**СОДЕРЖАНИЕ**

[**ВВЕДЕНИЕ** 5](#_Toc155894110)

[**1** **Организация поиска информации в сети Интернет** 7](#_Toc155894111)

[**1.1** **Поиск информации** 7](#_Toc155894112)

[**1.2** **Анализ полученной информации** 12](#_Toc155894113)

[**9.** **ЗАКЛЮЧЕНИЕ** 15](#_Toc155894114)

[**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ** 16](#_Toc155894115)

# **ВВЕДЕНИЕ**

Интернет — это глобальная сеть, которая объединяет устройства по всему миру и дает пользователям множество возможностей: работать, учиться, вести бизнес и так далее. В наши дни от работоспособности интернета зависят буквально все процессы жизнедеятельности. «РБК Тренды» публикуют перевод материала Internal Pointers, где приводится подробный разбор архитектуры Интернета.

Компьютерные сети — это сложные объекты, которые различаются по размеру, доступности и применяемым технологиям. Внутридомовая сеть с точки зрения размера, структуры, оборудования и функциональности сильно отличается от той, которая соединяет два города. Такое разнообразие сетей затрудняет их совместную работу. Однако решение кроется в интерсетях или интернете.

Межсетевые соединения — это тип сети, который соединяет несколько сетей и нивелирует их различия, чтобы они могли взаимодействовать как совместимые.

Главная общедоступная сеть называется интернетом. Любой может получить доступ к интернету с помощью устройства — компьютера, телефона, планшета и так далее — для обмена информацией и совместного использования ресурсов.

Интернет состоит из трех физических компонентов: кабелей, маршрутизаторов и сетей. Технические специалисты определяют интернет-инфраструктуру как сочетание интернет-оборудования и программного обеспечения, которое на нем работает.

Интернет образуется из множества оптоволоконных кабелей, которые соединяют вместе несколько небольших сетей. Эти кабели передают информацию из одной сети в другую в виде оптических сигналов. Вся наша планета покрыта интернет-кабелями, кроме Антарктиды. При этом многие расположены под водой, чтобы покрывать огромные расстояния. Например, длина кабеля Atlantic Crossing 1 (AC-1), который связывает США с Европой, составляет 14 тыс. км.

Еще одним важным элементом интернета являются маршрутизаторы, которые пересылают данные между точками доступа. Таким образом, маршрутизаторы отвечают за то, чтобы сообщение, отправленное с компьютера А, дошло до компьютера Б как можно быстрее и без сбоев.

Информация может передаваться по нескольким маршрутам, а функция маршрутизаторов состоит в том, чтобы выбрать лучший. Это особенно важно, когда сеть перегружена — в случае, если многие другие компьютеры одновременно пытаются обмениваться данными — или если кратчайший путь заблокирован из-за неисправности сети.

Целью данного реферата является поиск информации по теме магистерской работы. Темой магистерской работы является создание мобильного приложения для интеллектуального анализа данных.

Интеллектуальный анализ данных – неправильный термин, поскольку цель интеллектуального анализа данных не состоит в том, чтобы извлекать или анализировать сами данные. На самом деле уже имеется большой объем данных, из которых нужно извлечь значение или ценные знания. Ниже описан типичный процесс сбора, хранения, анализа и интеллектуального анализа данных.

* Сбор данных – это сбор данных из разных источников, таких как отзывы клиентов, платежи и заказы на покупку.
* Хранение данных – это процесс хранения этих данных в большой базе данных или [хранилище данных](https://aws.amazon.com/data-warehouse/).
* Аналитика данных – это дальнейшая обработка, хранение и анализ данных с использованием сложного программного обеспечения и алгоритмов.
* Интеллектуальный анализ данных – это направление аналитики данных или аналитическая стратегия, используемая для поиска скрытых или ранее неизвестных закономерностей в данных.

Интеллектуальный анализ данных является важной частью любой успешной аналитической инициативы. Компании могут использовать процесс обнаружения знаний, чтобы повышать доверие клиентов, находить новые источники дохода и привлекать новых клиентов. Эффективный интеллектуальный анализ данных помогает в различных аспектах бизнес-планирования и управления операциями.

