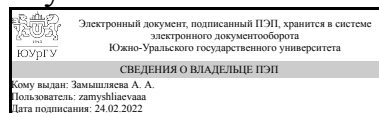


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Директор института
Институт естественных и точных
наук



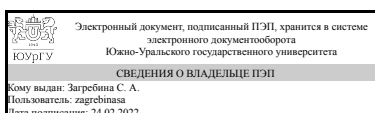
А. А. Замышляева

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.П1.19.02 Анализ и обработка больших массивов данных для направления 02.03.01 Математика и компьютерные науки
уровень Бакалавриат
профиль подготовки Компьютерное моделирование в инженерном и технологическом проектировании
форма обучения очная
кафедра-разработчик Математическое и компьютерное моделирование

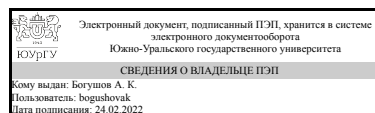
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 02.03.01 Математика и компьютерные науки, утверждённым приказом Минобрнауки от 23.08.2017 № 807

Зав.кафедрой разработчика,
д.физ.-мат.н., проф.



С. А. Загребина

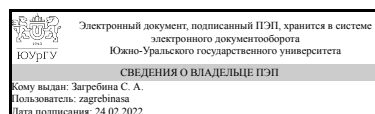
Разработчик программы,
старший преподаватель



А. К. Богушов

СОГЛАСОВАНО

Руководитель образовательной
программы
д.физ.-мат.н., проф.



С. А. Загребина

1. Цели и задачи дисциплины

Цель дисциплины: формирование у студентов знаний, умений и навыков в области разработки и использования систем обработки и анализа больших массивов данных: постановки задачи анализа данных, предварительной обработки данных, визуализации данных, разработки, реализации и применения методов интеллектуального анализа данных к большим массивам данных, представления результатов работы. - изучить задачи классификации и кластеризации больших объемов данных; - изучить математическое моделирование процессов и объектов на базе стандартных пакетов автоматизированного проектирования и исследований; - сформировать навыки работы с большими массивами данных; - изучить языки программирования, удобные для работы с большими объемами данных.

Краткое содержание дисциплины

Введение в большие данные. Высокопроизводительная обработка данных. Алгоритм Map Reduce. Поиск похожих объектов. Анализ потоков данных. Алгоритм PageRank. Поиск устойчивых групп событий. Алгоритм Argiori. Кластеризация. Рекомендательные системы. Анализ графов. Понижение размерности. Машинное обучение на больших данных.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-2 Способен демонстрировать базовые знания математических и естественных наук, основ программирования и информационных технологий	Знает: основные элементы процесса анализа больших данных, основные подходы к обработке больших массивов данных Умеет: визуализировать имеющиеся данные, отбрасывать несущественную информацию, структурировать информацию в рамках поставленной задачи Имеет практический опыт: использования современных высоконагруженных систем хранения и обработки больших данных

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Методы и средства научной визуализации, Теория оптимизации, Основы компьютерного моделирования, Современные технологии разработки программного обеспечения, Практикум по основам компьютерного моделирования, Вычислительная математика, Практикум по интерактивным графическим системам, Программирование на языке Java,	Параллельные и распределенные вычисления, Применение системы ANSYS к решению инженерных задач, Применение системы ANSYS к моделированию физических процессов, Производственная практика, преддипломная практика (8 семестр)

Офисные приложения и технологии, Производственная практика, научно-исследовательская работа (6 семестр)	
--	--

