

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КАЛИНИНГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

См. с. 10

А. Б. Тристанов

БОЛЬШИЕ ДАННЫЕ

Учебно-методическое пособие по изучению дисциплины
для студентов по направлению подготовки 38.04.01 Экономика

Калининград
Издательство ФГБОУ ВО «КГТУ»
2022

Рецензент

кандидат физико-математических наук, и.о. заведующего кафедрой
прикладной математики и информационных технологий ФГБОУ ВО
«Калининградский государственный технический университет» А. И. Руденко.

Тристанов, А.Б. Большие данные: учебно-методическое пособие по изучению дисциплины для студентов направления подготовки 38.04.01 Экономика / А.Б. Тристанов – Калининград: Изд-во ФГБОУ ВО «КГТУ», 2022. – 19с.

В учебно-методическом пособии приведен тематический план изучения дисциплины. Представлены методические указания по изучению дисциплины. Даны рекомендации по подготовке к промежуточной аттестации в форме зачета и по выполнению самостоятельной работы. Пособие подготовлено в соответствии с требованиями утвержденной рабочей программы модуля «Цифровая экономика» по дисциплине «Большие данные» направления подготовки 38.04.01 Экономика.

Рис. – 2, табл. - 2, , список лит. – 11 наименований

Учебно-методическое пособие рассмотрено и одобрено в качестве локального электронного методического материала на заседании кафедры прикладной математики и информационных технологий 02 ноября 2022 г., протокол № 11

Учебно-методическое пособие по изучению дисциплины рекомендовано к использованию в учебном процессе в качестве локального электронного методического материала методической комиссией института цифровых технологий ФГБОУ ВО «Калининградский государственный технический университет» 03 ноября 2022 г., протокол № 09

Учебно-методическое пособие по изучению дисциплины рекомендовано к использованию в учебном процессе в качестве локального электронного методического материала методической комиссией ИНОТЭКУ 10 ноября 2022 г., протокол № 11

© Федеральное государственное бюджетное образовательное учреждение высшего образования "Калининградский государственный технический университет", 2022 г.
© Тристанов А.Б., 2022 г.

ОГЛАВЛЕНИЕ

Введение.....	4
1.1. Тематический план для студентов очной формы обучения.....	5
Таблица 1	5
1.2. Тематический план для студентов заочной формы обучения.....	6
Таблица 2	6
2. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ И УКАЗАНИЯ К ИЗУЧЕНИЮ	7
2.1. Раздел 1. Инфраструктура больших данных.....	7
Тема 1.1 Введение	7
Тема 1.2 Интеграция и хранение данных.....	8
Тема 1.3 Предобработка данных.....	10
2.2. Раздел 2. Технологии работы с большими данными	11
Тема 2.1 Методы Data Mining	11
Тема 2.2 Технологии BigData.....	12
Тема 2.3 Визуализация данных.....	14
3. ТРЕБОВАНИЯ К АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ.....	17
3.1. Текущая аттестация	17
3.2. Условия получения положительной оценки	17
Литература	18

ВВЕДЕНИЕ

Данное учебно-методическое пособие предназначено для студентов направления подготовки 38.04.01 Экономика, изучающих дисциплину «Большие данные».

Цель освоения дисциплины:

В результате освоения дисциплины ожидается, что студенты получат целостное представление о месте технологий обработки и анализа больших данных в профессиональной области, а также освоят ряд инструментальных средств анализа данных.

Для успешного освоения дисциплины, в соответствии с учебным планом, ей предшествуют «Информационные технологии в профессиональной деятельности», «Методы исследований в экономике и менеджменте».

Далее в пособии представлен тематический план, содержащий перечень изучаемых тем, обязательных лабораторных/практических работ, мероприятий текущей аттестации и отводимое на них аудиторное время (занятия в соответствии с расписанием) и самостоятельную работу. При формировании личного образовательного плана на семестр следует оценивать рекомендуемое время на изучение дисциплины, возможно, вам потребует больше времени на выполнение отдельных заданий или проработку отдельных тем.

В разделе Содержание дисциплины приведены подробные сведения об изучаемых вопросах, по которым вы можете ориентироваться в случае пропуска каких-то занятий, а также методические рекомендации преподавателя для самостоятельной подготовки, каждая тема имеет ссылки на литературу (или иные информационные ресурсы), а также контрольные вопросы для самопроверки.