Межотраслевой стандартный процесс интеллектуального анализа данных (Cross-Industry Standard Process for Data Mining, CRISP-DM) является отличным руководством для начала процесса интеллектуального анализа данных. CRISP-DM – это методология и модель процессов, не зависящая от отрасли, инструментов и приложений.

* В качестве методологии CRISP-DM описывает типичные этапы проектов интеллектуального анализа данных, обозначает задачи, выполняемые на каждом из этапов, и объясняет взаимосвязи между ними.
* В качестве модели процессов CRISP-DM предоставляет обзор жизненного цикла интеллектуального анализа данных.

Цели выполнения реферата:

1. ознакомиться с всемирной компьютерной сетью Интернет и используемыми в ней технологиями;

2. освоить приемы эффективного поиска необходимой информации в сетевых ресурсах;

3. получить практический опыт поиска информации по заданной теме;

4. освоить методы создания Web- документов;

5. получить практический опыт создания Web- страниц на основе найденных в сети

документов;

6. получить более глубокую теоретическую и практическую подготовку по

специальности на основе самостоятельного изучения и обобщения научной,

учебной литературы и нормативных документов, а также формирования умений и навыков по работе в сети;

7. сделать обзор литературы по теме диссертации, используя ресурсы сети Интернет.

1. **Организация поиска информации в сети Интернет**
   1. **Поиск информации**

Информация, размещенная во Всемирной сети, исчисляется огромным количеством байт. Для поиска информации во Всемирной сети используются специальные веб-сайты – информационно-поисковые системы. Они позволяют по ключевым словам найти информационные ресурсы, связанные с ключевыми словами. Это может быть текст, содержащий ключевые слова, или графическое изображение одного из ключевых слов. Примерами информационно-поисковых систем являются системы Google и Yandex.

**Поиск информации** – одна из самых востребованных на практике задач, которую приходится решать любому пользователю Интернета.

**Существуют три основных способа поиска информации в Интернет:**

1. Указание адреса страницы.

2. Передвижение по гиперссылкам.

3. Обращение к поисковой системе (поисковому серверу).

**Способ 1:** Указание адреса страницы

Это самый быстрый способ поиска, но его можно использовать только в том случае, если точно известен адрес документа или сайта, где расположен документ.

Не стоит забывать возможность поиска по открытой в окне браузера web-странице (Правка-Найти на этой странице…).

**Способ 2:** Передвижение по гиперссылкам

Это наименее удобный способ, так как с его помощью можно искать документы, только близкие по смыслу текущему документу.

**Способ 3:** Обращение к поисковой системе

Пользуясь гипертекстовыми ссылками, можно бесконечно долго путешествовать в информационном пространстве Сети, переходя от одной web-страницы к другой, но если учесть, что в мире созданы многие миллионы web-страниц, то найти на них нужную информацию таким способом вряд ли удастся.

На помощь приходят специальные поисковые системы (ихеще называют поисковыми машинами). Адреса поисковых серверов хорошо известны всем, кто работает в Интернете. В настоящее время в русскоязычной части Интернет популярны следующие поисковые серверы:Яндекс (yandex.ru), Google (google.ru) и Rambler (rambler.ru

**Поисковая система** — веб-сайт, предоставляющий возможность поиска информации в Интернете.

Большинство поисковых систем ищут информацию на сайтах Всемирной паутины, но существуют также системы, способные искать файлы на ftp-серверах, товары в интернет-магазинах, а также информацию в группах новостей Usenet.

**По принципу действия** поисковые системы делятся на два типа: поисковые каталоги и поисковые индексы.

**Поисковые каталоги** служат для тематического поиска.

Информация на этих серверах структурирована по темам и подтемам. Имея намерение осветить какую-то узкую тему, нетрудно найти список web-страниц, ей посвященных.

Катало́г ресурсов в Интернете или каталог интернет-ресурсов или просто интернет-каталог — структурированный набор ссылок на сайты с кратким их описанием.

Каталог в котором ссылки на сайты внутри категорий сортируются по популярности сайтов называется рейтинг (или топ).

**Поисковые индексы** работают как алфавитные указатели. Клиент задает слово или группу слов, характеризующих его область поиска, — и получает список ссылок на web-страницы, содержащие указанные термины.

