

## Конспект к уроку «Структурирование и анализ данных»

Данный урок входит в серию уроков «Культура работы с данными» и является вторым в ней. Урок можно провести обособленно или в комплексе с другими уроками серии.

**Цель урока:** сформировать представление об основных понятиях по структурированию данных и их анализу.

### Задачи урока:

- познакомить обучающихся с понятиями структурированных и неструктурированных данных; задачами анализа данных;
- научить структурировать данные и проводить их анализ средствами табличного процессора;
- сформировать представление о специалистах сферы работы с данными.

### Сценарный ход урока

**Педагог (слайд №1):** Добрый день! Сегодня мы будем говорить о структурированных данных, что это такое и зачем данные нужно структурировать. Как вы понимаете термин структура? Вспомните, что мы называем данными? Кто-то может привести примеры структурированных данных?

Структура — [лат. structura] взаимоположение и связь составных частей чего-либо; строение; отношение порядка на множестве элементов.

Данные [data] - сведения, факты, показатели, выраженные как в числовой, так и любой другой форме. Информация – это продукт взаимодействия данных и адекватных им методов. Информация работает с данными, данные - составная часть информации.

**Обучающиеся отвечают на вопрос педагога, приводят примеры.**

**Педагог (слайд №2):** Данные называют структурированными, если они упорядочены таким образом, что можно применять к ним различные методы и инструменты обработки (математические, логические статистические).

Вопрос педагога: как вы думаете - где и как можно организовать данные?

Удобно структурировать данные в виде таблиц - организовывать их хранение и обработку по строкам и столбцам. Примерами такой организации являются таблицы, подобные файлам Excel и электронным таблицам Google Sheets, а также базы данных, используемые в СУБД.

Организация данных в табличном виде упорядочивает их по вертикале (в базах данных это столбцы, называется полями) и по горизонтали (строки, называемые записями). Строки и столбцы связаны между собой (структурированы), что позволяет легко получать доступ к хранимым данным для обработки и анализа.



### Обучающиеся отвечают на вопрос педагога, приводят примеры.

**Педагог (слайд №3):** Структурированные данные – правильно организованные и отформатированные данные. Их часто называют количественными данными, их объективный характер позволяет легко подсчитывать, измерять и выражать данные в числах или символах.

Для структурированных данных обычно требуется меньше места в памяти по сравнению с неструктурированными. Для обработки структурированных данных в 1970-х годах компанией IBM был разработан специальный язык программирования SQL (язык структурированных запросов). Которым пользуются до сих пор.

**Педагог (слайд №4):** Как вы думаете, в каких форматах представляются структурированные данные?

Структурированные данные чаще всего записываются в виде текста и чисел. Наиболее распространены форматы для хранения и обработки данных это CSV и XML.

Структурированными данными пользуются все современные люди – совершая покупки в интернете или обычном магазине, пользуясь услугами банков, поликлиник и других государственных и коммерческих учреждений. Данные о пациенте больницы или товаре в магазине структурированы в виде таблиц. Так, при покупке товара мы не задумываемся о том, как данные хранятся и обрабатываются, нас интересует только конечный результат – быстро совершить покупку нужной вещи, что возможно с использованием баз данных и SQL.

**Работа современной банковской структуры** (работа банкоматов, ведение бухгалтерского и налогового учетов) невозможны без структурированных данных и баз данных.

**Системы бронирования и покупки билетов, отелей** также работают благодаря стандартной структуре организации данных по строками и столбцами.

Любая отрасль, где есть обработка данных и их хранение работают с таблицами структурированных данных.

### Обучающиеся отвечают на вопрос педагога.

**Педагог (слайд №5):** Не всякие данные можно структурировать. Большинство данных, которыми мы пользуемся - неструктурированные данные.

Вопрос: какие данные считаются неструктурированными? Приведите примеры. Фото и видео, звук, документы в формате PDF, тексты – являются неструктурированными данными, их нельзя упорядочить (структурировать) заранее определенным образом (по строкам и столбцам), поэтому эти данные хранятся в своих собственных форматах.



Эти данные называют качественными данными. К ним нельзя (или сложно) применить традиционные (для структурированных данных) методы и инструменты обработки и анализа. Поэтому для обработки неструктурированных данных разработали базы данных NoSQL.

Для хранения неструктурированных данных в исходных форматах используют репозитории хранения - озера данных. Озеро данных – это хранилище или система, предназначенная для хранения огромных объемов данных в естественном / необработанном формате.