Раздел «Текущая аттестация» содержит описание обязательных мероприятий контроля самостоятельной работы и усвоения разделов или отдельных тем дисциплины. Далее изложены требования к завершающей аттестации – зачету и курсовой работе.

Помимо данного пособия, студентам следует использовать материалы, размещенные в соответствующем данной дисциплине разделе ЭИОС, в которые более оперативно вносятся изменения для адаптации дисциплины под конкретную группу.

В ходе изучения дисциплины, выполнения курсовой и лабораторных работ используется аналитическая платформа Loginom, бесплатная академическая версия которой может быть свободно загружена с сайта компании-разработчика ООО «Аналитические технологии» – <https://loginom.ru/download> и установлена на персональные компьютеры, отвечающие техническим требованиям, а также платформа визуализации Yandex.DataLens.

1. ТЕМАТИЧЕСКИЙ ПЛАН

1.1. Тематический план для студентов очной формы обучения

Таблица 1

	Раздел (модуль) дисциплины	Тема	Объем контактной работы, ч	Объем самост. работы, ч
Теоретическое обучение (лекции)				
1.1	Инфраструктура больших данных	Введение	2	2
1.2		Интеграция и хранение данных	2	2
1.3		Предобработка данных	2	2
2.1	Технологии работы с большими данными	Методы Data Mining	4	4
2.2		Вычислительные технологии BigData	4	4
2.3		Визуализация данных	2	2
			16	16
Практические (лабораторные занятия)				
1.1	Инфраструктура больших данных	ETL система Loginom	6	4
1.1	Инфраструктура больших данных	Предобработка и оценка качества больших данных	6	4
2.1	Технологии работы с большими данными	Визуализация данных	2	4
			14	12
Курсовая работа				
2.1	Инфраструктура больших данных	Контрольная точка 1. Раздел проекта 1	6 (ЭИОС)	9
3.1	Технологии работы с большими данными	Контрольная точка 2. Раздел проекта 2	6 (ЭИОС)	12
		Оформление проекта. Защита	4 (ЭИОС)	10
			16	31
Рубежный (текущий) и итоговый контроль				
2.1	Инфраструктура больших данных	Тестирование		5,80
		Итоговое тестирование (зачет)	x	33,2
			0	39,00
Всего			46	98

1.2. Тематический план для студентов заочной формы обучения

Таблица 2

	Раздел (модуль) дисциплины	Тема	Объем контактной работы, ч	Объем самост. работы, ч
Теоретическое обучение (лекции)				
1.1	Инфраструктура больших данных	Введение (установочная лекция)	2	5
1.2		Интеграция и хранение данных		6
1.3		Предобработка данных		6
2.1	Технологии работы с большими данными	Методы Data Mining	1	10
2.2		Вычислительные технологии Big Data	1	12
2.3		Визуализация данных		6
			4	45
Практические (лабораторные занятия)				
1.1	Инфраструктура больших данных	ETL система Loginom	6	4
1.1	Инфраструктура больших данных	Предобработка и оценка качества больших данных	6	4
2.1	Технологии работы с большими данными	Визуализация данных	4	4
			16	12
Курсовая работа				
2.1	Инфраструктура больших данных	Контрольная точка 1. Раздел проекта 1	1 (ЭИОС)	20
3.1	Технологии работы с большими данными	Контрольная точка 2. Раздел проекта 2	1 (ЭИОС)	20
		Оформление проекта. Защита		14
			2	54
Рубежный (текущий) и итоговый контроль				
2.1	Инфраструктура больших данных	Тестирование		4,60
		Итоговое тестирование (зачет)	x	6,4
			0	11,00
CP+контроль				
Всего			22	122

2. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ И УКАЗАНИЯ К ИЗУЧЕНИЮ

2.1. Раздел 1. Инфраструктура больших данных

Тема 1.1 Введение

Перечень изучаемых вопросов:

Определение больших данных. Источники больших данных. Управление на основе данных. Значение больших данных для бизнеса и государства. Лучшие практики. Цикл управления большими данными. Структурированные и неструктурированные данные. Метаданные. Задачи анализа больших данных. Технологии распределенных вычислений. Инструментальные средства работы с большими данными.

