



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Дальневосточный федеральный университет»
(ДВФУ)

ШКОЛА ЕСТЕСТВЕННЫХ НАУК

«СОГЛАСОВАНО»
Руководитель ОП


(подпись) Пустовалов Е.В.
« » 2018 г. (Ф.И.О. рук. ОП)



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Визуализация больших данных

Направление подготовки 09.04.02 Информационные системы и технологии

магистерская программа «Большие данные и облачные технологии»

Форма подготовки очная

курс 1 семестр 1
лекции 18 час.
практические занятия 18 час.
лабораторные работы час.
в том числе с использованием МАО лек. /пр. 14 /лаб. час.
всего часов аудиторной нагрузки 36 час.
в том числе с использованием МАО 14 час.
самостоятельная работа 72 час.
в том числе на подготовку к экзамену час.
контрольные работы (количество)
курсовая работа / курсовой проект семестр
зачет 1 семестр
экзамен не предусмотрен

Рабочая программа составлена в соответствии с требованиями образовательного стандарта, самостоятельно устанавливаемого ДВФУ, утвержденного приказом ректора от 07.07.2015 г. № 13-1282.

Рабочая программа обсуждена на заседании кафедры компьютерных систем, протокол № 18 от «16» июля 2018 г.

Заведующий кафедрой Пустовалов Е.В.
Составитель : к.ф.-м.н., ассистент Шевченко Ю.А.

Оборотная сторона титульного листа РПУД

I. Рабочая программа пересмотрена на заседании кафедры:

Протокол от «_____» _____ 20__ г. № _____

Заведующий кафедрой _____
(подпись) (И.О. Фамилия)

II. Рабочая программа пересмотрена на заседании кафедры:

Протокол от «_____» _____ 20__ г. № _____

Заведующий кафедрой _____
(подпись) (И.О. Фамилия)

АННОТАЦИЯ

Рабочая учебная программа дисциплины «Визуализация больших данных» разработана для магистров первого года обучения и предназначена для изучения в первом семестре по направлению 09.04.02 «Информационные системы и технологии», профиль «Большие данные и облачные технологии». В соответствии с требованиями собственных образовательных стандартов ДВФУ (ОС ВО ДВФУ) - Приказ ДВФУ № 12-13-1282 от 07.07.2015. Трудоемкость дисциплины составляет 3 зач.ед., 108 ак.часов.

Цель.

Целями освоения дисциплины «Визуализация больших данных» является развитие навыков программирования на языке Python, представления о сборе, обработке и анализе данных в интерактивной среде Jupyter, введение в автоматизированные методы работы с данными - машинное обучение и нейронные сети. Дисциплина является теоретической и практической частью подготовки магистров по использованию информационных и коммуникационных технологий (ИКТ) в профессиональной деятельности.

Задачи:

- Изучение основных положений анализа больших данных
- Изучение основных положений визуализации и обработки больших данных
- Изучение каналов связи и их характеристик
- Изучения методов кодирования и характеристик кодов

В результате изучения данной дисциплины у обучающихся формируются следующие общекультурные/ общепрофессиональные/ профессиональные элементы компетенций.

Код и формулировка компетенции	Этапы формирования компетенции	
ПК-12 умением проводить разработку и исследование теоретических и экспериментальных моделей объектов	Знает	Все фундаментальные теории из следующих областей науки: машиностроение, приборостроение, наука, техника, образование, медицина, административное управление, юриспруденция, бизнес, предпринимательство, коммерция, менеджмент, банковские системы, безопасность информационных систем,

<p>профессиональной деятельности в областях: машиностроение, приборостроение, наука, техника, образование, медицина, административное управление, юриспруденция, бизнес, предпринимательство, коммерция, менеджмент, банковские системы, безопасность информационных систем, управление технологическими процессами, механика, техническая физика, энергетика, ядерная энергетика, силовая электроника, металлургия, строительство, транспорт, железнодорожный транспорт, связь, телекоммуникации, управление инфокоммуникациями, почтовая связь, химическая промышленность, сельское хозяйство, текстильная и легкая промышленность, пищевая промышленность, медицинские и биотехнологии, горное дело, обеспечение безопасности подземных предприятий и производств, геология, нефтегазовая отрасль,</p>		<p>управление технологическими процессами, механика, техническая физика, энергетика, ядерная энергетика, силовая электроника, металлургия, строительство, транспорт, железнодорожный транспорт, связь, телекоммуникации, управление инфокоммуникациями, почтовая связь, химическая промышленность, сельское хозяйство, текстильная и легкая промышленность, пищевая промышленность, медицинские и биотехнологии, горное дело</p>
	Умеет	<p>Успешно применять полученные знания из одной области науки к другой</p>
	Владеет	<p>Основными навыками наездников сингулярности: как управлять в эпоху сложности, не привлекая при этом санитаров леса</p>

<p>геодезия и картография, геоинформационные системы, лесной комплекс, химико-лесной комплекс, экология, сфера сервиса, системы массовой информации, дизайн, медиаиндустрия, а также предприятия различного профиля и все виды деятельности в условиях экономики информационного общества</p>		
<p>ПК-13 умением проводить разработку и исследование методик анализа, синтеза, оптимизации и прогнозирования качества процессов функционирования информационных систем и технологий</p>	Знает	<p>Что такое интернет-браузер, всемирная паутина, интернет-сайт, поисковая система, wifi-сеть, электронная почта</p>
	Умеет	<p>Печатать тексты на клавиатуре, преобразовывать мысленный поток в последовательный и содержательный поисковый запрос, переходить по интернет-ссылкам, пользоваться текстовым редактором, почтовым клиентом</p>
	Владеет	<p>иностранном языком на уровне профессионального общения</p>

Для формирования вышеуказанных компетенций в рамках дисциплины «Обработка и классификация данных методами нейронных сетей» применяются следующие методы активного обучения: Работа в малых группах, Творческое задание по разработке и презентации одного из подразделов.

I. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

Раздел I. Знакомство с инструментами визуализации и обработки данных (9 час.)

Тема 1. (2 час.)

Введение в анализ данных на языке Python. Почему Python становится стандартом для работы с большими данными. Прикладные задачи политологов, для решения которых подходит язык Python.

Тема 2. (1 час.)

Математический аппарат для анализа данных: векторы, матрицы, функции и производные.

Тема 3. (1 час.)

Введение в модуль для работы с числовыми данными NumPy (Numerical Python).

Тема 4. (1 час.)

Введение в модуль для работы с табличным представлением данных Pandas. Преобразование словарей в табличный формат Pandas, загрузка данных из внешних источников. Особенности фильтрации и обращения к данным.

Тема 5. (1 час.)

Введение в визуализацию данных. Нюансы визуализации данных и принципы человеческого восприятия. Правила создания хороших визуализаций.

Тема 6. (1 час.)

Обзор различных видов графиков (гистограмма, бар-чарт, секторная диаграмма, линейные графики, график рассеяния, тепловая карта и т.д.). Особенности разных видов графиков и их использования.

Тема 7. (1 час.)

Работа с данными Gapminder в Pandas (<https://www.gapminder.org>): разведывательный анализ, преобразование данных. Воссоздание

визуализации Garminder (зависимость продолжительности жизни от ВВП на душу населения).

Тема 8. (1 час.)

Обзор языка разметки markdown. Экспорт данных и отчетов из Jupyter Notebook.

Раздел II. Изучение методов обработки данных (9 час.)

Тема 1. (2 час.)

Анализ текста. Латентное размещение Дирихле.

Тема 2. (1 час.)

Введение в анализ сетей. Основные метрики и параметры сетей.

Тема 3. (1 час.)

Введение в машинное обучение. Обучение с учителем и без учителя. Проблема переобучения.

Тема 4. (1 час.)

Регрессионные модели. Метод наименьших квадратов. Логистическая регрессия.

Тема 5. (1 час.)

Решение задач кластеризации. Меры расстояния. Обзор алгоритмов кластеризации (иерархические алгоритмы, алгоритмы квадратичной ошибки, выделение связных компонент).

Тема 6. (1 час.)

Введение в ансамбли моделей. Стэкинг и бэггинг. Случайный лес. Бустинг.

Тема 7. (1 час.)

Введение в нейронные сети. Перцептроны. Проблемы обучения нейронными сетями. Прикладное использование в физике.

Тема 8. (1 час.)

Библиотеки для работы с нейронными сетями - tensorflow и keras. Пример сверточной сети на наборе данных MNIST (распознавание рукописного текста).

II. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА

Практические занятия (18 час.)

Занятие 1. (2 час.)

Основы программирования на языке Python: типы данных и методы работы с ними (переменные, листы, словари, кортежи).

Занятие 2. (2 час.)

Основы программирования на языке Python: циклы (if и while), функции. Тест на знание синтаксиса языка.

Занятие 3. (2 час.)

Особенные типы данных в NumPy. Работа с векторами и матрицами. Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия).

Занятие 4. (1 час.)

Работа с табличными данными в Pandas на примере данных Всемирного банка по продолжительности жизни и данных о публикационной активности словацких экономистов.

Занятие 5. (1 час.)

Проведение разведывательного анализа данных в Pandas на примере данных о пассажирах “Титаника” (базовый тренировочный набор данных).

Занятие 6. (1 час.)

Создание различных видов визуализаций на синтетических данных и тренировочных наборах данных. Создание визуализаций для

разведывательного анализа данных для набора данных о пассажирах “Титаника”.

Занятие 7. (1 час.)

Работа с данными Gapminder в Pandas (<https://www.gapminder.org>): разведывательный анализ, преобразование данных. Воссоздание визуализации Gapminder (зависимость продолжительности жизни от ВВП на душу населения).

Занятие 8. (1 час.)

Создание интерактивных визуализаций и отчетов с помощью инструмента Plotly. Работа с данными о предвыборных кампаниях в США в 2008 и 2012 годах (создание тепловых географических карт, интерактивного пузырькового графика).

Занятие 9. (1 час.)

Сбор данных для текстового анализа с новостных сайтов. Подготовка данных для анализа. Использование моделей Word2vector.

Занятие 10. (1 час.)

Анализ текста, собранного из социальной сети (комментарии пользователей). Построение предсказательных моделей.

Занятие 11. (1 час.)

