

Документ подписан Министерством науки и высшего образования Российской Федерации  
Информация о владельце:  
ФИО: Макаренко Елена Николаевна  
Должность: Ректор  
Дата подписания: 02.04.2024 14:22:08  
Уникальный программный ключ:  
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

УТВЕРЖДАЮ  
Директор Института магистратуры  
Иванова Е.А.  
«01» июня 2023г.

**Рабочая программа дисциплины  
Обработка многомерных данных**

Направление 09.04.03 Прикладная информатика  
магистерская программа 09.04.03.01 "Информационные системы и технологии в  
бизнесе"

Для набора 2023 года

Квалификация  
магистр

## КАФЕДРА            **Информационных систем и прикладной информатики**

### Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	<b>3 (2.1)</b>		Итого	
Неделя	16 4/6			
Вид занятий	уп	рп	уп	рп
Лекции	14	14	14	14
Практические	28	28	28	28
Итого ауд.	42	42	42	42
Контактная работа	42	42	42	42
Сам. работа	26	26	26	26
Часы на контроль	4	4	4	4
Итого	72	72	72	72

### ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 28.03.2023 протокол № 9.

Программу составил(и): д.э.н, доц., Щербаков С.М.

Зав. кафедрой: д.э.н., доцент Щербаков С.М.

Методическим советом направления: д.э.н., доц., Щербаков С.М.

### 1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	освоить принципы, методы, технологии и инструменты использования больших данных в информационных системах в экономике.
-----	--

### 2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

<b>ПК-2:</b> Способен руководить проектами по созданию систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения со стороны заказчика
<b>ПК-4:</b> Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика
<b>ПК-6:</b> Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации

#### В результате освоения дисциплины обучающийся должен:

<b>Знать:</b>
новые методы и алгоритмы машинного обучения (соотнесено с индикатором ПК-2.1) основы аналитики больших данных (соотнесено с индикатором ПК-4.1) этапы жизненного цикла методологической и технологической инфраструктуры (соотнесено с индикатором ПК-6.1)
<b>Уметь:</b>
руководить разработкой архитектуры комплексных систем искусственного интеллекта со стороны заказчика (соотнесено с индикатором ПК-2.2) создавать комплексные системы на основе аналитики больших данных (соотнесено с индикатором ПК-4.2) управлять получением, хранением, передачей, обработкой больших данных (соотнесено с индикатором ПК-6.2)
<b>Владеть:</b>
навыками руководства созданием комплексных систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения (соотнесено с индикатором ПК-2.3) навыками руководства проектами по построению комплексных систем на основе аналитики больших данных в различных сферах (соотнесено с индикатором ПК-4.3) навыками управления качеством больших данных (соотнесено с индикатором ПК-6.3)

### 3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетенции	Литература
	<b>Раздел 1. Основы построения и использования систем больших данных</b>				
1.1	Тема 1.1 "Основы систем больших данных" Понятие Больших данных. Особенности сбора, хранения, обработки и анализа больших массивов данных. Источники больших данных. Использование больших данных в науке, бизнесе, государственном управлении. /Лек/	3	2	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.2	Тема 1.2 "Методы работы с распределенными информационными системами" Использование фреймворка Map-Reduce в распределенной среде. Реализации Map-Reduce. Состав и возможности программного комплекса. Языки поисковых запросов. /Лек/	3	4	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.3	Тема 1.1 "Основы систем больших данных" Средства построения распределенных информационных систем для BigData. Обзор возможностей. Выполнение заданий с использованием Eclipse, Hadoop, MongoDB, Postgresql, GreenPlum. /Пр/	3	4	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.4	Тема 1.2 "Методы работы с распределенными информационными системами" Изучение и конфигурирование программного комплекса. Размещения набора данных по заданной тематике. Построение поисковых запросов. Выполнение заданий с использованием Eclipse, Hadoop, MongoDB, Postgresql, GreenPlum. /Пр/	3	8	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3