Методические указания к изучению:

Большие данные – термин, имеющий широкий спектр толкований, наиболее предпочтительным является понимание, определяющееся тремя наиболее характерными признаками: большой объем, большая скорость накопления, большая вариабильность (V³ – Volume, Velocity, Variable). Эти признаки и определяют специфику обращения с такими данными. Разбирая определение «большие данные» следует уделить особое внимание каждому из признаков и находить особенности применения методов обработки и анализа именно для больших данных.

Рассматривая вопрос о значимости больших данных для государства и бизнеса следует в первую очередь определить возможные задачи которые для своего решения требуют больших объемов накопленных данных. Следует познакомиться с основными направлениями политики государства в сфере цифровой экономики, рассмотреть опыт цифровой трансформации крупных предприятий, учреждений и организаций. Для этого изучите интернет-ресурсы, приведенные в списке рекомендованных источников.

Опираясь на собранную информацию, определите возможные источники больших данных. С какими сложностями и потенциальными опасностями связано накопление и использование данных? В чем особенности сбора, хранения и обработки персональных данных граждан и данных технического мониторинга.

Основываясь на имеющемся профессиональном опыте приведите примеры возможных задач, решение которых связано с накоплением и обработкой больших объемов данных.

Литература:

[1–3, 5].

Контрольные вопросы:

1. Раскройте понятия больших данных с точки зрения их объема, скорости накопления и вариабельности.
2. Перечислите не менее трех технологических проблем обработки больших данных.
3. Приведите примеры применения технологий больших данных в бизнесе, в государственном управлении. Почему именно «большие» данные дают эффект.

Тема 1.2 Интеграция и хранение данных

Перечень изучаемых вопросов:

Источники больших данных. Понятие базы данных. СУБД. Модели данных. Хранилища данных, витрины данных. Основы языка SQL.

Методические указания к изучению:

Прежде чем приступить к обработке и анализу данных, необходимо решить вопрос с подходами к сбору данных и их хранению. Приступая к изучению данной темы, следует вспомнить базовый материал курса «Информатика» (или аналогичного) программы бакалавриата, а также курса «Информационные технологии в профессиональной деятельности» данной программы магистратуры. Определите место процессов сбора и хранения в информационной технологии. Рассмотрите два подхода к хранению больших объемов: «локальный» – на сервере предприятия и «облачный» – на арендованных виртуальных вычислительных ресурсах. Определите преимущества и недостатки первого и второго способа. Ознакомьтесь с документацией облачных платформ, например, Яндекс.Облако [1] SberCloud [2].

К источникам больших данных традиционно относят [3]: интернет вещей (IoT) и подключенные к нему устройства; соцсети, блоги и СМИ; данные компаний: транзакции, заказы товаров и услуг, поездки на такси и каршеринге, профили клиентов; показания приборов: метеорологические станции, измерители состава воздуха и водоемов, данные со спутников; статистика городов и государств: данные о перемещениях, рождаемости и смертности; медицинские данные: анализы, заболевания, диагностические снимки.

Приведите примеры наборов данных, которые порождаются перечисленными выше источниками, а также сервисы (услуги, продукты), которые построены на базе анализа этих наборов. Не забудьте обозначить примерные объемы и скорость роста объемов данных. Почему их можно отнести к категории «больших данных».

Базы данных (БД) стали неотъемлемой частью повседневной жизни, поэтому их изучение является обязательным элементом информационной

культуры любого специалиста. Базу данных можно рассматривать как набор логически связанных данных, а систему управления базами данных (СУБД) - как программное обеспечение, которое управляет доступом к базе данных. Поскольку базовые понятия были изучены в соответствующих курсах бакалавриата, то, обобщая имеющийся опыт, опишите основные характеристики подхода, основанного на использовании БД, и сравните с характеристиками подходов, используемых в обычных файловых системах.

Дайте характеристику компонентам среды СУБД: аппаратное обеспечение программное обеспечение, данные, процедуры, пользователи. Вспомните трехуровневую архитектуру СУБД: внешний уровень, представляющий БД с точки зрения пользователя, концептуальный уровень, обобщающий представление базы данных и описывающий хранимые данные и связи между ними, и внутренний уровень физического представления БД в компьютере.

Моделью данных называют интегрированный набор понятий для описания данных, связей между ними и ограничений, накладываемых на данные в некоторой структуре. Можно выделить три категории моделей данных - объектную модель, модель на основе записей и физическую модель данных. Первые две используются для описания данных на концептуальном и внешнем уровне, последняя модель – для описания данных на внутреннем уровне. Обратите внимание на объектную модель данных типа сущность-связь, или ER-модель. К моделям данных на основе записей относят реляционную, сетевую и иерархическую модели.

