

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
**Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ
ПРАКТИЧЕСКИХ РАБОТ ПО ДИСЦИПЛИНЕ**

Обработка и анализ больших данных

Направление подготовки 39.03.01 Социология
Направленность (профиль) – «Аналитика и консалтинг»
Форма обучения – Очная, заочная
Год начала обучения – 2026 г.

Ставрополь, 2026

СОДЕРЖАНИЕ

Введение.....	3
Практическое занятие 1. Введение: определение, история вопроса, принципы работы с большими данными	
Практическое занятие 2. Прикладные инструменты для работы с Big Data	
Практическое занятие 3. Жизненный цикл анализа больших данных	
Практическое занятие 4. Технологии хранения больших данных	
Практическое занятие 5. Когнитивный анализ данных	
Практическое занятие 6. Принципы и подходы распределенной обработки больших данных	
Практическое занятие 7. Задачи Data Mining и способы их решения	
Практическое занятие 8. Методы анализа графовых данных	
Практическое занятие 9. Прикладные инструменты анализа больших данных	
Практическое занятие 10. Цифровые инструменты для анализа больших данных	
Практическое занятие 11. Готовые комплексные решения анализа больших данных	
Практическое занятие 12. Знакомство с программами, применяемыми при анализе больших данных	
Практическое занятие 13. Цифровые инструменты визуализации больших данных	
Практическое занятие 14. Распределенные базы данных	

ВВЕДЕНИЕ

Методические указания по выполнению практических работ разработаны на основе программы дисциплины «Обработка и анализ больших данных» и предназначены для студентов, обучающихся на бакалавриате по направлению «Социология».

Целью данных методических указаний является подготовка студентов к работе на практических занятиях, формирование набора профессиональных компетенций будущего бакалавра по направлению подготовки «Социология».

Цель освоения дисциплины направлена на формирование набора компетенций будущего бакалавра по направлению подготовки 39.03.01 Социология

Задачи освоения дисциплины:

- овладение методами, технологиями и инструментами анализа больших данных;
- овладение необходимым понятийным аппаратом.

Место дисциплины в структуре ОП бакалавриата. Дисциплина «Обработка и анализ больших данных» относится к дисциплинам части, формируемой участниками образовательных отношений направленности (профиля) Аналитика и консалтинг». Ее освоение происходит в 3 семестре.

На практических занятиях студенты углубляют и закрепляют теоретические знания, вырабатывают аналитические умения, приобретают навыки самостоятельной познавательной деятельности.

В результате изучения дисциплины магистрант должен приобрести следующие навыки:

- на основе использования технологий обработки и анализа больших данных, разрабатывать предложения по совершенствованию технологических процессов, методов сбора и анализа информации в фундаментальном или прикладном социологическом исследовании;
- анализируя и интерпретируя социологическую информацию, полученную в ходе социологических исследований, применять новые методы, технологии обработки анализа больших данных
- проводить отбор экспертов для проведения экспертизы результатов социологических исследований с использованием инструментария анализа больших данных
- владеть навыками анализа экспертных заключений по итогам фундаментального или прикладного социологического исследования, используя технологии анализа больших данных
- применяя технологии анализа больших данных, разрабатывать критерии и показатели, анализирует программы, стратегии управленческие решения в различных сферах жизнедеятельности общества.

СОДЕРЖАНИЕ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

Тема 1. Введение: определение, история вопроса, принципы работы с большими данными

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: Способен понимать принципы работы современных технологий (ОПК-1); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (УК-1).

Актуальность темы: определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть.

Большие данные (Big Data) — совокупность непрерывно увеличивающихся объемов информации одного контекста, но разных форматов представления, а также методов и средств для эффективной и быстрой обработки.

Принцип работы технологии big data основан на максимальном информировании пользователя о каком-либо предмете или явлении. Задача такого ознакомления с данными — помочь взвесить все «за» и «против», чтобы принять верное решение. В интеллектуальных машинах на основе массива информации строится модель будущего, а дальше имитируются различные варианты и отслеживаются результаты.

Источники больших данных — интернет вещей, соцсети, блоги, СМИ, показания приборов и датчиков (например, данные метеостанций), статистика, архивы, базы данных. Хранятся они в data lake — «озёрах данных» — в облачных хранилищах.

Анализ больших данных уже применяется во многих сферах — для обеспечения безопасности, диагностики и профилактики заболеваний, предсказания аварий и катастроф. В бизнесе и маркетинге на основании больших данных предсказывают поведение клиентов, оптимизируют расходы и процесс производства, оценивают платежеспособность, прогнозируют увольнение сотрудников

Вопросы и задания:

1. Эволюция термина «большие данные»
2. Большие данные в социальных науках
3. Источники данных, для которых необходимы методы работы с большими данными.
4. Принципы работы с большими данными.

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 2. Прикладные инструменты для работы с Big Data

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (УК-1); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (ОПК-1).

Актуальность темы (а): определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть.

Big Data — это серия подходов, инструментов и методов, используемых для обработки структурированных и неструктурированных данных огромных объемов и значительного разнообразия для получения результатов, воспринимаемых людьми, которые доказывают свою эффективность в условиях непрерывного роста. Большие данные служат альтернативой традиционным системам управления базами данных и решениям в рамках Business Intelligence.

Таким образом, большие данные не относятся к определенному количеству данных или даже к самим данным. Вместо этого термин означает методы обработки данных, которые позволяют распределенную обработку информации. Эти методы могут применяться как к огромным наборам данных (например, содержание всех страниц в Интернете), так и к небольшим (например, к содержанию этой статьи).

Большие данные имеют важное значение для глобального бизнеса, поскольку большее количество данных приводит к более точному анализу, который, в свою очередь, обеспечивает более эффективное принятие решений, повышение операционной эффективности и сокращению затрат.

Apache HTTP-сервер — свободный веб-сервер. Apache является кроссплатформенным ПО, поддерживает операционные системы Linux, BSD, Mac OS, Microsoft Windows, Novell NetWare, BeOS. Основными достоинствами Apache считаются надёжность и гибкость конфигурации

Вопросы и задания:

1

Языки программирования и инструменты для анализа больших данных

1. Языки программирования для анализа больших данных

2. Платные и бесплатные инструменты для обработки и анализа больших данных

2

Инструменты для работы с большими данными: сравнительный анализ

1. Инструменты для работы с большими данными: сравнительный анализ: Apache Hadoop. Apache Storm. RapidMiner. Tableau. Spark. Cassandra

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 3. Жизненный цикл анализа больших данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (ПК-2); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (ОПК-1).

Актуальность темы (а): определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть.

В сегодняшнем контексте больших данных предыдущие подходы являются либо неполными, либо неоптимальными. Например, методология SEMMA полностью игнорирует сбор данных и предварительную обработку различных источников данных. Эти этапы обычно составляют большую часть работы в успешном проекте больших данных.

Большой цикл анализа данных может быть описан следующими этапами:

Определение бизнес-проблемы

Исследование

Оценка человеческих ресурсов

Получение данных

Data Munging

Хранилище данных

Исследовательский анализ данных

Подготовка данных для моделирования и оценки моделирование

Реализация

Вопросы и задания:

1. Что является главным результатом процесса Business Intelligence
2. Что относится к средствам предоставления информации в Business Intelligence

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 4. Технологии хранения больших данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (ОПК-1); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (УК-1).

