

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

Дата подписания: 28.11.2023 12:50:36

Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a70e51517ff6d591

**Автономная некоммерческая образовательная организация высшего образования  
«Европейский университет в Санкт-Петербурге»**

УТВЕРЖДАЮ

Ректор  /Волков В.В./

« 31 » августа 2021 г.

Протокол Ученого Совета

№ 7 от 31 августа 2021 г.



**РАБОЧИЙ УЧЕБНЫЙ ПЛАН**

подготовки слушателей

дополнительная профессиональная программа

**«Прикладной анализ данных»**

**вид программы**

программа профессиональной переподготовки


Подразделение: факультет социологии

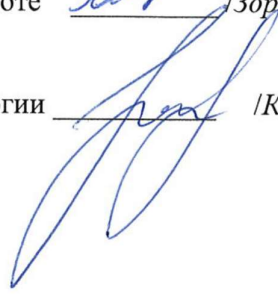
Срок обучения /объем программы: 734 часа

Форма обучения: очная

квалификация (степень) выпускника: аналитик данных

**Согласовано**

Проректор по учебной работе  /Зоринова Е.М./

Декан факультета социологии  /Костенко В.В./

## Учебный план

№ п/п	Наименование дисциплин	Всего часов	в том числе			Форма контроля
			Лекции	Практические/семинарские занятия <sup>1</sup>	СР <sup>2</sup>	
<b>Модуль 1</b>						
<b>Обязательные дисциплины</b>						
1.	Язык программирования Python	62	14	14	34	зачет
2.	Практический минимум	62	14	14	34	зачет
3.	Линейная алгебра	62	14	14	34	зачет
Всего модуль 1:		<b>186</b>	<b>42</b>	<b>42</b>	<b>102</b>	
<b>Модуль 2</b>						
<b>Обязательные дисциплины</b>						
1.	Программирование в R	62	14	14	34	зачет
2.	Основы статистики	62	14	14	34	зачет
<b>Дисциплины по выбору<sup>3</sup></b>						
1.	Административные и правовые данные	54	14	14	26	зачет
2.	Опросные данные	54	14	14	26	зачет
Всего модуль 2:		<b>232</b>	<b>56</b>	<b>56</b>	<b>120</b>	
<b>Модуль 3</b>						
<b>Обязательные дисциплины</b>						
1.	Машинное обучение: введение	62	14	14	34	зачет
<b>Дисциплины по выбору</b>						
1.	Разметка данных	54	14	14	26	зачет
2.	Текстовые данные	54	14	14	26	зачет
Всего по модулю 3:		<b>170</b>	<b>42</b>	<b>42</b>	<b>86</b>	
<b>Модуль 4</b>						
<b>Дисциплины по выбору</b>						
1.	Введение в обработку естественного языка (NLP)	54	14	14	26	зачет
2.	Введение в анализ изображений и видео	54	14	14	26	зачет
Всего по модулю 4:		<b>108</b>	<b>28</b>	<b>28</b>	<b>52</b>	
<b>Модуль 5</b>						

<sup>1</sup> Могут включать в себя: лабораторные работы, круглые столы, мастер-классы, мастерские, деловые игры, ролевые игры, тренинги, семинары по обмену опытом, выездные занятия, консультации

<sup>2</sup> СР – самостоятельная работа, включает в себя часы на текущий и промежуточный контроль

<sup>3</sup> Слушатель обязан выбрать из всех представленных в модулях блоков дисциплин по выбору суммарно четыре дисциплины

<b>Обязательные дисциплины</b>						
1.	Визуализация данных	62	14	14	34	зачет
<b>Дисциплины по выбору</b>						
1.	Геоданные	54	14	14	26	зачет
<b>Всего по модулю 5:</b>		<b>116</b>	<b>28</b>	<b>28</b>	<b>60</b>	
1.	Итоговая аттестация	84	-	32	52	зачет/ защита проекта
<b>Итого по программе профессиональной переподготовки<sup>4</sup>:</b>		<b>734</b>	<b>182</b>	<b>184</b>	<b>360</b>	

---

<sup>4</sup> Общее количество часов по программе профессиональной переподготовки не соответствует сумме часов по модулям, поскольку слушатель обязан выбрать суммарно только четыре из семи представленных в программе дисциплин по выбору (из всех модулей).

