

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования



**Пермский национальный исследовательский
политехнический университет**

УТВЕРЖДАЮ

Проректор по образовательной
деятельности

 А.Б. Петроченков

« 06 » марта 20 23 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Дисциплина: Интеллектуальный анализ Web-данных
(наименование)

Форма обучения: очная
(очная/очно-заочная/заочная)

Уровень высшего образования: магистратура
(бакалавриат/специалитет/магистратура)

Общая трудоёмкость: 180 (5)
(часы (ЗЕ))

Направление подготовки: 09.04.01 Информатика и вычислительная техника
(код и наименование направления)

Направленность: Информационные технологии интеллектуальной обработки
больших данных (Big Data)
(наименование образовательной программы)

1. Общие положения

1.1. Цели и задачи дисциплины

Формирование комплекса знаний, умений и навыков в области разработки методов и алгоритмов сбора и интеллектуального анализа web-данных и применения современных языков программирования для реализации предложенных методов и алгоритмов в информационных системах.

1.2. Изучаемые объекты дисциплины

Web-данные; модели, методы и алгоритмы сбора и интеллектуального анализа web-данных; язык программирования Python; модули и библиотеки анализа данных.

1.3. Входные требования

Не предусмотрены

2. Планируемые результаты обучения по дисциплине

Компетенция	Индекс индикатора	Планируемые результаты обучения по дисциплине (знать, уметь, владеть)	Индикатор достижения компетенции, с которым соотнесены планируемые результаты обучения	Средства оценки
ОПК-2	ИД-1ОПК-2	Знает порядок поиска и систематизации web-данных для решения научно-технической задачи в сфере профессиональной деятельности путем их интеллектуального анализа	Знает порядок поиска и систематизации информации об опыте решения научно-технической задачи в сфере профессиональной деятельности	Экзамен
ОПК-2	ИД-2ОПК-2	Умеет формулировать задачу интеллектуального анализа web-данных на основе знания проблем отрасли и опыта их решения	Умеет формулировать научно-техническую задачу в сфере профессиональной деятельности на основе знания проблем отрасли и опыта их решения	Отчёт по практическому занятию
ОПК-2	ИД-3ОПК-2	Владеет навыками выбора методов решения, установления ограничений к решениям задач по интеллектуальному анализу web-данных на основе нормативно-технической документации и знания проблем отрасли и опыта их решения	Владеет навыками выбора методов решения, установления ограничений к решениям научно-технической задачи в сфере профессиональной деятельности на основе нормативно-технической документации и знания проблем отрасли и опыта их решения	Защита лабораторной работы

3. Объем и виды учебной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		3	
1. Проведение учебных занятий (включая проведение текущего контроля успеваемости) в форме:	72	72	
1.1. Контактная аудиторная работа, из них:			
- лекции (Л)	18	18	
- лабораторные работы (ЛР)	24	24	
- практические занятия, семинары и (или) другие виды занятий семинарского типа (ПЗ)	26	26	
- контроль самостоятельной работы (КСР)	4	4	
- контрольная работа			
1.2. Самостоятельная работа студентов (СРС)	72	72	
2. Промежуточная аттестация			
Экзамен	36	36	
Дифференцированный зачет			
Зачет			
Курсовой проект (КП)			
Курсовая работа (КР)			
Общая трудоемкость дисциплины	180	180	

4. Содержание дисциплины

Наименование разделов дисциплины с кратким содержанием	Объем аудиторных занятий по видам в часах			Объем внеаудиторных занятий по видам в часах
	Л	ЛР	ПЗ	СРС
3-й семестр				
Введение в анализ веб-данных	2	0	2	0
Понятие веб-данных. Статические и динамические веб-данные. Проблема сбора, хранения и обработки веб-данных. Соотношение понятий «веб-данные» и «большие данные». Необходимость анализа веб-данных. Вопросы доступа к информации.				
Инструменты анализа веб-данных	4	6	6	18
Обзор методов и инструментов анализа веб-данных. Ограничения существующих инструментов. Анализ веб-данных средствами Python. Библиотеки и фреймворки для сбора, обработки и анализа веб-данных в экосистеме Python.				