Важной частью организации хранения данных является проектирование процесса загрузки данных. Ранее мы установили, что вариабельность структуры данных, форматов хранения требует использования специальных подходов к загрузке данных и процедуре интеграции. Рассмотрите понятие ETL (Extract, Transform, Load) процесса. Обозначьте действия, выполняемые с данными на каждом этапе: извлечение, преобразование, загрузка.

Далее переходите к изучению понятия «витрина данных», как специализированного предметно-ориентированного поднабора данных для решения частных задач обработки и анализа. Приведите примеры витрин данных в рамках одного предприятия в разрезе необходимых данных о его деятельности по должностям.

В завершение изучения темы рассмотрите основные понятия языка SQL как механизма извлечения и управления данными в реляционных базах. Язык SQL является первым и пока единственным стандартным языком работы с базами данных, который получил достаточно широкое распространение. Стандарты языка SQL используют такие термины, как таблица, столбец и строка. Язык SQL предоставляет пользователю возможности по созданию баз

данных и таблиц с полным описанием их структуры, позволяет выполнять основные операции манипулирования данными, такие как вставка, модификация и удаление данных из таблицы, выполнять простые и сложные запросы, осуществляющие преобразование необработанных данных в нужную информацию. Рассмотрите структуру основного оператора языка SQL - Select, предназначенного для выборки данных из БД. Ознакомьтесь с синтаксисом данного оператора, научитесь строить простые запросы (выборка столбцов из таблиц, выборка строк, удовлетворяющих определенному критерию (условию поиска)), освоите применение обобщающих функций языка SQL (count, sum, avg, min, max).

Литература:

[4, 1, 4, 5].

Контрольные вопросы:

1. Приведите примеры ETL платформ, укажите их основные функции.
2. Что такое витрина данных? Приведите не менее трех примеров.
3. В чем заключается основная идея хранения данных в виде таблиц?
4. Приведите примеры связей между таблицами для любой конкретной предметной области.
5. Для чего применяются обобщающие (агрегирующие) функции в запросах?

Тема 1.3 Предобработка данных.

Перечень изучаемых вопросов:

Понятие пространства признаков. Метрики. Введение в сокращение размерности. Сокращение числа признаков. Оценка качества данных. Технологии и методы оценки качества данных. Очистка и предобработка. Фильтрация данных. Обработка дубликатов и противоречий. Выявление аномальных значений. Восстановление пропущенных значений.

Методические указания к изучению:

Предобработка или первичная обработка данных играет важную роль в процессе анализа данных. Изучение этой темы целесообразно начинать с изучения способа представления объектов исследования как набора характеризующих его описаний. Обратите внимание на соответствии концептуальной модели базы данных и данного подхода. Каждому уникальному объекту может быть поставлена в соответствие точка в некотором n -мерном (в общем случае бесконечномерном) пространстве. Величина n — это размерность пространства. В этом пространстве можно ввести систему координат (базис), каждая ось соответствует множеству значений характеристик объекта — признакам. Такое пространство будем называть

признаковым. Теперь каждый уникальный объект может быть сопоставлен вектору, элементами которого являются признаки. Приведите примеры типов признаков, чем количественные признаки отличаются от качественных. В данном пространстве может быть введено понятие расстояния — метрики. Изучите различные варианты метрик: евклидова, манхеттенская, Ляпунова.

Литература:

[9].

Контрольные вопросы:

1. Приведите примеры признаков, которые могли бы использоваться для объектов профессиональной деятельности.
2. Дайте определение следующим математическим объектам: вектор, базис, векторное пространство, метрика.
3. Какими свойствами обладает метрика?

2.2. Раздел 2. Технологии работы с большими данными

Тема 2.1 Методы Data Mining

Перечень изучаемых вопросов:

Задачи интеллектуального анализа. Некоторые алгоритмы классификации (логистическая регрессия, искусственные нейронные сети), кластеризации (k-means), регрессии (линейная регрессия, искусственные нейронные сети), ассоциации (Apriori).

