

Подписано электронной подписью:
Вержицкий Данил Григорьевич
Должность: Директор КГПИ ФГБОУ ВО «КемГУ»
Дата и время: 2024-02-21 00:00:00
471086fad29a3b30e244c728abc3661ab35c9d50210dcf0e75e03a5b6fdf6436

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Кузбасский гуманитарно-педагогический институт
федерального государственного бюджетного образовательного учреждения
высшего образования
«Кемеровский государственный университет»

Факультет информатики, математики и экономики

УТВЕРЖДАЮ
Декан А.В. Фомина
«09» февраля 2023 г.

Рабочая программа дисциплины
К.М.07.03 Технологии работы с большими данными

Направление подготовки
01.03.02 Прикладная математика и информатика

Направленность (профиль) подготовки
ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Программа бакалавриата

Квалификация выпускника
бакалавр

Форма обучения
Очная

Год набора 2022

Новокузнецк 2023

Оглавление

1	Цель дисциплины	3
1.1	Формируемые компетенции.....	3
1.2	Индикаторы достижения компетенций.....	3
1.3	Знания, умения, навыки (ЗУВ) по дисциплине	3
2	Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации.....	4
3.	Учебно-тематический план и содержание дисциплины.....	4
3.1	Учебно-тематический план	4
3.2.	Содержание занятий по видам учебной работы.....	5
4	Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации.....	6
5	Материально-техническое, программное и учебно-методическое обеспечение дисциплины.....	7
5.1	Учебная литература	7
5.2	Материально-техническое и программное обеспечение дисциплины.....	7
5.3	Современные профессиональные базы данных и информационные справочные системы.....	8
6	Иные сведения и (или) материалы.....	8
6.1.	Примерные темы курсовых работ	8
6.2.	Примерные вопросы и задания / задачи для промежуточной аттестации	8

1 Цель дисциплины.

В результате освоения данной дисциплины у обучающегося должны быть сформированы компетенции основной профессиональной образовательной программы бакалавриата (далее - ОПОП): ПК-1

Содержание компетенций как планируемых результатов обучения по дисциплине см. таблицы 1 и 2.

1.1 Формируемые компетенции

Таблица 1 - Формируемые дисциплиной компетенции

Наименование вида компетенции (универсальная, общепрофессиональная, профессиональная)	Наименование категории (группы) компетенций	Код и название компетенции
профессиональная		ПК-1 Способен проводить аналитические исследования с применением технологий больших данных

1.2 Индикаторы достижения компетенций

Таблица 2 – Индикаторы достижения компетенций, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции по ОПОП	Дисциплины и практики, формирующие компетенцию ОПОП
ПК-1 Способен проводить аналитические исследования с применением технологий больших данных	ПК 1.1 Способен осуществлять выявление, формирование и согласование требований к результатам аналитических работ с применением технологий больших данных ПК 1.2 Способен планировать и организовывать аналитические работы с использованием технологий больших данных ПК 1.3 Способен подготавливать данные для проведения аналитических работ по исследованию больших данных ПК 1.4 Способен проводить аналитические исследования с применением технологий больших данных в соответствии с требованиями заказчика	К.М.07.01 Машинное обучение К.М.07.02 Аналитика данных К.М.07.03 Технологии работы с большими данными К.М.07.04 Технологии работы с открытыми данными К.М.07.ДВ.01.01 Машинное обучение с подкреплением/ К.М.07.ДВ.01.02 Глубокое обучение К.М.09.06(Пд) Преддипломная практика

1.3 Знания, умения, навыки (ЗУВ) по дисциплине

Таблица 3 – Знания, умения, навыки, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции, закрепленные за дисциплиной	Знания, умения, навыки (ЗУВ), формируемые дисциплиной
----------------------------	--	---

Код и название компетенции	Индикаторы достижения компетенции, закрепленные за дисциплиной	Знания, умения, навыки (ЗУВ), формируемые дисциплиной
ПК-1 Способен проводить аналитические исследования с применением технологий больших данных	ПК 1.3 Способен подготавливать данные для проведения аналитических работ по исследованию больших данных	<p>Знать:</p> <ul style="list-style-type: none"> - стандарты и требования, связанные с созданием и использованием SQL и NoSQL систем хранения и обработки больших данных. <p>Уметь:</p> <ul style="list-style-type: none"> - использовать современные программные средства для сбора и подготовки больших данных к анализу. <p>Владеть:</p> <ul style="list-style-type: none"> - навыками собирать и обрабатывать большие данные.

2 Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации.

