|  |
| --- |
| Пермский филиал федерального государственного автономного образовательного учреждения высшего образования  «Национальный исследовательский университет  «Высшая школа экономики»    *[Факультет социально-экономических и компьютерных наук](https://perm.hse.ru/scs/)* |
| Белов Егор Александрович  **РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ ВЫЧИСЛЕНИЯ МЕТРИК В СОЦИАЛЬНЫХ МЕДИА**  *Курсовая работа*  студента образовательной программы «Программная инженерия» по направлению подготовки *09.03.04 Программная инженерия*  Руководитель  к.ф.-м.н., доцент кафедры информационных технологий в бизнесе  Е.Б. Замятина |

Пермь, 2024 год

Аннотация

Автор: Белов Егор Александрович.

Название работы: Разработка приложения для вычисления метрик в социальных медиа.

В работе приведено описание алгоритмов и программных средств для вычисления метрик в социальных медиа. Вычисление метрик необходимо для того, чтобы дать оценки пользователям социальных сетей, выявить наиболее активных, количество их связей с другими пользователями и т.д. Метрики используют для проведения эффективных маркетинговых операций.

Данная работа содержит процесс анализа и разработки приложения для вычисления метрик в социальных медиа. Курсовая работа содержит 62 страницу и 6 глав. В данной работе содержится 25 рисунков, 3 приложения, 5 таблиц.

Оглавление

[Аннотация 2](#_Toc162033181)

[Оглавление 3](#_Toc162033182)

[Введение 5](#_Toc162033183)

[Глава 1. Анализ предметной области 8](#_Toc162033184)

[1.1. Использование социальных сетей для распространения информации 8](#_Toc162033185)

[1.2. Анализ метрик в социальных медиа 10](#_Toc162033186)

[1.3. Назначение системы 15](#_Toc162033187)

[1.4. Виртуальные и реальные сети 16](#_Toc162033188)

[1.5. Онтологии 18](#_Toc162033189)

[1.6. Анализ существующих решений 19](#_Toc162033190)

[Глава 2. Проектирование системы 23](#_Toc162033191)

[2.1. Бизнес-процессы и прецеденты 23](#_Toc162033192)

[2.2. Системы имитационного моделирования 28](#_Toc162033193)

[Глава 3. Моделирование системы виртуальной сети 32](#_Toc162033194)

[3.1. Требования к модели 32](#_Toc162033195)

[3.2. Реализация модели 34](#_Toc162033196)

[Глава 4. Реализация приложения 38](#_Toc162033197)

[4.1. Выбор средств реализации 38](#_Toc162033198)

[4.2. Реализация сервиса для сбора информации 39](#_Toc162033199)

[4.3. Реализация сервиса для вычисления метрик 46](#_Toc162033200)

[4.4. Реализация приложения для просмотра результатов 52](#_Toc162033201)

[Глава 5. Тестирование 55](#_Toc162033202)

[Глава 6. Внедрение 57](#_Toc162033203)

[Заключение 59](#_Toc162033204)

[Библиографический список 61](#_Toc162033205)

Введение

Работа посвящена созданию приложения для вычисления метрик в социальных медиа.

В современном мире социальные сети стали неотъемлемой частью жизни людей. Ежедневно миллионы пользователей проводят в них время, общаясь с друзьями, делясь новостями и интересами, а также потребляя разнообразный контент. В этом контексте социальные медиа выступают важной платформой для бизнеса, который активно использует их для продвижения своих товаров и услуг.

Реклама в социальных сетях, будучи мощным инструментом, требует точного таргетирования, чтобы достичь максимальной эффективности. Анализ поведения пользователей и вычисление релевантных метрик становятся необходимыми компонентами успешной рекламной стратегии. В данном контексте разработка приложения, способного автоматизировано выявлять перспективных пользователей для рекламы, представляет собой важный шаг в оптимизации маркетинговых усилий.

В этом свете, настоящая курсовая работа ориентирована на создание инструмента, который не только позволит вычислять метрики в социальных медиа, но и определять наиболее релевантных пользователей для целевой рекламы. Такой подход сделает рекламные кампании более таргетированными и эффективными, отвечая на вызовы современной маркетинговой парадигмы.

Оценка эффективности рекламных кампаний в социальных сетях невозможна без отслеживания ключевых метрик.

В контексте разработки приложения для вычисления метрик в социальных медиа представлены следующие направления научной новизны:

* Разработка алгоритма для поиска наиболее релевантных для рекламы пользователей.
* Использование модели виртуальной сети в системе информационного моделирования.
* Использование онтологического подхода, который позволяет сохранять знания о пользователях социальных сетей, их интересов и связей.

Объектом работы являются характеристики пользователей социальных медиа.

Предметом данной работы являются алгоритмы определения метрик.

Основная цель работы заключается в создании инструмента, способного автоматизировано собирать информацию о пользователях социальной сети, вычислять различные метрики и на основе полученных данных определять наиболее влиятельных пользователей для лучшего распространения рекламы. Этот подход позволит повысить эффективность маркетинговых кампаний, сделав их более таргетированными и направленными на целевую аудиторию.

Кроме того, в рамках работы предусмотрено создание модели виртуальной сети в системе информационного моделирования. Это дополнительный инструмент, способствующий более глубокому анализу взаимосвязей между пользователями, их интересами и поведенческими паттернами. Создание такой модели позволит более точно определить сегменты аудитории, на которые следует ориентироваться при проведении рекламных кампаний.

Цель данной работы достигается выполнением следующих задач:

* Провести анализ существующих методов сбора данных из социальных сетей.
* Разработать алгоритмы для вычисления ключевых метрик пользователей социальных сетей.
* Разработать алгоритм для поиска наиболее релевантных для рекламы пользователей.
* Использовать модель виртуальной сети в системе информационного моделирования и создать ее веб-интерфейс.
* Реализовать приложение, используя выбранные методы и алгоритмы.
* Провести тестирование приложения и анализ его эффективности.

Разработка приложения для автоматического сбора и анализа данных о пользователях социальных сетей позволит:

* Сократитьвремяитрудозатраты **-** приложение будет автоматически собирать данные, что позволит пользователям экономить время и силы.
* Повыситьэффективностьрекламныхкампаний **-** приложение позволит пользователям более точно таргетировать рекламные объявления, что приведет к повышению их эффективности.
* Снизитьрасходы **-** приложение будет бесплатным, что позволит пользователям экономить деньги на использовании сторонних инструментов.

Таким образом, данная работа представляет собой комплексный подход к оптимизации рекламных стратегий в социальных медиа, разделенный на несколько ключевых этапов. В первую очередь, в ней реализуется разработка приложения для вычисления метрик, необходимых для анализа эффективности рекламных кампаний. Далее, создается модель виртуальной сети в системе информационного моделирования, которая включает в себя визуализацию виртуальной сети. Кроме того, проводится анализ реальной сети для вычисления метрик на основе фактических данных. Наконец, осуществляется сбор информации в онтологии и отслеживание через журналы событий, что позволяет более полно и точно анализировать результаты и принимать обоснованные решения по оптимизации рекламных стратегий.

Глава 1. Анализ предметной области

Этот раздел фокусируется на первоначальном этапе исследования, который включает в себя анализ теоретических материалов, касающихся особенностей изучения процесса передачи информации в социальных сетях.

1.1. Использование социальных сетей для распространения информации

Интернет — это важная часть современной жизни. Он стал незаменимым средством коммуникации, источником информации, площадкой для самовыражения и многое другое. Большинство людей пользуются интернетом ежедневно и практически ни на что не могут обойтись без его помощи. Так что же делает интернет таким популярным?

Во-первых, интернет предоставляет доступ к огромному количеству информации. Он является огромной библиотекой знаний, где можно найти ответы на любые вопросы, изучить новые темы и получить актуальные новости. Люди используют интернет для образования, самообразования и развлечения. Благодаря интернету, информация стала легко доступной и распространяется во всемирном масштабе.

Во-вторых, интернет обеспечивает возможность связи и общения. Благодаря социальным сетям, мессенджерам и электронной почте, мы можем легко общаться с друзьями, родственниками и коллегами, независимо от расстояния. В интернете можно также найти единомышленников, присоединиться к группам или сообществам с общими интересами. Это открывает огромные возможности для социального взаимодействия и создания новых связей.

Социальные сети предоставляют пользователям возможность создавать персональные профили, заполнять их информацией о себе, загружать фотографии и видео, а также делиться своими мыслями и впечатлениями в формате постов.

Комментарии и лайки к постам позволяют людям выражать свое мнение, а также демонстрировать свою поддержку или несогласие с высказываниями других пользователей.

Общение через интернет позволяет преодолеть пространственные и временные ограничения. Люди могут общаться с друзьями и родственниками, находящимися в других городах или даже странах, без необходимости путешествовать и тратить деньги на звонки или почтовые отправления.

Социальные сети предоставляют возможность общаться не только с близкими людьми, но и находить интересные группы и сообщества, в которых можно обсуждать общие интересы, делиться полезной информацией, получать советы от экспертов и просто общаться с единомышленниками.

Реклама в социальных сетях предоставляет предпринимателям ряд возможностей. Размещение рекламы в социальных медиа способствует формированию благоприятного имиджа предприятия и эффективному расходованию ресурсов, поскольку онлайн-продвижение является наиболее экономичным методом рекламы товаров и услуг. Кроме того, такой вид рекламы позволяет точечно нацелиться на целевую аудиторию, которая уже заинтересована в предлагаемых продуктах или услугах. Преимущества размещения рекламы в социальных сетях включают возможность доступа к личной информации пользователей, такой как пол, возраст, уровень образования, увлечения и так далее. Это дает возможность исследовать профили пользователей, составляющих целевую аудиторию, и выяснять их мнение о продукции и услугах компании. Кроме того, пользователи могут активно обмениваться информацией друг с другом, что способствует органическому распространению рекламного контента, создавая так называемый "вирусный маркетинг".

Другие преимущества использования социальных сетей в качестве платформы для размещения рекламы включают следующее:

* Обширная аудитория пользователей, что обеспечивает широкий охват потенциальных клиентов.
* Более быстрое продвижение рекламы по сравнению с альтернативными методами.
* Низкие затраты или их полное отсутствие, что делает рекламу в социальных сетях привлекательной для бизнеса с любым бюджетом.
* Возможность непосредственного взаимодействия с потенциальными клиентами прямо на платформах социальных сетей, используя доступные всем пользователям инструменты коммуникации.

Среди пользователей социальных сетей можно выделить влиятельных лидеров мнений. Часто в маркетинге используется стратегия, включающая знаменитостей для распространения информации. Обычные пользователи чаще доверяют информации, полученной от таких лидеров. Это означает, что маркетинговое воздействие распространяется не только на самих знаменитостей, но и на их многочисленных последователей, увеличивая эффективность кампании. Для использования этого инструмента маркетинга можно применять различные методы влияния на лидеров мнений, такие как материальное вознаграждение или бартерные сделки.

Целью данной работы является вычисление метрик, которые будут определять наиболее влиятельного и релевантного пользователя, благодаря которому распространение рекламы будет наиболее оптимально и эффективно.

1.2. Анализ метрик в социальных медиа

Метрики позволяют отслеживать эффективность работы в социальных сетях и планировать выпуск контента или проведение рекламных кампаний [2].

Взаимная направленность — это свойство, которое указывает на то, является ли отношение между вершинами бинарным. Другими словами, это свойство показывает, является ли связь двунаправленной. В контексте социальных сетей, например, взаимная направленность может означать, что пользователь подписан на другого пользователя и этот другой пользователь также подписан на него. Это свойство может быть полезным для анализа связей и отношений между пользователями в социальной сети. Взаимная направленность между вершинами в графе может быть определена с использованием формулы коэффициента взаимной направленности (Reciprocity coefficient). Формула для вычисления коэффициента взаимной направленности (R) выглядит следующим образом:

*R=A+B/2⋅T,*

Где:

* *T* - количество двунаправленных связей (отношений), то есть связей, где вершина A связана с вершиной B и наоборот;
* *A* - количество связей, исходящих из вершины A;
* *B* - количество связей, входящих в вершину B.

Коэффициент взаимной направленности R принимает значения от 0 до 1. Значение ближе к 1 указывает на более высокую взаимную направленность между вершинами, что означает более сильные взаимосвязи и отношения в графе.

​Гомогенность — это свойство, которое указывает на степень появления связей между акторами, которые имеют схожие характеристики, такие как пол, возраст или интересы. В контексте социальных сетей, гомогенность может означать, что пользователи с похожими интересами или предпочтениями чаще взаимодействуют друг с другом, например, подписываются на одни и те же страницы или группы, комментируют и лайкают посты друг друга. Гомогенность в графе социальной сети может быть оценена с помощью коэффициента гомогенности (Homophily coefficient). Формула для расчета коэффициента гомогенности (H) выглядит следующим образом:

*H = T / N,*

Где:

* *T* - количество связей между акторами с схожими характеристиками (например, с одинаковыми интересами);
* *N* - общее количество связей в графе.