Наименование разделов дисциплины с кратким содержанием	Объем аудиторных занятий по видам в часах			Объем внеаудиторных занятий по видам в часах
	Л	ЛР	ПЗ	СРС
Сбор веб-данных	4	6	6	18
Понятие API. Методы и алгоритмы автоматизированного взаимодействия с API. Ограничения и особенности сбора веб-данных через API. Понятие web scraping (скрапинга). Инструменты скрапинга. Правовая сторона вопроса.				
Хранение веб-данных	4	6	6	18
Сериализация и десериализация веб-данных. Особенности веб-форматов. Хранение мультимедийных данных. Использование реляционных, объектно-ориентированных и документоориентированных СУБД для хранения веб-данных. Процедуры преобразования исходных данных.				
Интеллектуальная аналитика веб-данных	4	6	6	18
Постановка гипотез. Подготовка веб-данных для анализа. Библиотеки и фреймворки Python для интеллектуального анализа веб-данных. Построение информационных моделей на основе веб-данных. Применение методов машинного обучения в анализе веб-данных.				
ИТОГО по 3-му семестру	18	24	26	72
ИТОГО по дисциплине	18	24	26	72

Тематика примерных практических занятий

№ п.п.	Наименование темы практического (семинарского) занятия
1	Работа с BeautifulSoup
2	Работа с Selenium
3	Организация и оптимизация хранилища собранных веб-данных
4	Использование pandas и scikit-learn для анализа веб-данных

Тематика примерных лабораторных работ

№ п.п.	Наименование темы лабораторной работы
1	Использование Wikimedia API для получения веб-данных о предметной области
2	Сбор веб-данных со статического веб-сайта путем анализа HTML
3	Сбор веб-данных с SPA путем имитации пользовательского агента
4	Проверка статистической гипотезы путем анализа веб-данных

5. Организационно-педагогические условия

5.1. Образовательные технологии, используемые для формирования компетенций

Проведение лекционных занятий по дисциплине основывается на активном методе обучения, при котором учащиеся не пассивные слушатели, а активные участники занятия, отвечающие на вопросы преподавателя. Вопросы преподавателя нацелены на активизацию процессов усвоения материала, а также на развитие логического мышления. Преподаватель заранее намечает список вопросов, стимулирующих ассоциативное мышление и установление связей с ранее освоенным материалом.

Практические занятия проводятся на основе реализации метода обучения действием: определяются проблемные области, формируются группы. При проведении практических занятий преследуются следующие цели: применение знаний отдельных дисциплин и креативных методов для решения проблем и принятия решений; отработка у обучающихся навыков командной работы, межличностных коммуникаций и развитие лидерских качеств; закрепление основ теоретических знаний.

Проведение лабораторных занятий основывается на интерактивном методе обучения, при котором обучающиеся взаимодействуют не только с преподавателем, но и друг с другом. При этом доминирует активность учащихся в процессе обучения. Место преподавателя в интерактивных занятиях сводится к направлению деятельности обучающихся на достижение целей занятия.

При проведении учебных занятий используются интерактивные лекции, групповые дискуссии, ролевые игры, тренинги и анализ ситуаций и имитационных моделей.

5.2. Методические указания для обучающихся по изучению дисциплины

При изучении дисциплины обучающимся целесообразно выполнять следующие рекомендации:

1. Изучение учебной дисциплины должно вестись систематически.
2. После изучения какого-либо раздела по учебнику или конспектным материалам рекомендуется по памяти воспроизвести основные термины, определения, понятия раздела.
3. Особое внимание следует уделить выполнению отчетов по практическим занятиям, лабораторным работам и индивидуальным комплексным заданиям на самостоятельную работу.
4. Вся тематика вопросов, изучаемых самостоятельно, задается на лекциях преподавателем. Им же даются источники (в первую очередь вновь изданные в периодической научной литературе) для более детального понимания вопросов, озвученных на лекции.