1.5	Методы работы с распределенными информационными системами /Ср/	3	6	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
1.6	Базы данных NoSQL /Ср/	3	10	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
<b>Раздел 2. Разработка и использование приложений на основе распределенных баз данных</b>					
2.1	Тема 2.1 "Базы данных NoSQL" Варианты построения распределенных баз данных, репликация, фрагментация. Согласованность. CAP-теорема. Классы NoSQL баз данных. Примеры СУБД NoSQL. Семейства столбцов. Графовые СУБД. /Лек/	3	4	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.2	Тема 2.2 "Документно-ориентированные распределенные СУБД" Понятие агрегата. Современные документо-ориентированные СУБД. Запросы к СУБД. Использование фреймворка в документо-ориентированных СУБД. /Лек/	3	4	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.3	Тема 2.1 "Базы данных NoSQL" Изучение основных классов NoSQL СУБД, графовых, мультитолоночных, документо-ориентированных, типа "имя=значение". Проектирование и разработка графовой базы данных в СУБД Neo4j на заданную тему. Поисковые запросы. Выполнение заданий с использованием Eclipse, Hadoop, MongoDB, Postgresql, GreenPlum. /Пр/	3	8	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.4	Тема 2.2 "Документно-ориентированные распределенные СУБД" Запросы на выборку и модификацию. Использование драйверов. Настройка фрагментации. Выполнение заданий с использованием Eclipse, Hadoop, MongoDB, Postgresql, GreenPlum. /Пр/	3	8	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.5	Документно-ориентированные распределенные СУБД /Ср/	3	10	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3
2.6	/Зачёт/	3	4	ПК-2 ПК-4 ПК-6	Л1.1 Л1.2Л2.1 Л2.2 Л2.3

#### 4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущей и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

#### 5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

##### 5.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л1.1	Слюсаренко П. И.	Распределенные СУБД: практическое пособие	Москва: Лаборатория книги, 2012	<a href="https://biblioclub.ru/index.php?page=book&amp;id=142013">https://biblioclub.ru/index.php?page=book&amp;id=142013</a> неограниченный доступ для зарегистрированных пользователей
Л1.2	Пол Дейтел, Харви Дейтел	Python: Искусственный интеллект, большие данные и облачные вычисления	Санкт-Петербург: Питер, 2021	<a href="https://ibooks.ru/reading.php?short=1&amp;productid=371701">https://ibooks.ru/reading.php?short=1&amp;productid=371701</a> неограниченный доступ для зарегистрированных пользователей

##### 5.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Колич-во
--	---------------------	----------	-------------------	----------

	Авторы, составители	Заглавие	Издательство, год	Колич-во
Л2.1	Емельянов А. А.	Прикладная информатика: журнал	Москва: Синергия ПРЕСС, 2010	<a href="https://biblioclub.ru/index.php?page=book&amp;id=120321">https://biblioclub.ru/index.php?page=book&amp;id=120321</a> неограниченный доступ для зарегистрированных пользователей
Л2.2		Распределенные базы данных: учебное пособие	Ставрополь: Северо-Кавказский Федеральный университет (СКФУ), 2015	<a href="https://biblioclub.ru/index.php?page=book&amp;id=457594">https://biblioclub.ru/index.php?page=book&amp;id=457594</a> неограниченный доступ для зарегистрированных пользователей
Л2.3	Братченко, Н. Ю.	Распределенные базы данных: лабораторный практикум	Ставрополь: Северо-Кавказский федеральный университет, 2014	<a href="http://www.iprbookshop.ru/63129.html">http://www.iprbookshop.ru/63129.html</a> неограниченный доступ для зарегистрированных пользователей

### 5.3 Профессиональные базы данных и информационные справочные системы

ИСС «КонсультантПлюс»

ИСС «Гарант» <http://www.internet.garant.ru/>

Центральная база статистических данных <https://www.gks.ru/dbscripts/cbsd/dbinet.cgi>

Национальная электронная библиотека (НЭБ) - <https://rusneb.ru/>

### 5.4. Перечень программного обеспечения

Eclipse

Hadoop

MongoDB

Postgresql

GreenPlum

### 5.5. Учебно-методические материалы для студентов с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

## 6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:

- столы, стулья;

- персональный компьютер / ноутбук (переносной);

- проектор, экран / интерактивная доска.

Практические занятия проводятся в компьютерных классах, рабочие места в которых оборудованы необходимыми лицензионными и/или свободно распространяемыми программными средствами и выходом в Интернет.