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Дисциплина	Требования
Основы компьютерного моделирования	Знает: основные понятия и методы компьютерного моделирования динамических систем Умеет: применять методы компьютерного моделирования динамических систем Имеет практический опыт: реализации моделирующих алгоритмов для исследования характеристик и поведения динамических систем.
Практикум по основам компьютерного моделирования	Знает: Умеет: моделировать компьютерные изображения в пакете Math Works-MATLAB Имеет практический опыт: использовать средства моделирования компьютерных изображений в пакете Math Works-MATLAB
Практикум по интерактивным графическим системам	Знает: Умеет: применять интерактивную графику в информационных системах Имеет практический опыт: работы с инструментальными средствами компьютерной графики
Теория оптимизации	Знает: основы построения оптимизационных задач и алгоритмы их решения Умеет: использовать методы оптимизации в математическом моделировании Имеет практический опыт: навыками решения практических задач с использованием базовых методов математических и естественных наук, программирования и информационных технологий
Вычислительная математика	Знает: существующие стандартные пакеты прикладных программ Умеет: применить соответствующую процессу математическую модель и проверить ее адекватность, провести анализ результатов моделирования, принять решение на основе полученных результатов Имеет практический опыт: использования методов математических и естественных наук, программирования и информационных технологий
Программирование на языке Java	Знает: синтаксис, базовые классы библиотеки языка Java, возможности языка и области применения Java –приложений; основные пакеты и классы языка Java Умеет: создавать классы на языке Java для решения типовых задач по принципам объектно-ориентированного программирования, разрабатывать приложения с графическим интерфейсом Имеет практический опыт: применения инструментальных средств

	для разработки приложений, библиотек и пакетов программ на языке программирования Java в научной и практической деятельности, использования базовых методов математических и естественных наук, программирования и информационных технологий для разработки приложений
Офисные приложения и технологии	Знает: основные методы использования информационных технологий Умеет: работать с современными информационными технологиями Имеет практический опыт: использования современных информационных технологий
Современные технологии разработки программного обеспечения	Знает: основные технологии разработки программного обеспечения, современные технологии и методы программирования Умеет: работать с основными технологиями разработки программного обеспечения, формировать требования, спецификацию и структуру программы при решении прикладных задач, оценивать результаты тестирования, локализовать ошибки в коде Имеет практический опыт: использования основных технологий разработки программного обеспечения, использования современных CASE-средств, применяемых при проектировании, тестировании и командной разработке
Методы и средства научной визуализации	Знает: базовые принципы визуализации, особенности постановок задач, возникающих в разных предметных областях Умеет: Имеет практический опыт: применения современных средств визуализации для решения ряда актуальных прикладных задач
Производственная практика, научно-исследовательская работа (6 семестр)	Знает: Умеет: анализировать и систематизировать полученную информацию, выбирать приёмы и методы её обработки, прогнозировать и принимать обоснованные социально-экономические решения; грамотно планировать распределение финансов в различных областях жизнедеятельности Имеет практический опыт: применения методов проектирования и производства программного продукта, принципов построения, структуры и приемов работы с инструментальными средствами, поддерживающими создание программного продукта, использования базовых методов математических и естественных наук, программирования и информационных технологий, применения основных методов обработки информации для решения практических задач, самостоятельного принятия обоснованных экономических решений в профессионально деятельности, использования необходимой информации из текстов профессиональной направленности, определения и решения круга задач в рамках поставленной цели

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 4 з.е., 144 ч., 73,25 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		7	
Общая трудоёмкость дисциплины	144	144	
<i>Аудиторные занятия:</i>	64	64	
Лекции (Л)	32	32	
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	0	0	
Лабораторные работы (ЛР)	32	32	
<i>Самостоятельная работа (СРС)</i>	70,75	70,75	
с применением дистанционных образовательных технологий	0		
Выполнение курсовой работы	30	30	
Подготовка к промежуточной аттестации	16	16	
Подготовка к лабораторным работам, заполнение отчетов	24,75	24.75	
Консультации и промежуточная аттестация	9,25	9,25	
Вид контроля (зачет, диф.зачет, экзамен)	-	диф.зачет,КР	

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Введение в большие данные. Высокопроизводительная обработка данных. Алгоритм Map Reduce	8	8	0	0
2	Поиск похожих объектов	6	2	0	4
3	Анализ потоков данных	6	2	0	4
4	Алгоритм PageRank	8	4	0	4
5	Поиск устойчивых групп событий. Алгоритм Apriori	6	2	0	4
6	Кластеризация	6	2	0	4
7	Рекомендательные системы	8	4	0	4
8	Анализ графов	6	2	0	4
9	Понижение размерности	2	2	0	0
10	Машинное обучение на больших данных	8	4	0	4