**Автономная некоммерческая образовательная организация высшего образования  
«Европейский университет в Санкт-Петербурге»**

УТВЕРЖДАЮ

Ректор

/Волков В.В./

« 31 » августа 2021 г.

Протокол Ученого Совета

№ 7 от 31 августа 2021 г.



**УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН  
программы профессиональной переподготовки  
«Прикладной анализ данных»**

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
<b>Модуль 1</b>							
<b>Обязательные дисциплины</b>							
<b>Язык программирования Python</b>							
1.	Что такое Python и почему мы выбрали именно его	История языка. Области применения. Дистрибутивы. Редакторы кода. Установка. Настройка. Лицензии на использование. Простые программы.	5	1	1	3	зачет
2.	Что такое ноутбуки и как ими пользоваться	Что такое ноутбуки и как ими пользоваться. Интерактивный режим. Понятие алгоритма. Переменные. Запуск Jupyter. Архитектура. Настройка.	5	1	1	3	
3.	Типы данных	Числа, строки, списки, словари, множества. Объекты. Методы объекта. Взаимосвязь алгоритма и структур данных	6	2	1	3	
4.	Циклы, функции, генераторы, list comprehension	Структурное программирование. Утиная типизация. Императивное и декларативное программирование.	7	1	2	4	
5.	Работа файлами различных форматов	Работа с распространенными форматами файлов: лог-файлы, бинарные файлы, CSV, JSON, XML. Модули для работы с ними. API. Работа с большими файлами, событийные	9	2	2	5	

<sup>5</sup> Самостоятельная работа, включает в себя часы на текущий контроль

<sup>6</sup> Могут включать в себя: лабораторные работы, круглые столы, мастер-классы, мастерские, деловые игры, ролевые игры, тренинги, семинары по обмену опытом, выездные занятия, консультации

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
		обработчики. Деревья объектов файла с данными.					
6.	Функции и их свойства	Стандартная библиотека. Создание своих функций	6	2	1	3	
7.	Pandas.	Установка библиотеки. Работа с файлами данных как с таблицами. Индексация, селекция, агрегация. NumPy, SciPy и Matplotlib. Взаимосвязь библиотек, области применения, типичные приемы использование. Примеры.	13	3	3	7	
8.	Работа со случайными величинами (ipython notebook)	ГПСЧ, получение распределения с заданными характеристиками. Оценка распределения по выборке. Доверительный интервал. Параметры распределения случайной величины. Практика.	9	2	3	4	
9.	Промежуточная аттестация	Проект	2	-	-	2	
Всего:			62	14	14	34	
<b>Практический минимум</b>							
1.	<i>Системное и прикладное ПО</i>	Операционная система как вид ПО. Виды ОС. Linux-подобные ОС. Вычислительные ядра.	7	2	1	4	зачет
2.	<i>Архитектура клиент-сервер</i>	Архитектура клиент-сервер, взаимодействие между компьютерами через интернет. Протоколы.	8	2	2	4	
3.	Удаленный доступ.	Консоль управления на примере Яндекс.Облака.	7	1	2	4	
4.	<i>Работа в ОС Linux</i>	Команды. Утилиты. Настройки базовых компонент.	8	2	2	4	
5.	<i>Серверное ПО. Базы данных. Понятие бекенда.</i>	Серверное ПО. Базы данных. Понятие бекенда.	8	2	2	4	
6.	<i>Службы</i>	Превращение своей программы в сервер. Запуск по расписанию. Логирование. Базовые приемы администрирования. Установка и настройка программ. Практика.	8	2	2	4	
7.	<i>Объединение компонент в ИС</i>	Объединение компонент в ИС. Введение в проектирование ИС. Рассмотрение архитектур существующих ИС.	7	2	1	4	
8.	<i>Параллельная обработка данных</i>	Параллельная обработка данных. Практикум по созданию вычислительных кластеров.	7	1	2	4	
9.	Промежуточная аттестация	Разработка проекта	2	-	-	2	
Всего:			62	14	14	34	