Таблица 4 – Объем и трудоемкость дисциплины по видам учебных занятий

Общая трудоемкость и виды учебной работы по дисциплине, проводимые в разных формах	Объём часов по формам обучения
	ОФО
1 Общая трудоемкость дисциплины	144
2 Контактная работа обучающихся с преподавателем (по видам учебных занятий) (всего)	42
Аудиторная работа (всего):	42
в том числе:	
лекции	6
лабораторные работы	36
в интерактивной форме	
Внеаудиторная работа (всего):	
3 Самостоятельная работа обучающихся (всего)	102
4 Промежуточная аттестация обучающегося - зачет с оценкой (7 семестр)	

3. Учебно-тематический план и содержание дисциплины.

3.1 Учебно-тематический план

Таблица 5 - Учебно-тематический план очной формы обучения

№ недели п/п	Разделы и темы дисциплины по занятиям	Общая трудоёмкость (всего час.)	Трудоёмкость занятий (час.)			Формы текущего контроля и промежуточной аттестации успеваемости	
			Аудиторн. занятия		СРС		
			лекц.	практ.			лаб.
Семестр 7							
1	Хранилища данных	48	2		12	34	Отчеты по лабораторным работам
2	Нереляционные (NoSQL) базы данных	48	2		12	34	Отчеты по лабораторным работам Тест

№ недели п/п	Разделы и темы дисциплины по занятиям	Общая трудоём- кость (<i>всего час.</i>)	Трудоёмкость занятий (час.)				Формы текущего контроля и промежуточной аттестации успеваемости
			Аудиторн. занятия		СРС		
			лекц.	практ.	лаб.		
3	Распределенные базы данных	48	2		12	34	Отчеты по лабораторным работам
	Промежуточная аттестация – <i>зачет с оценкой</i>						зачет с оценкой
	Всего:	144	6		36	102	

3.2. Содержание занятий по видам учебной работы

Таблица 6 – Содержание дисциплины

№ п/п	Наименование раздела, темы дисциплины	Содержание занятия
Семестр 7		
<i>Содержание лекционного курса</i>		
1	Хранилища данных	Хранилища данных. Задача консолидации. Хранилища данных. Основные концепции хранилищ данных. OLAP и OLTP системы. Основные понятия и определения OLAP-анализа. Требования, предъявляемые к многомерному анализу данных. Стадии многомерного анализа данных. Сферы применения технологии OLAP-анализа. Методика обнаружения нового знания в хранилищах данных (KDD). Характеристика этапов KDD. Программное обеспечение в области анализа данных. Понятия OLAP-куба, операции, выполняемые над гиперкубом. Архитектура OLAP-систем. Визуальный анализ данных, его преимущества. Процесс визуализации данных (Visual Mining). Характеристики средств визуализации данных. Классификация продуктов OLAP по способу представления данных. Многомерный OLAP (MOLAP). Реляционный OLAP (ROLAP). HОLAP. Интеллектуальный анализ данных. Интеграция OLAP и ИАД. Критерии оценки существующих продуктов OLTP-системы. Требования, предъявляемые к OLTP-системам. Характеристики BigData и хранилищ данных. Преимущества перед реляционными СУБД. Требования к хранилищам данных.
2	Нереляционные (NoSQL) базы данных	Причины и условия применения нереляционных баз данных (баз данных NoSQL). Преимущества и недостатки нереляционных баз данных. Классификация NoSql хранилищ. Их особенности. Требования ACID. CAP-теорема. BASE архитектура. Типы нереляционных баз данных: ключ-значение, документно-ориентированные, колоночно-ориентированные, графовые. Примеры, возможности и особенности использования нереляционных баз данных.
3	Распределенные базы данных	Распределенные базы данных NoSQL. Типы NoSQL. Репликация и шардинг. Пример NoSQL БД: HBase, Cassandra, Neo4j, MongoDB.
<i>Содержание лабораторных занятий</i>		
1	Хранилища данных	Программная реализация одной из представленных OLAP-систем. Программная реализация многомерного анализа данных.

№ п/п	Наименование раздела, темы дисциплины	Содержание занятия
		Примеры задач интеллектуального анализа данных и OLTP-систем. Структура OLTP-систем.
2	Нереляционные (NoSQL) базы данных	Очистка данных. Избавление от шумов. Основы работы с NoSQL СУБД MongoDB: создание базы данных, выполнение команд в консоли, работа с MongoExplorer. Форматы обмена данными: JSON и BSON. Перевод модели данных реляционных БД в документно-ориентированный вид
3	Распределенные базы данных	Экспорт баз данных. Формирование хранилища данных из разных источников. Создание пользовательских приложений. Хранимые процедуры, функции.
Промежуточная аттестация - зачет с оценкой		

4 Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации.

