ТИТУЛЬНИК

**ОГЛАВЛЕНИЕ**

[ВВЕДЕНИЕ 4](#_Toc71838443)

[1 АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ 8](#_Toc71838444)

[1.1 Телеграмм-боты 8](#_Toc71838445)

[1.2 Рекомендательные системы 10](#_Toc71838446)

[1.3 Обзор готовых решений 13](#_Toc71838447)

[Выводы по главе 1 14](#_Toc71838448)

[2 СБОР ДАННЫХ И ЕЕ СТРУКТУРИЗАЦИЯ 15](#_Toc71838449)

[2.1 Классификация пользователей 15](#_Toc71838450)

[2.2 Сбор данных о г. Санкт-Петербурге 16](#_Toc71838451)

[2.3 Создание цифрового хранилища для данных 19](#_Toc71838452)

[2.3.1 Логическая модель базы данных 19](#_Toc71838453)

[2.3.2 Нормализация базы данных 20](#_Toc71838454)

[2.3.3 Создание базы данных 22](#_Toc71838455)

[Выводы по главе 2 24](#_Toc71838456)

[3 ПРОЕКТИРОВАНИЕ И СОЗДАНИЕ ТЕЛЕГРАММ БОТА 25](#_Toc71838457)

[3.1 Проектирование рекомендательной системы 25](#_Toc71838458)

[3.2 Обучение рекомендательной системы 25](#_Toc71838459)

[3.3 Проектирование логической модели телеграмм бота 25](#_Toc71838460)

[3.4 Создание телеграмм бота 25](#_Toc71838461)

[Выводы по главе 3 25](#_Toc71838462)

[ЗАКЛЮЧЕНИЕ 27](#_Toc71838463)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 28](#_Toc71838464)

[ПРИЛОЖЕНИЕ А 32](#_Toc71838465)

# ВВЕДЕНИЕ

В любое время существования человечества одним из двигателей прогресса была информация. Наступление информационной эры привнесло в жизнь человека одну из важнейших возможностей – глобальная коммуникация, которая позволяет любому желающему без особого труда получить интересующую его информацию.

Благодаря новой возможности люди стали активно делиться различной информацией, не задумываясь об ее актуальности и пользе. Ввиду активной коммуникации человека цифровое пространство стало было заполняться контентом. Ежедневно каждый пользователь в интернете генерирует примерно одну страницу сайта, а если учитывать общее количество активных пользователей интернета, примерно 50% от мирового населения, это составляет 3.8 миллиарда страниц в день. [1, 2]

Ввиду выше сказанного высокую ценность приобрели источники, или так называемые медиа-продукты, которые предоставляют действительно полезную информацию. Основным популярным медиа-продуктом до недавнего времени являлись социальные сети – сейчас же появились более продвинутые сервисы, именуемые мессенджерами. Они позволяют обмениваться информацией более безопасно и комфортно. Примерами мессенджеров могут являться Viber, WhatsApp, ТамТам и Telegram.

Помимо коммуникации в формате «one-to-one», которая означает возможность общения только между двумя пользователями, большинство мессенджеров позволяют осуществлять группировку некоторого количества пользователей для того, чтобы последние могли осуществлять обмен информацией коллаборативно. Мессенджеры обычно именуют их группами, каналами и комнатами. Люди создают группы, каналы и комнаты, производя фильтрацию предоставляемой там информации по определённым признаками. Так, например, пользователь может присоединиться к каналу под названием «Новости России» и получать информацию о новостях России в данной группе, канале или комнате.

Наиболее популярным из перечисленных мессенджеров является Telegram (далее телеграмм). Популярность данного мессенджера обусловлена некоторыми факторами, а именно: конфиденциальность, отсутствие цензуры и удобство. [3] Стоит отметить, что указанные факторы не являются конечными, существуют и другие, однако перечисленные выше факторы являются наиболее распространёнными факторами, на основе которых люди выбирают данный мессенджер. Популярность телеграмма распространяется как на территории России, так и на территории СНГ.

Телеграмм предоставляет три варианта медиа: каналы, чаты и боты. Каналы предоставляют собой чат, в котором распространять информацию может исключительно администратор, обычные пользователи являются так называемыми слушателями. В свою очередь обычные чаты предоставляют возможность обмена сообщениями от лица любых пользователей. Больший интерес представляют боты – роботы, которые могут обрабатывать различные запросы от пользователей автоматически и программируются людьми.

В наше время младшему поколению достаточно сложно ориентироваться в информационном пространстве для поиска какой-либо информации. Так, например, детям достаточно сложно найти информацию об интересных местах какого-либо города или факты о нем. Связать это можно с тем, что на сегодняшний день не существует единого ресурса, который позволял бы на основе своих интересов и предпочтений осуществить выборовку интересной информации о городе, фактах, связанных с ним, или каких-либо мест.

В качестве решения описанной выше проблемы можно рассмотреть создание телеграмм-бота, который предварительно бы осуществлял сбор информацию о пользователе на основе небольшого анкетирования, связанного с выбором конечного множества интересов, которые являются наиболее ближе к пользователю, а затем осуществлял бы фильтрацию конечного списка информации о городе и предоставлял их пользователю.