**Обучающиеся отвечают на вопрос педагога, приводят примеры.**

**Педагог (слайд №6):** Как вы думаете в каких файлах могут храниться неструктурированные данные?

Как пример рассмотрим сообщения в группе мессенджера или соцсети – это могут тексты сообщений, эмодзи, фото, голосовые сообщения, видео – это все неструктурированные данные. Но сообщениями можно поделиться с другими пользователями (сделать репост) или поставить хэштеги (ключевые слова) – это уже структурированные данные, выраженные количественным показателем. Неструктурированные данные хранятся в своих собственных форматах («исходных» форматах). Это могут быть аудио (WAV, MP3, OGG), видеофайлы (MP4, WMV), PDF-документы, изображения (JPEG, PNG).

Искусственный интеллект (ИИ) облегчает работу с неструктурированными данными, автоматизируя сбор и анализ данных. Как пример такой работы — это распознавание звука, текста и изображений.

**Обучающиеся отвечают на вопрос педагога.**

**Педагог (слайд №7):** У структурированных данных есть как преимущества, так и проблемы по сравнению с неструктурированными. Среди преимуществ структурированных данных можно выделить такие как:

**Доступность использования.** Простота в обработке и доступе к данным для непрофессионала. Достаточно пользовательских навыков работы с таблицами и базами данных.

**Масштабируемость хранения.** Структурированные данные масштабируются алгоритмически, то есть по мере увеличения объема данных добавляется хранилище и вычислительная мощность. Современные системы, обрабатывающие структурированные данные, могут масштабироваться до нескольких тысяч ТБ данных (1 терабайт (ТБ) — это 1000 гигабайт (ГБ) или 1 000 000 мегабайт (МБ)).

**Автоматизация аналитики.** Использование математического и статистического аппарата, применение языка структурированных запросов (SQL) позволяет автоматизировать процесс анализа, изменения и обслуживания данных. Просто



организовать обработку и анализ больших массивов данных. Алгоритмы машинного обучения могут анализировать структурированные данные и выявлять общие закономерности для любой сферы экономики.

**Педагог (слайд №8):** Проблема при работе с неструктурированными данными возникает из-за жестко созданной структуры. Отсутствие гибкости в настройке структуры не позволяет оперативно менять схему хранения данных.

**Ограничение в модернизации.** Заключается в созданной структуре данных. Чтобы ее изменить и добавить новые поля и отношения необходимо изменить структуру базы данных или таблицы, что является ресурсоемким и затратным процессом.

**Педагог (слайд №9):** Как мы узнали, структурированные данные проще обрабатывать средствами программного обеспечения. Давайте потренируемся на практике. Предлагаю поработать со структурированными и неструктурированными данными средствами табличного процессора.

### **Практическая работа к уроку «Структурирование и анализ данных»**

Практическая работа 1. Структурирование данных.

Задание 1.1. Трансформирование неструктурированных данных в структурированные.

Задание 1.2. Трансформация данных CSV в Excel.

*Учитель раздает ученикам файлы с исходными данными.*

*Файлы с данными для практической работы получены с официального сайта*

*Федеральной службы государственной статистики <https://rosstat.gov.ru/folder/10705>*

**Педагог (слайд №10):** А сейчас мы поговорим об анализе данных. Как вы понимаете термин анализ? Как вы думаете, что такое анализ данных и зачем он нужен? Существует большое количество различных определений понятия анализа данных. Само слово «анализ» пришло к нам из Древней Греции и означает «распутывать», «освободить».

Согласно Большому экономическому словарю, анализ данных – это направление статистических исследований, включающее комплекс методов обработки многомерной системы данных наблюдений. Методы анализа данных используют только сведения, зафиксированные в этих данных.

В бизнес-среде популярно определение анализа данных Марио Фариа (Mario Faria), вице-президента Gartner (Gartner, Inc. — американская компания, специализирующаяся на исследованиях рынка информационных технологий): «анализ данных — это преобразование данных в выводы, на основе которых будут приниматься решения и строиться действия с помощью людей, процессов и технологий. То есть анализ данных позволяет найти закономерности и получить новые знания в исследуемой сфере».

### Обучающиеся отвечают на вопрос педагога и приводят примеры.

**Педагог (слайд №11):** Анализ данных это не просто обработка данных после их получения и сбора, это средство проверки гипотез.

Цель анализа данных заключается в применении существующих статистических и вычислительных моделей и методов для понимания трендов или закономерностей в собранных данных. Результаты заранее понятны и их легко описать в самом начале работы. Например, можно на основе данных прогнозировать объемы продаж или потребности в ремонте оборудования.

Для достижения этой цели ставятся следующие задачи анализа данных:

- ☐ сбор информации
- ☐ структуризация информации
- ☐ выявление закономерностей, анализ
- ☐ прогнозирование и получение рекомендаций.