Первой поисковой системой для Всемирной паутины был «Wandex», уже не существующий индекс, разработанный Мэтью Грэйем из Массачусетского технологического института в 1993.

**Как работает поисковой индекс?**

Поисковые индексы автоматически, при помощи специальных программ(веб-пауков), сканируют страницы Интернета и индексируют их, то есть заносят в свою огромную базу данных.

**Поисковый робот** («веб-паук») — программа, являющаяся составной частью поисковой системы и предназначенная для обхода страниц Интернета с целью занесения информации о них (ключевые слова) в базу поисковика. По своей сути паук больше всего напоминает обычный браузер. Он сканирует содержимое страницы, забрасывает его на сервер поисковой машины, которой принадлежит и отправляется по ссылкам на следующие страницы.

В ответ на запрос, где найти нужную информацию, поисковый сервер возвращает список гиперссылок, ведущих web-страницам, на которых нужная информация имеется или упоминается. Обширность списка может быть любой, в зависимости от содержания запроса.

http://www.yandex.ru/

**Яндекс** — российская система поиска в Сети. Сайт компании, Yandex.ru, был открыт 23 сентября 1997 года. Головной офис компании находится в Москве. У компании есть офисы в Санкт-Петербурге, Екатеринбурге, Одессе и Киеве. Количество сотрудников превышает 700 человек.

Слово «Яндекс» (состоящее из буквы «Я» и части слова index; обыгран тот факт, что русское местоимение «Я» соответствует английскому «I») придумал Илья Сегалович, один из основателей Яндекса, в настоящий момент занимающий должность технического директора компании.

Поиск Яндекса позволяет искать по Рунету документы на русском, украинском, белорусском, румынском, английском, немецком и французском языках с учётом морфологии русского и английского языков и близости слов в предложении. Отличительная особенность Яндекса — возможность точной настройки поискового запроса. Это реализовано за счёт гибкого языка запросов.

По умолчанию Яндекс выводит по 10 ссылок на каждой странице выдачи результатов, в настройках результатов поиска можно увеличить размер страницы до 20, 30 или 50 найденных документов.

Время от времени алгоритмы Яндекса, отвечающие за релевантность выдачи, меняются, что приводит к изменениям в результатах поисковых запросов. В частности, эти изменения направлены против поискового спама, приводящего к нерелевантным результатам по некоторым запросам.

http://www.google.ru/

Лидер поисковых машин Интернета, **Google** занимает более 70 % мирового рынка. Cейчас регистрирует ежедневно около 50 млн поисковых запросов и индексирует более 8 млрд веб-страниц. Google может находить информацию на 115 языках.

По одной из версий, Google — искажённое написание английского слова googol. «Googol (гугол)» – это математический термин, обозначающий единицу со 100 нулями. Этот термин был придуман Милтоном Сироттой, племянником американского математика Эдварда Каснера, и впервые описан в книге Каснера и Джеймса Ньюмена «Математика и воображение»(Mathematics and the Imagination). Использование этого термина компанией Google отражает задачу организовать огромные объемы информации в Интернете.

Интерфейс Google содержит довольно сложный язык запросов, позволяющий ограничить область поиска отдельными доменами, языками, типами файлов и т. д.

http://www.rambler.ru/

Rambler Media Group — интернет-холдинг, включающий в качестве сервисов поисковую систему, рейтинг-классификатор ресурсов российского Интернета, информационный портал.

Rambler создан в 1996 году.

Поисковая система Рамблер понимает и различает слова русского, английского и украинского языков. По умолчанию поиск ведётся по всем формам слова.

Информационно-поисковая система (ИПС) — это cистема, обеспечивающая поиск и отбор необходимых данных в специальной базе с описаниями источников информации (индексе) на основе информационно-поискового языка и соответствующих правил поиска. Главной задачей любой ИПС является поиск информации релевантной информационным потребностям пользователя. Очень важно в результате проведенного поиска ничего не потерять, то есть найти все документы, относящиеся к запросу, и не найти ничего лишнего.

Доступность информации для читателей научной библиотеки обеспечивается через четко организованный справочно-библиографический аппарат, который многоаспектно раскрывает состав и содержание единого библиотечного фонда. Одной из главных составляющих справочно-библиографического аппарата является система библиотечных каталогов.

