


**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Воронежский государственный технический университет»

УТВЕРЖДАЮ
Декан факультета информационных техноло-
гий и компьютерной безопасности

 /П.Ю. Гусев/
31.08.2021 г.

РАБОЧАЯ ПРОГРАММА
дисциплины (модуля)
«Технологии обработки больших данных»

**Направление подготовки (специальность) 09.03.02 Информационные си-
стемы и технологии**

**Профиль (специализация) Системы автоматизации проектирования и разра-
ботки информационных систем**

Квалификация выпускника бакалавр

Нормативный период обучения 4 года

Форма обучения Очная

Год начала подготовки 2021 г.

Автор(ы) программы


_____ *подпись*

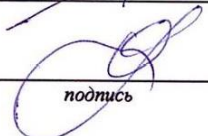
Ю.В. Минаева

**Заведующий кафедрой Системы
автоматизированного проектирования
и информационные системы**



Я.Е. Львович

Руководитель ОПОП


_____ *подпись*

О.Г. Яскевич

Воронеж 2021

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цель дисциплины

Целью дисциплины является освоение принципов, методов, технологий и инструментов обработки и использования больших данных в информационных системах

1.2. Задачи освоения дисциплины

- ознакомление студентов с основными классами задач представления и обработки больших данных;
- изучение методов и технологий подготовки, хранения, обработки и анализа больших данных;
- освоение методов аналитической обработки больших объёмов данных в информационных системах;
- приобретение навыков использования технологий и инструментов Big Data при решении практических задач.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Технологии обработки больших данных» относится к дисциплинам вариативной части блока Б1.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Технологии обработки больших данных» направлен на формирование следующих компетенций:

ПК-1 - Способен выполнять синтез требований к программному продукту и декомпозицию программного средства на компоненты

ПК-5 - Способен определять внешние и внутренние интерфейсы каждого из компонент информационной системы и осуществлять их разработку

Компетенция	Результаты обучения, характеризующие сформированность компетенции
ПК-1	Знать методы обработки и анализа больших данных
	Уметь использовать технологии обработки больших данных при проектировании и эксплуатации информационных систем
	Владеть инструментами Big Data
ПК-5	Знать технологии распределённой обработки данных
	Уметь разрабатывать компоненты программных систем для обработки и анализа больших данных
	Владеть навыками практического применения технологий Big Data в различных прикладных областях

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Технологии обработки больших данных» составляет 2 з.е.

Распределение трудоемкости дисциплины по видам занятий

очная форма обучения

Виды учебной работы	Всего часов	Семестры
		6
Аудиторные занятия (всего)	54	54
В том числе:		
Лекции	18	18
Лабораторные работы (ЛР)	36	36
Самостоятельная работа	18	18
Часы на контроль		
Виды промежуточной аттестации – Зачет	+	+
Общая трудоемкость: академические часы	72	72
зач.ед.	2	2

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий

очная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего, час
1	Введение в Big Data	Понятие больших данных. Особенности сбора, хранения, обработки и анализа больших массивов данных. Источники больших данных. Использование больших данных в информационных системах	2	2	2	6
2	Модели и методы анализа больших данных	Основные задачи анализа больших данных в информационных системах. Классификация современных методов анализа данных. Особенности технологии Data Mining. Использование методов Data Mining для анализа больших данных. Решение задач классификации, кластеризации, прогнозирования на основе больших данных. Нейросетевые технологии обработки информации. Технологии визуализации и трансформации данных. Многомерный анализ данных. OLAP-технологии.	6	8	4	18
3	Базы данных NoSQL	Варианты построения распределенных баз данных, репликация, фрагментация. Согласованность. CAP-теорема. Классы NoSQL баз данных. Семейства столбцов. Графовые СУБД. Документно-ориентированные распределённые СУБД. Понятие агрегата. Современные документно-ориентированные СУБД. Запросы к СУБД на языке JSON. Примеры СУБД NoSQL. Организация и принципы функционирования MongoDB.	4	8	4	16
4	Технологии распределённой обработки данных	Базовые архитектуры распределённой обработки данных. Параллельные и облачные вычисления. Использование фреймворка MapReduce в распределенной среде. Реализации MapReduce. Состав и возможности платформы распределённых вычислений Hadoop . Языки поисковых запросов для Hadoop. Распределённые файловые системы. Ар-	6	18	8	32

		хитектура и реализация распределённых файловых систем. HDFS. Распределенные базы данных. HBASE. Обзор современных инструментов BigData.				
Итого			18	36	18	72

5.2 Перечень лабораторных работ

1. Решение задач нейросетевого моделирования и прогнозирования в Matlab
2. Нейросетевая кластеризация и классификация данных в среде Matlab
3. Машинное обучение на больших данных.
4. Распределенные файловые системы;
5. Модель вычислений MapReduce;
6. Beyond MapReduce. Spark;
7. Key-value хранилища в больших данных.
8. Изучение возможностей и работа с СУБД MongoDB
9. Работа с документо-ориентированными БД на языке JSON.

6. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ) И КОНТРОЛЬНЫХ РАБОТ

В соответствии с учебным планом освоение дисциплины не предусматривает выполнение курсового проекта (работы) или контрольной работы.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

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

7.1.1 Этап текущего контроля

Результаты текущего контроля знаний и межсессионной аттестации оцениваются по следующей системе:

«аттестован»;

«не аттестован».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Аттестован	Не аттестован
ПК-1	Знать методы обработки и анализа больших данных	Знание особенностей технологии Big Data. Знание моделей и методов Data Mining, классов решаемых задач. Ответы на теоретические вопросы при защите лабораторных работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Уметь использовать технологии обработки больших данных при проектировании и эксплуатации информационных систем	Умение решать задачи классификации, распознавания и прогнозирования с использованием интеллектуальных методов анализа и обработки больших данных. Выполнение лабораторных работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	Владеть инструментами Big Data	Владение навыками использования инструментальных средств Big Data при решении практических задач. Выполнение лабораторных работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
ПК-5	Знать технологии	Знание методов и средств	Выполнение работ в	Невыполнение

распределённой обработки данных	распределённой обработки информации. Ответы на теоретические вопросы при защите лабораторных работ	срок, предусмотренный в рабочих программах	работ в срок, предусмотренный в рабочих программах
Уметь разрабатывать компоненты программных систем для обработки и анализа больших данных	Умение разрабатывать программное обеспечение для решения задач обработки больших данных. Выполнение лабораторных работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
Владеть навыками практического применения технологий Big Data в различных прикладных областях	Решение прикладных задач в конкретной предметной области, выполнение лабораторных работ	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах

7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 7 семестре для очной формы обучения по двухбалльной системе:

«зачтено»

«не зачтено»

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Зачтено	Не зачтено
ПК-1	Знать методы обработки и анализа больших данных	Тест	Выполнение теста на 70-100%	Выполнение менее 70%
	Уметь использовать технологии обработки больших данных при проектировании и эксплуатации информационных систем	Решение стандартных практических задач	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	Владеть инструментами Big Data	Решение прикладных задач в конкретной предметной области	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
ПК-5	Знать технологии распределённой обработки данных	Тест	Выполнение теста на 70-100%	Выполнение менее 70%
	Уметь разрабатывать компоненты программных систем для обработки и анализа больших данных	Решение стандартных практических задач	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	Владеть навыками практического применения технологий Big Data в различных прикладных областях	Решение прикладных задач в конкретной предметной области	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены

7.2 Примерный перечень оценочных средств (типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности)

7.2.1 Примерный перечень заданий для подготовки к тестированию

1. Снижение размерности - это

- уменьшение числа объектов
- уменьшение числа измерений
- **уменьшение числа признаков**
- уменьшение числа показателей

2. Кластерный анализ предназначен для:

- **группировки объектов**
- ранжирования объектов
- группировки показателей
- ранжирования показателей

3. Основное назначение систем интеллектуального анализа (Data Mining):

- **обнаружение в сырых данных скрытых знаний;**
- проведение статистического анализа;
- решения задач математического программирования;
- поиск агрегированных данных;

4. При проведении интеллектуального анализа из существующих данных извлекают:

- **шаблоны и тренды;**
- функциональные зависимости;
- свойства фактов;
- атрибуты измерений.

5. Перечислите правильную последовательность этапов Knowledge Discovery in Databases – процесса обнаружения знаний в базах данных:

- трансформация, интерпретация результатов, выборка, очистка, построение моделей.

- **построение моделей, выборка, очистка, трансформация, интерпретация результатов.**

- построение моделей, выборка, очистка, трансформация, интерпретация результатов,

- выборка, очистка, трансформация, построение моделей, интерпретация результатов.

6. OLAP-системы это:

- **информационные системы оперативной транзакционной обработки данных.**

- информационные системы оперативного анализа данных.

- информационные системы автоматической обработки данных.

- информационные системы алгоритмической обработки данных.