Помимо прочего, следует затронуть проблему объемности информации в интернете и ее вариативности. Большой объем неконтролируемой генерации информации привел к появлению так называемого «цифрового мусора» – ненужных и забытых данных, которые в дальнейшем никак не будут использоваться. [4] Действительно, если у нас есть необходимость в срочном порядке найти какую-либо информацию за короткий промежуток времени, то мы, как минимум, просмотрим 2–3 страницы веб-сайтов для того, чтобы найти нужную.

Для упрощения поиска нужной информации в интернете люди изобрели рекомендательные алгоритмы. Рекомендательные алгоритмы – это алгоритмы, которые позволяют на основе информации о человеке, его каких-либо признаков и факторов выделить из большого множества различной информации ту, которая потенциально может заинтересовать пользователя в данный момент время. [4] Данные системы, помимо упрощения поиска нужной информации, пользуются спросом и в предпринимательстве с целью получения прибыли.

Так, например, сервис, который предоставляет доступ к музыке, «Spotify» использует рекомендательную систему для создания уникальных плейлистов для пользователей, которые содержат ранее не прослушанные пользователем песни, которые могут его заинтересовать – данное применение рекомендательной системы положительно сказывается на популярности данного сервиса. Также рекомендательные системы активно применяются в интернет-магазинах, которые предоставляют различные товары; многие пользователи сталкивались при выборе какого-либо товара на интернет-сайте с блоком «Рекомендуем вам…», который содержит товары на основе какой-либо информации о нас –данный блок сгенерирован как раз таки с помощью рекомендательной системы. Перечисленные примеры не являются конечными, рекомендательные системы используются в различных других сферах деятельности человека.

Таким образом, для того, чтобы выдавать ребенку ту информацию о городе, которая его действительно заинтересует, необходимо использовать рекомендательную систему, которая на основе каких-либо данных о ребёнке будет осуществлять фильтрацию собранной ранее информации о городе.

Исходя из вышесказанного, можно определить основную цель данного проекта, а именно создание телеграмм-бота, который будет предоставлять различную информацию о городе Санкт-Петербург младшему поколению, используя рекомендательный алгоритм как способ ранжирования предоставляемых данных на основе интересов пользователя.

Для достижения поставленной цели составим список задач, которые необходимо решить в данной работе:

* Провести обзор предметной области;
* Провести сбор данных о городе Санкт-Петербург;
* Провести анализ собранных данных и их подготовку для дальнейшей работы;
* Спроектировать и создать базу данных для хранения собранных данных о городе;
* Спроектировать и создать модель машинного обучения для реализации рекомендательного алгоритма;
* Спроектировать и создать чат-бота для мессенджера Telegram;
* Сравнить работу модели при холодном старте и при обученной модели машинного обучения;
* Сделать выводы о проделанной работе.

# АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

## Телеграмм-боты

Телеграмм-боты, как было отмечено ранее, – это роботы, с которыми пользователи могут взаимодействовать, отправляя им различные сообщения-запросы, команды, а те в свою очередь различно реагировать на них. Поведение же бота определяется ранее написанным скриптом программистом, а его управление осуществляется посредству HTTPS[[1]](#footnote-1)-запросов к Telegram Bot-API[[2]](#footnote-2). [5] Скрипты для бота в свою очередь пишутся на различных языках программирования, например, на Python или C#. Стоит отметить, что телеграмм-боты могут применяться как в одиночных чатах, так и в групповых или в каналах.

Сегодня телеграмм-боты имеют широкий спектр использования. Они позволяют автоматизировать многие рутинные задачи такие, как: фильтрация отправляемых пользователями сообщений, проверка на наличие нецензурных слов в сообщениях, спама, оповещение пользователя о чем-либо, вычисление каких-либо функций, предоставление какого-либо контента, также можно осуществлять фильтрацию пользователей, блокируя или предупреждая пользователей на основе чего-либо. [6]

Так, например, сейчас существуют телеграмм-боты, связанные со следующими аспектами жизни: [7]

* Боты для админов телеграмм каналов – боты, которые позволяют составлять план размещаемого контента, хранилище рекламных постов, рассылки информации и оповещение пользователей; боты, которые позволяют создать красочные посты, которые будут содержать различные изображения, гифки, видео, кнопки[[3]](#footnote-3);
* Боты для чатов – это боты, которые осуществляют модерирование каких-либо аспектов чата, например, таких, как фильтрацию сообщений, удаление спама и гиперссылок, блокировку пользователей за использование определённого списка слов, запрет на какие-либо действия пользователя. Такие боты также могут вносить развлекательную нотку в чат пользователей, например, своей возможность рассказать какой-либо анекдот;
* Боты для знакомств, которые позволяют разместить свой профиль для поиска как друзей, так и второй половинки. Все, что вам требуется сделать – заполнить анкету о себе, а затем ожидать заявок на дружбу со стороны других пользователей, либо же самому осуществлять просмотр анкет других пользователей и отправлять им запросы на дружбу;
* Боты для поиска посылок позволяют осуществлять отслеживание за вашими посылками. Так, вы можете обратиться к специальному телеграмм-боту, который осуществляет отслеживание посылок Почты России, написав ему специальный трек-код, а затем телеграмм-бот будет вас оповещать об изменении статуса посылки;
* Боты для работы с музыкой – позволяют осуществлять поиск нужной музыки и дальнейшее её скачивание;
* Боты, позволяющие работать с видео и фильмами. С их помощью можно скачивать видео или аудио с популярного видео-сервиса «Ютуб», осуществлять поиск нужного фильма/сериала или же осуществить подписку на уведомление о выходе новых фильмов или серий в сериале; также могут данные боты могут отправить афишу кинотеатров;
* Боты из категории «Книги», позволяют осуществлять скачивание множества бесплатных книг, также они могут помочь при выборе книг на основе интересов, помочь с выбором места покупки книги, дабы сэкономить денежные средства;
* Новостные боты – боты, которые представляют собой подписку на обновления какого-либо СМИ. Данные боты будут оповещать пользователя о произошедших новостях, они также могут осуществлять поиск статьи по ключевым словам, проводить фильтрацию новостей по различным категориями, например, по популярности;
* Боты для работы с изображениями позволяют обрабатывать изображения. Они могут производить наложение водяных знаков, осуществлять различную коррекцию, обрезание изображения или же поиск нужных картинок;
* Боты для продаж по аналогии с ботами для знакомств позволяют осуществить размещение поста о продаже, который смогут увидеть другие пользователи данного бота, либо можно просматривать посты от других пользователей и что-либо приобретать;
* Боты для займа – данные боты позволяют осуществить краткосрочные займы;