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
<b>Линейная алгебра</b>							
1.	Векторы и матрицы	Определение вектора. Операции с векторами. Длина вектора и расстояние между векторами. Определение матрицы. Собственные векторы, собственные числа. Операции с матрицами - сложение, умножение, диагонализация. Обратная матрица	10	3	2	5	зачет
2.	Сингулярное разложение матрицы	Сингулярное разложение матрицы, область применимости	10	2	3	5	
3.	Функции и логарифмы	Функции, глобальные и локальные экстремумы функции. Определение логарифма. Натуральный, десятичный логарифм. Графическое представление	10	3	2	5	
4.	Производные	Определение предела. Определение производной. Геометрический смысл производной. Полная производная. Частная производная. Производная тригонометрических функций. Производная сложной функции. Градиент	10	2	3	5	
5.	Градиентный спуск	Знакомство с алгоритмом градиентного спуска, область применения, графическое представление	10	2	2	6	
6.	Обратное распространение	Знакомство с алгоритмом обратного распространения, область применения, графическое представление	10	2	2	6	
7.	Промежуточная аттестация	Ответ на вопросы (устно)	2	-	-	2	
Всего:			62	14	14	34	
Итого по Модулю 1:			<b>186</b>	<b>42</b>	<b>42</b>	<b>102</b>	
<b>Модуль 2</b>							
<b>Обязательные дисциплины</b>							
<b>Программирование в R</b>							
1.	Введение	R как язык программирования. Функциональное программирование vs. ООП. Знакомство с R-Studio. Основные типы данных в R (num, int, factor, char, bool, date&time). Основные типы структур данных (vector, data.frame, list). Основные типы библиотек в R и их структура. Объяснение принципов	10	3	2	5	зачет

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
		векторизации. Базовые функции в R. Как загружать данные разных типов из разных источников. Выполнение в R простейших операций, которые слушатели уже умеют на языке Python (ifelse, gsub, grep и проч.)					
2.	Основы R	Как работать с типом данных "data.frame"? Обращение к переменным и строкам, индексация. Фильтрация, выборка переменных. Преобразование данных, long2wide/wide2long: merge, rbind, cbind, aggregate, dcast. Практика объединений и преобразования датафреймов. Повторение пройденного на предыдущем занятии вместе с новым материалом (фильтрация и преобразование переменных с помощью функций ifelse, gsub и проч.).	10	2	3	5	
3.	Библиотека Dplyr	Объяснение общей логики работы в этой библиотеки. Повторение старого материала на dplyr. Основы визуализации данных в ggplot2	10	3	2	5	
4.	Data.table	Преимущества data.table. Основные принципы работы и специфические операторы data.table. Повторение предыдущего материала на базе этой библиотеки.	10	2	3	5	
5.	Циклы, специфика apply family	Повторение материала по data.table. Объяснения циклов в R (предполагается, что слушатели уже знают о них после курса по Python). Работа с разными типами apply family. Введение в работу с листами в R (иллюстрация работы типа apply family - lapply)	10	2	2	6	
6.	Методы кластеризации и кластеризация в R	Основные методы кластеризации и их применение в R: K-Means, Mean-Shift, DBSCAN.	10	2	2	6	
7.	Промежуточная аттестация	Проект	2	-	-	2	
Всего:			62	14	14	34	
<b>Основы статистики</b>							
1.	Описательные статистики и графическая	– классификация переменных: качественные и количественные, дискретные	8	2	2	4	зачет

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.			Форма промежуточной аттестации/ итоговой аттестации	
			Всего	Аудиторная работа по видам учебных занятий			СР <sup>5</sup>
				Л	СЗ <sup>6</sup>		
	репрезентация данных	и непрерывные; – алгебраические локация и разброс: среднее, дисперсия, среднеквадратичное отклонение; – структурные локация и разброс: мода, медиана, квартили, квантили, межквартильный размах; – графические методы представления данных: бокс-плот с выбросами, отличие столбиковой диаграммы от гистограммы, плотность частоты; скошенность данных и способы её измерить.					
2.	Теория вероятностей	– ключевые определения: эксперимент, исходы, вероятностное пространство, события, вероятность; – экспериментальная вероятность как частота, геометрическая интуиция вероятности; – репрезентация вероятностного пространства диаграммами Эйлера и Венна, разница между ними; – формула сложения вероятностей для двух событий; – формула условной вероятности; – независимость событий; – формула полной вероятности; формула Байеса, байесова ловушка и её интуитивное объяснение	8	2	2	4	
3.	Случайные величины	– дискретные и непрерывные случайные величины; – распределение дискретных случайных величин: определение, свойства, табличная запись, кумулятивная функция вероятности; – распределение непрерывных случайных величин: функция плотности вероятности и кумулятивная функция распределения; – математическое ожидание и дисперсия: расчёт для дискретных, общий концепт для непрерывных случайных	9	2	2	5	