## 7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ**

**1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания**

1.1. Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
ПК-2: Способен руководить проектами по созданию систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения со стороны заказчика			
З. новые методы и алгоритмы машинного обучения	Основные признаки BigData (3V)	полнота и содержательность ответа умение приводить примеры	ТЗ – тестовые задания (1-12), 3 – вопросы к зачету (1-12)
У. руководить разработкой архитектуры комплексных систем искусственного интеллекта со стороны заказчика	Хранение и обработка больших объемов данных в различных типах СУБД	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (1-2)
В. навыками руководства созданием комплексных систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения	инструментарий обработки и анализа BigData	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (1-2)
ПК-4. Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях со стороны заказчика			
З. основы аналитики больших данных	Виды NoSQL баз данных	полнота и содержательность ответа умение приводить примеры	ТЗ – тестовые задания (1-12), 3 – вопросы к зачету (1-12)
У. создавать комплексные системы на основе аналитики больших данных	Особенности работы в NoSQL СУБД	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (3-4)
В. навыками руководства проектами по построению комплексных систем на основе аналитики больших данных в различных сферах	СУБД MongoDB и Neo4j	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (3-4)
ПК-6. Способен управлять этапами жизненного цикла методологической и технологической инфраструктуры анализа больших данных в организации			
З. этапы жизненного цикла методологической и технологической инфраструктуры	языки доступа к данным	полнота и содержательность ответа умение приводить примеры	ТЗ – тестовые задания (1-12), 3 – вопросы к зачету (1-19)
У. управлять получением, хранением, передачей, обработкой больших данных	Взаимодействие СУБД с программным приложением	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (5-6)
В. навыками управления качеством больших данных	Средства построения распределенных информационных систем	полнота и содержательность ответа умение приводить примеры умение самостоятельно находить решение поставленных задач	ПЗ – практические задания (5-6)

## 1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале.

50-100 баллов (зачет);

0-49 баллов (незачет).

**2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы**

### Вопросы к зачету

- 1) Понятие Больших данных
- 2) Особенности сбора, хранения, обработки и анализа Больших данных
- 3) Требования к распределенным информационным системам
- 4) Средства построения распределенных информационных систем
- 5) Технология Map-Reduce
- 6) Система Apache Hadoop
- 7) Базы данных NoSQL. Особенности, классификация
- 8) Возможности NoSQL-баз данных по обеспечению целостности, доступности скорости обработки информации. CAP-теорема
- 9) Способы репликации и кластеризации баз данных
- 10) Документо-ориентированные базы данных
- 11) Возможности СУБД MongoDB
- 12) Работа с документо-ориентированными БД на языке JSON

***Зачетное задание включает два вопроса – один теоретический вопрос и одно практико-ориентированное задание из числа приведенных ниже практических заданий.***

### *Критерии оценивания:*

- 50-100 баллов («зачет») – изложенный материал фактически верен, наличие глубоких исчерпывающих знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; правильные, уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе, усвоение основной и знакомство с дополнительной литературой; наличие твердых и достаточно полных знаний в объеме пройденной программы дисциплины в соответствии с целями обучения, правильные действия по применению знаний на практике, четкое изложение материала, допускаются отдельные логические и стилистические погрешности, обучающийся усвоил основную литературу, рекомендованную в рабочей программе дисциплины; наличие твердых знаний в объеме пройденного курса в соответствии с целями обучения, изложение ответов с отдельными ошибками, уверенно исправленными после дополнительных вопросов; правильные в целом действия по применению знаний на практике;

- 0-49 баллов («незачет») – ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы.

### Тестовые задания

1. Data Mining — это процесс обнаружения в сырых данных

а) ранее сформулированных гипотез

б) неочевидных закономерностей

в) практических закономерностей

г) объективных закономерностей

д) большого количества закономерностей

2. В методе опорных векторов для классификации используется ...

а) не все множество образцов, а лишь их небольшая часть, которая находится на границах

б) все множество образцов

- в) часть образцов, которая находится вне границ
3. Группа синапсов нейрона – это ...
- а) однонаправленные входные связи, соединенные с выходами других нейронов
- б) выходная связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов
- в) один или несколько нейронов, на входы которых подается один и тот же общий сигнал
4. Сети без обратных связей - это ...
- а) перцептрон
- б) сети Хопфилда (задачи ассоциативной памяти)
- в) сети Кохонена (задачи кластерного анализа)
5. Спорный объект кластеризации — это объект, который по мере сходства ...
- а) может быть отнесен к нескольким кластерам
- б) не может быть отнесен ни к одному кластеру
- в) может быть отнесен более чем к двум кластерам
6. При использовании какого метода необходимо задавать количество кластеров?
- а) метод k-средних
- б) метод ближнего соседа
- в) вся группа иерархических методов
- г) все ответы неверны
7. Набор называют часто встречающимся (frequent), если:
- а) его поддержка выше определенного пользователем минимального значения
- б) его поддержка ниже определенного пользователем максимального значения
- в) его поддержка равна определенному пользователем значению
8. Традиционные методы визуализации могут находить следующее применение:
- а) представлять пользователю информацию в наглядном виде
- б) компактно описывать закономерности, присущие исходному набору данных
- в) снижать размерность или сжимать информацию
- г) упрощать расчеты в модели
- Д) восстановление пробелов в наборе данных
9. Существенными концепциями системы поддержки принятия решений являются:
- а) компьютерная интерактивная
- б) поддержка принятия решений
- в) слабоструктурированных и неструктурированных проблем
- г) структурированных и слабоструктурированных проблем
10. Для какого вида набора данных важно определение наличия сезонной компоненты:
- а) для упорядоченных данных
- б) для неупорядоченных данных
- в) для тех и других
11. Ошибки, которые возникают в процессе использования инструментов очистки (являющиеся двумя крайностями очистки данных) — это:
- а) решение инструментом очистки данных проблемы, которой на самом деле не существует
- б) ошибки, возникающие, когда инструменты очистки полностью упускают существующую проблему
- в) ошибки, возникающие, когда инструменты очистки не могут обнаружить существующую проблему
12. Данные представляют собой:
- а) факты и графики
- б) текст
- в) картинки, звуки, аналоговые или цифровые видео-сегменты
- г) все вместе