Коэффициент гомогенности *H* также принимает значения от 0 до 1. Значение ближе к 1 указывает на более высокую гомогенность в графе, что означает, что акторы с схожими характеристиками взаимодействуют чаще, чем акторы с различными характеристиками.

Транзитивность связей — это свойство, которое указывает на увеличение вероятности появления связей между акторами, у которых есть связи с одними и теми же вершинами. Другими словами, если у актора А есть связь с актором В, и у актора В есть связь с актором С, то вероятность того, что у актора А будет связь с актором С, увеличивается. В контексте социальных сетей, это может означать, что если пользователь А подписан на пользователя В, и пользователь В подписан на пользователя С, то вероятность того, что пользователь А также подпишется на пользователя С, возрастает. Транзитивность связей в графе социальной сети может быть измерена с помощью коэффициента транзитивности (Transitivity coefficient). Формула для расчета коэффициента транзитивности (T) выглядит следующим образом:

*T = 3\* количество замкнутых троек / количество троек*,

Где

* *Количество замкнутых троек* — это количество троек вершин (A, B, C), где существуют связи (A, B), (B, C) и (A, C);
* *Количество троек* - общее количество троек вершин в графе.

Коэффициент транзитивности T также принимает значения от 0 до 1. Значение ближе к 1 указывает на более высокую транзитивность связей в графе, что означает более вероятное появление связей между акторами, у которых уже есть общие связи.

Разница в распределении — это свойство, которое указывает на то, что у одних акторов есть большое количество связей, а у других - минимальное. Важным в данном случае является феномен "богатый становится богаче", который приводит к высокой дисперсии вершин. В контексте социальных сетей, это может означать, что некоторые пользователи имеют большое количество подписчиков и взаимодействий, в то время как другие пользователи имеют минимальное количество подписчиков и взаимодействий. Разница в распределении связей в графе социальной сети может быть оценена с помощью различных метрик, таких как коэффициент Гини (Gini coefficient) или коэффициент Робин Гуда (Robin Hood index).

Коэффициент Гини измеряет степень неравенства в распределении связей между акторами в графе. Чем ближе коэффициент Гини к 1, тем выше неравенство в распределении связей. В контексте социальных сетей, высокий коэффициент Гини может указывать на явление "богатый становится богаче", где некоторые пользователи имеют значительно большее количество связей, чем другие.

Коэффициент Робин Гуда, наоборот, оценивает, сколько связей необходимо перераспределить с богатых акторов (тех, у которых много связей) на бедных акторов (тех, у которых мало связей), чтобы достичь более равномерного распределения. Чем выше коэффициент Робин Гуда, тем больше связей необходимо перераспределить.

Использование этих метрик позволяет оценить степень неравенства в распределении связей между акторами в социальной сети и выявить феномен "богатый становится богаче".

Центральность — это метрика, которая позволяет определить значительность или влияние определенного узла или группы в сети. В контексте социальных сетей, центральность может указывать на то, насколько влиятельным является пользователь или группа пользователей. Например, пользователь с высокой центральной позицией может иметь большое количество подписчиков, получать много лайков и комментариев, а также иметь большое количество связей с другими пользователями. Центральность в графе социальной сети может быть оценена с помощью различных метрик:

* **Центральность по степени (degree centrality)**- эта метрика определяет, насколько узел важен, основываясь на количестве связей, которые он имеет с другими узлами. Пользователь с высокой центральностью по степени обычно имеет большое количество подписчиков и/или связей с другими пользователями.
* **Центральность по посредничеству (betweenness centrality)** - эта метрика оценивает, насколько узел важен для коммуникации между другими узлами в сети. Пользователь с высокой центральностью по посредничеству может играть ключевую роль в передаче информации или контроле над потоком коммуникации в сети.
* **Центральность по близости (closeness centrality)** - эта метрика измеряет, насколько близко узел к другим узлам в сети. Пользователь с высокой центральностью по близости обычно может быстро достичь других пользователей в сети и имеет прямой доступ к большому количеству информации.

Ассортативность — это свойство, которое указывает на склонность к образованию связей между вершинами большой степени. В контексте социальных сетей, ассортативность может означать, что пользователи с большим количеством подписчиков или взаимодействий чаще взаимодействуют друг с другом, чем с пользователями с меньшим количеством подписчиков или взаимодействий.

Диаметр сети — это метрика, которая указывает на максимальное количество шагов, необходимых для того, чтобы соединить две самые удаленные вершины в сети. В контексте социальных сетей, диаметр сети может указывать на то, насколько быстро информация может распространяться между пользователями. Например, если диаметр сети небольшой, то информация может быстро распространяться между пользователями, в то время как если диаметр сети большой, то информация может распространяться медленнее.

PageRank — это алгоритм, разработанный компанией Google, который используется для оценки важности веб-страниц на основе структуры ссылок между ними. Этот алгоритм также может быть применен к другим типам сетей, включая социальные сети, для определения значимости узлов в сети.

Формула PageRank выглядит следующим образом:

*PR*(*u*)=(1−*d)/N* ​ +*d* ∑*v*∈ *Bu* ​​ *PR*(*v*) / *L*(*v*),​

Где:

* *PR*(*u*) - PageRank узла *u*;
* *d* - коэффициент демпфирования, который обычно принимает значение около 0.85;
* *N* - общее количество узлов в сети;
* *Bu*​ - множество узлов, которые ссылается на узел *u*;
* *L*(*v*) - количество исходящих ссылок узла *v*.

Эта формула описывает итеративный процесс, в котором каждый узел получает оценку PageRank, основанную на оценке значимости узлов, ссылающихся на данный узел, и их собственной оценке PageRank. Алгоритм продолжает итерации до тех пор, пока изменения оценок PageRank между итерациями не станут достаточно малыми.

PageRank позволяет определить важность узлов в сети на основе их связей с другими узлами, что делает его мощным инструментом для анализа структуры сетей и идентификации ключевых узлов.

В этом разделе были представлены различные метрики и характеристики, позволяющие анализировать структуру и динамику социальных сетей.

Использование этих метрик может помочь в разработке эффективных маркетинговых стратегий, изучении распространения информации, обнаружении лидеров мнений и моделировании социальных сетей.

1.3. Назначение системы

Данная работа содержит три основные части, такие как создание модели виртуальной сети в системе информационного моделирования, создание базы знаний в онтологии и отслеживание информации в журнале событий, реализация сервиса по сбору информации о пользователях социальной сети и подсчет метрик.

**Создание модели виртуальной сети в системе информационного моделирования** предполагает разработку модели, которая абстрагирует и визуализирует виртуальную сеть пользователей в социальных медиа. Это включает в себя построение графа, где узлы представляют пользователей, а связи между узлами отображают связи и взаимодействия между пользователями (например, дружба, подписка, комментарии и т. д.). Использование систем информационного моделирования обеспечивает возможность анализа структуры сети, выявления ключевых узлов и понимания паттернов взаимосвязей между пользователями.

**В процессе создания базы знаний в онтологии и отслеживания информации в журнале событий** создается база знаний, использующая онтологии для описания концепций, связей и атрибутов пользователей и их взаимодействий в социальной сети. Журнал событий представляет собой систему, которая отслеживает и регистрирует действия и события, происходящие в социальной сети, такие как посты, комментарии, лайки и т. д.

Реализация сервиса по сбору информации о пользователях социальной сети и подсчет метрик фокусируется на создании сервиса, который собирает информацию о пользователях социальной сети, их взаимодействиях и активностях. После сбора данных сервис анализирует информацию и вычисляет различные метрики, такие как центральность по степени, посредничеству, PageRank и другие, которые помогают оценить важность и влиятельность пользователей в сети.

Для данной работы необходимо определить как функциональные, так и нефункциональные требования.

Функциональные требования:

* Создание модели виртуальной сети - возможность создания структуры виртуальной сети с помощью графического интерфейса, возможность добавления узлов и связей между ними для отображения взаимосвязей в сети, возможность проведения анализа виртуальной сети для выявления ключевых узлов и структур.
* Создание базы знаний в онтологии - возможность определения классов и свойств в онтологии, возможность добавления данных и отношений между классами в онтологию, возможность использования онтологии для описания концепций и взаимосвязей между ними.
* Отслеживание информации в журнале событий - возможность записи событий, происходящих в системе, в журнал событий, возможность просмотра и анализа журнала событий для выявления важных событий и трендов.
* Реализация сервиса по сбору информации о пользователях социальной сети и подсчет метрик - возможность сбора информации о пользователях из социальной сети с помощью API, возможность анализа данных о пользователях для вычисления различных метрик, таких как охват, вовлеченность и т.д., возможность предоставления результатов анализа в удобном формате для дальнейшего использования.

Нефункциональные требования:

* Система должна быть достаточно быстрой для обработки больших объемов данных.
* Надежность - система должна быть стабильной и надежной, чтобы обеспечить непрерывную работу.
* Масштабируемость - система должна быть способна масштабироваться для работы с растущим объемом данных и пользователей.
* Удобство использования - интерфейс системы должен быть интуитивно понятным и удобным для пользователей.

1.4. Виртуальные и реальные сети

В мире социальных медиа взаимодействие пользователей и формирование связей играют ключевую роль в распространении информации и влиянии. Для анализа этих связей используются концепции виртуальных и реальных сетей. В данном разделе мы рассмотрим различия и взаимодействие между виртуальными и реальными сетями в контексте социальных медиа.

Виртуальные сети представляют собой абстрактные модели социальных связей и взаимодействий между участниками. Они часто используются в системах информационного моделирования для визуализации и анализа социальных структур.

Характеристики виртуальных сетей:

* Виртуальные сети могут быть построены на основе различных алгоритмов, включая случайные графы, модели малого мира и т. д.
* Узлы виртуальной сети представляют собой участников социальной сети, а ребра - их взаимосвязи и взаимодействия.
* Виртуальные сети могут быть использованы для моделирования поведения пользователей, анализа структуры социальных групп и оценки влияния участников.

Виртуальные сети могут использоваться для моделирования и анализа структуры и динамики реальных социальных сетей. Они позволяют исследовать различные сценарии и предсказывать поведение пользователей.

Реальные данные социальных медиа могут быть интегрированы в виртуальные сети для более точного моделирования и анализа. Это позволяет учесть реальные взаимосвязи и влияние пользователей на основе фактических данных.

Реальные сети представляют собой фактические социальные связи между реальными пользователями социальных медиа. Они формируются на основе действительных взаимодействий и коммуникаций между участниками.

**Характеристики реальных сетей:**

* Реальные сети основаны на реальных данных, полученных из социальных медиа платформ, таких как Facebook, Twitter, Instagram и других.
* Узлы в реальных сетях представляют собой реальных пользователей социальных медиа, а связи между ними отображают дружеские связи, подписки, комментарии и другие формы взаимодействия.
* Анализ реальных сетей позволяет понять структуру и поведение пользователей в социальных медиа, а также оценить эффективность рекламных кампаний и распространение информации.

Виртуальные и реальные сети в социальных медиа играют важную роль в анализе структуры социальных взаимодействий и оптимизации рекламных стратегий. Понимание различий и взаимодействия между ними позволяет компаниям более эффективно использовать данные социальных медиа для достижения своих целей.

1.5. Онтологии

Для исследования социальной сети первоочередно требуется анализ информационных каналов, то есть, исследование того, как пользователи взаимодействуют друг с другом и откуда они получают информацию. Эффективным методом представления социальной сети является использование графа, где узлы представляют собой пользователей или сообщества, а рёбра отображают каналы связи между ними.

Такой подход обладает рядом преимуществ:

* Позволяет выявить различные типы взаимосвязей между участниками.
* Пользователи описываются с помощью атрибутов: каждый человек или группа имеют свои характеристики.
* Совокупность разнообразных связей между участниками формирует структуру сети.
* Некоторые участники могут быть более тесно связаны друг с другом, чем с остальными.

Таким образом, становится ясным, что важно сохранить информацию о структуре социальной сети. Каждый элемент этой сети может быть представлен как узел, который связан с другими элементами и имеет свои характеристики. Однако необходимо, чтобы эту информацию могли понять как люди, так и компьютеры. Здесь возникает вопрос: в каком формате следует хранить информацию о реальной социальной сети?

Для решения этой задачи было предложено применять онтологии [3].

Онтология — это детальное определение концепции или предметной области. Под детальностью понимается способ представления понятий и их определений таким образом, чтобы они могли быть однозначно поняты как человеком, так и компьютером. Современные онтологии можно рассматривать как систему, состоящую из набора <A, S, T>, где:

* T - набор концепций или понятий, принадлежащих предметной области;
* S - набор различных типов связей, поддерживаемых в этой области;
* A - отображение между элементами из набора T и S.