Вышеперечисленный список ботов не является конечным, но в большей степени охватывает перечень наиболее популярных.

Как было сказано, скрипты, на основе которых осуществляется контроль (поведение) бота пишутся на различных языка программирования таких, как Python, C#, Java, Node.JS и другие. Для хранения различной информации как о пользователях, так и о чем-либо другом используются цифровые хранилища, или же базы данных. Сам же скрипт вызывается посредству терминала. Для того, чтобы телеграмм-бот работал постоянно, скрипт загружают на определённый хостинг, или же осуществляют написание специального веб-приложения, который вызывает данный скрипт, а затем также загружается на хостинг. [8]

## Рекомендательные системы

Рекомендательная система – это система алгоритмов или единый алгоритм, который(-ые) на основе каких-либо признаков и факторов делает(-ют) выборовку из некоторого множества данных. Цель рекомендательной системы заключается в том, чтобы подобрать релевантный контент, основываясь на некоторых знаниях о пользователе, о том, какой ранее ему контент был интерес, о том, как он взаимодействовал с этим контентом. В наше время большинство популярных рекомендательных алгоритмов, использующихся в системах, основаны на машинном обучении. [9, 10]

В свою очередь машинное обучение заключается в формировании такого алгоритма для программы, на основе которого она будет самостоятельно находить решения определённых задач, комплексно используя некоторые статические данные, исходя из которых можно сформировать закономерности и на основе которых можно осуществить прогнозирование. [11]

Рассмотрим работу подборки подробнее на примере рекомендаций музыкальной системы Spotify:

1. Пользовательское приложение отправляет запрос на сервер Spotify о том, что необходимо составить плейлист с рекомендациями различных песен;
2. Сервер обращается к рекомендательной системе, которая выполняет несколько действий. Первым действием является получение вариантов объектов, в нашем случае песен. Сервер получает некоторый набор объектов, которые являются наиболее релевантными музыкальному вкусу пользователя в данный момент;
3. Допустим, на предыдущем этапе было получено 100 объектов-песен. Первые 40 песен, являющиеся лидерами списка, отправляются на следующий этап – анализ и прогнозирование;
4. Прогнозирование осуществляется на основе данных о пользователе, которые хранятся в базе данных пользователей. За счёт полученной информации о пользователе осуществляется перераспределение оценок в выбранной ранее группе лидирующих песен – отсекаются песни, которые прогнозируются низкой оценкой;
5. Как результат 4-ого пункта, остаются наиболее значимые песни. На данном этапе отобранные песни сортируются в порядке своих оценок, предпочтений пользователя;
6. Система возвращает конечный список песен;
7. Сервер выдаёт результат. [10]

В зависимости от необходимого результата логика рекомендательной системы может быть разной. На данный момент можно выделить четыре основные логики рекомендательных систем. Рассмотрим каждую логику отдельно:

1. Content-Based filtering – фильтрация, основанная на контенте. Данная фильтрация основана на текущих предпочтениях пользователя. К примеру, если в данный момент вы смотрели или приобрели мебель-кресло в интернете, то система, основываясь на этом, предложит вам ещё варианты кресел. [9, 12]
2. Collaborative filtering – коллаборативная фильтрация. Рассмотрим тот же вариант с креслом – вы купили его. Для того, чтобы вам не рекомендовать ещё варианты кресел, существуют коллаборативная фильтрация, а именно это система рассматривает подобных вам пользователей, то есть таких же, которые купили кресла, и исследует их дальнейшую историю покупок. Пользователи, приобрётшие кресла, будут искать различные аксессуары к данной мебель – основываясь на этом, коллаборативная система начнёт вам предлагать подобные аксессуары и прочие вещи (контент), которые были рассмотрены пользователями. [9, 12]
3. Knowledge-based filtering – фильтрация, основанная на знаниях. Данная система рекомендует объекты, базируясь на том, как определённые функции объекта удовлетворяют потребности и предпочтения пользователей; говоря другими словами, на том, как объект полезен пользователю. [13, 14, 9, 12] (Настроить сортировку ссылок)
4. Hybrid filtering – гибридные рекомендательные системы. Исходя из самого названия логической системы, можно сделать вывод, что данная система объединяет в себя несколько перечисленных выше систем. Используя гибридную логическую систему, можно ликвидировать недостатки «негибридных» систем. Например, объединяя content-based и collaborative системы, мы можем избавиться от проблемы рекомендации content-based товаров, которые мы уже приобрели ранее и в которых более не нуждаемся. [12, 14]