Система каталогов научной библиотеки представляет собой информационный ресурс, состоящий из электронного каталога (ЭК) и законсервированных карточных каталогов и картотек, дополняющих ЭК информацией о ретроспективной части библиотечного фонда.

ЭК научной библиотеки является ключевым звеном всей технологической цепочки автоматизации библиотечно-библиографических процессов. ЭК ведется с 1993 года и представляет собой единую базу данных, которая включает библиографическую и полнотекстовую информацию о документах различных видов.

ЭК содержит свыше 640 тысяч библиографических записей на, монографии, сборники научных трудов, учебники и учебные пособия, диссертации, авторефераты диссертаций, электронные ресурсы, учебно-методические пособия, сведения о периодических изданиях, аналитическую роспись статей, а также научно-образовательные ресурсы сотрудников ОГУ.

В процессе работы по формированию ЭК выявились его несомненные преимущества над традиционной системой карточных каталогов и картотек.

Наличие всех обязательных элементов библиографического описания, информационно-поисковых языков и других сведений, приведенных по правилам заполнения полей, необходимых для идентификации и характеристики документа в базе данных, позволяют ЭК выполнять функции всех видов каталогов:

1. по назначению (читательского, служебного, топографического);
2. по способу группировки (алфавитного, систематического, предметного);
3. по виду отражаемых документов (книги, статьи, нормативно-технические документы и др.);
4. по отражаемым фондам (ЭК одной библиотеки, или сводный ЭК).

Объединяя и сохраняя достоинства традиционных карточных каталогов, и значительно превосходя их по поисковым возможностям, ЭК выполняет целый ряд дополнительных функций:

1. электронный учет фонда;
2. электронный заказ изданий;
3. электронная книговыдача;
4. информация о местонахождении и доступности экземпляра в любой момент времени;
5. электронная проверка фонда (инвентаризация) и т. д.

ЭК доступен пользователям с любого автоматизированного рабочего места, подключенного к Интернет. Поиск в ЭК одновременно эффективен и прост. Его главное преимущество, состоит в том, что поиск документа может быть осуществлен по разным поисковым аспектам с максимальной полнотой и точной информацией о каждом экземпляре.

Программа предусматривает две формы поиска - основной (классический) и расширенный поиск.

Классический поиск - предназначен для задания простого поискового запроса (по автору и /или заглавию). Форма имеет одно поле ввода. Также форма содержит ограничения для поиска по виду издания.

Расширенный поиск - предназначен для задания расширенного запроса и содержит больше полей для ввода данных по сравнению с формой классического поиска. Эта форма полезна, если необходимо сформировать поисковый запрос на основе нескольких поисковых атрибутов. Форма расширенного поиска позволяет накладывать дополнительные фильтры на процесс поиска, тем самым, сужая и конкретизируя его область.

* 1. **Анализ полученной информации**

В открытом доступе нет информации об аналогах разрабатываемого программного обеспечения, так что ниже будет приведен анализ информации найденной по данной теме.

<https://learn.microsoft.com/ru-ru/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>

Интеллектуальный анализ данных представляет собой процесс обнаружения пригодных к использованию сведений в крупных наборах данных. В интеллектуальном анализе данных применяется математический анализ для выявления закономерностей и тенденций, существующих в данных. Обычно такие закономерности нельзя обнаружить при традиционном просмотре данных, поскольку связи слишком сложны, или из-за чрезмерного объема данных.

Эти закономерности и тренды можно собрать вместе и определить как модель интеллектуального анализа данных. Модели интеллектуального анализа данных могут применяться к конкретным сценариям, а именно:

* **Прогнозирование**: оценка продаж, прогнозирование нагрузки сервера или времени простоя сервера
* **Риск и вероятность**: выбор наиболее подходящих заказчиков для целевой рассылки, определение точки равновесия для рискованных сценариев, назначение вероятностей диагнозам или другим результатам
* **Рекомендации**: определение продуктов, которые с высокой долей вероятности могут быть проданы вместе, создание рекомендаций
* **Поиск последовательностей**: анализ выбора заказчиков во время совершения покупок, прогнозирование следующего возможного события
* **Группирование**: разделение заказчиков или событий на кластеры связанных элементов, анализ и прогнозирование общих черт

Построение модели интеллектуального анализа данных является частью более масштабного процесса, в который входят все задачи, от формулировки вопросов относительно данных и создания модели для ответов на эти вопросы до развертывания модели в рабочей среде.