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Введение в большие данные	2
2	1	Высокопроизводительная обработка данных. Распределенные файловые	2

		системы. Особенности построения вычислительного кластера	
3	1	Алгоритм Map Reduce. Теория сложности MapReduce. Алгоритмы с использованием Map Reduce	4
4	2	Поиск похожих объектов. Поиск ближайшего соседа. Метрики.	2
5	3	Анализ потоков данных. Выборка данных из потока. Фильтрация потоков. Подсчет различных элементов в потоке.	2
6	4	PageRank. Эффективное вычисление PageRank.	4
7	5	Поиск устойчивых групп событий. Алгоритм Apriori. Модель корзины покупок	2
8	6	Кластеризация. Иерархическая кластеризация. Алгоритм k средних	2
9	7	Рекомендательные системы. Рекомендации на основе фильтрации содержимого. Коллаборативная фильтрация	4
10	8	Анализ графов. Кластеризация графа социальной сети. Прямое нахождение сообществ. Нахождение пересекающихся сообществ.	2
11	9	Понижение размерности. Собственные значения и собственные векторы. Метод главных компонент.	2
12	10	Машинное обучение на больших данных. Нейронные сети. Метод опорных векторов.	4

5.2. Практические занятия, семинары

Не предусмотрены

5.3. Лабораторные работы

№ занятия	№ раздела	Наименование или краткое содержание лабораторной работы	Кол-во часов
1	2	Поиск похожих объектов. Хэширование с учетом близости.	4
2	3	Обработка потоков данных. Выборка данных из потока. Фильтрация потоков. Подсчет различных элементов в потоке.	4
3	4	Алгоритм PageRank	4
4	5	Поиск устойчивых групп событий. Алгоритм Apriori	4
5	6	Кластеризация. Алгоритм k-средних	4
6	7	Рекомендательные системы. Коллаборативная фильтрация	4
7	8	Анализ графов. Кластеризация графа социальной сети	4
8	10	Обучение нейронных сетей на больших данных	4

5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Выполнение курсовой работы	ЭУМД [1] - [4]	7	30
Подготовка к промежуточной аттестации	ЭУМД [3], главы 1-12	7	16
Подготовка к лабораторным работам, заполнение отчетов	ЭУМД [3], главы 1-12	7	24,75

6. Текущий контроль успеваемости, промежуточная аттестация

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учи-тыва-ется в ПА
1	7	Текущий контроль	Лабораторная работа: Поиск похожих объектов	1	5	Оценка суммируется из следующих оценок: 1) задание выполнено вовремя - 2 балла; оценка снижается на 1 балл за превышение сроков сдачи задания по неуважительной причине более, чем на 2 недели, или на 2 балла - более 2 недель; 2) выполнены все шаги лабораторной - 2 балла; 1-2 шага не выполнены или сделаны некорректно - 1 балл; больше 2 шагов не выполнены - 0 баллов; 3) ответы на контрольные вопросы удовлетворительны - 1 балл	дифференцированный зачет
2	7	Текущий контроль	Лабораторная работа: Обработка потоков данных	1	5	Оценка суммируется из следующих оценок: 1) задание выполнено вовремя - 2 балла; оценка снижается на 1 балл за превышение сроков сдачи задания по неуважительной причине более, чем на 2 недели, или на 2 балла - более 2 недель; 2) выполнены все шаги лабораторной - 2 балла; 1-2 шага не выполнены или сделаны некорректно - 1 балл; больше 2 шагов не выполнены - 0 баллов; 3) ответы на контрольные вопросы удовлетворительны - 1 балл	дифференцированный зачет