Рассмотрим также плюсы и минусы использования рекомендательных систем. Так рекомендательные системы несут несомненный плюс для различных онлайн-сервисов, которые помогают пользователю найти именно то, что ему интересно. [12]

С другой стороны, происходит некоторая деградация человека, а именно он начинает привыкать к рекомендациям, различным подсказкам рекомендательных систем. Именно это влияет на узкость видения человека, он перестаёт знакомиться с новыми тематиками и жанрами контента, потребляя при этом однотипный. Данная проблема явно отражается при использовании рекомендательных систем в новостных сервисах – человеческое мировоззрение фиксируется на одной позиции и никоем образом не развивается и не расширяется. [12]

Помимо проблемы узкости видения человека многие компании пренебрегают этическими ценностями. Так как основная цель компании при использовании рекомендательных систем – это повышение дохода, то компании намеренно создают рекомендательные системы, чтобы как можно дольше удержать внимание пользователя в электронном ресурсе. Тем самым рекомендательные системы могут негативно повлиять на психическое здоровье человека, вызывают различные зависимости и в целом тратят время человека. Так, например, всем известная проблема сервисов «Инстаграм» или «Тик-Ток», в которые заходишь буквально на 5 минут, а выходишь через несколько часов. [15]

## Обзор готовых решений

Для рассмотрения готовых решений необходимо предварительно определить потенциальных пользователей. Как было указано ранее, данная работа нацелена на детскую аудиторию, а именно дети с 5 до 13 лет. Проведём краткий обзор решений, направленных на работу с детьми данной возрастной категории или направленных на выдачу интересной информации на основе предпочтений пользователя. На текущий момент на рынке существуют следующие решения:

1. Бот Яндекс.Карт – бот, который находит интересные места в городе, а также сообщает дополнительно о пробках. Помимо прочего, бот подсказывает пользователю, в какое время поездка домой или на работу будет наиболее практичной, учитывая загруженность дорог.
2. Rss4you\_bot – телеграмм-бот, который создаёт персональную новостную ленту, сформированную на основе интересов пользователя. Стоит отметить, что использование данного бота является платным.

Базируясь на найденных решениях, можно сделать вывод, что на данный момент не существует решения, который в какой-либо мере был нацелен на детскую аудиторию, которые искали бы различную информацию в зависимости от предпочтений ребёнка. Таким образом, можно сказать, что необходимость на текущий день в написании бота существует.

## Выводы по главе 1

В первой главе был осуществлён анализ предметной области, а именно были описаны основные понятие такие, как: телеграмм-бот и рекомендательная система. Было рассказано про работу телеграмм-ботов, о том, что используется при их создании. Были рассмотрены категории жизни человека, в которых активно используются и применяются телеграмм-боты.

Также были рассмотрены рекомендательные системы, основные рекомендательные алгоритмы, пример работы рекомендательной системы; рассмотрены положительные и отрицательные стороны данной системы.

Помимо вышесказанного, был проведён поиск готовых решений на основе телеграмм-ботов по сформулированной ранее проблеме, связанной с поиском интересной информации для детей.

# СБОР ДАННЫХ И ЕЕ СТРУКТУРИЗАЦИЯ

## Классификация пользователей

Для того, чтобы качественно сформировать набор данных о городе Санкт-Петербурге, то есть наиболее валидную информацию, необходимо детально разобрать интересы будущих пользователей бота и провести классификацию пользователей.

Как было отмечено ранее, потенциальную аудиторию составляют дети с 5 до 7 и с 7 до 13 лет. В отдельности разберём интересы данных возрастных категорий и сформируем список интересов.

Для того, чтобы определить интересы человека, необходимо рассмотреть его восприятие и особенности мышления. Таким образом, основные моменты, на которые нужно обратить внимание при формировании конечного набора данных интересов ребёнка – это его эмоционально-волевая сфера, социальная сфера, интеллектуальное развитие, мышление и воображение. [16]

Для возрастной категории детей 5–7 лет характерно преобладание позитивных отношений с людьми, позитивное мироощущение, проявление интереса к познанию окружающего мира, наглядно-образное мышление. Также у таких детей постепенно расширяется кругозор, чему способствует появление нового вида общения, который предполагает разговоры о самих себе, других людях, человеческих отношениях, героях книг. [16]

Связи с этим можно сделать вывод, что большинству детей 5–7 лет оптимально подойдёт информация о различных сказках, мифологии или его мифических животных города.

Рассматривая категорию детей с 7 до 13 лет под углом интеллектуального развития, необходимо отметить, что в этом возрасте у ребёнка возникает механизм коррекции своего мышления со стороны логики и теоретического знания. Также ребёнок может хранить, извлекать из памяти информацию и обрабатывать её. Следовательно, в представленной возрастной категории ребёнок способен самостоятельно определять свои познавательные сферы, то бишь интересы. [17]

На основании вышесказанного обобщим интересы детей 7–13 лет, которые наиболее часто встречаются, самостоятельно:

* Интересные факты о городе и мифология;
* Музеи;
* Парки и аттракционы;
* Развлекательные центры;
* Арт-пространства и необычные кафе;
* Зоопарки;
* Театры;
* Цирки.