Методические указания к изучению:

Приведите определение технологии Data Mining. Первоочередной задачей при проработке данного раздела является понимание задач интеллектуального анализа, вытекающих из анализа работы биологического мозга. Рекомендуется ознакомиться с популярными лекциями о работе мозга В.А. Дубынина, К.В. Анохина, Т.В. Черниговской, А.В. Курпатова. Выделяют следующие задачи интеллектуального анализа: кластеризация, классификация и регрессия, а также ассоциация. Дайте определение каждой из задач, приведите примеры их решения мозгом, а также связь с профессиональными бизнес-задачами. Дайте классификацию алгоритмов по использованию маркированных данных — обучение с учителем и обучение без учителя. Чем данные подходы отличаются от алгоритмов обучения с подкреплением? Рассмотрите базовые алгоритмы решения задачи кластеризации — k-means, dbscan, классификации — логистическая регрессия и ассоциации — Apriori. Обратите внимание на алгоритмы классификации и регрессии на базе антропоморфного подхода — искусственные нейронные сети (ИНС). Объясните популярность этого подхода в настоящее время, приведите примеры использования ИНС в реальных

задачах. Укажите этапы предобработки при применении методов интеллектуального анализа данных. Сравните потенциальные результаты традиционного статистического анализа и интеллектуального анализа.

Литература:

[6-8].

Контрольные вопросы:

1. Как соотносятся понятия искусственный интеллект, машинное обучение, интеллектуальный анализ, нейронные сети, глубокое обучение?
2. В чем отличие обучения с учителем, без учителя и обучения с подкреплением?
3. В чем отличие задачи классификации и задачи регрессии?
4. В чем основные ограничения применения ИНС?

Тема 2.2 Технологии BigData

Перечень изучаемых вопросов:

Распределенные вычисления. Архитектура систем на основе BigData. Технология MapReduce. Виртуализация и облачные вычисления. Проблемы информационной безопасности.

Методические указания к изучению:

Участвуя в цифровой трансформации на предприятиях и в организациях специалисту не-IT-профиля необходимо быть информированным и в ключевых аспектах внедряемых технологий. Не исключением является и использование технологий больших данных.

Основной задачей данной технологии является повышение эффективности и скорости обработки значительных объемов структурированной и неструктурированной информации. В основе данной технологии лежит идея распределения процессов обработки между разными исполнителями. В данном случае под исполнителями понимаем вычислительные устройства реальные или виртуальные, хотя определенная аналогия с обработкой скажем большого потока первичной документации однотипными специалистами может также иметь место.

Система обработки больших данных должна иметь возможность горизонтального масштабирования, а именно возможность наращивания вычислительных ресурсов при увеличении объема данных и сокращении при уменьшении. Отказоустойчивость также должно рассматриваться в таких системах, т.к. чем больше элементов участвует в обработке данных тем более вероятен выход из строя части этих элементов, при этом информационных процесс не должен критично нарушаться. Следующий аспект – расположение обрабатываемых данных и их локализации близ вычислительного узла. Одним

из узких горлышек распределенных систем являются затраты на коммуникацию между вычислительными узлами.

Одной из распространенных технологий организации обработки больших данных является MapReduce. Технология заключается в распределении всего объема данных на параллельную обработку разными вычислительными узлами (этап Map), а затем консолидация результатов (этап Reduce). Рассмотрите применение данной технологии на модельном примере по подсчету слов в тексте (рис.1).