Для положительной оценки по результатам освоения дисциплины обучающемуся необходимо выполнить все установленные виды учебной работы. Оценка результатов работы обучающегося в баллах (по видам) приведена в таблицах 7-8.

Таблица 7 - Балльно-рейтинговая оценка результатов учебной работы обучающихся по видам (БРС) в 7 семестре

Учебная работа (виды)	Сумма баллов	Виды и результаты учебной работы	Оценка в аттестации	Баллы
Текущая учебная работа в семестре (Посещение занятий по расписанию и выполнение заданий)	80	Лабораторные работы (отчет о выполнении лабораторной работы) (24 работы).	1,5 балла - выполнение работы на 51-65% 2 балла – выполнение работы на 65,1-85% 3 балла – выполнение работы на 85,1-100%	36 – 72
		Тест	Тест Баллы за тест: 5 баллов (выполнено 51 - 65% заданий) 6,5 баллов (выполнено 66 - 85% заданий) 8 баллов (выполнено 86 - 100% заданий)	5-8
Итого по текущей работе в семестре				41 - 80
Промежуточная аттестация (зачет с оценкой)	20	Тест.	6 балла (пороговое значение) 10 баллов (максимальное значение)	6 - 10
		Решение задачи 1.	2 балла (пороговое значение) 5 баллов (максимальное значение)	2 - 5
		Решение задачи 2.	2 балла (пороговое значение) 5 баллов (максимальное значение)	2 - 5
Итого по промежуточной аттестации (зачету с оценкой)				10 – 20 б.
Суммарная оценка по дисциплине: Сумма баллов текущей и промежуточной аттестации				51 – 100 б.

В промежуточной аттестации оценка выставляется в ведомость в 100-балльной шкале и в буквенном эквиваленте (таблица 8)

Таблица 8 – Соотнесение 100-балльной шкалы и буквенного эквивалента оценки

Сумма набранных	Уровни освоения	Экзамен	Зачет
-----------------	-----------------	---------	-------

баллов	дисциплины и компетенций	Оценка	Буквенный эквивалент	Буквенный эквивалент
86 - 100	Продвинутый	5	отлично	Зачтено
66 - 85	Повышенный	4	хорошо	
51 - 65	Пороговый	3	удовлетворительно	
0 - 50	Первый	2	неудовлетворительно	Не зачтено

5 Материально-техническое, программное и учебно-методическое обеспечение дисциплины.

5.1 Учебная литература

Основная учебная литература

Парфенов, Ю. П. Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов ; под научной редакцией Н. В. Папуловской. — Москва : Издательство Юрайт, 2023. — 121 с. — (Высшее образование). — ISBN 978-5-534-09837-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/514724>.

Дополнительная учебная литература

Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2023. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511121>.

5.2 Материально-техническое и программное обеспечение дисциплины.

Учебные занятия по дисциплине проводятся в учебных аудиториях КГПИ ФГБОУ ВО «КемГУ»:

<p>404 Учебная аудитория для проведения:</p> <ul style="list-style-type: none"> - занятий лекционного типа; - групповых и индивидуальных консультаций; - текущего контроля и промежуточной аттестации. <p>Специализированная (учебная) мебель: доска меловая, кафедра, столы, стулья.</p> <p>Оборудование: <i>переносное</i> - ноутбук, экран, проектор.</p> <p>Используемое программное обеспечение: MSWindows (Microsoft Imagine Premium 3 year по лицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021 г.), LibreOffice (свободно распространяемое ПО), Яндекс.Браузер (отечественное свободно распространяемое ПО).</p> <p>Интернет с обеспечением доступа в ЭИОС.</p>	<p>Учебный корпус №4.</p> <p>654079, Кемеровская область, г. Новокузнецк, пр-кт Metallургов, д. 19</p>
<p>502 Компьютерный класс.</p> <p>Учебная аудитория (мультимедийная) для проведения:</p> <ul style="list-style-type: none"> - занятий лекционного типа; - занятий семинарского (практического) типа; - занятий лабораторного типа; - групповых и индивидуальных консультаций; - самостоятельной работы; - текущего контроля и промежуточной аттестации. <p>Специализированная (учебная) мебель: доска меловая, столы компьютерные, стулья.</p> <p>Оборудование для презентации учебного материала: <i>стационарное</i> - компьютер, экран, проектор, наушники.</p>	<p>Учебный корпус №4.</p> <p>654079, Кемеровская область, г. Новокузнецк, пр-кт Metallургов, д. 19</p>

<p>Лабораторное оборудование: стационарное – компьютеры для обучающихся (16 шт.).</p> <p>Используемое программное обеспечение: MSWindows (Microsoft Imagine Premium 3 year по лицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021 г.), LibreOffice (свободно распространяемое ПО), Firefox 14 (свободно распространяемое ПО), Яндекс.Браузер (отечественное свободно распространяемое ПО), Microsoft Visual Studio (Microsoft Imagine Premium 3 year по лицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021 г.), Среда статистических вычислений Rv.4.0.2 (свободно распространяемое ПО).</p> <p>Интернет с обеспечением доступа в ЭИОС.</p>	
---	--

5.3 Современные профессиональные базы данных и информационные справочные системы.