6. Перечень учебно-методического и информационного обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Печатная учебно-методическая литература

№ п/п	Библиографическое описание (автор, заглавие, вид издания, место, издательство, год издания, количество страниц)	Количество экземпляров в библиотеке
1. Основная литература		
1	Мыльников Л. А. Статистические методы интеллектуального анализа данных. Санкт-Петербург : БХВ-Петербург, 2021. 240 с. 19,35 усл. печ. л.	14
2. Дополнительная литература		
2.1. Учебные и научные издания		

1	Доусон М. Програмуем на Python : пер. с англ. Санкт-Петербург [и др.] : Питер, 2021. 414 с. 33,540 усл. печ. л.	6
2.2. Периодические издания		
	Не используется	
2.3. Нормативно-технические издания		
	Не используется	
3. Методические указания для студентов по освоению дисциплины		
	Не используется	
4. Учебно-методическое обеспечение самостоятельной работы студента		
	Не используется	

6.2. Электронная учебно-методическая литература

Вид литературы	Наименование разработки	Ссылка на информационный ресурс	Доступность (сеть Интернет / локальная сеть; авторизованный / свободный доступ)
Дополнительная литература	Бонцанини, М. Анализ социальных медиа на Python. Извлекайте и анализируйте данные из всех уголков социальной паутины на Python [Электронный ресурс]	https://e.lanbook.com/book/108129	локальная сеть; свободный доступ
Дополнительная литература	Сапрыкин, О. Н. Интеллектуальный анализ данных : учебное пособие [Электронный ресурс]	https://e.lanbook.com/book/188906	локальная сеть; свободный доступ
Дополнительная литература	Сузи Р.А. Язык программирования Python : учебное пособие [Электронный ресурс]	http://www.iprbookshop.ru/52211.html	локальная сеть; свободный доступ
Основная литература	Маккинни, У. Python и анализ данных [Электронный ресурс]	https://e.lanbook.com/book/131721	локальная сеть; свободный доступ
Основная литература	Митчелл, Р. Скрапинг веб-сайтов с помощью Python : руководство [Электронный ресурс]	https://e.lanbook.com/book/100903	локальная сеть; свободный доступ
Основная литература	Протодьяконов А.В. Алгоритмы Data Science и их практическая реализация на Python : учебное пособие [Электронный ресурс]	https://www.iprbookshop.ru/124000.html	локальная сеть; свободный доступ

6.3. Лицензионное и свободно распространяемое программное обеспечение, используемое при осуществлении образовательного процесса по дисциплине

Вид ПО	Наименование ПО
Операционные системы	Debian (GNU GPL)

Вид ПО	Наименование ПО
Офисные приложения.	LibreOffice 6.2.4. OpenSource, бесплатен.
Системы управления проектами, исследованиями, разработкой, проектированием, моделированием и внедрением	Protege
Среды разработки, тестирования и отладки	PIP (The Python Package Installer) Free
Среды разработки, тестирования и отладки	PostgreSQL (PostgreSQL License)

6.4. Современные профессиональные базы данных и информационные справочные системы, используемые при осуществлении образовательного процесса по дисциплине

Наименование	Ссылка на информационный ресурс
База данных Elsevier "Freedom Collection"	https://www.elsevier.com/
База данных Scopus	https://www.scopus.com/
База данных научной электронной библиотеки (eLIBRARY.RU)	https://elibrary.ru/
Научная библиотека Пермского национального исследовательского политехнического университета	http://lib.pstu.ru/
Электронно-библиотечная система Лань	https://e.lanbook.com/
Электронно-библиотечная система IPRbooks	http://www.iprbookshop.ru/
Информационные ресурсы Сети КонсультантПлюс	http://www.consultant.ru/
Информационно-справочная система нормативно-технической документации "Техэксперт: нормы, правила, стандарты и законодательства России"	https://техэксперт.сайт/

7. Материально-техническое обеспечение образовательного процесса по дисциплине

Вид занятий	Наименование необходимого основного оборудования и технических средств обучения	Количество единиц
Лабораторная работа	ПК с предустановленным интерпретатором Python версии 3.8 или выше, средой разработки, библиотеками Python для анализа web-данных	20
Лекция	Мультимедийный проектор	1
Практическое занятие	ПК с предустановленным интерпретатором Python версии 3.8 или выше, средой разработки, библиотеками Python для анализа web-данных	20

8. Фонд оценочных средств дисциплины

Описан в отдельном документе

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
**«Пермский национальный исследовательский политехнический
университет»**

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
для проведения промежуточной аттестации обучающихся по дисциплине
«Интеллектуальный анализ Web-данных»
Приложение к рабочей программе дисциплины