MapReduce

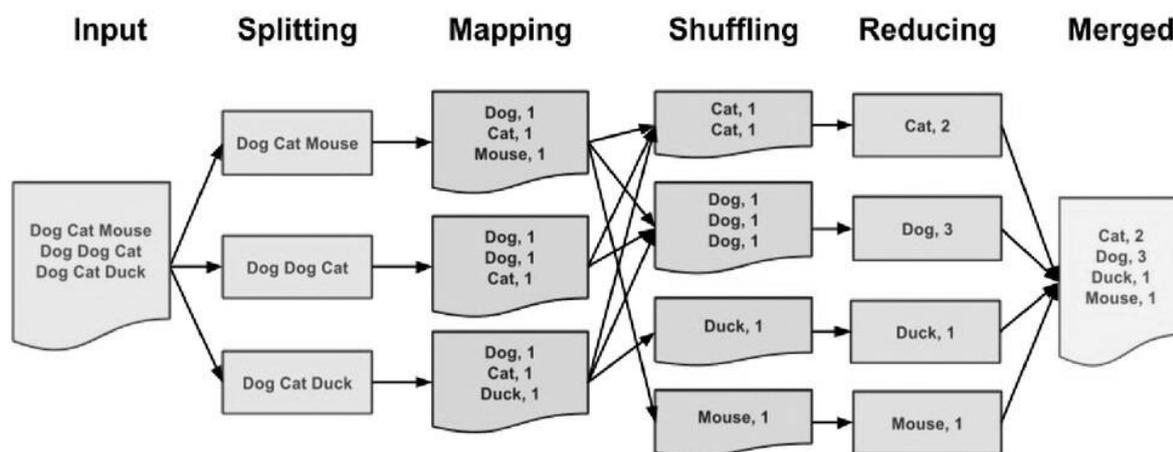


Рис. 1. Иллюстрация использования технологии MapReduce

Не для каждого предприятия целесообразно содержать собственную вычислительную инфраструктуру. В настоящее время широко распространена услуга по аренде вычислительных ресурсов. Изучите основные бизнес-модели данной услуги на примере крупных провайдеров (Yandex, Sber, VK). В основе данной услуги, как правило, лежит технологии виртуализации. Виртуализация – это предоставление набора вычислительных ресурсов или их логического объединения, абстрагированное от аппаратной реализации, и обеспечивающее при этом логическую изоляцию друг от друга вычислительных процессов, выполняемых на одном физическом ресурсе. По сути, виртуальный ресурс – это программная реализация различных физических объектов: компьютеров, хранилищ данных, сетей, серверов. Тем самым виртуальные машины, их элементы или совокупность машин представляются конечному пользователю как полноценный изолированный вычислительный ресурс, хотя физически это может быть часть ресурсов физического сервера или наоборот их совокупность.

Достаточно понятным примером могут служить облачные хранилища данных. Если на персональной ЭВМ диск имеет ограниченный объем, то для увеличения объема хранения необходимо подключить еще один диск, в облачной среде пользователь докупает услугу по увеличению объема, не вдаваясь в подробности как осуществляется физический процесс выделения дополнительной памяти. При этом если потребность в дополнительном объеме отпадает, то для реальной ЭВМ затраты на ее модернизацию уже осуществлены, а в облачном варианте от услуги, как правило, можно отказаться. Аналогично и с затратами машинного времени.

В основе виртуальной вычислительной среды лежит гипервизор — программа, которая распределяет аппаратные ресурсы между виртуальными машинами. Выделяют программную и аппаратную виртуализацию по позиции гипервизора в инфраструктуре.

В завершении изучения данной темы следует рассмотреть основные проблемы защиты вычислительной инфраструктуры больших данных.

Литература:

[11].

Контрольные вопросы:

1. Дайте определение облачной технологии.
2. В чем отличие программной и аппаратной виртуализации?
3. Приведите примеры компаний предоставляющие в аренду вычислительные ресурсы и их бизнес-модели. В чем причина различной стоимости их услуг?
4. В чем преимущества виртуализации ресурсов? Приведите не менее пяти преимуществ.
5. В чем проблемы информационной безопасности виртуальных ресурсов? Приведите примеры не менее трех угроз.

Тема 2.3 Визуализация данных

Перечень изучаемых вопросов:

Сторителлинг. Таблицы. Диаграмма. Инфографика. Дашборды.

Методические указания к изучению:

В бизнесе визуализация необходима для решения широкого спектра задач, начиная от кадровых вопросов и заканчивая предоставлением скидки конкретному покупателю. Инструменты для визуализации бизнес-данных относятся к классу BI-систем.

Одной из проблемы анализа больших данных является сложность представления результатов данного анализа в понятном для конечного пользователя виде. В рамках данного раздела дисциплины предлагается

изучить современные подходы к визуализации и представлению данных на основе платформы визуализации Yandex.DataLens. В ходе лабораторных работ изучите основные концепции платформы, используя документацию разработчика.

Изучите основные «чарты», применяемые в визуализации: линейные диаграммы, колонки и гистограммы, круговые диаграммы, полярные графики, точечные графики, карты, деревья, временные диаграммы, пр. Рекомендуется обращать внимание на английский эквивалент наименований чартов. Обозначьте области применения различных диаграмм и таблиц. Найдите не менее 10 примеров различных чартов и их комбинаций. Приведите примеры удачной визуализации и неудачной. Изучите приведенную ниже схему.

