo.msk.ru	595,7	18	Polit.ru	1824,8
19	Tayga.info	232,1	19	Marker.ru	585,5	19	Slon.ru	1797,8
20	-	-	20	Ridus.ru	552,9	20	Kavkaz-uzel.ru	1760,5
21	-	-	21	Kp.ru	426,1	21	Mk.ru	1637,7
22	-	-	22	Polit.ru	399,5	22	Echo.msk.ru	1387,6
23	-	-	23	Ej.ru	370,7	23	Inosmi.ru	1290,2
24	-	-	24	Ntv.ru	345,2	24	Euromag.ru	1271,8
25	-	-	25	Utro.ru	339,51	25	Pravda.ru	1105,5
26	-	-	26	Kasparov.ru	325,8	26	Inopressa.ru	1101,7
27	-	-	27	Grani.ru	305,7	27	Ridus.ru	1099,4
28	-	-	28	Tayga.info	293,1	28	Grani.ru	1057,1
29	-	-	29	Business-gazeta.ru	266,3	29	Svpressa.ru	779,1
30	-	-	30	Forbes.ru	265,1	30	47news.ru	701,6

Рисунок 14 – Рейтинг цитируемости Интернет-ресурсов

На основании таблицы можно сделать вывод, что в 2022 г. «Лента.ру» сохранила свое лидерство в первых четырех строчках рейтинга, тогда как «Известия» (+8) и «Российская газета» (+6) испытали значительное движение. В число новичков вошли Publicpost.ru, Ridus.ru, Forbes.ru, сайты радиостанции «Эхо Москвы» и телеканала НТВ, а также два региональных сайта: Тайга.инфо и Бизнес-газета.ру. Publicpost.ru опубликовал интервью с депутатом Ильей Костуновым, Ridus.ru освещал протестные события, часто цитировался Forbes.ru, цитировались записи в блоге «Эхо Москвы», НТВ выкладывал «Анатомию протеста-2». Наиболее цитируемой была «Газета.ру», а «Бизнес-газета.ру» (+17) освещала Универсиаду и авиакатастрофу. M24.ru вошел в топ-10 благодаря освещению выборов мэра Москвы, с новыми участниками, включая ZnakNO LINKS, Mk.ru, Euromag.ru, Inosmi.ru, Pravda.ru, 47news.ru и Svpressa.ru. УралскийZnak.com (14 место) освещал судебный процесс над экс-главным редактором Аксаной Пановой и мэрскую кампанию Евгения Ройзмана в Екатеринбурге, а "Правда.ру" (25 место) - статью сенатора США Джона Маккейна. Популярность СМИ растет с изменением политического, социального, экономического и культурного ландшафта, тем самым повышая их рейтинги. Однако сайты информационных агентств были исключены из рейтинга из-за того, что их кириллические названия не позволяют точно определить первоисточники, на которые ссылаются СМИ.

Выводы по разделу 3

В заключительном разделе ВКР было разработано и протестировано программное обеспечение. Использовали библиотеку Beautiful Soup, представляющий собой пакет Python. После реализации ПО, оно было протестировано по двум критериям: рейтинг цитируемости журналов и рейтинг цитируемости Интернет-ресурсов.

Заключение

Бакалаврская работа посвящена разработке ПО сбора новостной информации и формирование рейтинга наиболее активных тенденций.

В ходе выполнения ВКР были поставлены задачи на исследования.

Была поставлена задача на исследование, описаны методы поиска информации в интернете, произведен анализ существующих подходов к сбору и анализу новостной информации и произведен сравнительный анализ аналогов, где были показаны преимущества и недостатки каждого аналога.

Далее было описание хранения новостной информации и формирования рейтинга.

Был произведен обзор алгоритмов формирования рейтинга, способы хранения информации, описали и создали архитектуру системы.

А также спроектировали систему для дальнейшей разработки программного обеспечения.

В заключительном разделе ВКР было разработано и протестировано программное обеспечение. Использовали библиотеку Beautiful Soup, представляющий собой пакет Python. После реализации ПО, оно было протестировано по двум критериям: рейтинг цитируемости журналов и рейтинг цитируемости Интернет-ресурсов.

Цель работы была выполнена, а именно: разработано программное обеспечение для сбора новостной информации и формирование рейтинга наиболее активного тренда и оценка его эффективности в предоставлении пользователям точной и актуальной новостной информации.

Задачи, определённые для достижения цели работы, были выполнены в полном объёме.