3	7	Текущий контроль	Лабораторная работа: Алгоритм PageRank	1	5	Оценка суммируется из следующих оценок: 1) задание выполнено вовремя - 2 балла; оценка снижается на 1 балл за превышение сроков сдачи задания по неуважительной причине более, чем на 2 недели, или на 2 балла - более 2 недель; 2) выполнены все шаги лабораторной - 2 балла; 1-2 шага не выполнены или сделаны некорректно - 1 балл; больше 2 шагов не выполнены - 0 баллов; 3) ответы на контрольные вопросы удовлетворительны - 1 балл	дифференцированный зачет
4	7	Текущий контроль	Лабораторная работа: Поиск устойчивых групп событий	1	5	Оценка суммируется из следующих оценок: 1) задание выполнено вовремя - 2 балла; оценка снижается на 1 балл за превышение сроков сдачи задания по неуважительной причине более, чем на 2 недели, или на 2 балла - более 2 недель; 2) выполнены все шаги лабораторной - 2 балла; 1-2 шага не выполнены или сделаны некорректно - 1 балл; больше 2 шагов не выполнены - 0 баллов; 3) ответы на контрольные вопросы удовлетворительны - 1 балл	дифференцированный зачет
5	7	Текущий контроль	Лабораторная работа: Кластеризация	1	5	Оценка суммируется из следующих оценок: 1) задание выполнено вовремя - 2 балла; оценка снижается на 1 балл за превышение сроков сдачи задания по неуважительной причине более, чем на 2 недели, или на 2	дифференцированный зачет

						балла - более 2 недель; 2) выполнены все шаги лабораторной - 2 балла; 1-2 шага не выполнены или сделаны некорректно - 1 балл; больше 2 шагов не выполнены - 0 баллов; 3) ответы на контрольные вопросы удовлетворительны - 1 балл	
6	7	Текущий контроль	Лабораторная работа: Рекомендательные системы	1	5	Оценка суммируется из следующих оценок: 1) задание выполнено вовремя - 2 балла; оценка снижается на 1 балл за превышение сроков сдачи задания по неуважительной причине более, чем на 2 недели, или на 2 балла - более 2 недель; 2) выполнены все шаги лабораторной - 2 балла; 1-2 шага не выполнены или сделаны некорректно - 1 балл; больше 2 шагов не выполнены - 0 баллов; 3) ответы на контрольные вопросы удовлетворительны - 1 балл	дифференцированный зачет
7	7	Текущий контроль	Лабораторная работа: Анализ графов	1	5	Оценка суммируется из следующих оценок: 1) задание выполнено вовремя - 2 балла; оценка снижается на 1 балл за превышение сроков сдачи задания по неуважительной причине более, чем на 2 недели, или на 2 балла - более 2 недель; 2) выполнены все шаги лабораторной - 2 балла; 1-2 шага не выполнены или сделаны некорректно - 1 балл; больше 2 шагов не выполнены - 0 баллов; 3) ответы на	дифференцированный зачет

						контрольные вопросы удовлетворительны - 1 балл	
8	7	Текущий контроль	Лабораторная работа: нейронные сети	1	5	Оценка суммируется из следующих оценок: 1) задание выполнено вовремя - 2 балла; оценка снижается на 1 балл за превышение сроков сдачи задания по неуважительной причине более, чем на 2 недели, или на 2 балла - более 2 недель; 2) выполнены все шаги лабораторной - 2 балла; 1-2 шага не выполнены или сделаны некорректно - 1 балл; больше 2 шагов не выполнены - 0 баллов; 3) ответы на контрольные вопросы удовлетворительны - 1 балл	дифференцированный зачет
9	7	Проме- жуточная аттестация	Зачет	-	10	Студенту задаются 5 вопросов из разных тем курса. За каждый верный ответ студент получает 2 балла. За частично верный или неполный ответ 1 балл. За неверный ответ баллы не начисляются.	дифференцированный зачет
10	7	Курсовая работа/проект	Анализ предметной области	-	3	Необходимо найти хотя бы один источник (книгу или статью), описать предметную область и решаемую задачу. Критерии оценки: задание выполнено вовремя - 2 балла, иначе -0 баллов; рассмотрено несколько существующих решений данной задачи - 1 балл.	кур- совые работы
11	7	Курсовая работа/проект	Постановка задачи	-	5	Критерии оценки: - Задание выполнено вовремя - 3 балла - Сформулирована тема - 1 балл - Выбран набор данных для анализа - 1	кур- совые работы