№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.			Форма промежуточной аттестации/ итоговой аттестации	
			Всего	Аудиторная работа по видам учебных занятий			СР <sup>5</sup>
				Л	СЗ <sup>6</sup>		
		<p>величин;</p> <ul style="list-style-type: none"> <li>– Биномиальное распределение: расчет коэффициентов, интерпретация отдельных вероятностей и интервалов;</li> <li>некоторые часто употребляемые распределения с примерами</li> </ul>					
4.	Нормальное распределение	<ul style="list-style-type: none"> <li>– статистический бум конца XIX века и полидетерминистическая интерпретация нормального распределения;</li> <li>– понятие семейства распределений и нормировка: формула приведения любой нормально распределенной случайной величины к стандартному нормальному распределению;</li> <li>– задачи на вероятности интервалов для нормально распределенных величин с использованием его симметричности;</li> <li>– типы выборок, концепция статистического вывода о генеральной совокупности по выборке; Центральная предельная теорема в формулировке распределения выборочного среднего для случайных выборок.</li> </ul>	9	2	2	5	
5.	Тестирование гипотез	<ul style="list-style-type: none"> <li>– логика статистического теста, формулировка нулевой и альтернативной гипотез;</li> <li>– уровень статистической значимости;</li> <li>– ошибки 1 и 2 рода;</li> <li>– мощность теста;</li> <li>– двусторонние и односторонние тесты для дискретной биномиально распределенной случайной величины;</li> <li>– двусторонние и односторонние тесты для непрерывной нормально распределенной случайной величины;</li> <li>– стандартная ошибка среднего и её зависимость от размера выборки;</li> <li>– распределение</li> </ul>	9	2	2	5	

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
		Стюдента и учёт размера выборки при проведении т-теста; проверка гипотез для пропорций.					
6.	Доверительные интервалы	– построение доверительного интервала для среднего; – доверительная вероятность: почему степень уверенности нельзя интерпретировать как вероятность; расчёт размера выборки, необходимого для заданной точности оценки.	8	2	2	4	
7.	Зависимые случайные величины	– парные переменные и их отображение на диаграмме рассеяния; – понятие корреляции и её графическая репрезентация, зависимые и независимые переменные; – расчёт ковариации как меры линейной зависимости случайных величин; – нормированная ковариация — коэффициент корреляции Пирсона; – ограничения применимости коэффициента корреляции Пирсона (формальные ограничения метода, независимость от наклона прямой, нечувствительность к кластерам, спонтанные корреляции, квартет Энскомба) – линейная регрессия для двух переменных: расчёт уравнения, интерполяция и экстраполяция; – многомерная линейная регрессия; логистическая регрессия: сведение к линейной через логарифм отношения шансов, почему используется сигмоид для бинарной объясняемой переменной.	9	2	2	5	
8.	Промежуточная аттестация	Опрос по вопросам (устно)	2	-	-	2	
Всего:			62	14	14	34	

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
<b>Дисциплины по выбору<sup>7</sup></b>							
<b>Административные и правовые данные</b>							
1.	Административные данные	Административные данные: природа, возникновение, хранение	8	2	2	4	зачет
2.	Получение информации о существующих данных и получение доступа к ним	Получение информации о существующих данных и получение доступа к ним	8	2	2	4	
3.	Перспективы и особенности административных данных	Перспективы и особенности административных данных	8	2	2	4	
4.	Ограничения административных данных	Ограничения административных данных	8	2	2	4	
5.	Технология контроля качества и реконструкции смысла официальных данных.	Технология контроля качества и реконструкции смысла официальных данных. Интерпретация результатов	10	3	3	4	
6.	Постановка вопросов к данным	Постановка вопросов к данным и дизайны исследований на административных данных.	10	3	3	4	
5.	Промежуточная аттестация	Письменная работа	2	-	-	2	
Всего:			54	14	14	26	
<b>Опросные данные</b>							
1.	Источники опросных данных	Источники опросных данных. Сравнение доступных баз. Особенности опросов в России. Понятие о панельных и кросс-секционных данных.	7	2	2	3	зачет
2.	ГС и выборка, взвешивание.	ГС и выборка, взвешивание. Теоретические проблемы и ограничения опросных данных. CATI, CAPI.	8	2	2	4	
3.	Reliability, validity, and construction of measurement scales.	Reliability, validity, and construction of measurement scales.	8	2	2	4	
4.	Вопросы и шкалы	Вопросы и шкалы, построение индексов, шкалирование.	7	2	2	3	
5.	Пропущенные значения в опросных данных.	Методы работы с пропусками (множественная импутация, ограничения импутации)	7	2	2	3	
6.	Особенности межкультурных опросов	Особенности межкультурных опросов, измерительная инвариантность. Культурно-специфичные вопросы, интерпретация сензитивных	8	2	2	4	