Актуальность темы (а): определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть.

Служба хранилища Azure — управляемая облачная служба с высоким уровнем доступности, безопасности, надежности, масштабируемости и избыточности. Корпорация Майкрософт отвечает за обслуживание и решает критические проблемы, не требуя вашего участия. Служба хранилища Azure является самым универсальным решением для хранения данных в Azure, так как с ней можно сочетать множество служб и средств.

Существует несколько служб хранилища Azure, которые можно использовать для хранения данных. Самый гибкий вариант для хранения больших двоичных объектов из нескольких источников данных — это хранилище BLOB-объектов. Большие двоичные объекты по сути являются файлами. Это могут быть изображения, документы, HTML-файлы, виртуальные жесткие диски (VHD), большие данные (например, журналы или резервные копии баз данных), то есть практически любые файлы. BLOB-объекты хранятся в контейнерах, которые по принципу использования похожи на папки. Контейнер обеспечивает группирование набора больших двоичных объектов. Учетная запись хранения может содержать неограниченное количество контейнеров. В каждом контейнере может храниться неограниченное количество больших двоичных объектов.

Хранилище Azure хорошо подходит для больших данных и аналитики благодаря гибкости, высокому уровню доступности и низкой цене. Оно предоставляет уровни "горячего", "холодного" и архивного доступа для вариантов использования. Дополнительные сведения см. в разделе Хранилище BLOB-объектов Azure: "горячий", "холодный" и архивный уровни.

Хранилище BLOB-объектов Azure доступно из Hadoop (через HDInsight). HDInsight может использовать контейнер больших двоичных объектов в службе хранилища Azure в качестве файловой системы по умолчанию для кластера. С помощью интерфейса распределенной файловой системы Hadoop, который предоставляется драйвером WASB, все компоненты в HDInsight могут напрямую взаимодействовать со структурированными или неструктурированными данными, хранимыми в виде больших двоичных объектов. К хранилищу BLOB-объектов Azure можно также обращаться через Azure Synapse Analytics с помощью PolyBase.

Вопросы и задания для обсуждения:

1. Гетерогенные хранилища данных
2. Аналитические базы данных

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 5. Когнитивный анализ данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (ОПК-1); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (УК-1).

Теоретическая часть

Характерной чертой современного моделирования является симбиоз наук -психология, социология, экономика, математика, информатика — вот далеко не полный перечень. Исследование операций, общая теория систем, кибернетика и прогностика - четыре науковедческих дисциплины, давшие парадигму методологии современного моделирования, а на его основе - прогнозирования и планирования.

Анализ различных математических моделей применительно к развитию и функционированию социально-экономических систем показывает, что для

этих целей достаточно удобно использовать аппарат знаковых, взвешенных знаковых и функциональных знаковых графов. Аппарат позволяет работать с данными как качественного, так и количественного типа, причем степень использования количественных данных может увеличиваться в зависимости от возможностей количественной оценки взаимодействующих факторов в итерационном цикле моделирования. Аппарат знаковых графов позволяет формально строить прогнозы развития или траектории движения моделируемой системы в фазовом пространстве ее переменных (факторов) на основе информации о ее структуре и программах развития системы путем аппроксимации их кусками траекторий импульсных процессов на знаковых орграфах.

В основе технологии когнитивного анализа и моделирования лежит когнитивная (познавательная-целевая) структуризация знаний об объекте и внешней для него среде, причем объект и внешняя среда разграничиваются нечётко.

Цель такой структуризации состоит в выявлении наиболее существенных (базисных) факторов, характеризующих пограничный слой взаимодействия объекта и внешней среды, и установление качественных (причинно-следственных) связей между ними, т.е. какие взаимовлияния оказывают факторы друг на друга в ходе их изменения. Взаимовлияния факторов отображаются с помощью когнитивной карты (модели), которая представляет собой знаковый (взвешенный) ориентированный граф.

Вопросы и задания для обсуждения:

1. Структура когнитивного управления
2. Математический инструментальный анализа данных

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 6. Принципы и подходы распределенной обработки больших данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (УК-1); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (ОПК-1).

Актуальность темы (а) определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть

Исходя из определения Big Data, можно сформулировать основные принципы работы с такими данными:

1. Горизонтальная масштабируемость. Поскольку данных может быть сколь угодно много – любая система, которая подразумевает обработку больших данных, должна быть расширяемой. В 2 раза вырос объём данных – в 2 раза увеличили количество железа в кластере и всё продолжило работать.

2. Отказоустойчивость. Принцип горизонтальной масштабируемости подразумевает, что машин в кластере может быть много. Например, Hadoop-кластер Yahoo имеет более 42000 машин. Это означает, что часть этих машин будет гарантированно выходить из строя. Методы работы с большими данными должны учитывать возможность таких сбоев и переживать их без каких-либо значимых последствий.

3. Локальность данных. В больших распределённых системах данные распределены по большому количеству машин. Если данные физически находятся на одном сервере, а обрабатываются на другом – расходы на передачу данных могут превысить расходы на саму обработку. Поэтому одним из важнейших принципов проектирования BigData-решений является принцип локальности данных – по возможности обрабатываем данные на той же машине, на которой их храним.

Все современные средства работы с большими данными так или иначе следуют этим трём принципам. Для того, чтобы им следовать – необходимо придумывать какие-то методы, способы и парадигмы разработки средств разработки данных.

Вопросы и задания для обсуждения:

1. Понятие распределенной обработки данных
2. Технологии распределенной обработки больших данных
3. Инструменты распределенной обработки больших данных

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 7. Задачи Data Mining и способы их решения

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке

предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (ПК-2); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (ОПК-1).

Актуальность темы (а) определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть

Data Mining (добыча данных, интеллектуальный анализ данных, глубинный анализ данных или просто майнинг данных) — это процесс, используемый компаниями для превращения необработанных больших данных в полезную информацию. Также для этой технологии используется менее популярный термин «обнаружение знаний в данных» или KDD (knowledge discovery in databases).

Если термином Big Data обозначают все большие данные — как обработанные, так и нет, то Data Mining представляет собой процесс глубокого погружения в эти данные для извлечения ключевых знаний.

используя программное обеспечение для поиска закономерностей в больших пакетах данных, предприятия могут выстраивать маркетинговые стратегии, управлять кредитными рисками, обнаруживать мошенничество, фильтровать спам или даже выявлять настройки пользователей.

Интеллектуальный анализ данных зависит от эффективного сбора, хранения и компьютерной обработки данных. Data Mining считается отдельной дисциплиной в области науки о данных.

Термин «интеллектуальный анализ данных» фигурировал в академических журналах еще в 1970 году, но по-настоящему популярным он стал только в 1990-х после появления интернета. Тогда компаниям потребовалось анализировать большие объемы разнородных данных, чтобы отыскать нетривиальные паттерны и научиться предсказывать поведение клиентов. Обычные модели статистики оказались неспособны справиться с этой задачей.

Первые системы Data Mining предназначались для обработки данных о продажах в супермаркетах по нескольким параметрам, включая их объем по регионам и тип продукта.