## Сбор данных о г. Санкт-Петербурге

Проведём сбор данных о городе в соответствии с ранее определёнными интересами детей. Определим первичный набор данных о легендах, мифах и сказках о городе, которые будут интересны детям 5–7 лет. Для поиска необходимой информации будем использовать общедоступные ресурсы в Интернете. Ниже представлен перечень данных о городе:

* Мифические животные [18, 19]:
  + Грифоны у воды;
  + Грифоны в стене;
  + Драконы на колоннах «Пале рояль»;
  + Драконий мост в парке Царского села;
  + Нетопыри в здании на перекрёстке Садовой и Вознесенского;
  + Демоны на Лахтинской улице;
  + Китайские львы Ши-Цза на Петровской набережной;
  + «Метафизические сфинксы» напротив «Крестов»;
  + Тритоны;
  + Медуза Горгона.
* Петербургские сказки [20]:
  + Сказка о Медном всаднике;
  + Нева;
  + Как образовались сфинксы;
  + Маленький волшебник в городе на Неве;
  + Медный всадник;
  + Обыкновенное чудо;
  + Петербургское приведение;
  + Сказка о реке Мойке;
  + Приключения Малой Невки.

Также определим перечень предоставляемой информации для детей 7 – 13 лет, а также вместе с этим укажем в скобках их классификацию (в дальнейшем будем называть их «тегами») по интересам детей [21, 22, 23, 24, 25]:

* «Легенда о Михайловский замок» (Интересные факты о городе и мифология);
* «Легенда о Большеохтинском мосте» (Интересные факты о городе и мифология);
* «Легенда о подземном ходе из особняка Матильды Кшесинской» (Интересные факты о городе и мифология);
* «Дом Пиковой дамы» (Интересные факты о городе и мифология);
* «Ксения Петербургская и Мефистофель» (Интересные факты о городе и мифология);
* «Михайловский замок. Дух Павла I» (Интересные факты о городе и мифология);
* «Призрак Распутина» (Интересные факты о городе и мифология);
* «Дух быков» (Интересные факты о городе и мифология);
* «Эрмитаж» (Интересные факты о городе и мифология, музеи);
* «Кунсткамера» (Интересные факты о городе и мифология, музеи);
* «Петропавловская крепость» (Интересные факты о городе и мифология);
* «Екатерининский дворец» (Интересные факты о городе и мифология, музеи);
* «Музей игрушки» (Музеи);
* «Музей советских игровых автоматов» (Музеи);
* «Музей магии» (Музеи);
* «Музей ретро-автомобилей Route 66» (Музеи);
* «Музей кошки» (Музеи);
* «Аквапарк Родео Драйв» (Парки и аттракционы);
* «Crazy Park» (Парки и аттракционы);
* «Гагарин Парк» (Парки и аттракционы);
* «Открытый парк аттракционов Планета Лета» (Парки и аттракционы);
* «Парк развлечений «Волшебная миля»» (Развлекательные центры);
* «Крытые веревочные парки «Высотный город»» (Развлекательные центры);
* «Умный город безопасности «ТачкиГрад»» (Развлекательные центры);
* «Сказкин дом» (Развлекательные центры);
* «Антимузей PhysLand» (Развлекательные центры);
* «Музей Эмоции» (Развлекательные центры);
* «Студия Смешариков» (Развлекательные центры);
* «Музей Лего» (Развлекательные центры, музеи);
* «Ленинградский зоопарк» (Зоопарки);
* «Пушистики» (Зоопарки);
* «Экологический центр «Лесная сказка»» (Зоопарки);
* «Экзоопарк» (Зоопарки);
* «Большой театр кукол» (Театры);
* «Театр кукол «Бродячая Собачка»» (Театры);
* «Театр Karlsson Haus» (Театры);
* «Театр «За черной речкой»» (Театры);
* «Цирк на Фонтанке» (Цирк);
* «Цирк в Автово» (Цирк);
* «Упсала-Цирк» (Цирк);

## Создание цифрового хранилища для данных

### Логическая модель базы данных

Для хранения данных о пользователях, а именно их классификации и будущих коэффициентов для уникальной подборки рекомендаций, необходимо спроектировать и создать цифровое хранилище, или базу данных (в дальнейшем БД).

Определим первичные объекты и охарактеризуем их признаками, которые их идентифицируют (см. таблица 2.1).

Таблица 2.1 – Первичные объекты

|  |  |
| --- | --- |
| Объект | Признаки |
| Пользователь | 1. Уникальный идентификатор; 2. Логин (Телефон); 3. Набор интересов; 4. Набор просмотренных данных о городе; 5. Общий набор оценок как поставленных самим пользователем, так и прогнозированные рекомендательной системой. |
| Данные о городе | 1. Уникальный идентификатор; 2. Основная информация в виде текста; 3. Набор тегов-интересов, которым может соответствовать информация. |

Логическая модель представляет собой графическое представление логической структуры базы данных. Данная модель иллюстрирует набор сущностей, их характеристические признаки (далее атрибуты) и взаимоотношения между собой.

Сущности в свою очередь описывают объекты, которые мы определили ранее в таблице 2.1. Зависимости, или же взаимоотношения, между сущностями иллюстрируются с помощью связей. Правила и ограничения взаимоотношений описываются с помощью свойств связей.