<sup>7</sup> Слушатель обязан выбрать из всех представленных в модулях блоков дисциплин по выбору суммарно четыре дисциплины

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
		вопросов.					
7.	Совмещение опросных данных с другими типами данных	Совмещение опросных данных с другими типами данных	7	2	2	3	
8.	Промежуточная аттестация	Письменная работа	2	-	-	2	
Всего			54	14	14	26	
Итого по Модулю 2:			<b>232</b>	<b>56</b>	<b>56</b>	<b>120</b>	
<b>Модуль 3</b>							
<b>Обязательные дисциплины</b>							
<b>Машинное обучение: введение</b>							
1.	kNN. Общий вид метрических классификаторов, kNN, WkNN, отбор эталонов	Общий вид метрических классификаторов, kNN, WkNN, отбор эталонов. Оценка классификаторов, Precision, Recall, ROC, AUC. Валидация. Кластеризация. kMeans, Mean Shift, DBSCAN, Agglomerative Clustering. Метрики кластеризации. Кластеризация и обучение с частичным привлечением учителя (semi-supervised).	7	2	1	4	зачет
2.	Деревья. Построение деревьев, информационный выигрыш.	Энтропия, критерий Джини, девиация. CART. Прунинг. Ансамбли деревьев, бустинг деревьев. Случайный лес, адаптивный бустинг, градиентный бустинг.	8	2	2	4	
3.	Нейронные сети	Логистическая регрессия, градиентный спуск. Полносвязные сети, обратное распространение градиента. Сверточные нейронные сети и анализ изображений. Свертки, пулинг. LeNet, AlexNet, VGG, ResNet.	8	2	2	4	
4.	Векторные представления и анализ естественного языка	Векторное представление слова. Рекуррентные нейронные сети, LSTM, GRU. Механизм внимания, Трансформер, BERT.	7	1	2	4	
5.	SVM	SVM. Прямая и двойственная задача. Решение двойственной задачи. Линейно неразделимый случай. Ядерный трюк.	8	2	2	4	
6.	Байес. Байесовский классификатор	Оценка распределения признаков. EM-алгоритм. Регрессия. Линейная регрессия, гребневая регрессия, LASSO, LARS. Борьба с выбросами.	7	2	1	4	
7.	Локальный поиск	Байесовская оптимизация. Hill Climb, отжиг, генетический алгоритм.	7	1	2	4	

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
8.	Обучение с подкреплением	Постановка задачи. Функция ценности и функция качества. Q-обучение. Глубокое обучение с подкреплением, DQN, Actor-Critic.	8	2	2	4	
9.	Промежуточная аттестация	Опрос по вопросам (устно)	2	-	-	2	
Всего			62	14	14	34	
<b>Дисциплины по выбору<sup>8</sup></b>							
<b>Разметка данных</b>							
1.	Введение и общая концепция краудсорсинга	Краудсорсинг как особый способ построение бизнес-процессов. Платформа Толока. Ключевые этапы работы. Декомпозиция задачи.	8	2	2	4	зачет
2.	Декомпозиция исходной задачи	Декомпозиция исходной задачи: как перевести исследовательский дизайн в задачу разметки. Отбор исполнителей. Популярные модели агрегации	8	2	2	4	
3.	Инструкция для краудсорсингового задания	Почему инструкция важна. Типовая структура задания. Форматирование. Иллюстрации.	8	2	2	4	
4.	Контроль качества ассессоров: inter-rater reliability et al. (+ EY)	Контроль качества ассессоров: inter-rater reliability et al. (+ EY)	9	2	2	5	
5.	Динамическое перекрытие и прайсинг	Пул заданий. Механизм динамического перекрытия. Подготовка полей выходных данных.	8	2	2	4	
6.	Сценарий «Опрос»	Проблема смещенности ассессоров — можно ли генерализовать результаты.	4	1	1	2	
7.	Сценарий «Расшифровка аудио»	Специфика сценария. Шаблон "Транскрипция аудио". Поле входных данных "аудио".	4	1	1	2	
8.	Практикум: разработка дизайна проекта, использующего «Толоку»		8	2	2	4	
9.	Промежуточная аттестация	Выполнение проекта	2	-	-	2	
Всего			54	14	14	26	
<b>Текстовые данные</b>							
1.	Текстовые данные вокруг нас	Корпусы текстовых данных: КРЯ, Wikipedia, CommonCrawl, Taiga.	7	2	2	3	зачет
2.	Распознавание	Распознавание текстов.	8	2	2	4	