*Критерии оценивания:*

Из имеющегося банка тестов формируется тестовое задание, содержащее 10 тестовых вопросов для одного обучающегося. Каждый тестовый вопрос содержит 3-4 варианта ответов, один или несколько из которых – верные.

Правильный ответ на один тестовый вопрос – 1 балл, неправильный – 0 баллов.

**Максимальное количество баллов за тестовые задания – 10.**



## Практические задания

### Практическое задание №1

Средства построения распределенных информационных систем для BigData. Обзор возможностей

### Практическое задание №2

Изучение и конфигурирование программного комплекса Apache Hadoop. Размещения набора данных по заданной тематике. Построение поисковых запросов на языке Pig

### Практическое задание №3

Изучение основных классов NoSQL СУБД, графовых, мультиколоночных, документо-ориентированных, типа "имя=значение". Проектирование и разработка графовой базы данных в СУБД Neo4j на заданную тему. Поисковые запросы на языке Cypher

### Практическое задание №4

Работа с MongoDB. Запросы на выборку и модификацию. Использование драйверов. Настройка фрагментации. Использование Map-Reduce

### Практическое задание №5

Изучение основных классов NoSQL СУБД, графовых, мультиколоночных, документо-ориентированных, типа "имя=значение". Проектирование и разработка графовой базы данных в СУБД Neo4j на заданную тему. Поисковые запросы на языке Cypher

### Практическое задание №6

Работа с MongoDB. Запросы на выборку и модификацию. Использование драйверов. Настройка фрагментации. Использование Map-Reduce

*Критерии оценивания (для каждого задания):*

11-15 б. – задание выполнено верно;

7-10 б.– при выполнении задания были допущены неточности, не влияющие на результат;

4-6 б. – при выполнении задания были допущены ошибки;

1-3 б. – при выполнении задания были допущены существенные ошибки.

**Максимальное количество баллов за все практические задания – 90 (6 заданий по 15 баллов).**

## 3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

**Текущий контроль** успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

**Промежуточная аттестация** проводится в форме зачета.

Зачет проводится по расписанию промежуточной аттестации. Количество вопросов в зачетном задании – 2 (один теоретический вопрос и одно практико-ориентированное задание). Объявление результатов производится в день зачета. Результаты аттестации заносятся в зачетную ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать задолженность в установленном порядке.

## МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены следующие виды занятий:

- лекции;
- практические занятия.

В ходе лекционных занятий рассматриваются основные теоретические вопросы, даются рекомендации для самостоятельной работы и подготовке к практическим занятиям.

В ходе практических занятий углубляются и закрепляются знания студентов по ряду рассмотренных на лекциях вопросов, развиваются навыки практической работы.

При подготовке к практическим занятиям каждый студент должен:

- изучить рекомендованную учебную литературу;
- изучить конспекты лекций;
- подготовить ответы на все вопросы по изучаемой теме.

В процессе подготовки к практическим занятиям студенты могут воспользоваться консультациями преподавателя.

Вопросы, не рассмотренные на лекциях и практических занятиях, должны быть изучены студентами в ходе самостоятельной работы. Контроль самостоятельной работы студентов над учебной программой курса осуществляется в ходе занятий методом выполнения тестовых и практических заданий. В ходе самостоятельной работы каждый студент обязан прочитать основную и по возможности дополнительную литературу по изучаемой теме, дополнить конспекты лекций недостающим материалом, выписками из рекомендованных первоисточников, выделить непонятные термины, найти их значение в энциклопедических словарях.

Студент должен готовиться к предстоящему практическому занятию по всем обозначенным в рабочей программе дисциплины вопросам.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами. Также обучающиеся могут взять на дом необходимую литературу на абонементе университетской библиотеки или воспользоваться читальными залами.