Направление подготовки:	09.04.01 Информатика и вычислительная техника
Направленность (профиль) образовательной программы:	Информационные технологии интеллектуальной обработки больших данных (Big Data) Компьютерные системы и сети Автоматизированные системы обработки информации и управления Технологии искусственного интеллекта в социальных и экономических системах
Квалификация выпускника:	«Магистр»
Выпускающая кафедра:	Информационные технологии и автоматизированные системы
Форма обучения:	Очная
Курс: 2	Семестр: 3
Трудоёмкость:	
Кредитов по рабочему учебному плану:	5 ЗЕ
Часов по рабочему учебному плану:	180 ч.
Форма промежуточной аттестации:	
Экзамен: 3 семестр	

Пермь 2023 г.

Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине является частью (приложением) к рабочей программе дисциплины. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине разработан в соответствии с общей частью фонда оценочных средств для проведения промежуточной аттестации основной образовательной программы, которая устанавливает систему оценивания результатов промежуточной аттестации и критерии выставления оценок. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине устанавливает формы и процедуры текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине.

1. Перечень контролируемых результатов обучения по дисциплине, объекты оценивания и виды контроля

Согласно РПД, освоение учебного материала дисциплины запланировано в течение одного семестра (3-го семестра учебного плана) и разбито на 5 учебных модулей. В каждом модуле предусмотрены аудиторские лекционные, практические и лабораторные занятия, а также самостоятельная работа студентов. В рамках освоения учебного материала дисциплины формируются компоненты компетенций *знать, уметь, владеть*, указанные в РПД, которые выступают в качестве контролируемых результатов обучения по дисциплине (таблица 1.1).

Контроль уровня усвоенных знаний, освоенных умений и приобретенных владений осуществляется в рамках текущего, рубежного и промежуточного контроля при изучении теоретического материала, сдаче отчетов по лабораторным работам и экзамена. Виды контроля сведены в таблицу 1.1.

Таблица 1.1. Перечень контролируемых результатов обучения по дисциплине

Контролируемые результаты обучения по дисциплине (ЗУВы)	Вид контроля					
	Текущий		Промежуточный /рубежный		Итоговый	
	С	ТО	ОЛР	Т/КР		Экзамен
Усвоенные знания						
З.1 Знает порядок поиска и систематизации web-данных для решения научно-технической задачи в сфере профессиональной деятельности путем их интеллектуального анализа		ТО				ТВ
Освоенные умения						
У.1 Умеет формулировать задачу интеллектуального анализа web-данных на основе знания проблем отрасли и опыта их решения			ОЛР1 ОЛР4			КЗ
Приобретенные владения						
В.1 Владеет навыками выбора методов решения, установления ограничений к решениям задач по интеллектуальному анализу web-данных на основе нормативно-технической документации и знания проблем отрасли и опыта их решения			ОЛР2 ОЛР3			КЗ

С – собеседование по теме; *ТО* – коллоквиум (теоретический опрос); *КЗ* – кейс-задача (индивидуальное задание); *ОЛР* – отчет по лабораторной работе; *Т/КР* – рубежное тестирование (контрольная работа); *ТВ* – теоретический вопрос; *ПЗ* – практическое задание; *КЗ* – комплексное задание экзамена.

Итоговой оценкой достижения результатов обучения по дисциплине

является промежуточная аттестация в виде экзамена, проводимая с учетом результатов текущего и рубежного контроля.

2. Виды контроля, типовые контрольные задания и шкалы оценивания результатов обучения

Текущий контроль успеваемости имеет целью обеспечение максимальной эффективности учебного процесса, управление процессом формирования заданных компетенций обучаемых, повышение мотивации к учебе и предусматривает оценивание хода освоения дисциплины. В соответствии с Положением о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся по образовательным программам высшего образования – программам бакалавриата, специалитета и магистратуры в ПНИПУ предусмотрены следующие виды и периодичность текущего контроля успеваемости обучающихся:

- входной контроль, проверка исходного уровня подготовленности обучаемого и его соответствия предъявляемым требованиям для изучения данной дисциплины;

- текущий контроль усвоения материала (уровня освоения компонента «знать» заданных компетенций) на каждом групповом занятии и контроль посещаемости лекционных занятий;

- промежуточный и рубежный контроль освоения обучаемыми отдельных компонентов «знать», «уметь» заданных компетенций путем компьютерного или бланочного тестирования, контрольных опросов, контрольных работ (индивидуальных домашних заданий), защиты отчетов по лабораторным работам, рефератов, эссе и т.д.