Онтологии служат своего рода словарем для представления и обмена знаниями в определенной предметной области, а также определяют связи между терминами в этом словаре. Можно также представить онтологию в виде графа.

1.6.Анализ существующи**х** решений

Существует множество решений для анализа метрик пользователей социальной сети, каждое из которых предлагает свои уникальные возможности и подходы.

Facebook Analytics - платформа для анализа активности пользователей, поведения аудитории и эффективности рекламных кампаний на Facebook и Instagram. Предоставляет широкий спектр метрик, включая охват, вовлеченность, конверсии и другие. Имеет графический интерфейс и инструменты для создания отчетов и дашбордов.

Преимущества:

* Интегрирован с платформами Facebook и Instagram, что обеспечивает доступ к данным о целевой аудитории.
* Предоставляет обширный набор метрик для анализа, включая охват, вовлеченность, конверсии и другие.
* Имеет удобный интерфейс и инструменты для создания отчетов и дашбордов.

Недостатки:

* Ограничен доступом к данным только с платформ Facebook и Instagram.
* Не всегда предоставляет полную картину эффективности рекламных кампаний вне платформы Facebook.

Google Analytics предоставляет аналитику для веб-сайтов и приложений, включая социальные сети. Позволяет отслеживать трафик, взаимодействие пользователей, конверсии и другие метрики. Обладает функционалом для создания целевых аудиторий и настройки рекламных кампаний.

Преимущества:

* Предоставляет аналитику для веб-сайтов и приложений, включая социальные сети.
* Обладает обширным функционалом для отслеживания трафика, взаимодействия пользователей, конверсий и других метрик.
* Интегрирован с другими продуктами Google, такими как Google Ads.

Недостатки:

* Не обеспечивает доступ к некоторым данным социальных сетей, если они не связаны с веб-сайтом или приложением.
* Требует некоторого времени и опыта для полного освоения и настройки.

Sprout Social - инструмент для управления социальными медиа и анализа метрик. Позволяет отслеживать активность пользователей, эффективность публикаций, вовлеченность и рост аудитории. Имеет возможности для планирования и автоматизации публикаций.

Преимущества:

* Имеет удобный интерфейс и инструменты для управления социальными медиа.
* Позволяет анализировать активность пользователей, эффективность публикаций и рост аудитории.
* Обладает функционалом для планирования и автоматизации публикаций.

Недостатки:

* Может быть дорогим для небольших бизнесов или отдельных предпринимателей.
* Некоторые функции могут быть ограничены на более дешевых планах подписки.

Hootsuite - платформа для управления и анализа социальных медиа. Предоставляет инструменты для мониторинга активности пользователей, анализа конкурентов, планирования публикаций и управления рекламными кампаниями.

Преимущества:

* Позволяет управлять несколькими социальными медиа с одной платформы.
* Предоставляет инструменты для мониторинга активности пользователей, анализа конкурентов и управления рекламными кампаниями.

Недостатки:

* Некоторые функции доступны только на платных тарифах.
* Интерфейс может показаться сложным для новичков.

Buffer - инструмент для управления социальными медиа и анализа результатов. Позволяет планировать и публиковать контент, а также анализировать его эффективность по различным метрикам, таким как вовлеченность и рост аудитории.

Преимущества:

* Прост в использовании и предлагает базовый набор функций для управления социальными медиа.
* Позволяет анализировать эффективность публикаций по различным метрикам.

Недостатки:

* Может быть ограничен в функциональности для более крупных бизнесов или профессиональных маркетологов.
* Отсутствует поддержка для некоторых платформ социальных сетей.

Talkwalker - платформа для мониторинга социальных медиа и анализа репутации бренда. Предоставляет возможности для отслеживания упоминаний, анализа настроений, оценки эффективности кампаний и исследования конкурентов.

Преимущества:

* Позволяет мониторить упоминания бренда в социальных медиа и интернете в целом.
* Предоставляет аналитику для оценки эффективности кампаний и исследования настроений.

Недостатки:

* Может быть дорогим для малых и средних компаний.
* Некоторые функции могут быть избыточны для небольших бизнесов.

Общий вывод по рассмотренным решениям для анализа метрик пользователей социальной сети позволяет сделать следующие наблюдения:

* Каждое решение предлагает свой набор функций и инструментов для анализа метрик, что позволяет выбрать наиболее подходящее в зависимости от потребностей и целей бизнеса.
* Некоторые решения, такие как Buffer и Sprout Social, известны своим простым и интуитивно понятным интерфейсом, что делает их привлекательными для небольших компаний или предпринимателей.
* Решения типа AnyLogic и Google Analytics предлагают более широкий набор функций и гибкость в анализе данных, что делает их предпочтительными для крупных компаний с сложными потребностями в аналитике.
* Важным фактором при выборе решения является его стоимость. Некоторые решения, такие как Facebook Analytics и Google Analytics, предоставляют базовый функционал бесплатно, в то время как другие, например, Sprout Social и Talkwalker, требуют платных подписок.
* При выборе решения необходимо учитывать его возможность интеграции с другими инструментами и платформами, такими как социальные сети, CRM-системы и т.д.

Глава 2. Проектирование системы

2.1. Бизнес-процессы и прецеденты

Для успешной реализации приложения необходимо проанализировать основные бизнес-процессы, которые будут выполняться в системе. Ниже перечислены ключевые процессы.

Процесс сбора данных о пользователях социальной сети.

**Прецеденты данного процесса:**

1. **Автоматизированный сканинг профилей:**

* Описание: использование специализированных алгоритмов для сбора информации о профилях пользователей, включая их демографические данные, интересы и активность.
* Значимость: обеспечивает комплексную информацию о целевой аудитории для дальнейшего анализа и таргетирования рекламных кампаний.
* В таблице 2.1 описан прецедент **автоматизированный сканинг профилей.**

***Таблица 2.1 Описание прецедента сканнинг профилей***

|  |  |
| --- | --- |
| Акторы | Пользователь, Администратор |
| Краткое описание | **Автоматизированный сканинг профилей** |
| Триггер | Запуск сервиса по сканнингу |
| Основной поток | 1. После нажатия на кнопку, система сохранит информацию о пользователях в БД |
| Результат | Заполненная БД |

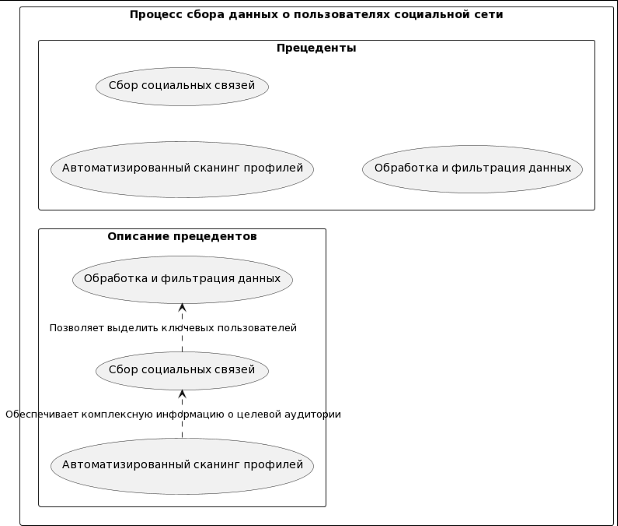
1. **Сбор социальных связей:**

* Описание: анализ связей между пользователями, таких как друзья, подписчики, лайки и комментарии, для выявления структуры социальной сети.
* Значимость: позволяет выделить ключевых пользователей и группы, а также оценить взаимодействие между ними.
* Таблица

1. **Обработка и фильтрация данных:**

* Описание: применение алгоритмов обработки данных для устранения дубликатов, ошибок и неполных записей, а также для классификации пользователей по различным критериям.
* Значимость: гарантирует высокое качество данных, необходимых для дальнейшего анализа и выявления ключевых пользователей.

Данный процесс представлен на рисунке 2.1.



***Рисунок 2.1. Процесс сбора данных о пользователях***

Процесс анализа метрик и оценки влияния пользователей.

Прецеденты:

1. Вычисление метрик активности:

* Описание: расчет различных метрик, таких как охват, вовлеченность, частота публикаций и прочие, для каждого пользователя на основе его активности в социальной сети.
* Значимость: позволяет оценить уровень активности и вовлеченности аудитории каждого пользователя.
* В таблице 2.2 описан прецедент Вычисление метрик активности

***Таблица 2.2 Описание прецедента Вычисление метрик активности***

|  |  |
| --- | --- |
| Акторы | Пользователь, Администратор |
| Краткое описание | Вычисление метрик активности |
| Триггер | Запуск сервиса по сканнингу |
| Основной поток | 1. После нажатия на кнопку, система сохранит информацию о метриках в БД |
| Результат | Заполненная БД |

1. Оценка влияния:

* Описание: анализ социальных связей и взаимодействия пользователей для определения степени их влияния в социальной сети.
* Значимость: идентифицирует ключевых пользователей, способных оказывать значительное влияние на свою аудиторию.

1. Применение методов машинного обучения:

* Описание: использование алгоритмов машинного обучения для анализа данных и выявления паттернов, связанных с влиянием пользователей.
* Значимость: предоставляет более точные прогнозы и оценки влияния пользователей на основе обширных объемов данных.

Данный процесс представлен на рисунке 2.2.

******

***Рисунок 2.2. Процесс анализа метрик и оценки влияния пользователей***

Процесс определения наиболее влиятельных пользователей.

**Прецеденты:**

1. Ранжирование пользователей:

* Описание: упорядочивание пользователей по уровню их влияния на основе рассчитанных метрик активности и оценки влияния.
* Значимость: выделяет наиболее значимых пользователей для использования в рекламных кампаниях и других маркетинговых стратегиях.

1. Анализ графа социальной сети:

* Описание: применение методов графового анализа для выявления центральных узлов и сообществ в социальной сети.
* Значимость: позволяет идентифицировать ключевых активистов и группы пользователей с наибольшим влиянием на остальных участников сети.

Данный процесс представлен на рисунке 2.3.



***Рисунок 2.3. Процесс определения наиболее влиятельных пользователей***

Процесс мониторинга и анализа результатов.

**Прецеденты:**

1. **Постоянный мониторинг активности:**

* Описание: регулярное отслеживание изменений в активности пользователей и влиянии их действий на аудиторию.
* Значимость: обеспечивает оперативную реакцию на изменения в сети и корректировку стратегий маркетинговых кампаний.
* В таблице 2.3 описан прецедент **постоянный мониторинг активности**

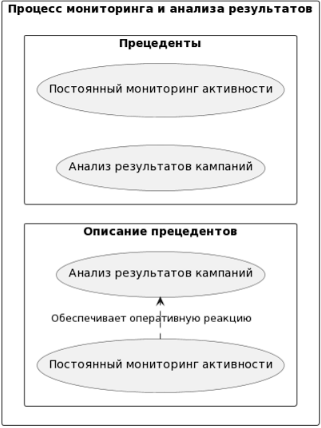
***Таблица 2.3 Описание прецедента постоянный мониторинг активности***

|  |  |
| --- | --- |
| Акторы | Пользователь, Администратор |
| Краткое описание | **Постоянный мониторинг активности** |
| Триггер | Запуск сервиса по мониторингу |
| Основной поток | 1. После нажатия на кнопку, система сохранит начинает отслеживать и сохранять информацию |
| Результат | Онтология и журнал событий |

1. **Анализ результатов кампаний:**

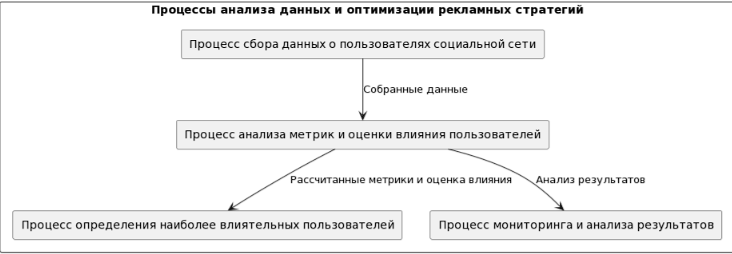
* Описание: оценка эффективности проведенных рекламных кампаний на основе собранных данных и метрик успеха.
* Значимость: предоставляет понимание результатов и выявление областей для улучшения стратегий маркетинга в будущем.

Данный процесс представлен на рисунке 2.4.



***Рисунок 2.4. Процесс мониторинга и анализа результатов***

На рисунке 2.5 представлена полная диаграмма бизнес-процессов.



***Рисунок 2.5. Бизнес-процессы***

Каждый из этих процессов и прецедентов играет важную роль в достижении цели повышения эффективности маркетинговых кампаний в социальных медиа.