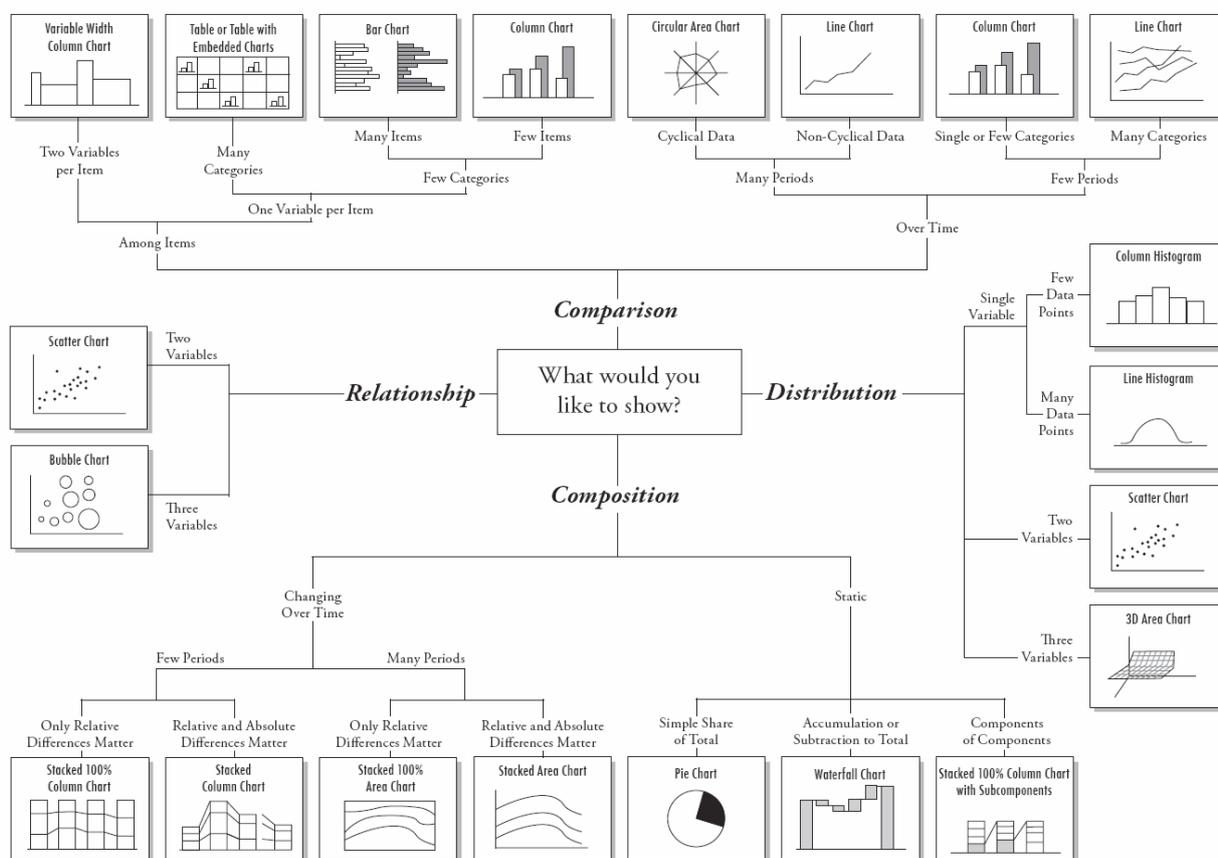


Рис. 2 Рекомендуемые типы «чартов» в зависимости от визуализируемой информации. (Заимствовано <https://www.labnol.org/software/find-right-chart-type-for-your-data/6523/>)

Далее рассмотрите понятие «сторителлинг» в анализе данных и понятие «инфорграфика». Раскройте смысл контекста при представлении результата, какие цели с точки зрения визуализации преследуются при подготовке докладов по результатам анализа данных? Дайте определения понятию «дашборд» (информационная панель), какова задача интерактивности при

представлении результатов. Приведите примеры не менее 10 доступных дашбордов, изучите их структуру.

Литература:

[9, 10].