7. Что входит в состав хранилища данных:

- **организационная структура, технические средства, базы или совокупности баз данных и программное обеспечение.**

- базы или совокупности баз данных и программное обеспечение.

- источники данных и программное обеспечение.

- организационная структура и программное обеспечение

8. Что является основными составляющими структуры хранилищ дан-

НЫХ:

- **таблица исходной информации и таблица запросов.**
- таблица базы данных и запросы.
- таблица фактов и таблица измерений.
- таблица запросов и таблица данных.

10. На основе чего реализуется концептуальная многомерная модель данных:

- **на основе представления данных в виде многомерного пространства, размерность которого определяется количеством измерений.**
- на основе представления данных в виде многомерного пространства, размерность которого определяется количеством граней куба.
- на основе представления данных в виде бесконечного пространства.
- на основе представления данных в виде пространства, ограниченного многомерным кубом.

7.2.4 Примерный перечень вопросов для подготовки к зачету

1. Основные задачи обработки больших данных в информационных системах.
2. Классификация современных методов анализа данных. Понятие о технологии Data Mining.
3. Понятие искусственной нейронной сети. Задачи, решаемые с использованием искусственных нейронных сетей.
4. Классификация искусственных нейронных сетей. Однослойные и многослойные сети.
5. Подходы к обучению искусственных нейронных сетей. Обучение с учителем. Обучение без учителя.
6. Решение задач прогнозирования на основе нейросетевых технологий.
7. Решение задач кластеризации на основе нейросетевых технологий.
8. Решение задач классификации и распознавания на основе нейросетевых технологий.
9. OLAP-технологии.
10. Требования к распределенным информационным системам
11. Средства построения распределенных информационных систем
12. Технология Map-Reduce
- 13 Система Apache Hadoop
14. Базы данных NoSQL. Особенности, классификация
- 15 Возможности NoSQL-баз данных по обеспечению целостности, доступности скорости обработки информации. CAP-теорема.
- 16 Способы репликации и кластеризации баз данных
- 17 Документо-ориентированные базы данных
- 18 Возможности СУБД MongoDB
- 19 Работа с документо-ориентированными БД на языке JSON.
20. Современные инструменты BigData

7.2.5 Примерный перечень заданий для подготовки к экзамену

Не предусмотрено учебным планом

7.2.6. Методика выставления оценки при проведении промежуточной аттестации

Зачет проводится по билетам, каждый из которых содержит 2 теоретических вопроса и задачу. Каждый правильный ответ на вопрос в билете оценивается 5 баллами, задача оценивается в 10 баллов (5 баллов за верное решение и 5 баллов за верный ответ). Максимальное количество набранных баллов – 20. В процессе ответа преподаватель задаёт студенту дополнительные вопросы по тематике излагаемого материала.

Критерии оценивания ответов на теоретические вопросы:

5 баллов – студент даёт полный развёрнутый ответ на вопрос, демонстрирует уверенное владение материалом, излагает ответ логично и последовательно, обосновывает все выводы и положения, теоретический материал иллюстрирует примерами, отвечает на дополнительные вопросы преподавателя, свободно владеет терминологией дисциплины.

4 балла – студент даёт достаточно полный ответ на вопрос, показывает умение пользоваться основными понятиями и определениями, отвечает на дополнительные вопросы, однако имеются незначительные неточности в ответах и затруднения в иллюстрации теоретического материала практическими примерами;

3 балла – студент демонстрирует понимание сути вопроса, отвечает на вопрос по существу, правильно применяет теоретические положения, однако ответ представлен недостаточно полно, выводы и положения недостаточно обоснованы, имеются затруднения в ответах на дополнительные вопросы;

2 балла – студент имеет обобщённые знания основных разделов изучаемой дисциплины и содержания лекционных курсов, но не усвоил его деталей. Ответ представлен на обобщённом уровне, содержит существенные неточности, недостаточно правильные формулировки, нарушения логичности и последовательности изложения материала. Отсутствуют ответы более чем на 50% дополнительных вопросов.

1 балл – студент демонстрирует поверхностные знания материала, ответ представлен неполно (на уровне основных понятий) и содержит существенные ошибки, отсутствует логичность и последовательность в изложении, студент не отвечает на большинство дополнительных вопросов.

0 баллов – ответ на вопрос отсутствует или является полностью неверным.

Критерии оценивания результатов экзамена:

1. Оценка «Незачет» ставится в случае, если студент набрал менее 12 баллов.