**Педагог (слайд №12):** Чтобы из всего массива данных выделить необходимые для конкретного анализа данные, необходима их обработка. Небольшие массивы данных можно обрабатывать средствами табличных процессоров Excel. При большом объеме данных анализ должен происходить в профессиональных аналитических приложениях:

1. Аналитика Плюс.
2. Logiном — аналитическая low-code платформа, которая позволяет проводить анализ данных любого уровня сложности без программирования.
3. АСУ-аналитика.
4. IBM SPSS Statistics.

**Педагог (слайд №13):** Анализ данных подразумевает использование статистических инструментов, методов и технологий для получения осмысленных выводов на основе данных. Искусственный интеллект – это более высокий уровень, на котором данные применяются для решения когнитивных задач, традиционно относившихся к человеческому интеллекту: обучение, распознавание тенденций и самовыражение. Это набор сложных алгоритмов, которые «учатся» по ходу дела и постепенно решают задачи все лучше и лучше.

Примерами использования ИИ при автоматизации сбора и анализа данных можно привести распознавание звука, текста и изображений.

**Распознавание звука ИИ** (речь человека, звуки птиц и животных, звуки от механических устройств и аппаратов) позволяет идентифицировать людей (например, в телефонных и страховых компаниях, банках, службах спасения и соцзащиты), животных (в научных и природоохранительных мероприятиях), механические устройства и аппараты (поиск нарушений в их работе) и принимать оперативные управленческие, маркетинговые и бизнесберегающие решения.



**Распознавание изображений ИИ** используется в медицине для расшифровки снимков и постановке диагноза; в службах безопасности для идентификации сотрудников по биометрии (распознавание лица, радужка глаза и отпечатки пальцев) и поиске нарушителей.

**Анализ текста ИИ** при помощи чат-ботов, использующих обработку естественного языка (NLP) для анализа текстов в электронных письмах, в новостных лентах, в социальных сетях. Собранные данные анализируются для принятия различных решений. Например, для выявления фейковых новостей или получения данных об отзывах клиентов о продукции или услугах.

**Педагог (слайд №14):** Работа с данными — это удел профессионалов. Наиболее востребованными являются такие профессии как:

**Аналитик данных** (Data Analyst или дата-аналитик) — это специалист по анализу больших данных: он собирает их, обрабатывает и делает выводы. На основании его отчетов в компаниях принимают важные решения. Профессия аналитика данных находится на стыке IT, менеджмента и математики.

**Дата-сайентист** (data scientist) — это программист, который создаёт модели, предсказывающие результат. Для этого он сначала ищет в массивах данных связи и закономерности, на основе которых и строит модель. Разница между дата-сайентистом и дата-аналитиком в том, что аналитик не строит модели, а занимается анализом данных.

**Педагог (слайд №15): Архитектор данных** — это IT-специалист экспертного уровня, который анализирует информационную инфраструктуру компаний, ведет корпоративную бизнес-модель и жизненный цикл данных, использует различные цифровые инструменты для создания инновационных решений в области хранения и управления большими данными.

Архитектор баз данных (БД, Database – DB) — опытный технический специалист, который руководит целой группой сотрудников, обслуживающих систематизированный массив четко структурированной взаимосвязанной информации — базы данных, хранящейся в цифровом формате. Database Architect занимается выбором технологии для хранения данных, составляет план разработки и ТЗ для подчиненных, может выполнять проектирование и оптимизацию БД, следит за ее безопасностью.

**Разработчик баз данных (Database Developer)** создает, настраивает, оптимизирует, модернизирует и обслуживает базы данных (БД), которые входят в информационные системы.

Разработчик баз данных (Database Developer или Database Programmer) отвечает за создание, администрирование и устранение неполадок компьютерных баз данных,



которые могут обрабатывать большие объемы информации и обеспечивать ее безопасность. В его обязанности входит использование кода и веб-архитектуры для проектирования систем данных, анализа и поддержки существующих баз данных и внедрения новых пользовательских функций. Роль разработчика БД состоит в том, чтобы максимально облегчить пользователям баз данных доступ к необходимой им информации и сохранить информационные системы для будущих разработок компании.

Предлагаю на практике познакомиться с работой аналитика данных.

### **Практическая работа к уроку «Структурирование и анализ данных»**

Практическая работа 2. Анализ статистических данных для принятия управленческих решений.

Задание 2.1. Упорядочение и анализ данных при помощи фильтров и сортировки (расширенный режим).

Задание 2.2. Упорядочение и анализ данных при помощи сводных таблиц.

*Учитель раздает ученикам файлы с исходными данными.*

*Файлы с данными для практической работы получены с официального сайта Федеральной службы государственной статистики <https://rosstat.gov.ru/folder/10705>*

**Педагог (слайд №16):** Наше занятие подходит к концу. Надеюсь, вы смогли по-новому взглянуть на данные и методы их анализа. Сегодня мы познакомились с понятиями структурированных и неструктурированных данных, на практике поработали с задачами по анализу данных.