2.2. Системы имитационного моделирования

Информационное моделирование систем (SIM) представляет собой процесс создания моделей сложных взаимосвязанных систем. Информационные модели систем — это цифровые аналоги связанных систем, таких как приборы и системы управления, электроснабжение и связь. Объекты, смоделированные в SIM, соответствуют объектам физической системы в пропорции 1:1. Компоненты, связи и функции определяются и моделируются так, как они существуют в реальном мире.

SIM играет важную роль в инженерии, науке и различных отраслях промышленности по нескольким причинам:

* Системы информационного моделирования позволяют инженерам и дизайнерам создавать цифровые модели сложных систем до их физической реализации. Это помогает сократить время и затраты на разработку новых продуктов и технологий, а также идентифицировать потенциальные проблемы ещё на этапе проектирования.
* Используя СИМ, исследователи могут анализировать поведение системы в различных условиях и оптимизировать её работу. Это позволяет предвидеть и управлять различными аспектами системы, такими как производительность, энергопотребление, надежность и безопасность.
* SIM позволяет проводить виртуальные тесты и симуляции, которые могут быть недоступны или слишком дороги в реальной среде. Это позволяет тестировать новые идеи, стратегии и процессы без риска повреждения оборудования или угрозы для безопасности.
* Путем создания моделей различных сценариев и прогнозирования их последствий, SIM помогает оценивать и управлять рисками, связанными с различными аспектами системы. Это может быть особенно важно в критических отраслях, таких как медицина, авиация и энергетика.
* SIM также широко используется в образовательных целях для обучения студентов и профессионалов в различных областях науки, техники и управления. Виртуальные симуляции позволяют студентам получать практический опыт работы с реальными системами без необходимости доступа к дорогостоящему оборудованию.

Таким образом, Системы информационного моделирования являются мощным инструментом, который помогает улучшить процессы проектирования, анализа, тестирования и управления различными системами, что способствует развитию науки, технологии и промышленности.

Далее будет приведена таблица для сравнений различных систем имитационного моделирования.

***Таблица 2.1 Сравнение СИМ***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Характеристика** | **AnyLogic** | **Simul8** | **Arena** | **ExtendSim** |
| Графический интерфейс | Да | Да | Да | Да |
| Язык моделирования | Java, Python, C++ | Встроенный | Встроенный | Встроенный |
| Мультипарадигменность | Да | Нет | Нет | Нет |
| Поддержка агентного моделирования | Да | Нет | Нет | Нет |
| Гибкость моделирования | Высокая | Средняя | Средняя | Высокая |
| Поддержка различных секторов промышленности | Да | Да | Да | Да |
| Встроенные аналитические инструменты | Да | Нет | Да | Да |
| Поддержка разработки расширений | Да | Нет | Нет | Нет |
| Поддержка облачного моделирования | Да | Нет | Нет | Нет |
| Скорость выполнения моделирования | Высокая | Высокая | Высокая | Высокая |

Для данной работы была выбрана система информационного моделирования AnyLogic.

AnyLogic — это мощная платформа для создания многопроцессорных динамических системных моделей, используемая для симуляции и анализа процессов в различных областях, включая производство, логистику, транспорт, здравоохранение, финансы и многое другое.

Преимущества AnyLogic:

* Бесплатная лицензированная версия: Доступна Personal Learning Edition.
* Применение агентного моделирования - возможность построения модели социальной сети с использованием агентного подхода.
* Возможность создания модели случайной безмасштабной сети с использованием алгоритма, основанного на принципе предпочтительного присоединения.
* Моделирование процесса распространения контента - возможность использования моделей эпидемии, таких как SI, SIR и SEIR, для анализа распространения информации.
* Включение параметров активности и влияния - возможность учета публикационной активности и уровня воздействия в модели.
* Визуализация динамики распространения информации - возможность наглядного отображения процесса распространения контента в социальной сети.

Глава 3. Моделирование системы виртуальной сети

В данной главе будет описан процесс разработки и моделирования модели виртуальной сети в системе имитационного моделирования AnyLogic.

3.1. Требования к модели

Для изучения и анализа социальной сети, а также для создания наглядной графической модели её структуры, наиболее предпочтительным подходом является использование методов графового моделирования. Этот подход позволяет наглядно представить связи между участниками сети и выявить их основные свойства. Для создания такой модели эффективно использовать специализированное программное обеспечение, способное анализировать и визуализировать социальные графы, как было выбрано в разделе 2.2 в данная работа будет реализовываться в AnyLogic [5].

На первом этапе моделирования были определены требования к модели. Основное внимание было уделено тому, чтобы модель точно отражала процесс распространения контента в социальной сети, учитывая разнообразные параметры. Также было выявлено, что для достижения этой цели модель должна включать в себя агентов (пользователей) и связи между ними, которые отражают структуру социальной сети [8]. В качестве основного типа сети был выбран безразмерный вариант, что обеспечивает гибкость в моделировании различных социальных сетей.

С целью изучения возможности распространения информации в социальных сетях, с учетом активности пользователей и степени реализации возможностей влияния на социальные сети, в данной работе применяется совместная разработка агентского подхода (Agent-based modelling) и математического моделирования. Данный графовый подход, который можно назвать моделью Барабаши-Альберта или же моделью агентного моделирования, применяется в совокупности с концепцией графового подхода, которая позволяет наглядно представить взаимодействие и взаимосвязи между агентами социальной сети, а также визуализировать динамику распространения информации.

В данной исследовательской работе в качестве основы для моделирования процесса распространения контента используется модель SEIR, которая базируется на принципах эпидемиологии. Такой выбор обоснован тем, что эти модели считаются наиболее соответствующими для анализа социальных сетей, поскольку механизм передачи информации в таких сетях аналогичен механизму распространения инфекций.

Согласно поставленной задаче исследования, моделирование процесса распространения на основе модели SEIR должно быть реализовано следующим образом:

* Начальным состоянием всех пользователей моделируемой сети является состояние "Susceptible". В этом состоянии агенты сети считаются восприимчивыми к контенту и могут стать его получателями.
* Cостояние Exposed учитывает тот факт, что в реальных социальных сетях пользователи могут находиться в состоянии нерешительности в связи с личными убеждениями или факторами социального воздействия. Агенты, находящиеся в этом состоянии, уже подверглись влиянию распространяемого контента, однако еще не являются распространителями.
* Состояние "Infective" относится к пользователям-распространителям, которые уже подверглись воздействию распространяемой информации. Они могли увидеть контент, опубликованный для всех, или получить информацию от связанного агента, который уже находится в состоянии "Infective".
* Для исследования влияния параметра публикационной активности на процесс распространения включается состояние "InfectivePublisher". Это дополнительное состояние, модификация стандартной модели SI.
* Состояние "Recovered" соответствует пользователям, которые больше не заинтересованы в распространении информации. Они перестают быть распространителями.

Переход пользователя из состояния восприимчивого (Susceptible) в состояние подвергшегося влиянию информации (Exposed) должен выполняться аналогично переходу Susceptible–Infective, то есть пользователь переходит в состояние Exposed либо увидев контент, публикуемый для всех, либо получив информацию от связанного агента (друга), который уже находится в состоянии зараженного (Infective).

Переход агента в состояние "InfectivePublisher" происходит следующим образом: при запуске симуляции выбирается агент, который начнет процесс распространения информации, используя таргетинг для показа контента всем пользователям моделируемой социальной сети, представляющим заданную целевую аудиторию. Частота показа контента определяется параметром публикационной активности. Для корректной работы модели добавляется параметр, определяющий вероятность получения публикуемой информации.

Переход пользователя из состояния "Infective" в состояние "Recovered" выполняется с определенной интенсивностью, зависящей от среднего времени, в течение которого пользователь заинтересован в распространении увиденной информации.

Поскольку в социальной сети лидерство определяется главным образом числом подписчиков, центральность по степени выглядится наиболее подходящей метрикой. Это ключевой индикатор в исследовании, который выявляет самых влиятельных участников в социальном графе. Она опирается на количество связей, и чем их больше у пользователя, тем больше его влияние на поведение и мнения других пользователей.

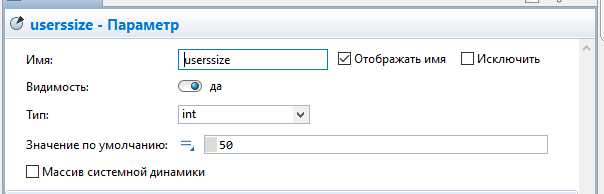
Выбор этой метрики обоснован её простотой и легкостью в понимании, а также связью между числом связей и влиятельностью пользователя. У пользователя с большим числом связей больше возможностей оказывать воздействие на других пользователей и распространять свой контент. Следовательно, центральность по степени позволяет выделить наиболее значимых и влиятельных участников в социальной сети.

3.2. Реализация модели

Будет реализована максимально простая модель. Целью реализации является посмотреть, как проходит распространение информации в сети.

Необходимо создать агентов Main и User.

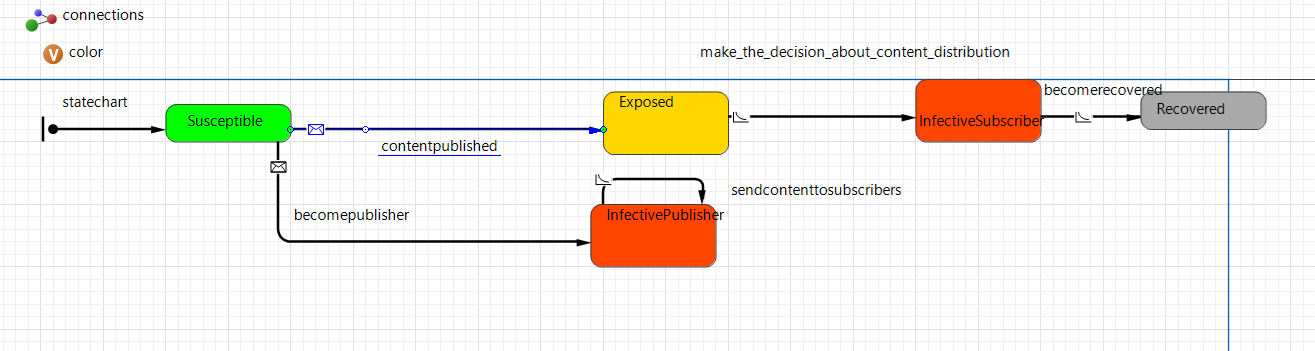
После настройки отображения агента User, необходимо добавить популяцию агентов User в агента Main. После чего, добавить агенту Main параметр usersize, который будет отвечать за размер популяции. Параметр usersize представлен на рисунке 3.1.



***Рисунок 3.1. Настройка параметра usersize***

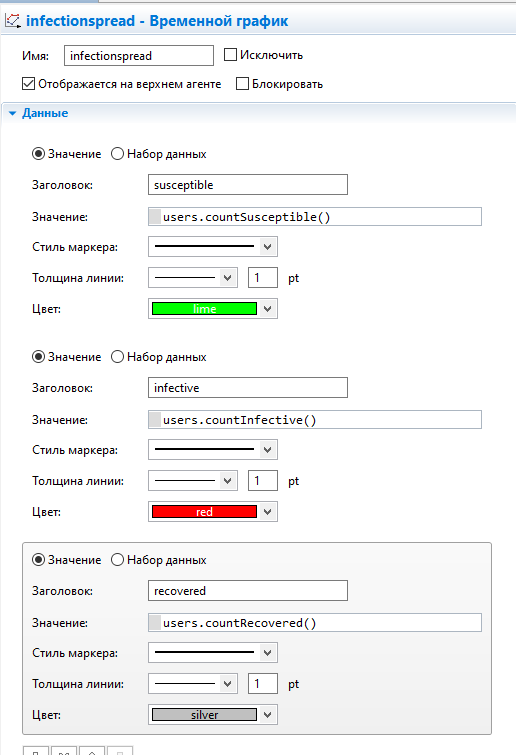
Далее для реализации переходов между состояниями необходимо добавить параметры infectionduration и latencyduration. Infectionduration будет отвечать за переход из состояния Infective в состояние Recovered, то есть это длительность заражения или то, сколько пользователь распространяет информацию. Параметр latencyduration отвечает за переход из состояния Exposed в состояние Infective и означает длительной задержки, то есть сколько пользователь раздумывает над становление распространителем.

Диаграмма состояний представлена на рисунке 3.2.