2. Оценка «Зачтено» ставится в случае, если студент набрал от 12 до 20 баллов

7.2.7 Паспорт оценочных материалов

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Введение в Big Data	ПК-1, ПК-5	Тест, защита лабораторных работ, вопросы к зачету
2	Модели и методы анализа больших данных	ПК-1, ПК-5	Тест, защита лабораторных работ, вопросы к зачету
3	Базы данных NoSQL	ПК-1, ПК-5	Тест, защита лабораторных работ, вопросы к зачету
4	Технологии распределённой обработки данных	ПК-1, ПК-5	Тест, защита лабораторных работ, вопросы к зачету

7.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Тестирование осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных тест-заданий на бумажном носителе. Время тестирования 30 мин. Затем осуществляется проверка теста экзаменатором и выставляется оценка согласно методики выставления оценки при проведении промежуточной аттестации.

Решение стандартных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

Решение прикладных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ)

8.1 Перечень учебной литературы, необходимой для освоения дисциплины

1. Питолин, А.В. Искусственные нейронные сети. Теория и практика : Учеб. пособие. - Воронеж : ВГТУ, 2007. – 129 с.
2. Пальмов С.В. Интеллектуальный анализ данных [Электронный ресурс]: учеб. пособие. – Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. – 127 с. – Режим доступа <http://www.iprbookshop.ru/75376>
3. Минаева Ю.В. Методы статистического и интеллектуального анализа данных: учеб. пособие. – Воронеж: ВГТУ, 2017. 90 с.
4. Чеканов В.С., Кандаурова Н.В. Технологии обработки информации: учеб. пособие. – Ставрополь: Северо-Кавказский Федеральный университет, 2014. – 175 с. <http://www.iprbookshop.ru>
5. Новикова Н.М. Обработка экспериментальных данных: Учеб. пособие. – Воронеж: ВГТУ, 2010. – 119 с.

8.2 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения, ресурсов информационно-телекоммуникационной сети «Интернет», современных профессиональных баз данных и информационных справочных систем:

Matlab

Hadoop

MongoDB

<http://bigor.bmstu.ru/>

<http://www.exponenta.ru>

<http://www.intuit.ru>

<http://ibooks.ru/>

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Специализированная лекционная аудитория, оснащенная оборудованием для лекционных демонстраций и проекционной аппаратурой
--

Дисплейный класс, оснащенный компьютерными программами для проведения лабораторного практикума
--

10. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

По дисциплине «Технологии обработки больших данных» читаются лекции, проводятся лабораторные работы.

Основой изучения дисциплины являются лекции, на которых излагаются наиболее существенные и трудные вопросы, а также вопросы, не нашедшие отражения в учебной литературе.

Лабораторные работы выполняются на лабораторном оборудовании в соответствии с методиками, приведенными в указаниях к выполнению работ.

Вид учебных занятий	Деятельность студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; помечать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначение вопросов, терминов, материала, которые вызывают трудности, поиск ответов в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на лекции или на практическом занятии.
Лабораторная работа	Лабораторные работы позволяют научиться применять теоретические знания, полученные на лекции при решении конкретных задач. Чтобы наиболее рационально и полно использовать все возможности лабораторных для подготовки к ним необходимо: следует разобрать лекцию по соответствующей теме, ознакомиться с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и выполнить другие письменные задания.

Самостоятельная работа	Самостоятельная работа студентов способствует глубокому усвоения учебного материала и развитию навыков самообразования. Самостоятельная работа предполагает следующие составляющие: - работа с текстами: учебниками, справочниками, дополнительной литературой, а также проработка конспектов лекций; - выполнение домашних заданий и расчетов; - работа над темами для самостоятельного изучения; - участие в работе студенческих научных конференций, олимпиад; - подготовка к промежуточной аттестации.
Подготовка к промежуточной аттестации	Готовиться к промежуточной аттестации следует систематически, в течение всего семестра. Интенсивная подготовка должна начаться не позднее, чем за месяц-полтора до промежуточной аттестации. Данные перед экзаменом три дня эффективнее всего использовать для повторения и систематизации материала.

11 Лист регистрации изменений

№ п/п	Перечень вносимых изменений	Дата внесения изменений	Подпись заведующего кафедрой, ответственной за реализацию ОПОП
1	Актуализирован раздел 8.1 Перечень учебной литературы, необходимой для освоения дисциплины	31.08.2020	
2	Актуализирован раздел 8.2 в части состава используемого лицензионного программного обеспечения, современных профессиональных баз данных и справочных информационных систем	31.08.2020	
3	Актуализирован раздел 8.1 Перечень учебной литературы, необходимой для освоения дисциплины	31.08.2021	