Перечень СПБД и ИСС по дисциплине

CITForum.ru - on-line библиотека свободно доступных материалов по информационным технологиям на русском языке - <http://citforum.ru>

Научная электронная библиотека eLIBRARY.RU – крупнейший российский информационный портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты - www.elibrary.ru

База данных Science Direct (более 1500 журналов издательства Elsevier, среди них издания по математике и информатике), режим доступа :<https://www.sciencedirect.com>

6 Иные сведения и (или) материалы.

6.1. Примерные вопросы и задания / задачи для промежуточной аттестации

Теоретические вопросы

1. Хранилища данных. Задача консолидации. Хранилища данных. Основные концепции хранилищ данных.
2. OLAP и OLTP системы. Основные характеристики. Сходство и различие.
3. Основные понятия и определения OLAP анализа. Требования, предъявляемые к многомерному анализу данных. Стадии многомерного анализа данных. Сферы применения технологии OLAP-анализа.
4. Методика обнаружения нового знания в хранилищах данных (KDD). Характеристика этапов KDD.
5. Программное обеспечение в области анализа данных.
6. Понятия OLAP-куба, операции, выполняемые над гиперкубом.
7. Архитектура OLAP-систем.
8. Визуальный анализ данных, его преимущества. Процесс визуализации данных (Visual Mining). Характеристики средств визуализации данных.
9. Классификация продуктов OLAP по способу представления данных.
10. Многомерный OLAP (MOLAP).
11. Реляционный OLAP (ROLAP).
12. Гибридный OLAP (HOLAP).
13. Интеллектуальный анализ данных. Интеграция OLAP и ИАД. Критерии оценки существующих продуктов.
14. OLTP системы. Требования, предъявляемые к OLTP системам.
15. Характеристики BigData и хранилищ данных. Преимущества перед реляционными СУБД. Требования к хранилищам данных.
16. Причины и условия применения нереляционных баз данных (баз данных NoSQL). Преимущества и недостатки нереляционных баз данных. Классификация NoSql хранилищ. Их

особенности.

17. Требования ACID. CAP-теорема. BASE архитектура.

18. Типы нереляционных баз данных: ключ-значение, документно-ориентированные, колоночно-ориентированные, графовые. Примеры, возможности и особенности использования нереляционных баз данных.

19. Распределенные базы данных NoSQL. Типы NoSQL. Репликация и шардинг.

20. Пример NoSQL БД: HBase, Cassandra, Neo4j, MongoDB.

Практические задания

1. Составить схему РБД в соответствии с вашим вариантом. Создать БД по созданной схеме в любой СУБД. Заполнить базу данных произвольными данными. Представить созданную на предыдущем шаге БД в нереляционном виде, записанную в JSON.

Примерные варианты:

- Фирма, занимающаяся производством USB-гаджетов.
- Предприятие, производящее аудио-системы.
- Магазин, торгующий ноутбуками.
- Предприятие, содержащее парк самолетов.
- Магазин программного обеспечения.
- Производитель кухонной техники.
- Дистрибьютор мотоциклов.
- Магазин компьютерной периферии.
- Магазин одежды.
- Интернет-магазин наручных часов.

2. Создать базу данных и коллекцию, в которой будут храниться документы. Наполнить коллекцию документами в соответствии с вариантом. Создать не менее 10 документов. Документы в коллекции должны содержать, по крайней мере, один массив, и один вложенный документ. Создать составной и полнотекстовый индексы для документов в коллекции. Создать запрос, производящий замену документа. Создать запрос, производящий обновление любого поля документа и запрос, добавляющий в него новые поля. Создать запрос для получения всех документов коллекции и определить, используется ли для него индекс.

3. Создать несколько запросов для вставки данных в массив. Создать запросы, производящие обновление данных в массиве: как по позиции элемента в массиве, так и по его значению. Создать запросы, удаляющие элементы из массива: по позиции элемента в массиве и по его значению.

Составитель (и): старший преподаватель кафедры МФММ Гаврилова Ю.С.
(фамилия, инициалы и должность преподавателя (ей))