<https://aws.amazon.com/ru/what-is/data-mining/>

Последовательность этапов CRISP-DM строго не определена, поэтому команды по работе с данными могут по необходимости перемещаться вперед и назад между этапами. Кроме того, программные технологии могут выполнять некоторые из этих задач или поддерживать их.

**1.Понимание бизнеса**

Сначала специалисты по обработке и анализу данных определяют цель и объем проекта. Они сотрудничают с заинтересованными сторонами бизнеса, чтобы получить определенную информацию, например:

* проблемы, которые требуют решения;
* ограничения проекта;
* воздействие потенциальных решений на бизнес.

Затем они используют эту информацию, чтобы определить цели интеллектуального анализа данных и ресурсы, необходимые для обнаружения знаний.

**2. Понимание данных**

Определив бизнес-проблему, специалисты по работе с данными приступают к предварительному анализу данных. Они собирают наборы данных из различных источников, получают права доступа и готовят отчет с описанием данных. В отчете содержится информация о типах и количестве данных, требованиях к аппаратному и программному обеспечению для обработки данных. После утверждения плана компанией они приступают к исследованию и проверке данных. Они обрабатывают данные, используя базовые статистические методы, оценивают качество данных и выбирают итоговый набор данных для следующего этапа.

**3. Подготовка данных**

Специалисты по анализу данных тратят больше всего времени на этот этап, поскольку ПО для интеллектуального анализа данных требует высококачественных данных. В бизнес-процессах данные собираются и хранятся по причинам, отличным от интеллектуального анализа, поэтому специалисты по анализу должны подготовить данные, прежде чем использовать их для моделирования. Подготовка данных состоит из указанных ниже процессов.

#### **Очистка данных**

Например, обработка отсутствующих данных, ошибок и стандартных значений, а также корректировка данных.

#### **Интеграция данных**

Например, объединение двух разных наборов данных для получения итогового целевого набора.

#### **Форматирование данных**

Например, преобразование типов данных или настройка данных для конкретной используемой технологии интеллектуального анализа.

1. **Моделирование данных**

Специалисты по анализу данных вводят подготовленные данные в ПО для интеллектуального анализа данных и изучают результаты. Для этого они могут использовать один из множества методов и инструментов интеллектуального анализа данных. Они также должны написать тесты для оценки качества результатов интеллектуального анализа данных. Для моделирования данных специалисты по работе с данными могут:

* обучать модели машинного обучения на небольших наборах данных с известными результатами;
* использовать модели для дальнейшего анализа неизвестных наборов данных;
* настраивать или перенастраивать ПО для интеллектуального анализа данных, чтобы получать удовлетворительные результаты.

1. **Оценка**

После создания моделей специалисты по анализу данных начинают сопоставлять их с первоначальными бизнес-целями. Они делятся результатами с бизнес-аналитиками и собирают обратную связь. Модель может предоставить ответ на первоначальный вопрос или показать новые и ранее неизвестные закономерности. Специалисты по анализу данных могут изменить модель, скорректировать бизнес-цель или пересмотреть данные, полагаясь на обратную связь компании. Непрерывная оценка, обратная связь и внесение изменений являются частью процесса обнаружения знаний.

**9. Развертывание**

Во время развертывания другие заинтересованные стороны используют рабочую модель для получения бизнес-аналитики. Специалисты по работе с данными планируют процесс развертывания, который включает в себя обучение других функциям модели, постоянный мониторинг и поддержку приложения интеллектуального анализа данных. Бизнес-аналитики используют приложение для создания отчетов для руководства, обмена результатами с клиентами и улучшения бизнес-процессов.

1. **ЗАКЛЮЧЕНИЕ**

В заключение можно сказать, что интеллектуальный анализ данных — это метод, используемый в науке о данных для извлечения ценной информации и знаний из больших наборов данных. Он включает использование алгоритмов и статистических моделей для анализа и выявления закономерностей, тенденций и взаимосвязей в данных.