Давайте подведем итоги урока в форме синквейна. Выразите свои знания, эмоции, успехи от урока в виде синквейна. Синквейн — это пятистрочные строфы.

- 1-я строка – одно ключевое слово, определяющее содержание синквейна;
- 2-я строка – два прилагательных, характеризующих данное понятие;
- 3-я строка – три глагола, обозначающих действие в рамках заданной темы;
- 4-я строка – короткое предложение, раскрывающее суть темы или отношение к ней;
- 5-я строка – синоним ключевого слова (существительное).

Например:

- 1) Данные
- 2) Сложные, структурированные
- 3) Собираем, храним, анализируем
- 4) Данные нужны для принятия решений
- 5) Информация

Всем спасибо за занятие!



Вклад  
в будущее  
СБЕР



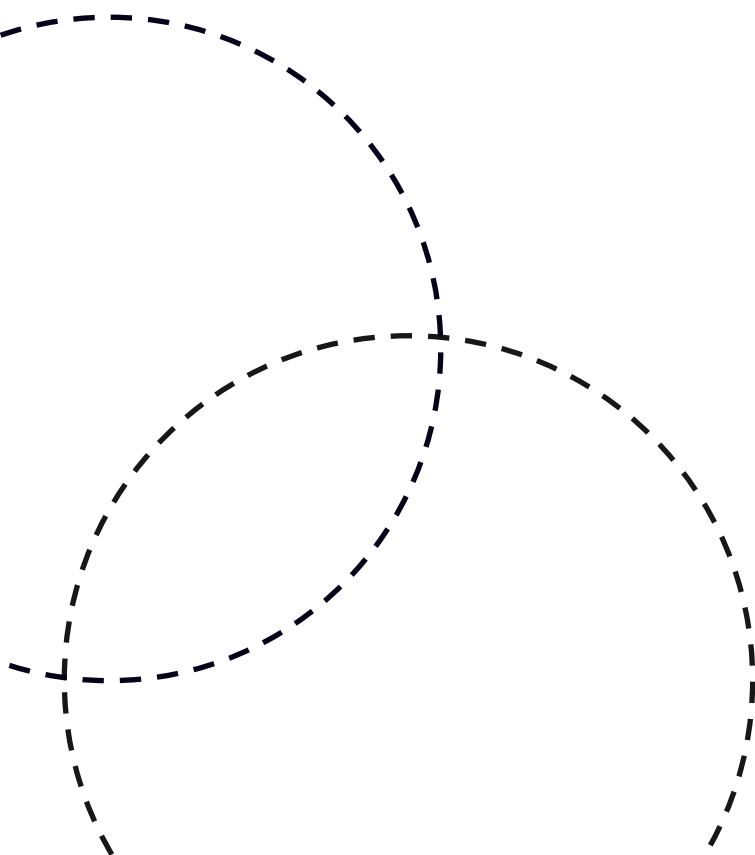
АКАДЕМИЯ  
искусственного интеллекта  
для школьников

**Урок подготовлен экспертами Академии искусственного интеллекта для школьников Благотворительного фонда Сбербанка «Вклад в будущее».**

Академия ИИ для школьников – Всероссийский образовательный проект, реализуемый Благотворительным фондом «Вклад в будущее» при поддержке Сбера с 2018 года. Цель проекта – формирование интереса и развитие прикладных навыков у школьников в сфере искусственного интеллекта. В рамках Академии ИИ проводятся просветительские активности, организуются соревнования, создаются образовательные курсы и вводные уроки, а также ведется работа с сообществом педагогов и амбассадоров Академии ИИ.

Все материалы опубликованы в открытом доступе на сайте Академии искусственного интеллекта для школьников и могут быть использованы для образовательных некоммерческих целей.

**Ссылка: <https://ai-academy.ru>**





Вклад  
в будущее  
СБЕР



АКАДЕМИЯ  
искусственного интеллекта  
для школьников

### Используемые источники:

1. ПрофГид. ИТ-профессии. Режим доступа  
<https://www.profguide.io/professions/category/it/>
2. Росстат. Федеральная служба государственной статистики.  
<https://rosstat.gov.ru/statistic>
3. Анализ данных: определение, цели и задачи. <https://analytikaplus.ru/chto-takoe-analiz-dannyh-opredelenie/>
4. Структурированные и неструктурированные данные: сравнение и объяснение.  
<https://asu-analitika.ru/strukturirovannye-i-nestrukturirovannye-dannye-sravnenie-i-objasnenie>