Введение в модуль NetworkX. Подготовка данных для анализа сетей. Примеры визуализации сетей на примере данных из сети Twitter.

Занятие 12. (1 час.)

Библиотека sklearn. Постановка гипотезы, преобразование данных и построение регрессионных моделей для задачи с пассажирами “Титаника”.

Занятие 13. (1 час.)

Построчный разбор алгоритма кластеризации. Проверка его работы на синтетических данных (ирисы, случайные формы). Применение алгоритма кластеризации к набору данных о политических режимах.

Занятие 14. (1 час.)

Разбор примеров. Построение ансамблей моделей на наборе данных “Титаник”.

Занятие 15. (1 час.)

Разбор примеров применения нейронных сетей в политологии. Разбор статьи Anastasopoulos et al. (2017) “Political image analysis with deep neural networks?”

III. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине «Визуализация больших данных» представлено в Приложении 1 и включает в себя:

план-график выполнения самостоятельной работы по дисциплине, в том числе примерные нормы времени на выполнение по каждому заданию;

характеристика заданий для самостоятельной работы обучающихся и методические рекомендации по их выполнению;

требования к представлению и оформлению результатов самостоятельной работы;

критерии оценки выполнения самостоятельной работы.

IV. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций	Оценочные средства		
			текущий контроль	промежуточная аттестация	
1	Раздел I. Знакомство с инструментами визуализации и обработки	ПК-12 ПК-13	знает	ПР-7	Зачет, вопросы 1-29
			умеет	ПР-13	Зачет, задания тип 1-20
			владеет	УО-3	<i>Решение задач</i>
	Раздел II. Изучение	ПК-12 ПК-13	знает	ПР-7	Зачет, вопросы 1-29

методов обработки данных	умеет	ПР-13	Зачет, задания тип 1-20
	владеет	УО-3	<i>Решение задач</i>

Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков и (или) опыта деятельности, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 2.

V. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

(электронные и печатные издания)

№	Название	Ссылка в ЭК НБ ДВФУ	Внешняя ссылка
1	Адлер, Ю. П. Статистическое управление процессами. «Большие данные» [Электронный ресурс] : учебное пособие / Ю. П. Адлер, Е. А. Черных. — Электрон. текстовые данные. — М. : Издательский Дом МИСиС, 2016. — 52 с. — 978-5-87623-969-3.	http://lib.dvfu.ru:8080/lib/item?id=IPRbooks:IPRbooks-64199&theme=FEFU	
2	Вернеке, Д. Язык географической разметки KML [Электронный ресурс] : справочник / Д. Вернеке. — Электрон. дан. — Москва : ДМК Пресс, 2010. — 288 с.	http://lib.dvfu.ru:8080/lib/item?id=Lan:Lan-1259&theme=FEFU	
3	Потопахин, В. Искусство алгоритмизации [Электронный ресурс] / В. Потопахин. — Электрон. дан. — Москва : ДМК Пресс, 2011. — 320 с.	http://lib.dvfu.ru:8080/lib/item?id=Lan:Lan-1269&theme=FEFU	
4	Сошников, Д.В. Функциональное программирование на F# [Электронный ресурс] / Д.В.	http://lib.dvfu.ru:8080/lib/item?id=Lan:	

	Сошников. — Электрон. дан. — Москва : ДМК Пресс, 2011. — 192 с.	Lan-1274&theme=FEFU	
5	Баженова, И. Ю. SQL и процедурно-ориентированные языки [Электронный ресурс] / И. Ю. Баженова. — 2-е изд. — Электрон. текстовые данные. — М. : Интернет-Университет Информационных Технологий (ИНТУИТ), 2016. — 166 с. — 5-94774-539-9.	http://lib.dvfu.ru:80/80/lib/item?id=IPRbooks:IPRbooks-57532&theme=FEFU	
6	Игнатъев, С. А. Построение базы данных в Microsoft Access 2010 [Электронный ресурс] : учебное пособие / С. А. Игнатъев. — Электрон. текстовые данные. — Саратов : Саратовский государственный технический университет имени Ю.А. Гагарина, ЭБС АСВ, 2012. — 129 с. — 978-5-7433-2602-0.	http://lib.dvfu.ru:80/80/lib/item?id=IPRbooks:IPRbooks-76500&theme=FEFU	
7	Волков, Д. А. Базы данных [Электронный ресурс] : учебно-методическое пособие / Д. А. Волков. — Электрон. текстовые данные. — М. : МИСИ-МГСУ, Ай Пи Эр Медиа, ЭБС АСВ, 2018. — 77 с. — 978-5-7264-1883-4.	http://lib.dvfu.ru:80/80/lib/item?id=IPRbooks:IPRbooks-79883&theme=FEFU	
8	Гурвиц Д. Просто о больших данных. М. : Сбербанк, 2015. 395 с.	https://lib.dvfu.ru:8443/lib/item?id=chamo:826169&theme=FEFU	

Дополнительная литература
(печатные и электронные издания)

№	Название	Ссылка в ЭК НБ ДВФУ	Внешняя ссылка

1	Яглом А.М., Яглом И.М. Вероятность и информация / М.:Наука, 1973. - 512с.		
2	Орлов В.А., Филиппов Л.И. Нейронные сети в упражнениях и задачах: [Учебное пособие для втузов]/ М.: Высшая школа, 1976.- 134 с.		
3	Галагер Р. Нейронные сети и надежная связь / Пер. с англ., под ред. М.С.Пинскера и Б.С.Цыбакова, М.: Сов. радио, 1974.- 720с.		
4	Липкин А.И. Основы статистической радиотехники, нейронных сетей и кодирования/ М.: Сов. радио, 1978. - 240с.		