Цель интеллектуального анализа данных — выявить скрытые закономерности и идеи в данных, которые можно использовать для принятия лучших решений и получения конкурентного преимущества. Интеллектуальный анализ данных можно использовать в различных приложениях, включая маркетинг, здравоохранение, финансы и обнаружение мошенничества.

Очистка данных является важным этапом в процессе анализа данных. Он включает выявление и исправление ошибок, несоответствий и неточностей в данных, чтобы гарантировать их точность, полноту и актуальность для анализа.

Важность очистки данных в анализе данных можно резюмировать следующим образом:

1. Точные результаты: очистка данных помогает гарантировать, что анализ основан на точных и надежных данных. Неверные или неполные данные могут привести к ошибочным выводам и неточным результатам, что может иметь серьезные последствия для принятия решений.
2. Экономия времени: очистка данных может сэкономить много времени в долгосрочной перспективе за счет уменьшения необходимости ручных исправлений и повторного анализа. Очищая данные в самом начале, аналитики могут сосредоточиться на анализе данных, а не на исправлении ошибок.
3. Улучшенное качество данных. Очистка данных помогает улучшить общее качество данных. Это особенно важно при работе с большими наборами данных, которые могут содержать ошибки и несоответствия, которые трудно выявить и исправить без надлежащей очистки.

В целом, интеллектуальный анализ данных играет решающую роль в науке о данных, помогая извлекать ценные идеи и знания из больших и сложных наборов данных. Это позволяет организациям принимать обоснованные решения, улучшать свою деятельность и получать конкурентные преимущества.

Таким образом, очистка данных является важным шагом в процессе анализа данных. Это помогает обеспечить точность, согласованность и актуальность данных, что имеет решающее значение для принятия обоснованных решений на основе надежных данных.

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. TimePad [Электронный ресурс] / Портал выбора технологий и поставщиков. – Режим доступа: https://www.tadviser.ru/index.php/ TimePad – Дата доступа: 10.10.2023

2. Enternet [Электронный ресурс] / Википедия. Свободная энциклопедия. – Режим доступа: https://ru.wikipedia.org/wiki/enternet – Дата доступа: 11.11.2022

3. AI[Электронный ресурс] / Интеллектуальный поиск. – Режим доступа: https://livetyping.com/ru/blog/na-chem-pishut-prilozhenija-pod-android – Дата доступа: 10.11.2023

4. TensorFlow и PyTorch [Электронный ресурс] /. Сравнение фреймворков глубого обучения. – Режим доступа: https://habr.com/ru/company/ru\_mts/blog/565456/ – Дата доступа: 12.11.2023

5. Android Studio [Электронный ресурс] / Первая встреча с Android Studio. – Режим доступа: https://developer.android.com/studio/intro – Дата доступа: 13.11.2023

6. IntelliJ IDEA [Электронный ресурс] /. Документация и справочная. – Режим доступа: https://www.jetbrains.com/ru-ru/idea/ – Дата доступа: 13.11.2023

7. Eclipse [Электронный ресурс] / Платформа Eclipsed. – Режим доступа: https://hightech.in.ua/content/art-eclipse-platform – Дата доступа: 13.11.2023

8. Visual Studio Code [Электронный ресурс] / Visual Studio Code Docs. – Режим доступа: https://code.visualstudio.com/docs – Дата доступа: 14.04.2023

9. Sublime Text [Электронный ресурс] / Платформа Sublime. – Режим доступа: https://www.sublimetext.com/– Дата доступа: 15.11.2023

10. Python IDLE [Электронный ресурс] / Платформа IDLE. – Режим доступа: https://docs.python.org/3/library/idle.htmlplatform – Дата доступа: 15.04.2023

11. Gradle [Электронный ресурс] /. Пользовательский учебник по Gradle – Режим доступа: https://docs.gradle.org/current/userguide/userguide.html – Дата доступа: 25.11.2023

12. VGG16 [Электронный ресурс] /. VGG16 and VGG19. – Режим доступа: https://keras.io/api/applications/vgg/ – Дата доступа: 20.10.2023

13. Арлоу Д., Нейшталд А. UML и Унифицированный процесс: практический объектно-ориентированный анализ и проектирование, 2-е издание. –

Пер. с англ. – СПб: Символ-Плюс, 2007. – 624 с.