В данном случае логическая модель была построена на основе методологии IDEF1X. Сущности в IDEF1X описывает собой совокупность или набор экземпляров похожих по свойствам, но однозначно отличаемых друг от друга по одному или нескольким признакам. Связи в IDEF1X представляют собой ссылки, соединения и ассоциации между сущностями. Связи – это глаголы, которые показывают, как соотносятся сущности между собой.

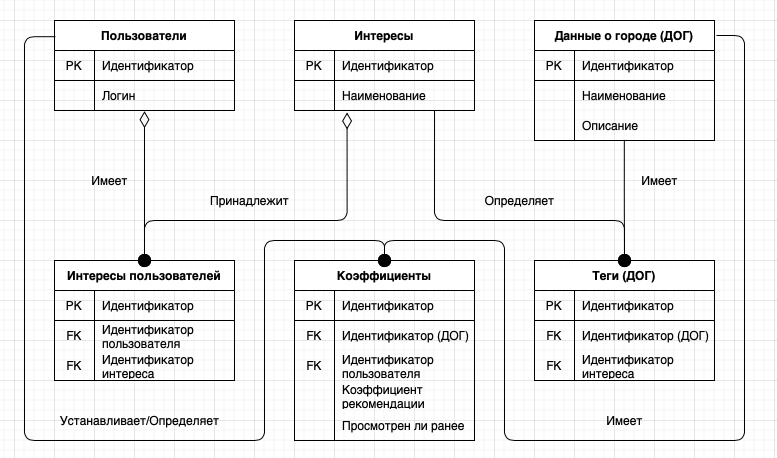
Ниже представлена диаграмма логических связей будущей базы данных на основе методологии IDEF1X.

Диаграмма 2.1 – Логическая модель базы данных

Добавить пояснения по логической модели, что за сущность, ее описание и зачем она нужна

### Нормализация базы данных

Нормализация – это метод организации реляционной базы данных с целью сокращения избыточности. Избыточность в свою очередь – это превышение хранимой информации в базе данных, которая описывает ту или иную сущность, которая влечёт за собой целый ряд проблем. Например, из-за избыточности база данных может занимать слишком много лишнего места на накопителе, или же появятся так называемые аномалии при работе с БД – аномалии – это проблемы, связанные с обработкой информации, которые могут возникать при удалении информации, её модифицировании или добавлении. Таким образом, нормализация данных позволяет на стадии проектировании базы данных минимизировать появление возможных аномалий. Устранение избыточности осуществляется не только за счёт разумного назначения размерности объектам таблицы, а также за счёт формирования эффективных взаимоувязанных табличных структур, объединяющих несколько отношений. [26]

Для того, чтобы провести нормализацию базы данных, вводится понятие нормальной формы (далее НФ) – это определённое состояние базы данных, структура которой соответствует определённым критериям, которые определяет НФ. Если база данных соответствует критериям НФ, то её называют полной по этой форме. [26]

Всего существует шесть нормальных форм. Нормализация базы данных начинается с первой нормальной формы. Чтобы перейти к следующей НФ, необходимо чтобы текущая структура БД уже была полной по предыдущим нормальным формам.

Зачастую для исключения большинства проблем, которые могут возникнуть с БД, достаточно, чтобы БД была, как минимум, полной по третьей форме. Таким образом, проведём нормализацию БД до третьей формы.

Первая нормальная форма, или 1НФ, требует:

* Любая запись таблицы должна содержать описание одной сущности;
* Значение в ячейке должно быть неделимым и не содержать группы значений;
* В таблице не должно быть повторяющихся групп поля, содержащие одинаковые по смыслу значения.

2НФ требует:

* Столбцы таблиц должны зависеть от полного первичного ключа, а не от его ключей;
* Поля таблиц, которые зависят только от части первичного ключа, должны быть выделены в составе отдельных таблиц;

3НФ предполагает отсутствие транзитивной зависимости, то есть зависимости значений не ключевых полей от значений других не ключевых полей.

Проведя анализ и сравнение существующей структуры базы данных с нормальными формами, можно сделать вывод, что текущая структура БД соответствует критериям третей нормальной форме, что за собой влечёт соответствие второй и первой формам. Таким образом, нормализация БД проведена до 3НФ.

### Создание базы данных

Для того, чтобы работать с базами данных существуют так называемые системы управления базами данных (далее СУБД). СУБД предоставляет большой набор функционала для работы с базой данных, начиная с достаточно очевидных: предоставление доступа к данным, хранящимся в БД, их модификация, добавление, удаление, заканчивая созданием статистикой, написанием триггеров и прочего.

В свою очередь клиенты могут работать с БД со стороны программного обеспечения, отправляя запросы СУБД. Большинство СУБД используют язык SQL (Structured Query Language) для работы с запросами. С помощью SQL можно осуществить следующие действия:

* Создание БД, таблиц и их структуры;
* Осуществление основных операций манипулирования данными, то есть добавление, модификация, удаление данных из таблиц;
* Выполнение простых и сложных запросов.

Техническом заданием ранее определена СУБД для работы с БД – SQLite. Осуществим реализацию спроектированной БД в данной СУБД. Таблицами ниже описаны свойства, типа данных и ограничения объектов, входивших в структуру БД.