Перечень информационных технологий и программного обеспечения

Аудитория с мультимедиапроектором, ПК с MS Office или LibreOffice.

VI. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

При изучении дисциплины 36 часов отводится на аудиторные занятия и 72 часа на самостоятельную работу студентов. Из 36 часов аудиторных занятий 18 лекционных, 18 практических, 14 с использованием методов активного обучения. С целью более эффективного освоения материала рекомендуется практические занятия проводить в конце семестра, когда студенты будут знакомы с теоретическим содержанием курса.

При подготовке к зачету рекомендуется просмотреть материалы лекций и собственные конспекты, разбить вопросы по разделам и темам, затем определить содержание ответов на вопросы.

Выполнение заданий в малых группах и творческого задания предусматривает выбор подраздела курса и согласование его с преподавателем.

VII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Для проведения исследований, связанных с выполнением задания по практике, а также для организации самостоятельной работы студентам доступно следующее лабораторное оборудование и специализированные кабинеты, соответствующие действующим санитарным и противопожарным нормам, а также требованиям техники безопасности при проведении учебных и научно-производственных работ:

Наименование оборудованных помещений и помещений для самостоятельной работы	Перечень основного оборудования
<p>Владивосток, о. Русский, п. Аякс д.10, корпус L, ауд. L 502</p> <p>учебная аудитория для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации</p>	<p>Мультимедийное оборудование: ЖК-панель 47", Full HD, LG M4716 CCBA - 1 шт. Парты и стулья</p>
<p>г. Владивосток, о. Русский, п. Аякс д.10, корпус D, ауд. D734</p> <p>учебная аудитория для проведения занятий семинарского типа, практических занятий: компьютерный класс</p>	<p>Моноблок HPP-B0G08ES#ACB/8200E AIO i52400S 500G 4.0G 28 PC - 15 шт</p> <p>Мультимедийное оборудование: Экран проекционный ScreenLine Trim White Ice 50 см черная кайма сверху, размер рабочей области 236x147 см</p> <p>Документ-камера Avervision CP355AF ЖК-панель 47", Full HD, LG M4716 CCBA Мультимедийный проектор Mitsubishi EW330U, 3000 ANSI Lumen, 1280x800 Сетевая видеочка Multipix MP-HD718</p>
<p>Читальные залы Научной библиотеки ДВФУ с открытым доступом к фонду (корпус А - уровень 10)</p> <p>Аудитории для самостоятельной работы</p>	<p>Читальные залы Научной библиотеки ДВФУ с открытым доступом к фонду: Моноблок Lenovo C360G-i34164G500UDK – 15 шт.</p> <p>Интегрированный сенсорный дисплей Polymedia FlipBox - 1 шт.</p> <p>Копир-принтер-цветной сканер в e-mail с 4 лотками Xerox WorkCentre 5330 (WC5330C – 1 шт. Скорость доступа в Интернет 500 Мбит/сек. Рабочие места для людей с ограниченными возможностями здоровья оснащены дисплеями и принтерами Брайля; оборудованы: портативными устройствами для чтения плоскочечатных текстов, сканирующими</p>

	и читающими машинами видеоувеличителем с возможностью регуляции цветовых спектров; увеличивающими электронными лупами и ультразвуковыми маркировщиками
--	--

В целях обеспечения специальных условий обучения инвалидов и лиц с ограниченными возможностями здоровья в ДВФУ все здания оборудованы пандусами, лифтами, подъемниками, специализированными местами, оснащенными туалетными комнатами, табличками информационно-навигационной поддержки.



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Дальневосточный федеральный университет»
(ДФУ)

ШКОЛА ЕСТЕСТВЕННЫХ НАУК

**УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ
РАБОТЫ ОБУЧАЮЩИХСЯ**

по дисциплине «Визуализация больших данных»

**Направление подготовки 09.04.02 Информационные системы и
технологии**

магистерская программа «Большие данные и облачные технологии»

Форма подготовки очная

**Владивосток
2018**

План-график выполнения самостоятельной работы по дисциплине

№ п/п	Дата/сроки выполнения	Вид самостоятельной работы	Примерные нормы времени на выполнение	Форма контроля
1	1-18 неделя обучения, еженедельно (3 часа в неделю)	Домашнее задание	54	Проверка ДЗ
2	12-18 неделя обучения	Подготовка творческого задания	18	Выступление по результатам

Самостоятельная работа по выполнению домашнего задания должна включать в себя повторение лекционного материала, повторение формул по разделу, повторение решенных задач по разделу, решение задач домашнего задания по разделу. Выполненное задание должно быть оформлено в соответствии с требованиями по оформлению решения задач, текст, формулы легко читаемы.