Вопросы и задания для обсуждения:

1. Задачи классификации Data Mining
2. Задачи кластеризации Data Mining
3. Методы прогнозирования в Data Mining

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 8. Методы анализа графовых данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (ПК-2); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (УК-1).

Актуальность темы (а) определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть

Граф - математический объект, который изображает отношения между сущностями. Граф состоит из вершин (объектов) и ребер (связей). С помощью графов можно представить разных ситуации: например, пользователей соцсети, которые находятся друг у друга в друзьях, клиентов банка, которые переводят друг другу денежные средства, географические объекты и пути между ними.

Графовый (также сетевой) анализ - это набор методов, направленных на изучение связей между сущностями. При помощи этих методов исследуется структура графа и выявляются неочевидные зависимости.

Извлекать из графов полезную информацию позволяют графовые алгоритмы, которые можно условно поделить на несколько семейств. Рассмотрим эти семейства на примере социальной сети:

Обнаружение сообществ (community detection) — это выделение тесно связанных между собой групп людей (к примеру, через большое количество общих связей). При этом совсем не обязательно, чтобы все участники взаимодействовали друг с другом напрямую. Это семейство алгоритмов является своего рода аналогом кластеризации.

Алгоритмы центральности (centrality algorithms) поможет выявить лидеров мнений и влиятельных людей в сообществах. Под центральностью мы подразумеваем некоторую меру значимости вершины или ребра. Алгоритмы центральности и сообществ можно применять для создания новых предикторов в ML-pipeline.

Предсказание связей (link prediction) оценивает вероятность наличия связи между двумя отдельными людьми в том случае, если её не существует на графе. Связи, подобранные таким образом, могут помочь в рекомендации друзей.

Алгоритмы сходства (similarity algorithms) пригодятся, чтобы найти похожие группы людей. Это может быть полезно, чтобы собрать аудиторию для рекламы по принципу lookalike или выявить поддельные учетные записи,

основываясь на свойствах их окружения.

Поиск путей (path detection) окажется полезен для того, чтобы найти кратчайшую цепочку знакомств между людьми. В качестве меры расстояния (весов ребер) можно использовать характеристики взаимодействия пользователей - например, частоту их общения.

Вопросы и задания для обсуждения:

1. Понятие и виды графов
2. Инструменты обработки графовых данных

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 9. Прикладные инструменты анализа больших данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (ПК-2); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (ОПК-1).

Актуальность темы (а) определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть

Hadoop — это программная структура, способная к распределенной обработке больших объемов данных. Но Hadoop обрабатывается надежным, эффективным и масштабируемым образом. Hadoop надежен, так как предполагает, что вычислительные элементы и хранилище выйдут из строя, поэтому он поддерживает несколько копий рабочих данных, чтобы обеспечить возможность их перераспределения для неисправных узлов. Hadoop эффективен, поскольку работает параллельно и ускоряет обработку благодаря параллельной обработке. Hadoop также масштабируется и может обрабатывать петабайты данных. Кроме того, Hadoop полагается на серверы сообщества, поэтому его стоимость относительно низкая, и любой может его использовать.

В Ktime процесс программирования логики осуществляется через создание Workflow. Workflow состоит из узлов, которые выполняют ту или иную функцию (например чтение данных из БД, трансформация,

визуализация). Узлы, соответственно, соединяются между собой стрелочками, которые показывают направление движения данных.

Аналитическая система Orange – это программа с открытым исходным кодом для машинного обучения и визуализации данных, обладающая большим набором исследовательских функций.

Программный продукт Orange (рус. Оранж), разрабатываемая Лабораторией биоинформатики Люблянского университета, предназначена для интеллектуального анализа данных (ИАД), статистических исследований и визуализации данных. Компоненты аналитической платформы называются виджетами, и они варьируются от минималистичной визуализации данных, выбора подмножеств и предварительной обработки до эмпирической оценки алгоритмов обучения и прогностического моделирования. Система станет эффективным инструментом в руках аналитика данных, исследователя и учёного.

В программном обеспечении Orange Data Mining применяется визуальное программирование, которое реализуется удобным графическим интерфейсом. В рамках визуального программирования аналитические процедуры создаются путём связывания predetermined или разработанных пользователем блоков (виджетов), в то время как продвинутые пользователи могут использовать Orange в качестве программной библиотеки Python для манипулирования данными и создания новых блоков (виджетов).

Вопросы и задания для обсуждения:

1. Архитектура Hadoop
2. Основные компоненты Hadoop
3. Архитектура KNIME
4. Основные компоненты KNIME
5. Архитектура Orange
6. Основные компоненты Orange

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 10. Цифровые инструменты для анализа больших данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (УК-1); способен осуществлять социологическую экспертизу стратегий,

мероприятий, качества исследований (ОПК-1).

Актуальность темы (а) определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть

Платформа для обработки больших данных — это решение, которое объединяет различные инструменты, необходимые специалистам по data science. Такие платформы существенно упрощают их работу, охватывая весь жизненный цикл data science проектов: от идеи и исследования данных до построения и развертывания аналитических моделей.

Они позволяют решить так называемую проблему «последней мили»: интегрировать результаты анализа данных в операционную деятельность, чтобы они влияли на принятие решений и трансформировали бизнес-процессы. Это может быть реализовано в виде API предиктивной модели, к которой обращаются другие системы, веб-приложения, которым могут пользоваться сотрудники, или просто ежедневного отчета, отправляемого на почту.

Каждый data science-проект проходит жизненный цикл, состоящий из трех этапов:

Сбор данных и исследование.

Экспериментирование и разработка модели.

Развертывание и интеграция.

На каждом этапе специфические задачи, которые помогает выполнять платформа. И есть более общие задачи, включающие управление данными, управление процессами обработки и масштабирования.

Для решения всех этих задач платформы обработки данных предлагают такой технический функционал: прием, подготовка и исследование данных, генерация признаков, создание, обучение, тестирование и деплой моделей, мониторинг и обслуживание системы.

Также платформа должна обеспечивать безопасность данных и их хранение, каталогизацию источников, предоставлять инструменты для визуализации и формирования отчетов. Облачные платформы дополнительно дают большой объем хранилища и вычислительных мощностей.

Вопросы и задания для обсуждения:

1. Понятие цифровых инструментов обработки данных
2. Характеристика свободно распространяемого цифрового ПО для обработки и анализа больших данных

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 11. Готовые комплексные решения анализа больших данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (ОПК-1); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (УК-1).

Актуальность темы (а) определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть

Современные компании используют большие данные для принятия решений в режиме реального времени с целью получения конкурентных преимуществ. Периферийно-облачная платформа HPE GreenLake помогает ускорить анализ больших данных за счет снижения сложности и стоимости развертывания сред Nadoor и управления ими:

Предварительное интегрированное аппаратное и программное обеспечение установлено в вашей среде или на объектах совместного размещения.

Услуги по измерению, мониторингу и управлению ресурсами с опережением спроса. Переменные платежи на основании фактического использования. HPE GreenLake Central с функциями самообслуживания для полного контроля над затратами, ресурсами и многим другим.

Вопросы и задания для обсуждения:

1. Понятие комплексных решений в области анализа больших данных
2. Классификация цифровых комплексных решения для анализа больших данных.