<sup>8</sup> Слушатель обязан выбрать из всех представленных в модулях блоков дисциплин по выбору суммарно четыре дисциплины

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
	текстов	Создание процедуры превращения изображений или PDF в корпус. Tesseract + ImageMagick.					
3.	Чем текстовые данные отличаются от других типов данных	Закон Ципфа. Издержки токенизации.	7	2	2	3	
4.	Извлечение сущностей из текстовых данных	Извлечение сущностей из текстовых данных. Natasha, Pullenti.	7	2	2	3	
5.	Разреженное векторное представление текстовых данных	Разреженное векторное представление текстовых данных. Строковые расстояния. Модель «мешок слов». Byte Pair Encoding. Анализ коллокаций.	8	2	2	4	
6.	Уплотненное векторное представление текстовых данных	Уплотненное векторное представление текстовых данных. Факторизация матриц. SVD, LSA, LDA, BigARTM. Дистрибутивная семантика. От word2vec к контекстуальным эмбедингам. RusVectōrēs.	7	2	2	3	
7.	Поиск по представлениям текстовых данных	Поиск по представлениям текстовых данных. Метод (приближенного) поиска ближайших соседей. Возможности и ограничения семантического поиска	8	2	2	4	
8.	Промежуточная аттестация	Проект	2	-	-	2	
Всего:			54	14	14	26	
Итого по Модулю 3:			<b>170</b>	<b>42</b>	<b>42</b>	<b>86</b>	
<b>Модуль 4</b>							
<b>Дисциплины по выбору<sup>9</sup></b>							
<b>Введение в обработку естественного языка (NLP)</b>							
1.	Введение	Варианты названия дисциплины, особенности, задачи, приложения, методы. Краткая история мировой и советской/российской ОЕЯ. Содержание и "идеология" курса. Ресурсы.	6	2	1	3	зачет
2.	Морфологический анализ	Чем занимается морфология, для чего нужна машинная морфология. Стемминг, лемматизация, частеречная (ЧР) разметка. Алгоритм Портера для английского. Лемматизация для русского языка: mystem, rufomorphy2, АОТ. Словарь	7	2	2	3	

<sup>9</sup> Слушатель обязан выбрать из всех представленных в модулях блоков дисциплин по выбору суммарно четыре дисциплины

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
		Зализняка как источник данных для лемматизации. Как строить гипотезы для незнакомых слов. Задача ЧР-разметки. Скрытые марковские модели для ЧР-разметки. Источники данных. Алгоритм Витерби.					
3.	Классический сценарий информационного поиска, особенности задач информационного поиска	Базовые метрики качества поиска: точность, полнота, F1. Морфология в поиске. Частотные свойства терминов в коллекции: законы Хипса и Ципфа. Модель векторного пространства. Взвешивание терминов: подход tf.idf. Другие задачи ОЕЯ в поиске. Автоматическое реферирование. Постановка задачи, области применения, примеры. Типы рефератов. Методы реферирования текстовых документов. Постобработка рефератов. Оценка, инициативы по оценке методов автоматического реферирования. Веб-сниппеты (рефераты с учетом запроса): особенности задачи, методы. Оценка веб-сниппетов.	7	2	2	3	
4.	Языковые модели (Language Models)	Области применения, ограничения. N-граммы, оценка вероятностей, доступные большие коллекции n-грамм. Оценка языковых моделей, перплексия. Сглаживание (smoothing), откат (backoff) и интерполяция (interpolation). Технические приемы при работе с языковыми моделями масштаба Веба, языковые модели с "памятью". Сглаживаие Лапласа, Гудатьюринга, Кнезера-Нея.	6	1	2	3	
5.	Синтаксические анализ (syntactic parsing)	Два формализма описания синтаксиса: система составляющих (constituency) и дерево зависимостей (dependency). Контекстно-свободные грамматики (КСГ): возможности и ограничения. Вероятностные КСГ. Алгоритм вероятностного синтаксического разбора. Оценка результатов разбора. Лексикализация ВКСГ.	7	2	2	3	
6.	Извлечение	Особенности задачи, источники	6	2	1	3	