Самостоятельная работа по подготовке творческого задания должна включать в себя поиск информации в сети Интернет и рекомендуемых источниках, обсуждение основных характеристик, подготовка черновиков презентаций и текста выступления. Презентация должна быть не менее 15 слайдов, выступление продолжительностью 15-18 минут. Каждая малая группа должна подготовить вопросы для остальных групп по их темам.

Самостоятельная работа по подготовке к экзамену должна включать повторение теоретического материала, подготовку ответов на вопросы с использованием лекций и рекомендуемых источников.

Оценка результатов самостоятельной работы по подготовке творческого задания выполняется по следующим критериям:

5 баллов выставляется, если студент выразил своё мнение по сформулированной проблеме, аргументировал его, точно определив ее содержание и составляющие. Приведены данные отечественной и зарубежной литературы, статистические сведения, информация нормативно-правового характера. Продемонстрированы знания и владения навыками самостоятельной исследовательской работы по теме исследования; методами и приемами анализа теоретических и/или практических аспектов изучаемой области. Фактических ошибок, связанных с пониманием проблемы, нет; графически работа оформлена правильно

4 балла - работа характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Для аргументации приводятся данные отечественных и зарубежных авторов. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет. Допущены одна-две ошибки в оформлении работы

3 балла - Студент провел достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимает базовые основы и теоретическое обоснование выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы, оформлении работы

2 балла - если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая темы. Допущено три или более трех ошибок в смысловом содержании раскрываемой проблемы, в оформлении работы.

Оценка результатов самостоятельной работы в малых группах выполняется по следующим критериям:

5 баллов выставляется, если студент/группа выразили своё мнение по сформулированной проблеме, аргументировали его, точно определив ее содержание и составляющие. Приведены данные отечественной и зарубежной литературы, статистические сведения, информация нормативно-правового характера. Продемонстрировано знание и владение навыком самостоятельной исследовательской работы по теме исследования; методами и приемами анализа международно-политической практики. Фактических ошибок, связанных с пониманием проблемы, нет

4 балла - работа студента/группы характеризуется смысловой цельностью, связностью и последовательностью изложения; допущено не более 1 ошибки при объяснении смысла или содержания проблемы. Для аргументации приводятся данные отечественных и зарубежных авторов. Продемонстрированы исследовательские умения и навыки. Фактических ошибок, связанных с пониманием проблемы, нет.

3 балла - проведен достаточно самостоятельный анализ основных этапов и смысловых составляющих проблемы; понимание базовых основ и теоретического обоснования выбранной темы. Привлечены основные источники по рассматриваемой теме. Допущено не более 2 ошибок в смысле или содержании проблемы

2 балла - если работа представляет собой пересказанный или полностью переписанный исходный текст без каких бы то ни было комментариев, анализа. Не раскрыта структура и теоретическая составляющая темы. Допущено три или более трех ошибок смыслового содержания раскрываемой проблемы



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Дальневосточный федеральный университет»
(ДФУ)

ШКОЛА ЕСТЕСТВЕННЫХ НАУК

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине «Визуализация больших данных»
Направление подготовки 09.04.02 Информационные системы и
технологии
магистерская программа «Большие данные и облачные технологии»
Форма подготовки очная

Владивосток
2018

Паспорт ФОС

Код и формулировка компетенции	Этапы формирования компетенции	
ПК-12 умением проводить разработку и исследование теоретических и экспериментальных моделей объектов профессиональной деятельности в областях: машиностроение, приборостроение, наука, техника, образование, медицина, административное управление, юриспруденция, бизнес, предпринимательство, коммерция, менеджмент, банковские системы, безопасность информационных систем, управление технологическими процессами, механика, техническая физика, энергетика, ядерная энергетика, силовая электроника, металлургия, строительство, транспорт, железнодорожный транспорт, связь, телекоммуникации, управление инфокоммуникациями, почтовая связь, химическая промышленность, сельское хозяйство, текстильная и легкая промышленность,	Знает	Все фундаментальные теории из следующих областей науки: машиностроение, приборостроение, наука, техника, образование, медицина, административное управление, юриспруденция, бизнес, предпринимательство, коммерция, менеджмент, банковские системы, безопасность информационных систем, управление технологическими процессами, механика, техническая физика, энергетика, ядерная энергетика, силовая электроника, металлургия, строительство, транспорт, железнодорожный транспорт, связь, телекоммуникации, управление инфокоммуникациями, почтовая связь, химическая промышленность, сельское хозяйство, текстильная и легкая промышленность, пищевая промышленность, медицинские и биотехнологии, горное дело
	Умеет	Успешно применять полученные знание из одной области науки к другой
	Владеет	Основными навыками наездников сингулярности: как управлять в эпоху сложности, не привлекая при этом санитаров леса