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 12. Знакомство с программами, применяемыми при анализе больших данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения

темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (ОПК-1); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (УК-1).

Актуальность темы (а): определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть

Weka — библиотека алгоритмов машинного обучения для решения задач интеллектуального анализа данных (data mining). Система позволяет непосредственно применять алгоритмы к выборкам данных, а также вызывать алгоритмы из программ на языке Java.

Weka расшифровывается как Waikato Environment for Knowledge Analysis — среда для анализа знаний, разработанная в университете Вайкато (Новая Зеландия).

Целью проекта являлось создание современной среды разработки средств машинного обучения и применения их к реальным данным, создание доступных для повсеместного применения средств машинного обучения. Предполагается, что с помощью данной среды специалист в прикладной области сможет использовать методы машинного обучения для извлечения полезных знаний непосредственно из данных, возможно, очень большого объёма. Пользователями WEKA являются исследователи в области машинного обучения и прикладных наук. Она также широко используется в учебных целях.

RapidMiner — это мощная и многопользовательская платформа, она служит для создания, передачи и обслуживания наукоемких данных. Платформа RapidMiner предлагает больше функций, чем любое другое визуальное решение, плюс она открыта и расширяема для поддержки всех потребностей научных данных.

Унифицированная платформа RapidMiner ускоряет создание полных аналитических рабочих процессов — от подготовки данных до моделирования до развертывания бизнеса — в единой среде, значительно повышая эффективность и сокращая время, необходимое для проектов в области данных.

Аналитическая система Orange — это программа с открытым исходным кодом для машинного обучения и визуализации данных, обладающая большим набором исследовательских функций.

Программный продукт Orange (рус. Оранж), разрабатываемая Лабораторией биоинформатики Люблянского университета, предназначена для интеллектуального анализа данных (ИАД), статистических исследований и визуализации данных. Компоненты аналитической платформы называются виджетами, и они варьируются от минималистичной визуализации данных, выбора подмножеств и предварительной обработки до эмпирической оценки алгоритмов обучения и прогностического моделирования. Система станет

эффективным инструментом в руках аналитика данных, исследователя и учёного.

Аналитическая платформа KNIME предоставляет пользователям возможности визуально создавать потоки данных (конвейеры), выборочно выполнять отдельные или все шаги анализа, а затем проверять результаты, модели, используя интерактивные виджеты и представления.

Аналитическая система KNIME Analytics Platform может быть использована как в виде отдельно-стоящего приложения на базе дистрибутива с открытым исходным кодом, так и в форме аренды облачной версии системы.

Вопросы и задания:

1. Функционал программы WEKA
2. Возможности программы WEKA
3. Функционал программы RapidMiner
4. Возможности программы RapidMiner
5. Преимущества и недостатки Orange
6. Виджеты визуализации данных Orange
7. Функционал программы KNIME Analytics Platform
8. Возможности программы KNIME Analytics Platform

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 13. Цифровые инструменты визуализации больших данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (ПК-2); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (УК-1).

Актуальность темы (а) определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть

Визуализация больших данных относится к областям как научной, так и информационной визуализации. В первом случае «большие данные» возникают в результате сложного компьютерного моделирования различных объектов и процессов. Во втором - имеет место визуальное описание и

представление абстрактной информации, получаемой в результате процесса сбора и обработки многокатегориальных данных, для анализа которых необходимо применение нескольких количественных и качественных мер оценки. Работы, описывающие результаты по визуализации больших данных, появились сравнительно недавно. Среди них можно выделить «Белую книгу» компании Intel, посвященную визуализации результатов «большого счета», установочную публикацию известного специалиста в области компьютерной визуализации и человеко-компьютерного взаимодействия Б. Шнейдермана и введение к спецвыпуску, посвященному визуализации больших данных. Отметим, что один из соавторов работы, профессор Калифорнийского Университета в Дэвисе К.-Л. Ма (Kwan-Liu Ma), в настоящее время является ведущим специалистом по визуализации больших данных. Он является соавтором большого числа научных работ, а также организатором конференций, ов и лекций по данному направлению.

Вопросы и задания для обсуждения:

1. Корпоративные инструменты визуализации больших данных
2. Бесплатные инструменты для визуализации больших данных и их характеристика.

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Тема 14. Распределенные базы данных

Цель: формирование универсальной компетенции социологов (будущих социологов) в сфере применения цифровых инструментов сбора, обработки и анализа больших данных.

Знания и умения, приобретаемые студентом в результате освоения темы (а), формируемые компетенции или их части: способен к разработке предложений по совершенствованию и разработке методов сбора и анализа данных фундаментальных и прикладных социологических исследований (УК-1); способен осуществлять социологическую экспертизу стратегий, мероприятий, качества исследований (ПК-3).

Актуальность темы (а) Актуальность темы (а) определяется необходимостью формирования аналитических компетенций и умений анализировать большие данные у социологов (будущих социологов).

Теоретическая часть

Базы данных NoSQL специально созданы для определенных моделей данных и обладают гибкими схемами, что позволяет разрабатывать современные приложения. Базы данных NoSQL получили широкое распространение в связи с простотой разработки, функциональностью и производительностью при любых масштабах. Ресурсы, представленные на

этой странице, помогут разобраться с базами данных NoSQL и начать работу с ними.

Базы данных NoSQL используют разнообразные модели данных для доступа к данным и управления ими. Базы данных таких типов оптимизированы для приложений, которые работают с большим объемом данных, нуждаются в низкой задержке и гибких моделях данных. Все это достигается путем смягчения жестких требований к непротиворечивости данных, характерных для других типов БД.

Вопросы и задания для обсуждения:

1. Функции нереляционных баз данных
2. Влияние технологий NoSQL на распределенные облачные системы

Список литературы, рекомендуемый к использованию по данной теме:

Основная литература: 1-4

Дополнительная литература: 1-3

Интернет-ресурсы: 1-10

Список литературы, рекомендуемый к использованию

Перечень основной литературы:

1. Воронова, Л. И. Machine Learning: регрессионные методы интеллектуального анализа данных Электронный ресурс: Учебное пособие / Л. И. Воронова, В. И. Воронов. - Machine Learning: регрессионные методы интеллектуального анализа данных, 2024-02-26. - Москва: Московский технический университет связи и информатики, 2018. - 82 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 2227-8397, экземпляров неограничено

2. Храмов, А. Г. Методы и алгоритмы интеллектуального анализа данных Электронный ресурс / Храмов А. Г.: учебное пособие. - Самара: СамГУ, 2019. - 176 с. - Рекомендовано редакционно-издательским советом федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» в качестве учебного пособия для обучающихся по основной образовательной программе высшего образования по направлению подготовки 01.04.02 Прикладная математика и информатика. - ISBN 978-5-7883-1414-3, экземпляров неограниченно

3. Агалаков, С. А. Статистические методы анализа данных: учебное пособие / С.А. Агалаков; Министерство образования и науки РФ; Омский государственный университет им. Ф. М. Достоевского. - Омск: ОмГУ им. Ф.М. Достоевского, 2017. - 92 с.: табл., граф., схем., ил. - <http://biblioclub.ru/>. - Библиогр. в кн. - ISBN 978-5-7779-2187-1, экземпляров неограниченно