Таблица 2.2 – «users» (Пользователи)

|  |  |  |  |
| --- | --- | --- | --- |
| № | Название поля | Тип данных | Ограничение |
| 1 | id | int | Primary key |
| 2 | login | text |  |

Таблица 2.3 – «interests» (Интересы)

|  |  |  |  |
| --- | --- | --- | --- |
| № | Название поля | Тип данных | Ограничение |
| 1 | id | int | Primary key |
| 2 | name | text |  |

Таблица 2.4 – «docs» (Данные о городе)

|  |  |  |  |
| --- | --- | --- | --- |
| № | Название поля | Тип данных | Ограничение |
| 1 | id | int | Primary key |
| 2 | name | text |  |
| 3 | description | text |  |

Таблица 2.5 – «user\_interests» (Интересы пользователей)

|  |  |  |  |
| --- | --- | --- | --- |
| № | Название поля | Тип данных | Ограничение |
| 1 | id | int | Primary key |
| 2 | user\_id | int | Foreign key |
| 3 | interest\_id | int | Foreign key |

Таблица 2.6 – «coeffs» (Коэффициенты)

|  |  |  |  |
| --- | --- | --- | --- |
| № | Название поля | Тип данных | Ограничение |
| 1 | id | int | Primary key |
| 2 | doc\_id | int | Foreign key |
| 3 | user\_id | int | Foreign key |
| 4 | coeff | real |  |
| 5 | is\_shown | int |  |

Таблица 2.7 – «tegs» (Теги данных о городе)

|  |  |  |  |
| --- | --- | --- | --- |
| № | Название поля | Тип данных | Ограничение |
| 1 | id | int | Primary key |
| 2 | doc\_id | int | Foreign key |
| 3 | interest\_id | int | Foreign key |

Связи между таблицами не претерпели каких-либо изменений и реализованы в точности по диаграмме 2.1.

## Выводы по главе 2

Во второй главе был осуществлён сбор данных и её структуризация.

Для того, чтобы осуществить сбор валидной данных о городе, предварительно была проведена классификация пользователей по возрасту и интересам. Проанализировав открытые источники информации касательно интересов детей определённых возрастов, были сформированы группы интересов для каждой возрастной категории детей.

Проведя классификацию пользователей, позже был осуществлён поиск наиболее валидной информации о городе, соответствующей интересам детей, из электронных ресурсов.

Для хранения найденных данных был было создано цифровое хранилище, а именно была создана логическая модель базы данных, проведена её нормализация до 3НФ и создана с помощью СУБД SQLite.

# ПРОЕКТИРОВАНИЕ И СОЗДАНИЕ ТЕЛЕГРАММ БОТА

## Проектирование рекомендательной системы

Задача рекомендательной системы в данной работе заключается в предоставлении максимально интересной информации для пользователя на основе его интересов. Ранжировать предоставляемую информацию не составит труда ввиду задания тегов-интересов. Так, например, информация о «Музее кошек» помечается тегом «музеи», что позволяет выделить его на фоне другой информации, если пользователь определил в своих интересах элемент «музеи». Такой подход называется, как было отмечено ранее, knowledge-based filtering.

Помимо выборовки информации на основе схожести интересов и тегов с пользователями, также можно воспользоваться коллаборативной фильтрацией (сollaborative filtering). Данный вариант предусматривает выявление наиболее похожих между собой пользователей. Так, например, если известно, что ближайший сосед поставил положительную оценку информации о «Музее кошек», мы можем сделать некоторый вывод о том, что данная информация может заинтересовать и другого схожего пользователя. Схожесть в нашем случае можно определять на основе интересов.