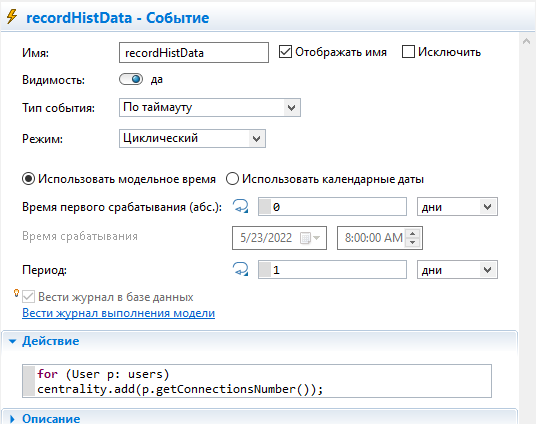
***Рисунок 3.2. Диаграмма состояний***

Для визуализации статистических данных, необходимо добавить временной график, настройки временного графика представлены на рисунке 3.3.



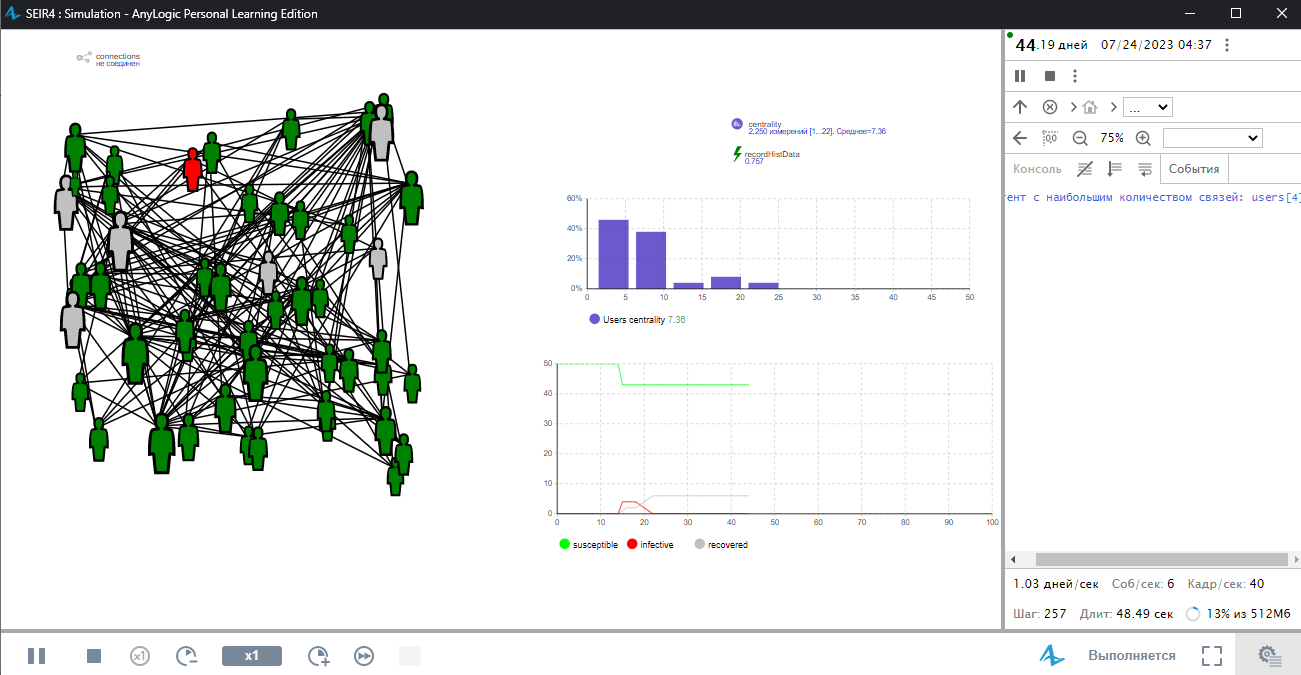
***Рисунок 3.3. Настройки временного графика***

Также необходимо отслеживать центральность, то есть количество связей между users. Для этого добавим гистограмму с данными centrality и Event. Event отвечает за подсчет количества связей, его свойства представлены на рисунке 3.4.



***Рисунок 3.4. Свойства события***

Работа модели представлена на рисунке 3.5.



***Рисунок 3.5. Работа модели***

Ссылка на модель в AnyLogic Cloud представлена в Приложении Б.

Глава 4. Реализация приложения

В данной главе будут описаны процессы реализации сервисов для сбора информации и вычисления метрик, а также приложения для просмотра результатов.

4.1. Выбор средств реализации

В качестве СУБД для хранения информации о пользователя была выбрана PostgreSQL. PostgreSQL — это реляционная система управления базами данных (РСУБД) с открытым исходным кодом, которая заслужила признание благодаря своей надежности, стабильности и гибкости. Широкое использование PostgreSQL в различных сферах делает ее предпочтительным выбором для множества разработчиков и организаций. Ниже приведена сравнительная таблица нескольких СУБД, в которой видно, что PostgreSQL является качественным вариантом для данной работы.

***Таблица 4.1 Сравнение СУБД***

|  |  |  |  |
| --- | --- | --- | --- |
| **Характеристика** | **PostgreSQL** | **MySQL** | **Oracle** |
| Лицензия | BSD | GPL | Проприетарная |
| Тип лицензии | Open Source | Open Source | Проприетарная |
| Распространение | Широкое | Широкое | Широкое |
| Тип хранилища данных | Реляционная | Реляционная | Реляционная |
| Поддержка распределенных транзакций | Да | Да | Да |
| Производительность | Высокая | Средняя | Высокая |
| Сложность настройки | Средняя | Низкая | Высокая |
| Сообщество разработчиков | Активное | Активное | Активное |
| Поддержка языков SQL | Полная | Полная | Полная |
| Поддержка репликации | Да | Да | Да |
| Поддержка JSON | Да | Да | Да |
| Сложные запросы | Высокая | Средняя | Высокая |
| Профессиональная поддержка | Да | Да | Да |
| Стоимость | Бесплатно | Бесплатно | Платно |

В качестве языка программирования был выбран Python, так имеет множество полезных возможностей:

* В экосистеме Python существует множество библиотек и фреймворков для веб-разработки, включая Flask, Django, Pyramid и другие. Это позволяет разработчикам выбирать инструменты в зависимости от требований и предпочтений проекта.
* Python имеет простой и понятный синтаксис, что делает его доступным для новичков и облегчает работу опытным разработчикам.
* Python широко применяется в различных областях разработки, включая веб-разработку, науку о данных, искусственный интеллект, автоматизацию и другие. Это обеспечивает разработчикам возможность использовать один язык для различных задач и проектов.
* Python предоставляет разработчикам множество удобных инструментов для работы с API и базами данных, что делает его предпочтительным выбором для создания приложений, взаимодействующих с внешними сервисами и хранилищами данных.

Также для вывода данных и удобного развертывания системы был выбрал фреймворк Flask.

Для взаимодействия и отображения онтологий будет использоваться Protégé - Бесплатный редактор онтологий с открытым исходным кодом и фреймворк для создания интеллектуальных систем.

4.2. Реализация сервиса для сбора информации

В данном разделе представлена реализация скрипта, который представляет собой Python-скрипт, который взаимодействует с социальной сетью ВКонтакте (VK) и выполняет различные задачи, такие как сбор данных о пользователях и группах, анализ действий пользователей и обновление онтологии полученной информацией.

Изначально необходимо определить сущности, которые будут отслеживаться, то есть пользователь, сообщество и пост, то есть класс Person, Community и Post.

Следующее, что необходимо сделать сформировать методы для взаимодействия с VK API [4], для этого создан класс VK. Данный класс обрабатывает взаимодействие с VK API, включая получение данных о пользователях, сообществах и записях, проверку статуса онлайн пользователя и получение понравившихся записей.

Для реализации класса VK будет использоваться официальная билиотека для работы с методами данной социальной сети – vk\_api. Необходимые для работы методы:

* users.get - этот метод позволяет получить полную информацию о пользователе по его идентификатору. В скрипте используется для извлечения данных о пользователе, таких как имя, фамилия, пол, дата рождения, фотография и другие атрибуты.
* groups.get - используется для получения списка идентификаторов групп, в которых состоит пользователь. Эта информация может быть полезна для анализа социальной сети пользователя и его активности в различных группах.
* groups.getById - этот метод получает полную информацию о группе по ее идентификатору. В скрипте используется для извлечения данных о сообществе, таких как название, активность и другие характеристики.
* friends.get - позволяет получить список идентификаторов пользователей, с которыми выбранный пользователь состоит в дружеских отношениях. В скрипте этот метод используется для получения списка друзей пользователя.
* wall.get - используется для получения списка последних постов на стене пользователя или сообщества. В скрипте этот метод используется для извлечения последних постов и анализа активности пользователей и групп.
* likes.isLiked - этот метод проверяет, добавил ли пользователь конкретный пост в список понравившихся. В скрипте он используется для проверки, понравился ли определенный пост пользователю.

Далее реализуем класс Ontology, который управляет созданием и сохранением онтологии, а также предоставляет методы для сохранения информации о пользователях, сообществах и записях в онтологию. Данный класс содержит следующие методы:

1) Метод \_\_init\_\_:

* Инициализирует объект класса Ontology.
* Принимает путь к файлу онтологии.
* Загружает онтологию из указанного файла и инициализирует объект блокировки threading.Lock().

2) Метод create\_ontology:

* Создает структуру онтологии, определяя классы и свойства для хранения информации о пользователях, сообществах и их взаимодействии.
* Определяет классы **Person**, **Community**, **Activity** и **Post**, а также свойства для хранения информации о них, такие как имя, фамилия, пол, фотография, активность и т.д.
* Сохраняет онтологию в файл по указанному пути.

3) Метод save\_person:

* Сохраняет информацию о пользователе в онтологию.
* Создает экземпляр класса **Person** и заполняет его атрибуты данными о пользователе (имя, фамилия, пол, дата рождения, фотография и т.д.).
* Добавляет информацию о друзьях и подписках пользователя.
* Сохраняет изменения в онтологию.

4) Метод save\_community:

* Сохраняет информацию о сообществе в онтологию.
* Создает экземпляр класса **Community** и заполняет его атрибуты данными о сообществе (название, активность и т.д.).
* Сохраняет изменения в онтологию.

5) Метод save\_post:

* Сохраняет информацию о посте в онтологию.
* Создает экземпляр класса **Post** и заполняет его атрибуты данными о посте (владелец, теги, репосты и т.д.).
* Проверяет, является ли владелец поста пользователем или сообществом, и сохраняет соответствующую информацию.
* Сохраняет изменения в онтологию.

6) Методы post\_viewed и post\_liked:

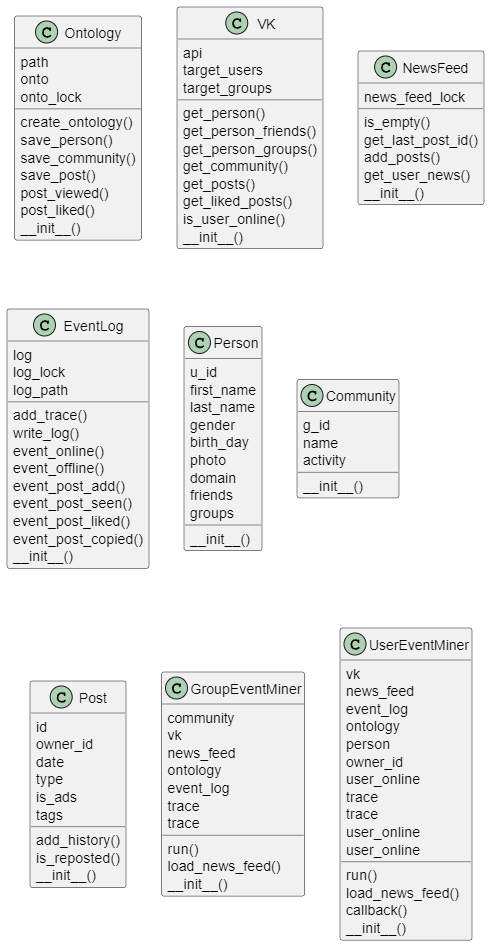
* Добавляют информацию о просмотре и лайке поста в онтологию.
* Создают связи между пользователями и постами, указывая, кто посмотрел пост и кто поставил лайк.
* Сохраняют изменения в онтологию.

Класс NewsFeed управляет глобальной лентой новостей, храня записи для каждого владельца (пользователя или сообщества) и предоставляет методы для добавления и извлечения записей.

Класс EventLog управляет ведением журнала событий, включая добавление трассировок, запись журналов в файлы и определение различных типов событий, таких как статус онлайн, добавление, просмотр, лайк и копирование записей.

Классы GroupEventMiner и UserEventMiner наследуются от класса Thread. Метод run() предоставляет возможность запуска асинхронного сбора информации для определенного пользователя или сообщества.

Диаграмма классов представлена на рисунке 4.1.