4. Тарасов, И. Е. Статистический анализ данных в информационных системах Электронный ресурс / Тарасов И. Е.: учебно-методическое пособие. - Москва: РТУ МИРЭА, 2020. - 96 с., экземпляров неограниченно

Перечень дополнительной литературы:

1. Глебов, В. И. Практикум по математической статистике. Проверка гипотез с использованием Excel, MatCalc, R и Python: учебное пособие / В. И. Глебов, С. Я. Криволапов. - Практикум по математической статистике. Проверка гипотез с использованием Excel, MatCalc, R и Python, 2026-04-02. - Электрон. дан. (1 файл). - Москва: Прометей, 2019. - 86 с. - электронный. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-907100-66-4, экземпляров неограниченно

2. Лихтенштейн, В. Е. Информационные технологии в бизнесе. Том 2. Применение системы Decision в решении прикладных экономических задач Электронный ресурс: Учебное пособие / В. Е. Лихтенштейн, Г. В. Росс. - Информационные технологии в бизнесе. Том 2. Применение системы Decision в решении прикладных экономических задач, 2028-01-26. - Саратов: Ай Пи Эр Медиа, 2018. - 420 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-4486-0283-2, экземпляров неограниченно

3. Мартышенко, С. Н. Автоматизация анализа данных в исследовании социально-экономических процессов. Монография Электронный ресурс / Мартышенко С. Н. - Владивосток: ВГУЭС, 2019. - 164 с. - ISBN 978-5-9736-0583-4, экземпляров неограниченно

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Сайт Института социологии РАН – <http://www.isras.ru/socis.html>.
2. Журнал «Социологические исследования» // [Электрон. ресурс]. – Режим доступа: <http://socis.isras.ru>
3. Гуманитарный Интернет портал // [Электрон. ресурс]. – Режим доступа: www.auditorium.ru
4. Санкт-Петербургский государственный университет. Факультет социологии [Электрон. ресурс]. – Режим доступа: <http://www.soc.spb.ru/archive/>
5. «Фонд общественное мнение» (ФОМ) // [Электрон. ресурс]. – Режим доступа: <http://www.fom.ru/>
6. ВЦИОМ [Электрон. ресурс]. – Режим доступа: <https://wciom.ru/>
7. Левада-центр [Электрон. ресурс]. – Режим доступа: <http://www.levada.ru/>
8. Электронно-библиотечная система «Университетская библиотека онлайн» - www.biblioclub.ru;
9. «Фолиант» - <http://catalog.ncstu.ru>;
10. Электронная библиотечная система «IPRbooks»; База данных SCOPUS и др.

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
**Федеральное государственное автономное образовательное учреждение
высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО
ОРГАНИЗАЦИИ И ПРОВЕДЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ
ПО ДИСЦИПЛИНЕ**

Обработка и анализ больших данных

Направление подготовки 39.03.01 Социология
Направленность (профиль) – «Аналитика и консалтинг»
Форма обучения – Очная, заочная
Год начала обучения – 2026 г.

Ставрополь, 2026

СОДЕРЖАНИЕ

Введение.....	41
1. Общая характеристика самостоятельной работы студента.....	43
2. Методические рекомендации по изучению теоретического материала.....	43
3. Методические рекомендации по работе с литературой.....	46
4. Методические рекомендации подготовке к дискуссии, круглому столу.....	47
5. Методические рекомендации по написанию и оформлению доклада.....	48
6. Методические рекомендации по составлению глоссария.....	52
Список литературы.....	53

ВВЕДЕНИЕ

Методические указания для самостоятельной работы разработаны на основе программы дисциплины «Обработка и анализ больших данных» и предназначены для студентов, обучающихся на бакалавриате по направлению «Социология».

Цель освоения дисциплины направлена на формирование набора компетенций будущего бакалавра по направлению подготовки 39.03.01 Социология.

Задачи освоения дисциплины:

- овладение методами, технологиями и инструментами анализа больших данных;
- овладение необходимым понятийным аппаратом.

Место дисциплины в структуре ОП магистратуры. Дисциплина «Обработка и анализ больших данных» относится к дисциплинам части, формируемой участниками образовательных отношений направленности (профиля) «Аналитика и консалтинг». Ее освоение происходит в 3 семестре.

В результате изучения дисциплины будущий бакалавр должен приобрести следующие навыки:

— разработки предложений по совершенствованию технологических процессов, методов сбора и анализа информации в фундаментальном или прикладном социологическом исследовании в сфере государственного и муниципального управления.

— применять новые методы, процессы, технологии обработки социологической информации при проведении комплексных исследований государственно-управленческой деятельности и проблем муниципального управления.

— проводить отбор экспертов для проведения экспертизы результатов внедрения управленческих стратегий в сфере государственного и муниципального управления с использованием инструментария социологических исследований.

— владеть навыками анализа экспертных заключений по итогам фундаментального или прикладного социологического исследования, разрабатывает структуру экспертно-аналитического сопровождения деятельности органов государственной и муниципальной власти

применять знания в области осуществления социологической экспертизы, разрабатывает системы критериев, показателей, норм в соответствии экспертизы в сфере государственного и муниципального управления, стратегии, управленческих решений.

1. Общая характеристика самостоятельной работы студента

Задачами самостоятельной работы являются:

- развитие и привитие навыков самостоятельной учебной работы и формирование потребностей в самообразовании и социализации;
- освоение, углубление содержания и основных положений курса, выносимых на самостоятельное изучение студентом; в ходе составления опорного конспекта;
- использование материала, собранного и полученного в ходе самостоятельных занятий, при подготовке докладов.

Основными формами реализации самостоятельной работы являются:

- изучение отдельной темы;
- составление словаря научных понятий;
- выполнение письменных заданий;
- подготовка реферативного обзора литературы по темам курса;
- подготовка тематического доклада по избранной проблеме.

Текущий контроль осуществляется в процессе анализа самостоятельной работы. Образовательные продукты (доклад, словарь научных понятий, выполнение письменные задания) составляют портфолио, которое представляется в оформленном виде на последнем занятии.

2. Методические рекомендации по изучению теоретического материала

Лекционные занятия, проводимые в интерактивной форме, предполагают совместное обсуждение методологических и теоретических положений учебного курса, отработку умения работать с научной литературой, анализируя и обобщая социологические факты, личные примеры и предложенные ситуации. Цель таких занятий – усвоение системы теоретических знаний по основным разделам науки. Для этого по каждому занятию указана литература, непосредственно относящаяся к содержанию темы. С помощью нее студент заранее готовится к занятию по предложенному плану с тем, чтобы в ходе лекции участвовать в коллективном анализе различных подходов к категориям и проблемам социологии, в формулировании дефиниций и в научном обосновании собственных позиций по проблемам. Подготовка к лекционным занятиям предусматривает ознакомление с основными положениями по теме, усвоение нормативной лексики, предложенной к разделу через критическую работу с литературой и научными социологическими текстами. Данная работа необходима студенту для ведения научной дискуссии на занятии.

Работа с конспектом лекций

Просмотрите конспект сразу после занятий. Отметьте материал конспекта лекций, который вызывает затруднения для понимания. Попытайтесь найти ответы на затруднительные вопросы, используя предлагаемую литературу. Если самостоятельно не удалось разобраться в

материале, сформулируйте вопросы и обратитесь на текущей консультации или на ближайшей лекции за помощью к преподавателю.