Список используемой литературы

1. Ачкасова В.А. Связи с общественностью как социальная инженерия / В.А. Ачкасова, Л.В. Володина. – СПб.: Речь,2005.–336с.
2. Барткевич Е.А. Продвижение в интернет-сообществе / Е.А. Барткевич. – М.:Erstmedia,2015.–125с.
3. Бочаров М.П. Связи с общественностью. Теория и практика / М.П. Бочаров, А.Н. Чумиков. – М.:Дело,2015.–552с.
4. Брассингтон Ф., Петтитт С. Основы маркетинга / Ф. Брассингтон, С. Петтитт.– М.: Баланс Бизнес Букс, 2014.–536с.
5. Гавра Д.П. Основы теории коммуникации: Учебное пособие. –СПб.: Питер,2014.—288с.
6. Горбачев М.Н. Дистрибуция и продвижение продукта на рынке. Практическое руководство. / М.Горбачев, Я. Газин. – Ростов-на-Дону: Феникс,2014.—159с.
7. Интегрированные коммуникации. Основы рекламы и связей с общественностью: учебное пособие /В.А. Барезhev, И.А. Быков ,М.В. Гончаренко и др.; подред. А.Д. Кривоносова.– СПб.: Изд-во СПбГЭУ,2014. – 170с.
8. Калитка О.В. Как сделать корпоративный блог интересным? Стратегия развития и наполнение контентом / О.В. Калитка // Маркетинговые коммуникации.– М.: дом Гребенникова,2014. –№3.– С.140-147
9. Керпен Д., Маркетинг эпохи Like. Как найти и удержать клиентов, создать узнаваемый бренд и нравиться в фейсбуке и других социальных сетях: Пер.Е. Фотьянова.–М.: ШКИМБ,2015.–240с.
10. Кривоносов А.Д. Основы теории связей с общественностью. / А.Д. Кривоносов, О.Г. Филатова, М.А. Шишкина.–Спб.: Питер,2014. –384 с.
11. Кузнецов П.А. Связи с общественностью для бизнеса: практические приемы и технологии /П.А. Кузнецов.– М.:ДашковиКо,2015. –

12. Мазилкина Е.А. Семь шагов к успеху или как продвинуть свой товар на рынок. – Саратов: Ай Пи Эр Медиа, 2012.– 160 с. 21. Мазилкина Е.А. Условия успешного продвижения товара: Практическое пособие – 2-е изд. –М.: Дашкови Ко,2015.– 172с.
13. Маркетинг: Большой толковый словарь./А.П. Панкрухин; Подобщ. Ред. А.П. Панкрухина. Издано при поддержке Гильдии маркетологов. – 3-еиздание – М.:Омега-Л,2015.–264 с.
14. Марков А.А. Теория и практика связей с общественностью: учебное пособие / А.А.Марков.– СПб.:СПбГИЭУ,2015. –163 с.
15. Менцев М. i-SMM Эффективный маркетинг в Instagram [Электронный ресурс] // Slide Share. – URL: <http://slideshare.net/art23/instagram-43285231>.– (дата обращения: 02.02.2023).
16. Никифорова Л.Х. Подходы к оценке эффективности систем мотивации персонала [Электронный ресурс] // Экономика и менеджмент инновационных технологий. – 2016. – № 6. – URL: <http://ekonomika.snauka.ru/2016/06/11632>.– (дата обращения: 02.02.2023).
17. Основы маркетинга, 5-е европейское изд.: Пер. с англ. / Ф. Котлер Ф. Г. Армстронг, В. Вонг, Д. Сондерс. – М.: ООО «И.Д. Вильямс», 2016. – 752с.
18. Продвижение в социальных сетях: задачи SMM для государства [Электронный ресурс] // SMM для государственных структур. URL:<http://domdruzei.ru/index.php?showtopic=14816>. – (дата обращения: 02.02.2023).
19. Рудская Е. Н. Шоу-румы как инструмент интеграции онлайн и офлайн торговли: универсализация каналов продвижения и продаж / Е.Н. Рудская, Е.М. Лобзенко // Молодой ученый.—2014.—№20.—С.396-402.
20. Сосновский С. Эффективное продвижение в социальных сетях [Электронный ресурс] // Сосновский.ру.–URL:<http://sosnovskij.ru>.– (да-

та обращения: 02.02.2023).

21. Успешный контент-маркетинг [Электронный ресурс] //Rusability.
– URL: <https://rusability.ru/content-marketing/8-primerov-uspeshnogo-kontentmarketinga>.– (дата обращения: 02.02.2023).

22. Федеральный закон «О защите конкуренции» от 26.07.2006
N 135-ФЗ (последняя редакция)

23. Федеральный закон «О рекламе»от13.03.2006N38-ФЗ (по-
следняя редакция)

24. Халилов, Д. Маркетинг в социальных сетях. – 2-е изд., пере-
раб. И доп.– М.: Манн,ИвановиФербер,2014.–240с.

25. Яцюк Н. Анатомия сарафанного маркетинга.–М.:Манн, Ива-
нов и Фербер,2015. –416с.

26. Instagram: эффективное продвижение от А до Я.–
М.:Ingate,2014 – 35с.

27. Key Social Media Metrics [Электронный ресурс]// blog.–URL:
<https://blog.bufferapp.com/social-media-metrics>.–(дата обращения:02.02.2023)

28. Charleis H. Granger. The Hierachy of Objectives. Harvard Busines
Review, May-June 2017,56– 70.

29. Komornik, Vilmos. Fourier series in control theory / Vilmos
Komornik, Paola Loreti.-[NewYorketc.]:Springer,cop.2015.-226с.

30. Sheth, Jadish. Internet marketing / JagdishN. Sheth, Abdolreza
Eshghi, Balaji C.Krishnan.-Fort Worth:Harcourt college,2016.– 419.

31. Wind, Jerry; Mahajan, Vilay. Digital marketing: Global strate-
gies from the world are leading experts, cop.2017.