***Рисунок 4.1. Диаграмма классов***

Для обеспечения асинхронной и распределенной обработки задач или сообщений будет использоваться RabbitMQ. RabbitMQ является брокером сообщений, который позволяет различным компонентам приложения (например, микросервисам или рабочим потокам) обмениваться данными или задачами через очереди сообщений.

При запуске определяются параметры, такие как интервалы времени для различных действий, целевые пользователи, целевые группы, а также инициализируются онтология, API VK, журнал событий и лента новостей.

Настраивается подключение к RabbitMQ и объявляются необходимые обменники и очереди.

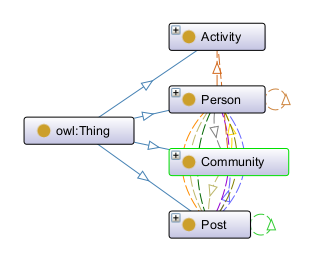
Определяется функция обратного вызова для обработки входящих сообщений от RabbitMQ, которая получает данные о пользователе или группе из VK, запускает добытчики событий для пользователей и групп, обновляет онтологию полученными данными и публикует сообщения для обновления сущностей.

Наконец, начинается потребление сообщений от RabbitMQ и запускается поток записи журнала событий.

Для запуска агента используются данные из файлов users.txt и groups.txt. В каждом из этих файлов содержатся уникальные идентификаторы пользователей и групп социальной сети. При работе агента можно расширить список групп, добавив новый идентификатор в очередь сообщений.

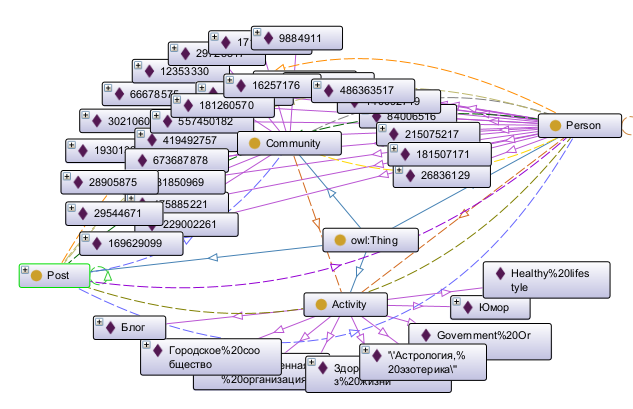
Скрипт читает данные из двух текстовых файлов: users.txt и groups.txt. Каждый файл содержит идентификаторы пользователей или групп, которые будут отправлены в RabbitMQ для обработки. Далее создается соединение с сервером RabbitMQ и открывается канал для обмена сообщениями. После чего происходит отправка сообщений в очередь RabbitMQ. Каждый элемент из списка groups и users преобразуется в формат JSON и отправляется в очередь с указанным обменником ("social\_data") и пустым ключом маршрутизации.

Результатом работы программы является онтология и журнал событий. Структура онтологии представлена на рисунке 4.2.

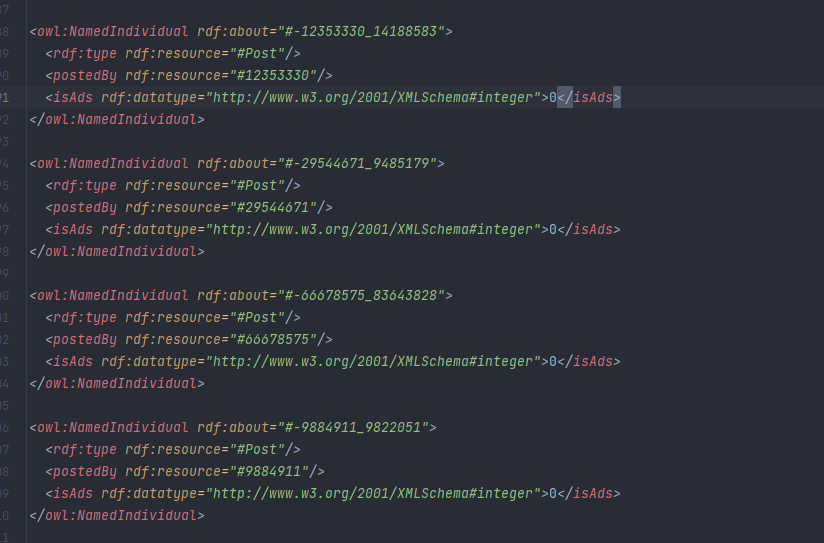


***Рисунок 4.2. Структура онтологии***

Полученная онтология представлена на рисунках 4.3 и 4.4.

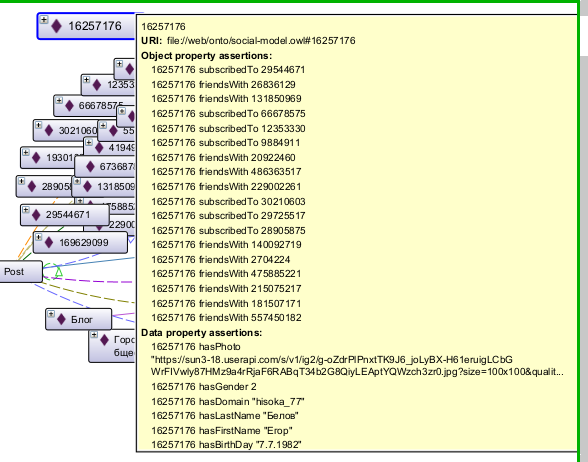


***Рисунок 4.3. Расширенная структура онтологии***



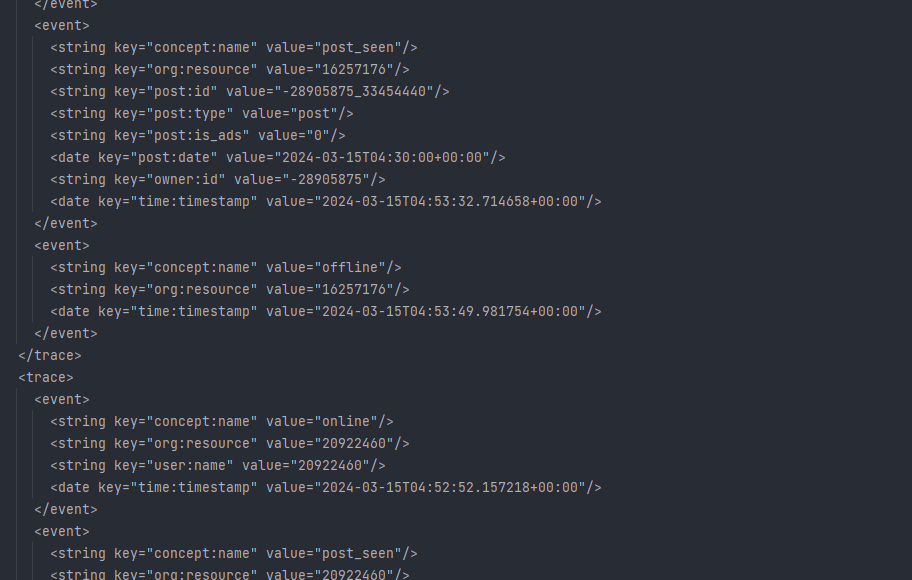
***Рисунок 4.4. OWL код полученной онтологии***

На рисунке 4.5 представлена индивидуальная информация о пользователе и его связях с другими сущностями.



***Рисунок 4.5. Индивидуальная информация***

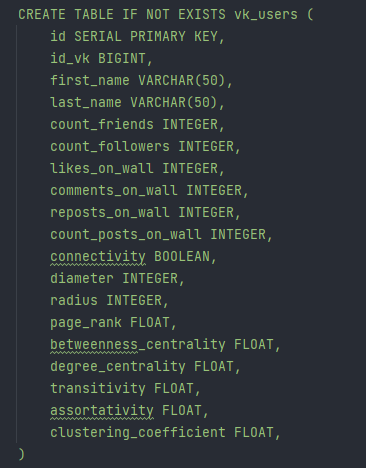
Также был сформирован журнал событий, представленный на рисунке 4.6.



***Рисунок 4.6. Журнал событий***

4.3. Реализация сервиса для вычисления метрик

Сначала необходимо создать базу данных, в которой будет храниться информация о пользователях и его метриках. SQL-запрос на создание БД представлен на рисунке 4.7.



***Рисунок 4.7. Создание БД***

База данных будет содержать информацию о пользователях, такую как: ID ВК, Имя, Фамилия, Количество друзей и подписчиков, Количество лайков, репостов и комментариев на стене пользователя, Общее количество постов на стене, а также метрики: Связность, Диаметр, Радиус, PageRank, Центральность, Транзитивность, Ассортативность, Коэффициент кластеризации.

Далее необходимо реализовать методы для взаимодействия с VK APi, такие как:

* request\_count\_followers – метод, для подсчета количества подписчиков, использующий метод API users.getFollowers;
* request\_count\_friends – подсчет количества друзей с помощью friends.get;
* request\_count\_likes\_from\_wall – метод для расчета количества лайков, комментариев, репостов на стене пользователя и общее количество постов, используя wall.get;
* request\_users\_get – получение информации о пользователях с помощью users.get.

Далее заполняем БД с помощью данных методов.

Следующим шагом является создание графа друзей определенного пользователя, то есть графа реальной сети. С помощью этого графа будет возможно вычислить метрики, встроенные Python библиотеки [7].

Определяем класс VkFriends. Этот класс содержит методы для работы с VK API и получения информации о пользователе, его друзьях и общих друзьях.

**Метод** request\_url - этот метод формирует URL-адрес запроса к API VK на основе переданных параметров.

**Метод** base\_info - этот метод используется для получения основной информации о пользователе, такой как его имя, фамилия и фото.

**Метод** friends - этот метод используется для получения списка друзей пользователя.

**Метод** common\_friends - этот метод используется для поиска общих друзей текущего пользователя с его друзьями. Метод делит список друзей на части по 25 человек и для каждой части запрашивает общих друзей с помощью метода friends.getMutual API VK. Результаты обработки сохраняются в список result в виде кортежей с информацией о пользователе и его общих друзьях. Если возникает ошибка при запросе общих друзей, она выводится на экран, и процесс продолжается. Метод возвращает список кортежей с информацией о пользователе и его общих друзьях.

Класс D3 является подклассом класса VkFriends. Он используется для генерации JSON-файла, который можно использовать для визуализации социального графа с помощью библиотеки D3.js.

В конструкторе вызывается конструктор родительского класса VkFriends, чтобы инициализировать экземпляр с использованием переданных аргументов.

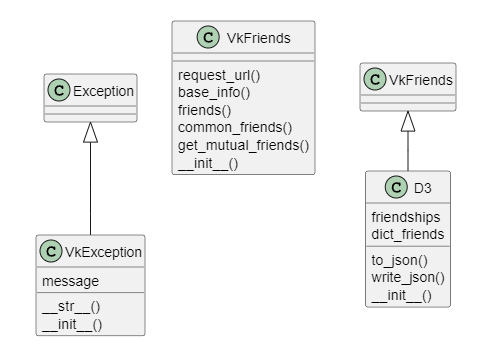
self.friendships инициализируется результатом вызова метода common\_friends, который находит общих друзей текущего пользователя с его друзьями.

self.js инициализируется пустым словарем, который будет содержать узлы и ребра графа.

self.dict\_friends инициализируется результатом вызова метода to\_json(), который преобразует данные о друзьях в JSON-формат.

JSON-данные записываются в файл с помощью метода write\_json().