Каждую неделю отводите время для повторения пройденного материала, проверяя свои знания, умения и навыки по контрольным вопросам и тестам.

Подготовка к практическим занятиям

Каждую неделю отводите время для повторения пройденного материала, проверяя свои знания, умения и навыки по контрольным вопросам и тестам.

На первом занятии получите у преподавателя задания по курсу, планы подготовки к практическим занятиям. Обзаведитесь всем необходимым методическим обеспечением.

Практическое занятие – одна из форм организации учебного процесса – самостоятельной работы обучающихся, которая регламентируется требованиями документа. Практическое занятие представляет собой коллективное обсуждение обучающимися теоретических вопросов под руководством преподавателя.

Практическое занятие () – вид аудиторной совместной деятельности, осуществляемой в форме практики мышления, речи и письма. На е студент отвечает на вопросы; уточняет, дополняет и обобщает ответы других; участвует в ролевых играх, дискуссиях; выступает с сообщением, докладом. Достоинством а является дискуссионная форма усвоения материала, сотворчество в обсуждении проблем, интеллектуальная состязательность.

При подготовке к практическому занятию сначала следует прочесть конспект лекций по теме занятия, затем изучить учебную и справочную литературу (учебник, учебные пособия, словари, энциклопедии), после чего – первоисточники.

Литературу необходимо конспектировать и давать сжатые письменные ответы на вопросы практического занятия. Такие ответы пригодятся при подготовке к итоговой аттестации. В них следует отразить содержание главных понятий и связь между ними, дать собственную оценку тем или иным концепциям, выявить связь изучаемого вопроса с современностью, с будущей профессиональной деятельностью. По неясным вопросам необходимо обращаться к преподавателю.

Текст выступления на практическом занятии должен быть написан самим студентом (уже переработанный). Выступления, на которых студент ограничивается только чтением уже опубликованного текста, преподавателем не учитываются.

По пропущенным занятиям студент отчитывается в устной или письменной форме.

Преподаватель оценивает подготовку студента к практическому занятию по следующим критериям:

- наличие конспектов, лекций;
- понимание главных понятий, их связей;

- умению мыслить самостоятельно, логично, последовательно, доказательно;

- мнениями оценивать идеи, концепции, обосновывать свою теоретическую позицию, связывать теорию с практикой, приводить примеры, приводить аналогии.

На практические занятия выносятся узловые темы, запланированные в программе учебной дисциплины, освоение которых определяет качество профессиональной подготовки будущих специалистов.

Цель занятия – формирование знаний по теме, развитие самостоятельности мышления и творческой активности обучающихся путем равноправного и активного их участия в обсуждении рассматриваемых вопросов.

Задачи практического занятия:

- закрепление, расширение и углубление знаний по психологии общения и конкретной теме;

- развитие навыков постановки проблем и решения задач;

- отработка навыков ведения дискуссии, аргументации и отстаивания своей точки зрения;

- демонстрация обучающимися достигнутого уровня знаний;

- формирование навыков самостоятельной работы.

Функции практического занятия (а):

- познавательная,

- воспитательная,

- развивающая,

- контрольная.

3. Методические рекомендации по работе с литературой

При изучении дисциплины у студентов должен вырабатываться рационально-критический подход к изучаемым проблемам и явлениям. Это включает понимание того, что со временем ряд информационных и теоретических материалов устаревает, требуя критического отношения. С другой стороны, каждый текущий вопрос имеет свою историю, которую тоже полезно знать. Каждое событие может иметь разные интерпретации, поэтому слова, сказанные много лет назад, могут иметь важное значение.

Чтобы понять содержание материала, нужно уметь его прочитывать. Начинать следует с предварительного просмотра, в ходе которого ознакомиться с названием работы, с аннотацией, оглавлением, предисловием. Часто замысел работы ясен уже при ознакомлении с ее названием. Но особенно интересен просмотр оглавления, в результате которого становится ясным развитие мысли автора. Неплохо было бы появившиеся при этом мысли зафиксировать на бумаге.

Просматривая текст оглавления, нужно остановиться на тех главах, которые представляют для вас особенный интерес, бегло ознакомиться с ними,

составляя в общих чертах свое представление о них. Цель этого действия – найти места, относящиеся к искомой теме, определив при этом, что ценного в каждом из них.

Следующий этап – прочтение выделенных мест с фиксацией самых главных сведений. При этом надо четко и ясно осознавать цель чтения, постоянно держа ее перед собой: по какому вопросу нужна информация, для чего нужна, ее характер и т.д. необходимо менять режим чтения – от беглого вдумчивого – в зависимости от ценности информации, останавливаясь там, где это требуется для глубокого понимания текста.

Следует научиться определять структуру текста по соподчиненности его частей, учитывая взаимосвязь текста с рисунками, сносками, примечаниями и таблицами. Все это поможет пониманию текста при беглом ознакомлении с ним. Так вырабатывается способность при прочтении сразу понимать смысл и значение новой информации.

Многие книги и статьи имеют в своем аппарате списки литературы, которые дают возможность пополнить информационную осведомленность о дополнительной литературе по данному вопросу.

Отдельный этап прочтения – ведение записей прочитанного. Существует несколько видов записей: план, выписки, тезисы, аннотация, резюме, конспект.

Планом удобно пользоваться при подготовке к устному выступлению по выбранной теме. Каждый пункт плана должен раскрывать одну из сторон избранной темы, а весь план должен охватывать ее целиком.

Тезисы предполагают сжатое изложение основных положений текста в форме утверждения или отрицания. Они являются более совершенной формой записей и представляют основу для дискуссии. К тому же их легко запомнить.

Аннотация – краткое изложение содержания – дает общее представление о работе.

Резюме кратко характеризует выводы, главные итоги произведения.

Конспект является наиболее распространенной формой ведения записей. Основную ткань конспекта составляют тезисы, дополненные доказательствами и рассуждениями. Конспект может быть текстуальным, свободным или тематическим. Текстуальный представляет собой цитатник с сохранением логики работы и структуры текста. Свободный конспект основан на изложении материала в том порядке, который более удобен автору. В этом смысле конспект представляет собирание воедино мыслей, разбросанных по всей книге. Тематический конспект может быть составлен по нескольким источникам, где за основу берется тема, интерпретируемая по-разному.

Экономия времени дает использование при записях различного рода сокращений, аббревиатуры и т.д. многие используют для регистрации исследуемых тем систему карточек. Преимущество карточек в том, что тема там излагается очень сжато, и они очень удобны в использовании, т.к. их можно разложить на столе, перегруппировать и без труда найти искомую тему.

4. Методические рекомендации подготовке к дискуссии, круглому столу

Так как занятие предусматривает проведение дискуссии, следует, то необходимо:

- знать материал темы проводимой дискуссии.
- уметь подбирать дискуссионные проблемы.
- владеть навыками участия в работе «круглого стола» (выдвижения положения, его аргументация).
- знать не только учебную, но и монографическую литературу по теме занятия, материалы научных периодических изданий.
- знать методику подготовки и проведения «круглых столов» разных видов,
- уметь разработать сценарий «круглого стола» и вступить в роли модератора «круглого стола».
- владеть навыком анализа и обобщения позиций участников «круглого стола».