<p>пищевая промышленность, медицинские и биотехнологии, горное дело, обеспечение безопасности подземных предприятий и производств, геология, нефтегазовая отрасль, геодезия и картография, геоинформационные системы, лесной комплекс, химико-лесной комплекс, экология, сфера сервиса, системы массовой информации, дизайн, медиаиндустрия, а также предприятия различного профиля и все виды деятельности в условиях экономики информационного общества</p>		
<p>ПК-13 умением проводить разработку и исследование методик анализа, синтеза, оптимизации и прогнозирования качества процессов функционирования информационных систем и технологий</p>	Знает	<p>Что такое интернет-браузер, всемирная паутина, интернет-сайт, поисковая система, wifi-сеть, электронная почта</p>
	Умеет	<p>Печатать тексты на клавиатуре, преобразовывать мысленный поток в последовательный и содержательный поисковый запрос, переходить по интернет-ссылкам, пользоваться текстовым редактором, почтовым клиентом</p>
	Владеет	<p>иностранном языком на уровне профессионального общения</p>

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Раздел I. Знакомство с	ПК-12	ПК-13	знает	Зачет, вопросы 1-29

	инструментами визуализации и обработки		умеет	ПР-13	Зачет, задания тип 1-20
			владеет	УО-3	<i>Решение задач</i>
	Раздел II. Изучение методов обработки данных	ПК-12 ПК-13	знает	ПР-7	Зачет, вопросы 1-29
			умеет	ПР-13	Зачет, задания тип 1-20
			владеет	УО-3	<i>Решение задач</i>

Шкала оценивания уровня сформированности компетенций

Код и формулировка компетенции	Этапы формирования компетенции		критерии	показатели
ПК-12, умением проводить разработку и исследование теоретических и экспериментальных моделей объектов профессиональной деятельности в областях: машиностроение, приборостроение, наука, техника, образование, медицина, административное управление, юриспруденция, бизнес, предпринимательство, коммерция, менеджмент,	знает (пороговый уровень)	методы моделирования прикладных ИС и бизнес-процессов предприятия и организации; методы реинжиниринга прикладных и информационных процессов предприятия и организации	воспроизводить и объяснять учебный материал с требуемой степенью научной точности и полноты	способность показать базовые знания и основные умения в использовании: - принципов моделирования прикладных и информационных бизнес-процессов; - графических нотаций моделирования прикладных и информационных бизнес-процессов для задач реинжиниринга; - общесистемные принципы реинжиниринга прикладных и информационных процессов предприятия и организации; - инструментальных CASE-средств реинжиниринга прикладных и информационных бизнес-процессов
	умеет (продвинутый)	применять методы моделирования прикладных ИС	выполнять типичные	способность применить знания и практические умения в задачах, связанных с выбором и

<p>банковские системы, безопасность информационных систем, управление технологическими процессами, механика, техническая физика, энергетика, ядерная энергетика, силовая электроника, металлургия, строительство, транспорт, железнодорожный транспорт, связь, телекоммуникации, управление инфокоммуникациями, почтовая связь, химическая промышленность, сельское хозяйство, текстильная и легкая промышленность, пищевая промышленность, медицинские и биотехнологии, горное дело, обеспечение безопасности подземных предприятий и производств, геология, нефтегазовая отрасль, геодезия и картография,</p>		<p>и бизнес-процессов предприятия и организации; применять методы реинжиниринга прикладных и информационных процессов предприятия и организации</p>	<p>задачи на основе воспроизведения стандартных алгоритмов решения</p>	<p>обоснованием стратегии и проектных решений по моделированию прикладных ИС и реинжинирингу прикладных и информационных процессов предприятия и организации, используя инструментальные CASE средства, принципы реинжиниринга управления бизнес-процессами</p>
	<p>владеет (высокий)</p>	<p>навыками моделирования прикладных ИС и бизнес-процессов предприятия и организации; инструментальными средствами обеспечения работ по реинжинирингу прикладных и информационных процессов предприятия и организации</p>	<p>решать усложненные задачи в нетипичных ситуациях на основе приобретенных знаний, умений и навыков</p>	<p>способность применить фактическое и теоретическое знание, практические умения по разработке проектных материалов в профессиональной области прикладной информатики, связанных с выбором и обоснованием стратегии и проектных решений по моделированию прикладных ИС и реинжинирингу прикладных и информационных процессов предприятия и организации, используя инструментальные CASE средства, принципы реинжиниринга управления бизнес-процессами</p>

<p>геоинформационные системы, лесной комплекс, химико-лесной комплекс, экология, сфера сервиса, системы массовой информации, дизайн, медиаиндустрия, а также предприятия различного профиля и все виды деятельности в условиях экономики информационного общества</p>				
<p>ПК-13, умением проводить разработку и исследование методик анализа, синтеза, оптимизации и прогнозирования качества процессов функционирования информационных систем и технологий</p>	<p>знает (пороговый уровень)</p>	<p>критерии и требования построения ИС; методологии построения ИС: (MRP, MRPII, ERP и CSRP); методы управления информационными ресурсами ИС</p>	<p>воспроизводить и объяснять учебный материал с требуемой степенью научной точности и полноты</p>	<p>способность показать базовые знания и основные умения в использовании: - общесистемных принципов управления информационными ресурсами и информационными системами; - принципов управления на основе ИС; - принципов построения ИС и управления информационными ресурсами предприятий</p>
	<p>умеет (продвинутый)</p>	<p>выбирать аппаратно-программную платформу для ИС; применять методы управления информационными ресурсами</p>	<p>выполнять типичные задачи на основе воспроизведения</p>	<p>способность применить знания и практические умения в задачах, связанных с выбором и обоснованием проектных решений по автоматизации прикладных и информационных процессов предприятий информатизации</p>