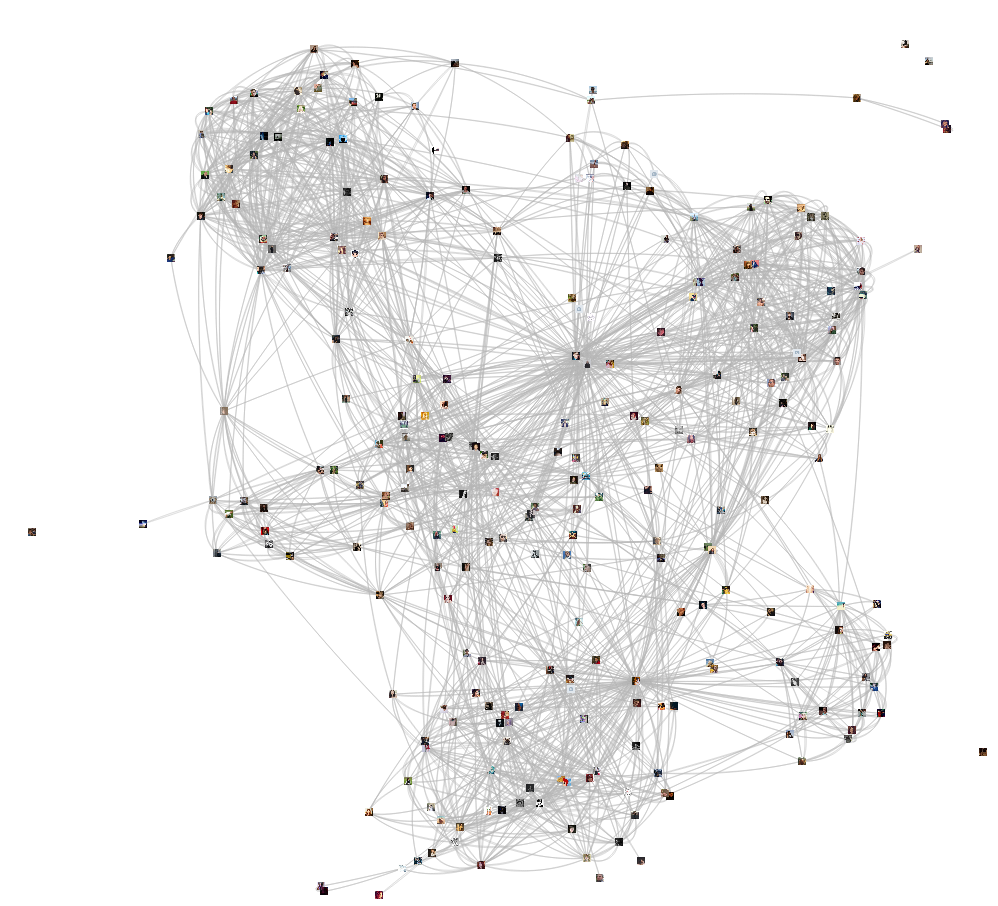
Диаграммы классов представлены на рисунке 4.8.



***Рисунок 4.8. Диаграмма классов***

С помощью D3.js (Data-Driven Documents), библиотеки JavaScript для создания интерактивных и динамических визуализаций данных в веб-браузере, создадим веб-страницу для отображения графа друзей. Она позволяет связывать данные с элементами DOM и применять к ним различные визуальные эффекты с использованием стандартных веб-технологий, таких как HTML, SVG и CSS.

Граф представлен на рисунке 4.9.



***Рисунок 4.9. Граф друзей***

Также было принято решение реализации графа друзей друзей. Выбор графа друзей друзей может быть обоснован тем, что он учитывает не только непосредственные связи пользователя, но и расширенную сеть его влияния. Это позволяет более точно оценить его потенциальное воздействие на целевую аудиторию, включая людей, которые могут быть недоступными через прямые связи. Такой подход также учитывает транзитивность связей и масштабируется для более широкого анализа влияния пользователя в социальной сети.

Необходимо реализовать несколько функций, связанных с работой с API ВКонтакте и созданием графа социальной сети на основе данных о друзьях и друзьях друзей в этой сети.

get\_user\_info - эта функция отправляет запрос к API ВКонтакте для получения информации о пользователе по его ID. Она формирует параметры запроса (user\_id, fields, name\_case, v, access\_token) и отправляет запрос методом GET. После получения ответа, функция возвращает словарь с данными о пользователе.

make\_dict\_from\_user\_info - эта функция преобразует ответ от сервера ВКонтакте в удобочитаемый словарь с информацией о пользователе. Она извлекает различные поля из словаря API-ответа и создает новый словарь, который будет содержать нужную информацию.

make\_node\_from\_user\_info - эта функция преобразует словарь с информацией о пользователе в строку, которая будет представлять вершину в графе.

write\_to\_csv - эта функция записывает информацию в CSV файл. Она принимает имя файла, заголовок и строки информации, а затем записывает их в указанный файл.

request\_friends - эта функция отправляет запрос к API ВКонтакте для получения списка друзей пользователя. Она также формирует параметры запроса (user\_id, count, fields) и отправляет запрос методом GET. После получения ответа, функция возвращает словарь с данными о друзьях пользователя.

is\_mutual - эта функция проверяет, является ли человек другом взаимным (находится ли он в списке друзей). Она принимает два аргумента - человека для проверки и список друзей. Если человек находится в списке друзей, функция возвращает True, в противном случае - False.

Также определена функция create\_graph, которая является основной функцией. Она выполняет следующие действия:

* Получает информацию о пользователе (ID, список друзей).
* Запрашивает список друзей каждого друга пользователя.
* Строит граф друзей и друзей друзей, формируя списки вершин и рёбер.
* Записывает вершины и рёбра в CSV файлы.
* Выводит сообщения о ходе выполнения процесса.

**csv\_to\_networkx\_dict** — это функция, которая преобразует информацию из CSV файлов в словарь, который используется для создания графа в библиотеке NetworkX. Она создает пустой граф, затем обрабатывает информацию о узлах и добавляет их в граф, а затем обрабатывает информацию о рёбрах и добавляет их в граф.

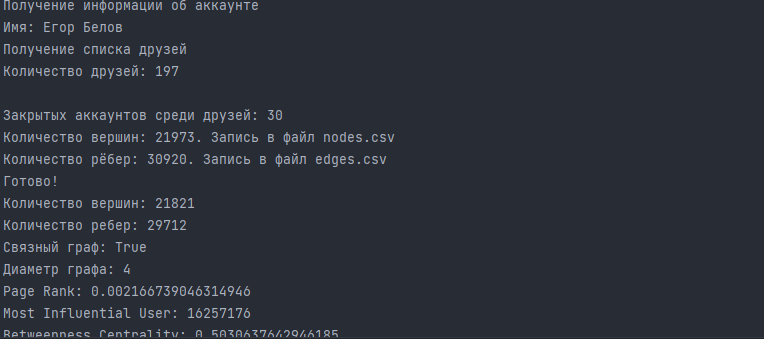
**calculate\_graph\_properties** — это функция, которая вычисляет различные свойства графа, такие как количество вершин и рёбер, связность графа, диаметр, коэффициент кластеризации, PageRank, центральность и т.д [6]. Она возвращает список метрик графа.

**Main** — это основная функция программы. Она считывает информацию из CSV файлов, вызывает функцию для создания графа и вычисления его свойств, а затем выводит результаты на экран.

Общий поток программы:

* Сначала в основной функции main() открываются CSV файлы с информацией о узлах и рёбрах.
* Используя функцию csv\_to\_networkx\_dict(), данные из CSV файлов преобразуются в структуру данных, понятную для библиотеки NetworkX.
* Затем вызывается функция calculate\_graph\_properties() для анализа графа и вычисления его свойств.
* Результаты выводятся на экран и возвращаются из функции main().
* После чего метрики сохраняются в базу данных.

Результат работы программы представлен на рисунке 4.10.

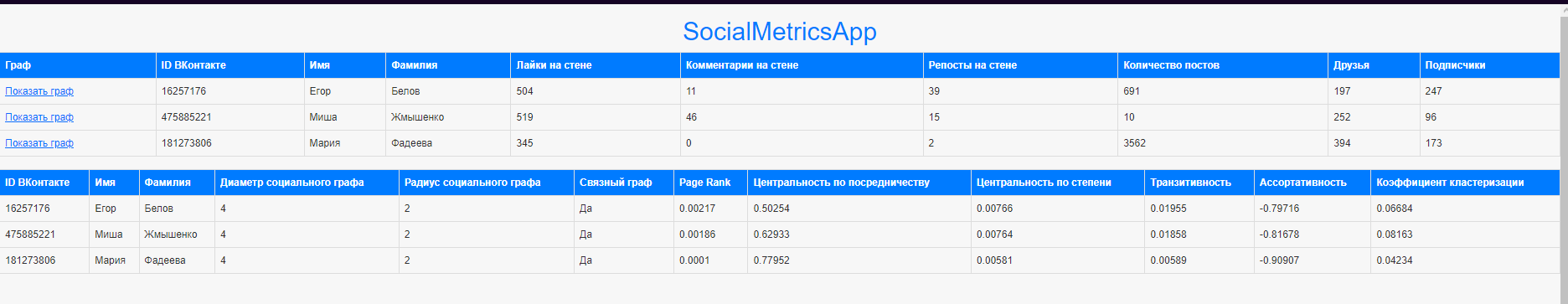


***Рисунок 4.10. Результат работы программы***

4.4. Реализация приложения для просмотра результатов

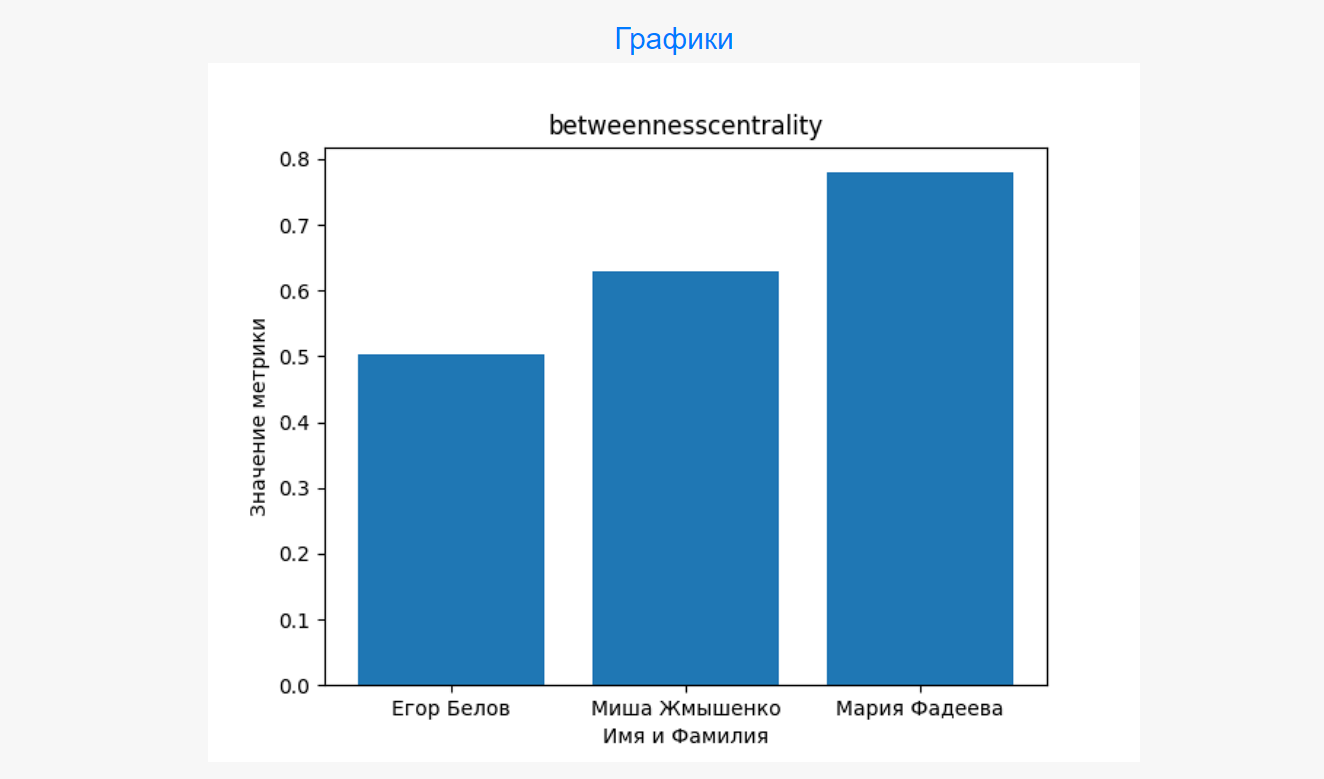
Веб приложение отвечает за просмотр собранной информации.

Первый элемент – это таблица, собранной информации о пользователях. Отображаются основные данные о пользователях, такие как ID VK, имя, фамилия, количество лайков, комментариев, репостов, количество постов, количество друзей и подписчиков. Каждый пользователь имеет ссылку "Показать граф", при нажатии на которую генерируется граф друзей, реализация которого представлена в разделе 4.3. Полученная таблица пользователей представлена на рисунке 4.11.