						балл	
12	7	Курсовая работа/проект	Выполнения анализа	-	15	Выполнение анализа Критерии оценки - Понятность - есть комментарии ко всем этапам аналитики - 5 баллов - Анализ содержит визуализации - 5 баллов - Анализ содержит выводы - 5 баллов	кур- совые работы
13	7	Курсовая работа/проект	Защита курсовой	-	15	Подготовлена презентация - 5 баллов В устном докладе студент показывает знания о проектных решениях в КР, свободно оперирует терминами применительно к рассматриваемой задаче - 5 баллов Студент может ответить на большинство вопросов членов комиссии - 5 баллов	кур- совые работы

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
курсовые работы	<p>За две недели до окончания семестра студент предоставляет руководителю пояснительную записку на проверку. При отсутствии замечаний руководитель допускает студента к защите, что подтверждается подписью на титульном листе пояснительной записки с указанием даты допуска. Защита курсовой работы проводится публично перед комиссией. На защите студент делает устный доклад, который сопровождается презентацией, и отвечает на вопросы комиссии.</p> <p>Отдельные этапы курсовой работы оцениваются в течение семестра. Итоговая оценка выставляется после защиты.</p>	В соответствии с п. 2.7 Положения
дифференцированный зачет	<p>Студент может улучшить свой рейтинг, пройдя контрольное мероприятие промежуточной аттестации, которое не является обязательным. Контрольное мероприятие промежуточной аттестации проводится во время зачета в виде устного опроса. Студенту задаются 5 вопросов из разных тем курса. Студенту дается 60 минут на подготовку ответов. Затем студент озвучивает свои ответы.</p>	В соответствии с пп. 2.5, 2.6 Положения

6.3. Оценочные материалы

Компетенции	Результаты обучения	№ КМ												
		1	2	3	4	5	6	7	8	9	10	11	12	13
ПК-2	Знает: основные элементы процесса анализа больших данных, основные подходы к обработке больших массивов данных	+	+	+	+	+	+	+	+	+				
ПК-2	Умеет: визуализировать имеющиеся данные, отбрасывать несущественную информацию, структурировать информацию в рамках поставленной задачи	+	+	+	+	+	+	+	+	+				
ПК-2	Имеет практический опыт: использования современных высоконагруженных систем хранения и обработки больших данных	+	+	+	+	+	+	+	+	+	+	+	+	+

Фонды оценочных средств по каждому контрольному мероприятию находятся в приложениях.

7. Учебно-методическое и информационное обеспечение дисциплины

Печатная учебно-методическая документация

а) *основная литература:*

Не предусмотрена

б) *дополнительная литература:*

Не предусмотрена

в) *отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:*

Не предусмотрены

г) *методические указания для студентов по освоению дисциплины:*

1. Общие рекомендации по организации самостоятельной работы
2. ТЕХНОЛОГИИ И ИНФРАСТРУКТУРА BIG DATA

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Общие рекомендации по организации самостоятельной работы
2. ТЕХНОЛОГИИ И ИНФРАСТРУКТУРА BIG DATA

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Дополнительная литература	Электронно-библиотечная система издательства Лань	Форман Дж., Много цифр: Анализ больших данных при помощи Excel. [Электронный ресурс] — Электрон. дан. — М. : Альпина Паблишер, 2016. — 461 с. https://e.lanbook.com/book/87871
2	Дополнительная литература	Электронно-библиотечная система издательства Лань	Симчера, В.М. Методы многомерного анализа статистических данных. [Электронный ресурс] — Электрон. дан. — М. : Финансы и статистика, 2008. — 400 с. https://e.lanbook.com/book/87871
3	Основная литература	Электронно-библиотечная система	Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. — ISBN

		издательства Лань	978-5-97060-190-7. https://e.lanbook.com/book/93571
4	Основная литература	Электронно-библиотечная система издательства Лань	Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 188 с. — ISBN 978-5-8114-9690-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/198599

Перечень используемого программного обеспечения:

1. -Python(бессрочно)

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Лабораторные занятия	405 (1)	ПК, проектор
Лекции	405 (1)	ПК, проектор