Критерии оценивания компетенций

Оценка «отлично» выставляется студенту, если теоретическое содержание курса освоено полностью, без пробелов, необходимые практические компетенции сформированы, исчерпывающе, последовательно, четко и логически стройно излагает материал курса, умеет увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, использует в ответе дополнительный материал (монографии, статьи, исследования), все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному.

Оценка «хорошо» выставляется студенту, если теоретическое содержание курса освоено полностью, необходимые практические компетенции в основном сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения достаточно высокое. Студент твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения.

Оценка «удовлетворительно» выставляется студенту, если теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические компетенции в основном сформированы, большинство предусмотренных программой обучения учебных задач выполнено, но в них имеются ошибки, при ответе на поставленный вопрос студент допускает неточности, недостаточно правильные формулировки, наблюдаются нарушения логической последовательности в изложении программного материала.

Оценка «неудовлетворительно» выставляется студенту, если он не знает

значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет практические работы, необходимые практические компетенции не сформированы, большинство предусмотренных программой обучения учебных заданий не выполнено, качество их выполнения оценено числом баллов, близким к минимальному.

5. Методические рекомендации по написанию и оформлению доклада

Написание доклада по дисциплине «Инструменты и технологии анализа больших данных» является одним из видов самостоятельной работы студентов по изучению социологии. Данный вид работы способствует повышению качества усвоения программного материала, углубленному пониманию наиболее сложных вопросов курса, расширению круга рассматриваемых проблем, овладению научными методами анализа социологических явлений.

Работая с литературными источниками, не следует ограничиваться простым пересказом содержания прочитанного. Необходимо выделить наиболее важные теоретические положения и обосновать их, раскрыть особенности различных точек зрения на один и тот же вопрос, оценить практическое и теоретическое значение результатов реферируемой работы, а также выразить собственное отношение к идеям и выводам автора, подкрепив его определенными аргументами (личным опытом, высказываниями других исследователей и пр.).

Реферируемый источник, списки использованной литературы, а также все ссылки на литературные работы должны быть оформлены по алфавиту с указанием фамилии и инициалов автора, название источника, места и года издания; для журнальных статей необходимо указать фамилию и инициалы автора, название статьи, журнала, год издания и номер.

*Доклад по дисциплине « _____ »
студента ... курса (фамилия, имя, отчество)
факультета, группы
Тема: ".....".
год выполнения.*

1. Общие положения

1.1. Доклад, как вид самостоятельной работы в учебном процессе, способствует формированию навыков исследовательской работы, расширяет познавательные интересы, учит критически мыслить.

1.2. При написании доклада по заданной теме студент составляет план, подбирает основные источники.

1.3. В процессе работы с источниками систематизирует полученные сведения, делает выводы и обобщения.

1.4. К докладу по крупной теме могут привлекать несколько студентов,

между которыми распределяются вопросы выступления.

2. Выбор темы доклада

2.1. Тематика доклада обычно определяется преподавателем, но в определении темы инициативу может проявить и студент.

2.2. Прежде чем выбрать тему доклада, автору необходимо выявить свой интерес, определить, над какой проблемой он хотел бы поработать, более глубоко ее изучить.

3. Этапы работы над докладом

3.1. Формулирование темы, причем она должна быть не только актуальной по своему значению, но и оригинальной, интересной по содержанию.

3.2. Подбор и изучение основных источников по теме (как правильно, при разработке доклада используется не менее 8-10 различных источников).

3.3. Составление списка использованных источников.

3.4. Обработка и систематизация информации

3.5. Разработка плана доклада.

3.6. Написание доклада.

3.7. Публичное выступление с результатами исследования.

4. Структура доклада:

- титульный лист

- оглавление (в нем последовательно излагаются названия пунктов доклада, указываются страницы, с которых начинается каждый пункт);

- введение (формулирует суть исследуемой проблемы, обосновывается выбор темы, определяются ее значимость и актуальность, указываются цель и задачи доклада, дается характеристика используемой литературы);

- основная часть (каждый раздел ее, доказательно раскрывая отдельную проблему или одну из ее сторон, логически является продолжением предыдущего; в основной части могут быть представлены таблицы, графики, схемы);

- заключение (подводятся итоги или дается обобщенный вывод по теме доклада, предлагаются рекомендации);

- список использованных источников

5. Структура и содержание доклада

5.1. Введение - это вступительная часть научно-исследовательской работы. Автор должен приложить все усилия, чтобы в этом небольшом по объему разделе показать актуальность темы, раскрыть практическую значимость ее, определить цели и задачи эксперимента или его фрагмента.

5.2. Основная часть. В ней раскрывается содержание доклада.

Как правило, основная часть состоит из теоретического и практического разделов.

В теоретическом разделе раскрываются история и теория исследуемой проблемы, дается критический анализ литературы и показываются позиции автора.

В практическом разделе излагаются методы, ход, и результаты самостоятельно проведенного эксперимента или фрагмента.

В основной части могут быть также представлены схемы, диаграммы, таблицы, рисунки и т.д.

5.3. В заключении содержатся итоги работы, выводы, к которым пришел автор, и рекомендации. Заключение должно быть кратким, обязательным и соответствовать поставленным задачам.

5.4. Список использованных источников представляет собой перечень использованных книг, статей, фамилии авторов приводятся в алфавитном порядке, при этом все источники даются под общей нумерацией литературы. В исходных данных источника указываются фамилия и инициалы автора, название работы, место и год издания.

5.5. Приложение к докладу оформляются на отдельных листах, причем каждое должно иметь свой тематический заголовок и номер, который пишется в правом верхнем углу, например: «Приложение 1»

6. Требования к оформлению доклада

6.1. Объем доклада может колебаться в пределах 5-15 печатных страниц; все приложения к работе не входят в ее объем.

6.2. Доклад должен быть выполнен грамотно, с соблюдением культуры изложения.

6.3. Обязательно должны иметься ссылки на используемую литературу.

6.4. Должна быть соблюдена последовательность написания библиографического аппарата.

7. Критерии оценки доклада

Процедура проведения данного оценочного мероприятия включает в себя: заслушивание доклада студента, уточнение отдельных вопросов.

Для подготовки к данному оценочному мероприятию необходимо выбрать тему доклада, на основе изучения различного рода источников написать текстовую часть доклада, подготовить презентацию и выступление, раскрывающее суть темы.

При подготовке к ответу студенту предоставляется право пользования библиотеки и информационных ресурсов СКФУ.

При проверке задания, оцениваются:

- последовательность, четкость и логичность изложения материала;
- восприятие аудиторией представленной информации;
- соблюдение временного регламента выступления;
- продемонстрированная глубина владения теоретическим материалом по теме доклада;
- полнота, аргументированность ответов на вопросы аудитории;
- обоснованность собственной позиции по возникающим в ходе дискуссии проблемным вопросам;
- степень сформированности компетенций.