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.			Форма промежуточной аттестации/ итоговой аттестации	
			Всего	Аудиторная работа по видам учебных занятий			СР <sup>5</sup>
				Л	СЗ <sup>6</sup>		
	информации из текстовых документов	данных. Именованные сущности (NEs), отношения. Основные подходы. Оценка. Машинный перевод: краткая история, сложности, подходы, приложения. Перевод, основанный на правилах (RBMT) и статистический машинный перевод (СМП, STM). Источники данных для статистического машинного перевода. Выравнивание параллельного корпуса. IBM Models 1, 2. Фразовый СМП. Оценка систем МП: ручная, автоматическая (BLUE).					
7.	Анализ тональности (sentiment analysis): источники данных и различные постановки задачи, примеры приложений.	Подход на основе классификации. Сложности анализа тональности: разнообразие форм выражения, ирония и сарказм, порядок и связя оценок. Использование словарей тонально окрашенных слов. Автоматическое пополнение/создание словарей. Обработка отрицаний. Фразы vs. отдельные слова. Аспекты/атрибуты: автоматическое выделение и списки, составленные вручную. Ресурсы для русского языка.	6	1	2	3	
8.	Семантика, разные подходы и определения	Семантика, разные подходы и определения: логика высказываний, семантический веб, базы знаний, онтологии предметных областей, тезаурусы. Лексическая семантика: омонимы, многозначные слова, синонимы, антонимы, гипонимы/гиперонимы. Семантические словари -- тезаурусы. WordNet: синсеты и отношения. Методы определения семантической близости слов на основе тезауруса. Дистрибутивная семантика: семантическая близость на основе корпуса. Положительная поточечная взаимная информация (PPMI). Вычисления близости на основе структуры словосочетаний. Векторное представление семантики слов на основе на основе нейронных сетей.	7	2	2	3	



№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
9.	Промежуточная аттестация	Письменная работа	2	-	-	2	
Всего			54	14	14	26	
<b>Введение в анализ изображений и видео</b>							
1.	Введение.	Задачи компьютерного зрения. История появлений графических снимков. Устройство цветковых пространств.	6	2	1	3	зачет
2.	Анализ Изображений	Выделение компонент связности. Выделение краев. Математическая морфология.	7	2	2	3	
3.	Зачем сравнивать изображения?	Зачем сравнивать изображения? Поиск. Классификация. Аннотирование.	7	2	2	3	
4.	Локальные признаки	Свойства особых точек. Угловые точки в качестве особых. Геометрические и фотометрические трансформации изображений. Детектор Харриса. Антиэлайзинг. SIFT descriptor.	6	1	2	3	
5.	Параметрические модели	Выбор модели. Оптимальные параметры модели. Метод наименьших квадратов. М-оценки, RANSAC. преобразование Хафа.	7	2	2	3	
6.	Классификация изображений	Бинарная и многоклассовая классификация. Категории объектов. Колоризация. Визуальные слова. Извлечение фрагментов. Методы машинного обучения для предсказания категории.	6	2	1	3	
7.	Нейросетевая классификация изображений	Стохастический градиентный спуск. Теорема Новикова. Приближение функций нейросетью. Нейрон как линейный фильтр. Многослойная нейросеть. Softmax-преобразование. Minibatch SGD. Адаптивные методы подбора шага. Свёрточные сети.	6	1	2	3	
8.	Представления и современные архитектуры сверточных сетей	Пакетная нормализация. Метод обратного распространения ошибки. Эволюция признаков. 1x1 свёртки. AlexNet (2012). GoogLeNet. SqueezeNet.	7	2	2	3	
9.	Промежуточная аттестация	Письменная работа	2	-	-	2	
Всего			54	14	14	26	
Итого по Модулю 4:			<b>108</b>	<b>28</b>	<b>28</b>	<b>52</b>	