Контрольные вопросы:

1. Что такое BI-система? Приведите примеры не менее 3 BI-систем.
2. В чем преимущество интерактивных панелей (дашбордов) перед статической презентацией?
3. В чем отличие «подключения» от «датасета» на платформе Yandex.DataLens.
4. В чем преимущества использования облачных платформ визуализации? Приведите не менее пяти преимуществ.
5. Предложите схему организации взаимодействия пользователей (аналитика, системного администратора, клиента, специалиста отдела продаж) платформы Yandex.DataLens при подготовке и использовании информационных панелей для некоторого модельного предприятия. Рассмотрите реальные кейсы внедрения.

3. ТРЕБОВАНИЯ К АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

3.1. Текущая аттестация

В ходе изучения дисциплины студентам предстоит пройти следующие этапы текущей аттестации:

Преподаватель вправе выбрать методику оценивания знаний студентов: традиционная зачетно-экзаменационная либо балльно-рейтинговая.

В рамках балльно-рейтинговой системы выставляется оценка за качество выполнения и защиту лабораторных и контрольных работ.

Виды деятельности и соотношение трудоемкости.

Таблица 3

Вид деятельности	Доля	Кол-во ед.	Макс. балл за ед.	Всего
Обязательные виды деятельности				
1-й семестр				
Посещаемость занятий	20%	N1	=200/N1	200
Выполнение лаб. работ (защита)	40%	3	200	400
Курсовая работа	40%	1	400	400
Итого:	100%			1000
Дополнительные задания (по выбору студента в каждом семестре)				
Подготовка реферата (видео-доклада)	20%		200	200
Выполнение задания в рамках НИРС	50%		500	500

3.2. Условия получения положительной оценки

Завершающим этапом изучения дисциплины является промежуточная аттестация в форме зачета. Оценка зачтено выставляется в случае если студентом набрано не менее 600 баллов, в случае применения БРС, либо выполнены и защищены все лабораторные работы и представлена к проверке курсовая работа, в случае традиционного подхода к оцениванию.

ЛИТЕРАТУРА

1. Надежное облако для вашего бизнеса — Yandex Cloud [Электронный ресурс]. URL: <https://cloud.yandex.ru/> (дата обращения: 28.05.2022).
2. SberCloud — облачный провайдер IaaS/PaaS и ML сервисов // SberCloud — облачный провайдер IaaS/PaaS и ML сервисов [Электронный ресурс]. URL: <https://sbercloud.ru/ru> (дата обращения: 28.05.2022).
3. Что такое Big Data и почему их называют «новой нефтью» // РБК Тренды [Электронный ресурс]. URL: <https://trends.rbc.ru/trends/innovation/5d6c020b9a7947a740fea65c> (дата обращения: 28.05.2022).
4. Источники данных Loginom Wiki [Электронный ресурс]. URL: <https://wiki.loginom.ru/data-sources.html> (дата обращения: 28.05.2022).
5. Простор для данных // Хабр [Электронный ресурс]. URL: <https://habr.com/ru/post/650237/> (дата обращения: 28.05.2022).
6. Паклин, Н. Б. Бизнес-аналитика: от данных к знаниям / Н. Б. Паклин, В. И. Орешков. - Санкт-Петербург: Питер, 2013. 704 с.
7. Visualizing K-Means Clustering [Электронный ресурс]. URL: <https://www.naftaliharris.com/blog/visualizing-k-means-clustering/> (дата обращения: 24.09.2022).
8. Visualizing DBSCAN Clustering [Электронный ресурс]. URL: <https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/> (дата обращения: 24.09.2022).
9. Документация Yandex Cloud | Yandex DataLens | Yandex DataLens [Электронный ресурс]. URL: <https://cloud.yandex.ru/docs/datalens/> (дата обращения: 05.06.2022).
10. Истории успеха [Электронный ресурс]. URL: <https://cloud.yandex.ru/cases> (дата обращения: 24.09.2022).
11. Документация Yandex Cloud | Облачная терминология | Что такое виртуализация [Электронный ресурс]. URL: <https://cloud.yandex.ru/docs/glossary/virtualization> (дата обращения: 24.09.2022).

Локальный электронный методический материал

Александр Борисович Тристанов

БОЛЬШИЕ ДАННЫЕ

Редактор Г. А. Смирнова

Уч.-изд. л. 1,4. Печ. л. 1,25

Издательство федерального государственного бюджетного
образовательного учреждения высшего образования
«Калининградский государственный технический университет».
236022, Калининград, Советский проспект, 1