		ИС	стандартных алгоритмов решения	прикладных задач, используя инструментарий управления информационными ресурсами и информационными системами, критерии и требования к информационным системам
	владеет (высокий)	инструментам и настройки и конфигурирования ИС по выбору; инструментам и управления информационными ресурсами ИС	решать усложненные задачи в нетипичных ситуациях на основе приобретенных знаний, умений и навыков	способность применить фактическое и теоретическое знание, практические умения по разработке проектных материалов в профессиональной области прикладной информатики, связанных с выбором и обоснованием проектных решений по автоматизации прикладных и информационных процессов предприятий информатизации прикладных задач, используя инструментарий управления информационными ресурсами и информационными системами, критерии и требования к информационным системам

Методические материалы, определяющие процедуры оценивания знаний, умений и навыков

Методическим материалом является Положение о ФОС ДВФУ

Вопросы к промежуточной аттестации (зачет)

1. Понятие Большие данные. Роль цифровой информации в 21 веке.
2. Виды массивов данных.
3. Базовые принципы обработки больших данных.

4. Технологии обработки больших данных: NoSQL, MapReduce, Hadoop, R.
5. Технологии Business Intelligence и реляционные системы управления базами данных.
6. Прогнозирование и предвидение: общее и особенное.
7. Виды прогнозов
8. Общие методы анализа социально-политических и медиа процессов.
9. Специальные методы анализа социально-политических и медиа процессов.
10. Предварительный анализ данных.
11. Проверка гипотез о законе распределения случайной величины.
12. Статистические оценки параметров. Доверительные области.
13. Теория моментов.
14. Корреляционный анализ.
15. Использование модели множественной линейной регрессии для прогнозирования экономических показателей.
16. Доверительные интервалы для зависимой переменной.
17. Сглаживание временных рядов. Динамические модели с распределенными лагами.
18. Стационарные временные ряды. Тестирование стационарности.
19. Коинтеграция. Анализ временных рядов.
20. Адаптивные и мультипликативные методы прогнозирования. Экспоненциальное сглаживание.
21. Авторегрессионные модели. Модели скользящего среднего.
22. Интегрированные процессы. Идентификация авторегрессионной модели скользящего среднего.
23. Прогнозирование с моделями временных рядов. Доверительные интервалы прогноза.
24. Предсказание и прогнозирование социально-экономических прогнозов.

25. Дисперсионный анализ влияния качественных факторов. Ранговые методы.
26. Факторный анализ. Метод главных факторов.
27. Многомерное шкалирование. Классическая модель многомерного шкалирования.
28. Неметрические методы. Кластерный анализ. Дискриминантный анализ.
29. Многомерный статистический анализ.

Критерии оценки (зачет)

приводить примеры современных проблем изучаемой области.

4 балла - ответ, обнаруживающий прочные знания основных процессов изучаемой предметной области, отличается глубиной и полнотой раскрытия темы; владение терминологическим аппаратом; умение объяснять сущность, явлений, процессов, событий, делать выводы и обобщения, давать аргументированные ответы, приводить примеры; свободное владение монологической речью, логичность и последовательность ответа. Однако допускается одна - две неточности в ответе.

3 балла - оценивается ответ, свидетельствующий в основном о знании процессов изучаемой предметной области, отличающийся недостаточной глубиной и полнотой раскрытия темы; знанием основных вопросов теории; слабо сформированными навыками анализа явлений, процессов, недостаточным умением давать аргументированные ответы и приводить примеры; недостаточно свободным владением монологической речью, логичностью и последовательностью ответа. Допускается несколько ошибок в содержании ответа; неумение привести пример развития ситуации, провести связь с другими аспектами изучаемой области.

2 балла - ответ, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы; незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов; неумением давать аргументированные ответы, слабым владением монологической речью, отсутствием логичности и последовательности. Допускаются серьезные ошибки в содержании ответа; незнание современной проблематики изучаемой области.

Типовые задания для текущего контроля

1. Имеется два игральных кубика, определите вероятность выпадения числа семь и энтропию такого опыта.

2. Ключ шифрования содержит 256 бит, сколько времени понадобится для его вскрытия методом перебора, если известно, что за одну секунду компьютер перебирает 10000 вариантов.
3. 1 сентября на первом курсе одного из факультетов запланировано по расписанию три лекции по разным предметам. Всего на первом курсе изучается 7 предметов. Студент, не успевший ознакомиться с расписанием, пытается его угадать. Какова вероятность успеха в данном эксперименте, если считать, что любое расписание из четырех предметов равновероятно? Какова неопределенность такого опыта?
4. В подъезде дома установлен замок с кодом. Дверь автоматически отпирается, если в определенной последовательности набрать три цифры из имеющихся десяти. На набор одной комбинации уходит 20 секунд. Какова вероятность открыть дверь за 40 минут? Какова неопределенность такого опыта?
5. Из набора домино (28 штук) выбирают 7 штук. Какова неопределенность данного события?
6. Из колоды карт (52 карты) извлекают три карты. Какова неопределенность такого события?
7. Бросают две игральные кости. Какова неопределенность данного события?
8. Имеется 12 монет одного достоинства, 11 из них имеют одинаковый вес, а одна — фальшивая — отличается по весу от остальных (причем неизвестно, легче ли она или тяжелее настоящих). Каково наименьшее число взвешиваний на чашечных весах без гирь, которое позволяет обнаружить фальшивую монету и выяснить, легче ли она, чем остальные монеты, или тяжелее. Решить тот же вопрос для случая 13 монет.
9. Какова вероятность того, что взятое наугад целое положительное число до 1000, окажется целой степенью другого целого числа с показателем, больше единицы? Вычислите энтропию такого события.