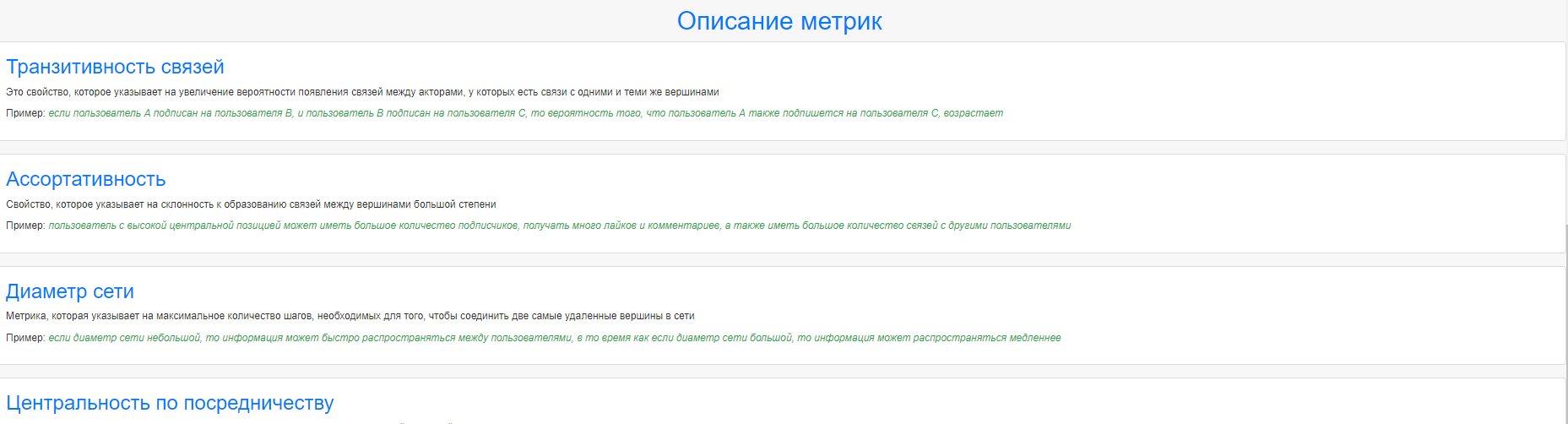
***Рисунок 4.11. Таблица с информацией о пользователях***

Следующим элементом является слайдер графиков для сравнения различных метрик у пользователя. Для реализации слайдера использовался Bootstrap. Bootstrap — это бесплатный набор инструментов для создания веб-сайтов и веб-приложений. Он содержит готовые CSS и JavaScript компоненты для быстрой и удобной разработки интерфейсов с использованием сеток, форм, кнопок, модальных окон и других элементов. Bootstrap обеспечивает адаптивную верстку, что позволяет сайтам корректно отображаться на различных устройствах и экранах. Слайдер представлен на рисунке 4.12.



***Рисунок 4.12. Слайдер с графиками метрик***

Третий элемент информационный, он содержит описания основных метрик социального графа, такие как транзитивность связей, ассортативность, диаметр сети и т.д. Описания метрик представлены на рисунке 4.13.



***Рисунок 4.13. Описание метрик***

Глава 5. Тестирование

Для обеспечения корректной работы приложения для вычисления метрик, модели виртуальной сети, а также анализа реальной сети был разработан комплексный план тестирования. Этот раздел описывает методику, используемую для тестирования каждой составляющей системы, включая сбор данных, анализ и визуализацию, а также проверку точности результатов.

Тестирование модели виртуальной сети включает в себя несколько ключевых этапов. В первую очередь, осуществляется проверка правильности визуализации. Это означает, что визуализация должна адекватно отображать все элементы виртуальной сети и их взаимосвязи. Для этого проводятся тесты на различных уровнях сложности моделей, чтобы удостовериться в корректности представления данных.

Далее, проводится тестирование сценариев использования модели. Это важный аспект, так как модель должна быть гибкой и поддерживать различные сценарии использования, такие как добавление новых узлов, изменение связей между ними, а также анализ сети в различных контекстах. В ходе тестирования оценивается функциональность и удобство использования модели для различных целей.

Для анализа реальной сети также необходимо проводить тщательное тестирование. Это включает проверку точности вычисления метрик эффективности рекламных кампаний, которые используются для оценки результатов рекламных стратегий. Тестирование алгоритмов анализа позволяет убедиться в их эффективности и точности при обработке различных сценариев данных.

Наконец, проводится тестирование сбора информации в онтологии и отслеживания через журналы событий. Это важный этап, так как онтология должна структурировать и хранить собранные данные, а журналы событий должны точно регистрировать происходящие события. Точность анализа данных через онтологии и журналы событий позволяет принимать обоснованные решения по оптимизации рекламных стратегий на основе фактических данных.

Все эти этапы тестирования совместно гарантируют качество и надежность всех компонентов системы, а также правильность вычисления и интерпретации метрик, необходимых для анализа эффективности рекламных кампаний.

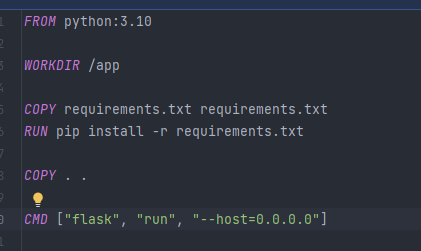
Для тестирования функциональности веб-приложения был использован метод, исключающий применение автоматизированных средств и включающий активное участие группы пользователей. Этот подход предусматривает, чтобы участники группы имели доступ к веб-приложению, развернутому на удаленном сервере, и использовали его в реальных условиях, исполняя различные действия и операции. По завершении сессии использования каждый участник делился своими впечатлениями и отзывами о продукте. Этот процесс позволил выявить и устранить обнаруженные недочеты и ошибки в работе приложения.

Глава 6. Внедрение

Для успешного развертывания системы используется Docker. Docker позволяет упаковать приложение и его зависимости в контейнер для обеспечения единообразной среды выполнения на различных платформах.

Для Создание Docker образа приложения необходимо создать Dockerfile в корневой директории приложения. Dockerfile определяет шаги по сборке образа, включая установку зависимостей, копирование исходного кода и установку необходимых настроек.

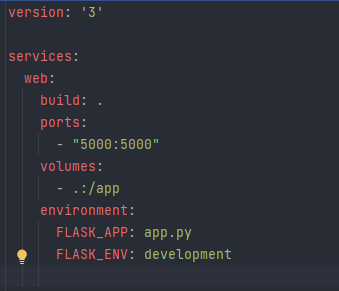
В данном случае, Dockerfile содержит инструкции по установке Python, копированию файлов приложения и установке зависимостей из файла requirements.txt. Содержание Dockerfile представлено на рисунке 6.1.



***Рисунок 6.1. Содержание Dockerfile***

Следующим шагом является Настройка Docker Compose для контейнеризации приложения. Docker Compose используется для определения и запуска многоконтейнерных Docker приложений. В файле docker-compose.yml определяются сервисы, их зависимости и настройки.

В файле docker-compose.yml для данной системы определен сервис web, который собирается из Dockerfile, пробрасывает порт 5000 и монтирует текущую директорию в контейнер для возможности внесения изменений в реальном времени. Содержание docker-compose.yml представлено на рисунке 6.2.



***Рисунок 6.2. Содержание docker-compose.yml***

После создания Dockerfile для всех компонентов системы и настройки Docker Compose файлов, можно запустить все сервисы одной командой. Docker Compose позволяет управлять контейнерами, масштабировать сервисы и управлять их зависимостями.

Таким образом, использование Docker и Docker Compose упрощает развертывание системы, обеспечивает ее изолированность и надежность в различных окружениях, а также упрощает управление контейнеризированными приложениями и сервисами.

Модель виртуальной сети была развернута с использованием платформы AnyLogic Cloud. AnyLogic Cloud предоставляет возможность разработки, тестирования и развертывания моделей виртуальных сетей в облаке, что позволяет получить доступ к модели из любого места, где есть подключение к интернету. Использование AnyLogic Cloud обеспечивает удобство в работе с моделью, а также обеспечивает масштабируемость и гибкость в ее управлении. Благодаря облачной платформе, модель виртуальной сети может быть доступна для использования и совместной работы с различными участниками проекта без необходимости установки специального программного обеспечения или ограничений по доступу.

Заключение

В рамках данной работы был предложен комплексный подход к оптимизации рекламных стратегий в социальных медиа, включающий несколько этапов.

Исследование социальных сетей является важной областью для понимания взаимодействия людей, формирования сообществ и распространения информации. Для этой цели существуют два основных метода исследования: исследование реальной и виртуальной социальной сети.

Была создана модель виртуальной сети в системе информационного моделирования, включающая в себя визуализацию виртуальной сети. Этот шаг позволил визуализировать структуру социальной сети и выявить ключевые узлы и взаимосвязи между ними.

Было проведен анализ реальной сети с целью реализации сервиса для вычисления метрик на основе фактических данных. Это позволило оценить реальную эффективность рекламных кампаний.

И, наконец, в рамках исследования была проведена работа по сбору информации в онтологии и отслеживанию через журналы событий, что позволило более полно и точно анализировать результаты и принимать обоснованные решения по оптимизации рекламных стратегий.

Данная работа имеет значительный потенциал для дальнейшего развития в следующих направлениях:

* **Улучшение методов анализа**. Можно провести дополнительные исследования для улучшения методов анализа данных и вычисления метрик эффективности рекламных кампаний. Это может включать в себя разработку новых алгоритмов или улучшение существующих.
* **Интеграция новых источников данных**. Работа может быть расширена путем интеграции новых источников данных, таких как данные из социальных сетей, блогов, форумов и т. д. Это позволит получить более полное представление об аудитории и ее поведении.
* **Разработка персонализированных рекомендаций**. На основе данных об аудитории и их поведении можно разработать систему персонализированных рекомендаций для оптимизации рекламных стратегий. Это поможет компаниям предлагать более релевантные и привлекательные рекламные сообщения.
* **Анализ влияния внешних факторов**. Исследования могут быть расширены для включения анализа влияния внешних факторов, таких как события в мире или изменения в экономической среде, на эффективность рекламных кампаний. Это поможет компаниям адаптировать свои стратегии к изменяющимся условиям рынка.
* **Более глубокое применение машинного обучения и искусственного интеллекта**. Применение методов машинного обучения и искусственного интеллекта может значительно улучшить процесс анализа данных и прогнозирования результатов рекламных кампаний. Это также может помочь автоматизировать процессы принятия решений и оптимизации стратегий.

Библиографический список

1. Торопов Б.А.  Модель независимых каскадов распространения репоста в онлайновой социальной сети // Кибернетика и программирование.  2016. № 5.  С. 199-205. DOI: 10.7256/2306-4196.2016.5.20624 URL: https://nbpublish.com/library\_read\_article.php?id=20624 .
2. Avinash Kaushik **“**Best Social Media Metrics: Conversation, Amplification, Applause, Economic Value” [Электронный ресурс] URL: https://www.kaushik.net/avinash/best-social-media-metrics-conversation-amplification-applause-economic-value/ (дата обращения: 11.11.2023).
3. Событийная онтология vs объектная [Электронный ресурс] URL: https://habr.com/ru/articles/706916/# (дата обращения: 15.12.2023).
4. Документация VK API [Электронный ресурс] URL: https://dev.vk.com/ru/reference (дата обращения: 02.01.2024).
5. Документация AnyLogic [Электронный ресурс] URL: https://www.anylogic.ru/getting-started/ (дата обращения: 03.02.2024).
6. Документация Networkx [Электронный ресурс] URL: https://networkx.org/documentation/networkx-1.9/reference/algorithms.html (дата обращения: 05.01.2024).
7. Анализ дружеских связей VK с помощью Python [Электронный ресурс] URL: https://habr.com/ru/articles/221251/ (дата обращения: 05.01.2024).
8. Dmitriev I., Zamyatina E. How to Prevent Harmful Information Spreading in Social Networks Using Simulation Tools // Analysis of Images, Social Networks and Texts. 8th International Conference, AIST 2019, Kazan, Russia, July 17–19, 2019, Revised Selected Papers. Communications in Computer and Information Science. Springer. 2020. Vol. 1086. P. 201–213. DOI: 10.1007/978-3-030-39575-9\_21.
9. Степанов П.П.  Решение актуальных проблем предприятий нефтегазовой отрасли с применением методов теории игр // Кибернетика и программирование.  2016. № 4.  С. 11-17. DOI: 10.7256/2306-4196.2016.4.20162 URL: https://nbpublish.com/library\_read\_article.php?id=20162.
10. Sebin Jose, P.Victer Paul A survey on identification of influential users in social media networks using bio inspired algorithms. 2023. C 2111-2121.
11. Tonnies F. Community and society. [1887] in Community and Society, C.P. Loomis (trans. and ed.) 1963. New York: Harper and Row. Public Domain // Lin J., 214. Mele C. (Eds.), The urban sociology reader. 2nd ed. L.: Routledge, 2013. -P. 18-33.
12. Гладченко И.А. Модели распространения политического мобилизационного контента в социальных медиа. 2020.
13. Документация Flask [Электронный ресурс] URL: https://flask.palletsprojects.com/en/latest/ (дата обращения: 21.02.2024) .
14. Имитационное моделирование в среде AnyLogic [Электронный ресурс] URL: https://itcloud-edu.ru/courses/ITCloud-IMAL/ (дата обращения: 15.01.2024).
15. Бурко, Р. А. Социальные сети в современном обществе / Р. А. Бурко, Т. В. Терёшина. — Текст : непосредственный // Молодой ученый. — 2014. — № 7 (66). — С. 607-608. — URL: https://moluch.ru/archive/66/11009/ (дата обращения: 03.11.2023).