Примерные темы докладов:

1. Языки программирования для анализа больших данных: сравнительный анализ

2. Платные и бесплатные инструменты для обработки и анализа больших данных

1. Результаты процесса Business Intelligence

2. Средства предоставления информации в Business Intelligence

3. Структура когнитивного управления

4. Математический инструментарий анализа данных

5. Data Mining: особенности, возможности применения в социологии.

6. Инструменты обработки графовых данных

7. Понятие цифровых инструментов обработки данных

8. Характеристика свободно распространяемого цифрового ПО для обработки и анализа больших данных

9. WEKA: функционал, возможности и особенности применения программы в социологии.

10. RapidMiner: функционал, возможности и особенности применения программы в социологии.

11. Orange: функционал, возможности и особенности применения программы в социологии.

12. Knime: функционал, возможности и особенности применения программы в социологии.

13. Tableau: функционал, возможности и особенности применения.

14. Функции нереляционных баз данных

15. Влияние технологий NoSQL на распределенные облачные системы

6. Методические рекомендации по составлению глоссария

Составление глоссария – вид самостоятельной работы студента, выражающейся в подборе и систематизации терминов, непонятных слов и выражений, встречающихся при изучении темы.

Развивает у студентов способность выделять главные понятия темы и формулировать их. Оформляется письменно, включает название и значение терминов, слов и понятий в алфавитном порядке.

Студенту необходимо:

– прочитать материал источника, выбрать главные термины, непонятные слова;

– подобрать к ним и записать основные определения или расшифровку понятий;

– критически осмыслить подобранные определения и попытаться их модифицировать (упростить в плане устранения избыточности и повторений);

– оформить работу и представить в установленный срок.

Вопросы к зачету

1. Сущность понятия «большие данные»

2. Методики анализа больших данных

3. Принципы анализа больших данных
4. Особенности хранения больших данных
5. Структурированные, неструктурированные и квази-структурированные данные
6. Понятие «Business Intelligence»
7. Цели и задачи Data Science
8. Характеристика жизненного цикла аналитики данных
9. Технология Apache Hadoop
10. Технология MapReduce, ее достоинства и недостатки
11. Хранилища NoSQL главные отличия от обычных БД
12. Основные задачи, требования и типы визуализации
13. Социологические задачи анализа больших данных
14. Задачи классификации больших данных
15. Задачи кластеризации больших данных
16. Понятие искусственного интеллекта
17. Социологические задачи для искусственного интеллекта
18. Понятие высокопроизводительных вычислений
19. Методы прогнозирования в аналитике больших данных
20. Анализ социально-политических процессов с помощью больших данных
21. Роль больших данных в формировании общественного мнения
22. Электоральный процесс и большие данные
23. Цифровые инструменты обработки и анализа больших данных
24. Регрессионные методы в аналитике больших данных
25. «Жизненный цикл» проекта по аналитике больших данных
26. Методы построения деревьев классификации
27. Достоинства и недостатки использования Weka при обработке больших данных
28. Достоинства и недостатки использования Orange при обработке больших данных
29. Достоинства и недостатки использования RapidMiner при обработке больших данных
30. Принципы визуализации больших данных
31. Современные способы хранения больших данных
32. Структурированные и неструктурированные данные
33. Цифровые инструменты визуализации больших данных
34. Роль новых медиа в популяризации больших данных
35. Применение методов классификации к большим данным в области социологии
36. Применение методов классификации к большим данным в области политологии
37. Требования к безопасности больших данных
38. Суть когнитивного анализа данных
39. Цифровые инструменты визуализации графовых данных
40. Цифровые инструменты визуализации текстовых данных

Список литературы

Основная литература:

1. Воронова, Л. И. Machine Learning: регрессионные методы интеллектуального анализа данных Электронный ресурс: Учебное пособие / Л. И. Воронова, В. И. Воронов. - Machine Learning: регрессионные методы интеллектуального анализа данных, 2024-02-26. - Москва: Московский технический университет связи и информатики, 2018. - 82 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 2227-8397, экземпляров неограничено

2. Храмов, А. Г. Методы и алгоритмы интеллектуального анализа данных Электронный ресурс / Храмов А. Г.: учебное пособие. - Самара: СамГУ, 2019. - 176 с. - Рекомендовано редакционно-издательским советом федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» в качестве учебного пособия для обучающихся по основной образовательной программе высшего образования по направлению подготовки 01.04.02 Прикладная математика и информатика. - ISBN 978-5-7883-1414-3, экземпляров неограниченно

3. Агалаков, С. А. Статистические методы анализа данных: учебное пособие / С.А. Агалаков; Министерство образования и науки РФ; Омский государственный университет им. Ф. М. Достоевского. - Омск: ОмГУ им. Ф.М. Достоевского, 2017. - 92 с.: табл., граф., схем., ил. - <http://biblioclub.ru/>. - Библиогр. в кн. - ISBN 978-5-7779-2187-1, экземпляров неограниченно

4. Тарасов, И. Е. Статистический анализ данных в информационных системах Электронный ресурс / Тарасов И. Е.: учебно-методическое пособие. - Москва: РТУ МИРЭА, 2020. - 96 с., экземпляров неограниченно

Дополнительная литература:

1. Глебов, В. И. Практикум по математической статистике. Проверка гипотез с использованием Excel, MatCalc, R и Python: учебное пособие / В. И. Глебов, С. Я. Криволапов. - Практикум по математической статистике. Проверка гипотез с использованием Excel, MatCalc, R и Python, 2026-04-02. - Электрон. дан. (1 файл). - Москва: Прометей, 2019. - 86 с. - электронный. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-907100-66-4, экземпляров неограниченно

2. Лихтенштейн, В. Е. Информационные технологии в бизнесе. Том 2. Применение системы Decision в решении прикладных экономических задач Электронный ресурс: Учебное пособие / В. Е. Лихтенштейн, Г. В. Росс. - Информационные технологии в бизнесе. Том 2. Применение системы Decision в решении прикладных экономических задач, 2028-01-26. - Саратов: Ай Пи Эр Медиа, 2018. - 420 с. - Книга находится в премиум-версии ЭБС IPR BOOKS. - ISBN 978-5-4486-0283-2, экземпляров неограниченно

3. Мартышенко, С. Н. Автоматизация анализа данных в исследовании социально-экономических процессов. Монография Электронный ресурс / Мартышенко С. Н. - Владивосток: ВГУЭС, 2019. - 164 с. - ISBN 978-5-9736-0583-4, экземпляров неограниченно

Интернет-ресурсы:

1. Сайт Института социологии РАН – <http://www.isras.ru/socis.html>.
2. Журнал «Социологические исследования» // [Электрон. ресурс]. – Режим доступа: <http://socis.isras.ru>
3. Гуманитарный Интернет портал // [Электрон. ресурс]. – Режим доступа: www.auditorium.ru
4. Санкт-Петербургский государственный университет. Факультет социологии [Электрон. ресурс]. – Режим доступа: <http://www.soc.pu.ru/archive/>
5. «Фонд общественное мнение» (ФОМ) // [Электрон. ресурс]. – Режим доступа: <http://www.fom.ru/>
6. ВЦИОМ [Электрон. ресурс]. – Режим доступа: <https://wciom.ru/>
7. Левада-центр [Электрон. ресурс]. – Режим доступа: <http://www.levada.ru/>
8. Электронно-библиотечная система «Университетская библиотека онлайн» - www.biblioclub.ru;
9. «Фолиант» - <http://catalog.ncstu.ru>;
10. Электронная библиотечная система «IPRbooks»; База данных SCOPUS и др.