## Обучение рекомендательной системы

## Проектирование логической модели телеграмм бота

## Создание телеграмм бота

## Выводы по главе 3

# ЗАКЛЮЧЕНИЕ

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

|  |  |
| --- | --- |
| [1] | Д. Хоружий, «Хайтер. Цифровой мусор.,» [В Интернете]. Available: https://hightech.fm/2020/12/22/dirty-garbage. [Дата обращения: 17 04 2021]. |
| [2] | «Список стран по числу пользователей Интернета,» [В Интернете]. Available: https://ru.wikipedia.org/wiki/Список\_стран\_по\_числу\_пользователей\_Интернета. [Дата обращения: 17 04 2021]. |
| [3] | «Почему популярен Telegram?,» [В Интернете]. Available: https://zen.yandex.ru/media/id/5daa9819a660d700ad15b553/pochemu-populiaren-telegram-5-prichin-ustanovit-ego-sebe-na-telefon-5daa9874d5bbc300ae437f6b. [Дата обращения: 18 04 2021]. |
| [4] | «Рекомендательные системы: как алгоритмы подбирают нам товары и контент,» [В Интернете]. Available: https://hub.forklog.com/rekomendatelnye-sistemy-kak-algoritmy-podbirayut-nam-tovary-i-kontent/. [Дата обращения: 17 04 2017]. |
| [5] | «Bots: An introduction for developers,» [В Интернете]. Available: https://core.telegram.org/bots. [Дата обращения: 18 04 2021]. |
| [6] | «What is Telegram Bot? The complete Guide On Telegram Bot,» [В Интернете]. Available: https://www.blogsgeek.com/telegram-bot/. [Дата обращения: 18 04 2021]. |
| [7] | П. Покровский, «Боты для Телеграм - ТОП-100 самых полезных и интересных ботов,» [В Интернете]. Available: https://tgrm.su/blog/faq/boty-telegram/#17. [Дата обращения: 13 05 2021]. |
| [8] | «Как написать Telegram бота: практическое руководство,» [В Интернете]. Available: https://khashtamov.com/ru/create-telegram-bot-in-python/. [Дата обращения: 13 05 2021]. |
| [9] | «Как работают рекомендательные системы,» [В Интернете]. Available: https://neurohive.io/ru/osnovy-data-science/rekomendatelnye-sistemy-modeli-i-ocenka/. [Дата обращения: 19 04 2021]. |
| [10] | К. Фальк, Рекомендательные системы на практике, Москва: ДМК Пресс, 2020, p. 36. |
| [11] | «Введение в машинное обучение,» [В Интернете]. Available: https://habr.com/ru/post/448892/. [Дата обращения: 20 04 2021]. |
| [12] | «Что такое рекомендательные системы и как они работают,» [В Интернете]. Available: https://skillbox.ru/media/code/chto\_takoe\_rekomendatelnye\_sistemy\_i\_kak\_oni\_rabotayut/. [Дата обращения: 19 04 2021]. |
| [13] | F. Ricci, L. Rokarch и B. Shapira, «Recommender System Handbook,» [В Интернете]. Available: https://www.researchgate.net/publication/227268858\_Recommender\_Systems\_Handbook. [Дата обращения: 19 04 2021]. |
| [14] | overmes, «Рекомендательные системы: You can (not) advise,» [В Интернете]. Available: https://habr.com/ru/post/176549/. [Дата обращения: 18 04 2021]. |
| [15] | Skillbox, «Философия алгоритма: как рекомендательные системы изменили человечество,» [В Интернете]. Available: https://vc.ru/future/204385-filosofiya-algoritma-kak-rekomendatelnye-sistemy-izmenili-chelovechestvo. [Дата обращения: 13 05 2021]. |
| [16] | «Особенности развития ребенка в 5-7 лет,» [В Интернете]. Available: https://kukuriku.ru/razvitie/kalendar/ot-5-do-6-let/osobennosti-razvitiya-doskholnikov/. [Дата обращения: 26 04 2021]. |
| [17] | Л. С. Выготский, «Культурно-историческая теория Л. С. Выготского,» [В Интернете]. Available: https://cyberpedia.su/4x5b8b.html. [Дата обращения: 27 04 2021]. |
| [18] | «Восемь мифических существ Петербурга,» [В Интернете]. Available: https://www.fiesta.ru/spb/places/vosem-mificheskih-suschestv-peterburga/. [Дата обращения: 27 04 2021]. |
| [19] | «Мифологические существа в Санкт-Петербурге,» [В Интернете]. Available: https://www.spb-guide.ru/page\_18342.htm. |
| [20] | «Петербургские сказки,» [В Интернете]. Available: https://deti.spb.ru/catalog/ubilei/tvorchestvo/. [Дата обращения: 27 04 2021]. |
| [21] | «Легенды и мифы Санкт-Петербурга,» [В Интернете]. Available: https://4traveler.ru/collections/legendy-i-mify-sankt-peterburga. [Дата обращения: 27 04 2021]. |
| [22] | «От Мефистофеля до "Крестов": Карта самых интересных мифов Петербурга,» [В Интернете]. Available: https://www.the-village.ru/city/the-village-guide/224001-spooky-spb. [Дата обращения: 27 04 2021]. |
| [23] | «Что лучше всего рассказать детям о Петербурге,» [В Интернете]. Available: https://razvivashka.online/poznavatelnoe/detyam-o-peterburge. [Дата обращения: 27 04 2021]. |
| [24] | «Достопримечательности Санкт-Петербурга с описанием для детей,» [В Интернете]. Available: https://arslogistik.ru/bez-rubriki/dostoprimechatelnosti-sankt-peterburga-s-opisaniem-dlya-detej.html. [Дата обращения: 27 04 2021]. |
| [25] | «100 самых интересных мест для детей в Петербурге,» [В Интернете]. Available: https://zen.yandex.ru/media/bedareva\_life/100-samyh-interesnyh-mest-dlia-detei-v-peterburge-spisok-so-ssylkami-5db95d06fc69ab00acbae446. [Дата обращения: 27 04 2021]. |
| [26] | О. Д.Л., Технологии проектирования баз данных, ДМК Пресс ред., 2019, pp. 131-145. |
| [27] | «Информацинной мусор: что это и как от него избавиться?,» [В Интернете]. Available: https://mentamore.com/socium/informacionnyj-musor.html. [Дата обращения: 17 04 2021]. |
| [28] | «Хобби и увлечения ребенка: помогаем выбрать,» [В Интернете]. Available: https://www.analogi.net/detyam/razvitie-i-vospitanie/hobbi-i-uvlecheniya-rebenka. [Дата обращения: 19 04 2021]. |

# ПРИЛОЖЕНИЕ А

1. HTTPS (аббр, от англ, HyperText Transfer Protocol Secure) – расширение протокола HTTP для поддержки шифрования в целях повышения безопасности. [↑](#footnote-ref-1)
2. API –(аббр, от англ, Application Programming Interface) – описание способов, которыми одна программа может взаимодействовать с другой программой. [↑](#footnote-ref-2)
3. Кнопки в телеграмм-боте – это активные функциональные объекты, нажатие на которые производит отправку определённого запроса боту, или же содержащее иную функциональную нагрузку такую, как гиперссылку. [↑](#footnote-ref-3)