10. Какую неопределенность содержит сообщение о событие – сдача экзамена студентом, если по опыту предыдущих экзаменов известно, что вероятность того, что студент сдал экзамен $7/8$.

11. Пусть X и Y два случайных опыта; $Z=X+Y$. Чему равна условная энтропия $H(x|z)$, если: а) X и Y независимы; б) X и Y зависимы; в) $X \equiv Y$.

12. Определить среднее количество информации, приходящееся на один символ сообщения 01001000101001, при условии, что источник эргодический, а последовательность типичная.

13. Имеются два дискретных источника с независимыми и равновероятными элементами: двоичный и троичный. На выходе первого зафиксированы два символа, на выходе второго три. Чему равны неопределенности полученных последовательностей букв, образованных парами символов первого источника и тройками символов второго?

14. Символы азбуки Морзе могут появиться в сообщении с вероятностями: для точки - 0.51, для тире - 0.31, для промежутка между буквами - 0.12, между словами - 0.06. Определить среднее количество информации в сообщении из 500 символов данного алфавита, считая, что связь между последовательными символами отсутствует.

15. Известно, что из 100 изготовленных деталей в среднем 10 деталей имеют дефекты. Для выявления брака используется метод, дающий всегда отрицательный эффект, если деталь изготовлена с браком. Если брак отсутствует, то деталь признается годной лишь в 80% случаев. Какое количество информации о качестве детали можно получить в среднем по результату такого метода отбраковки?

16. Орудие стреляет по удаленной цели. При каждом выстреле она поражается с вероятностью $p = 0.1$. Разведка может только один раз проверить, поражается ли цель хоть один раз. Через некоторое количество выстрелов k следует провести проверку, чтобы она дала максимальное количество информации?

17. Алфавит состоит из двух букв А, Б, В появляющихся в тексте с вероятностями 0.6, 0.3 и 0.1 соответственно. Закодировать отдельные буквы

равномерным кодом. Закодировать пары и тройки букв равномерным кодом. Повторить кодирование одной буквы с не равномерным кодом. Сравнить эффективность кодов. Построить кодовое дерево для не равномерного кода

18. По линии связи передаются сообщения из 5-ти равновероятных букв. Закодировать буквы равномерным кодом. Закодировать тройки букв равномерным кодом. Сравнить эффективности кодов.

19. Имеются два дискретных троичных источника с независимыми элементами. На выходе каждого источника появляются сообщения одинаковой длины по 25 элементов. Количество различных элементов в сообщении каждого источника постоянно. Сообщения каждого источника отличаются только порядком элементов. Зафиксированы два типичных сообщения: 0212021202120211201120200 – первого источника и 0121012011012012210200120 – второго. Элемент какого источника несет в среднем большее количество информации?

20. Сообщения с вероятностями 0,5; 0,25; 0,0625; 0,0625; 0,0625; 0,0625 кодируется одним из шести различных кодов: 1) 0-10-110-1110-1011-1101; 2) 1-011-010-001-000-110; 3) 0-10-110-1110-11110; 4) 111-110-101-100-011-010; 5) 0-01-011-0111-01111-01111; 6) 1-01-0011-0010-0001-0000. Определить, какие коды являются разделимыми (мгновенными). Вычислить характеристики кодов.

Критерии оценки результатов текущего контроля

5 баллов - если ответ показывает глубокое и систематическое знание всего программного материала и структуры конкретного вопроса, а также основного содержания и новаций лекционного курса по сравнению с учебной литературой. Студент демонстрирует отчетливое и свободное владение концептуально-понятийным аппаратом, научным языком и терминологией соответствующей научной области. Знание основной литературы и знакомство с дополнительно рекомендованной литературой. Логически корректное и убедительное изложение ответа.

4 - балла - знание узловых проблем программы и основного содержания лекционного курса; умение пользоваться концептуально-понятийным

аппаратом в процессе анализа основных проблем в рамках данной темы; знание важнейших работ из списка рекомендованной литературы. В целом логически корректное, но не всегда точное и аргументированное изложение ответа.

3 - балла - фрагментарные, поверхностные знания важнейших разделов программы и содержания лекционного курса; затруднения с использованием научно-понятийного аппарата и терминологии учебной дисциплины; неполное знакомство с рекомендованной литературой; частичные затруднения с выполнением предусмотренных программой заданий; стремление логически определено и последовательно изложить ответ.

2 балла - незнание, либо отрывочное представление о данной проблеме в рамках учебно-программного материала; неумение использовать понятийный аппарат; отсутствие логической связи в ответе.