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
<b>Модуль 5</b>							
<b>Обязательные дисциплины</b>							
<b>Визуализация данных</b>							
1.	Визуализация данных	Что такое визуализация данных, история визуализации. Гештальтовский принципы восприятия, восприятие форм, восприятие цветов (цвета и культур, цветовые ассоциации).	8	2	2	4	зачет
2.	Где взять данные	TAGS. DataMiner. Простой скрипт на R с парсингом страниц Визуализация данных в онлайн сервисах.	8	2	2	4	
3.	Misleading diagrams	Misleading diagrams, выбор графиков, какой лучше выбрать, убрать лишнее с графиков (data ink ratio). Практика графиков в Tableau.	8	2	2	4	
4.	Сети и карты, онлайн-сервисы	Сети и карты, онлайн-сервисы	9	2	2	5	
5.	Текст в визуализации	Текст в визуализации	9	2	2	5	
6.	Изменяемые графики	Изменяемые графики. Plotly R.	9	2	2	5	
7.	Создание приложений для визуализации	Создание приложений для визуализации, Shiny R.	9	2	2	5	
8.	Промежуточная аттестация	Проект	2	-	-	2	
Всего			62	14	14	34	
<b>Дисциплины по выбору<sup>10</sup></b>							
<b>Геоданные</b>							
1.	Введение. Чему посвящен курс. Используемые программы	Знакомство с основными программами, которые будут использоваться при обучении. Настройка окружения: ▪ Настройка git ▪ Настройка VSC • Настройка QGIS 3	8	3	2	3	зачет
2.	Тема 2 из консоли внутри QGIS 3	Работа с векторными слоями • Создание векторного слоя • Работа с различными форматами данных • Работа с таблицей атрибутов • Редактирование объектов • Добавление/удаление объектов • Добавление/Удаление полей	15	4	4	7	

<sup>10</sup> Слушатель обязан выбрать из всех представленных в модулях блоков дисциплин по выбору суммарно четыре дисциплины

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины час.				Форма промежуточной аттестации/ итоговой аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>5</sup>	
				Л	СЗ <sup>6</sup>		
		<ul style="list-style-type: none"> <li>• Работа с геометрией</li> <li>• Добавление/удаление геометрии</li> <li>• Редактирование геометрии</li> <li>• Работа с выборкой</li> <li>Работа с отображением векторных слоев</li> <li>• Работа с точечными символами</li> <li>• Работа с линейными символами</li> <li>• Работа с полигональными символами</li> <li>Работа с растровыми слоями</li> <li>Работа с интерфейсом программы</li> <li>• Добавление/Удаление пунктов меню</li> <li>• Управление канвасом</li> </ul>					
3.	Написание кастомных выражений в QGIS 3	Основы логических операторов. Проверка выражений. Основные виды выражений. Выражения для фильтрации. Обработка ошибок	14	3	4	7	
4.	Написание кастомных плагинов в QGIS 3 плагина	Структура файлов плагина. Метаданные плагина. Документация/Справка. Структура написания основного кода плагина. Дополнительные инструменты для создания плагина. Plugin Reloader. Публикация	15	4	4	7	
5.	Промежуточная аттестация	Проект	2	-	-	2	
Всего			54	14	14	26	
Итого по Модулю 5:			<b>116</b>	<b>28</b>	<b>28</b>	<b>60</b>	
Итоговая аттестация		Проектная работа	24	-	2	22	экзамен
<b>ИТОГО<sup>11</sup></b>			<b>734</b>	<b>-</b>	<b>-</b>	<b>-</b>	

<sup>11</sup> Общее количество часов по программе профессиональной переподготовки не соответствует сумме часов по модулям, поскольку слушатель обязан выбрать суммарно только четыре из семи представленных в программе дисциплин по выбору (из всех модулей). Общее количество часов по ППП складывается из суммы часов по обязательным дисциплинам, суммы часов дисциплин по выбору (из всех представленных в модулях блоков дисциплин по выбору суммарно выбирается четыре дисциплины) и суммы часов на итоговую аттестацию