Рубежный контроль по дисциплине проводится на следующей неделе после прохождения модуля дисциплины, а промежуточный – во время каждого контрольного мероприятия внутри модулей дисциплины;

- межсессионная аттестация, единовременное подведение итогов текущей успеваемости не менее одного раза в семестр по всем дисциплинам для каждого направления подготовки (специальности), курса, группы;

- контроль остаточных знаний.

2.1. Текущий контроль усвоения материала

Текущий контроль усвоения материала в форме собеседования или выборочного теоретического опроса студентов проводится по каждой теме. Результаты по 4-балльной шкале оценивания заносятся в книжку преподавателя и учитываются в виде интегральной оценки при проведении промежуточной аттестации.

2.2. Рубежный (промежуточный) контроль

Рубежный (промежуточный) контроль для комплексного оценивания усвоенных знаний, усвоенных умений и приобретенных владений (таблица 1.1) проводится в форме защиты лабораторных работ и выполнения кейс-задач.

2.2.1. Защита лабораторных работ

Всего запланировано 4 лабораторных работы. Типовые темы лабораторных работ приведены в РПД.

Защита лабораторной работы проводится индивидуально каждым студентом или группой студентов. Типовые шкала и критерии оценки приведены в общей части ФОС образовательной программы.

2.3. Промежуточная аттестация (итоговый контроль)

Допуск к промежуточной аттестации осуществляется по результатам текущего и рубежного контроля. Условиями допуска являются успешная сдача всех лабораторных работ и положительная интегральная оценка по результатам текущего и рубежного контроля.

Промежуточная аттестация, согласно РПД, проводится в виде экзамена по дисциплине устно по билетам. Билет содержит теоретические вопросы (ТВ) для проверки усвоенных знаний и комплексные задания (КЗ) для контроля уровня освоенных умений и приобретенных владений всех заявленных компетенций.

Билет формируется таким образом, чтобы в него попали вопросы и практические задания, контролирующие уровень сформированности всех заявленных компетенций. Форма билета представлена в общей части ФОС образовательной программы.

2.3.1. Типовые вопросы и задания для экзамена по дисциплине

Типовые вопросы для контроля усвоенных знаний:

1. Понятие веб-данных. Статические и динамические веб-данные.
2. Понятие API.
3. Понятие web scraping (скрапинга).
4. Сериализация и десериализация веб-данных.
5. Использование реляционных СУБД для хранения веб-данных.
6. Использование объектно-ориентированных СУБД для хранения веб-данных.
7. Использование документоориентированных СУБД для хранения веб-данных.
8. Подготовка веб-данных для анализа.
9. Построение информационных моделей на основе веб-данных.
10. Применение методов машинного обучения в анализе веб-данных.
11. Библиотеки и фреймворки Python для сбора и обработки веб-данных.
12. Библиотеки и фреймворки Python для интеллектуального анализа веб-данных.

Типовые комплексные задания для контроля освоенных умений и владений:

1. Описать методику сбора веб-данных с динамического SPA-приложения
2. Описать план проверки статистической гипотезы с использованием интеллектуального анализа веб-данных
3. Предложить методику автоматизированного описания заданной предметной области при помощи анализа веб-данных

Полный перечень теоретических вопросов и комплексных заданий в форме

утвержденного комплекта экзаменационных билетов хранится на выпускающей кафедре.

2.3.2. Шкалы оценивания результатов обучения на экзамене

Оценка результатов обучения по дисциплине в форме уровня сформированности компонентов *знать, уметь, владеть* заявленных компетенций проводится по 4-х балльной шкале оценивания путем выборочного контроля во время экзамена.

Типовые шкала и критерии оценки результатов обучения при сдаче экзамена для компонентов *знать, уметь и владеть* приведены в общей части ФОС образовательной программы.

3. Критерии оценивания уровня сформированности компонентов и компетенций

3.1. Оценка уровня сформированности компонентов компетенций

При оценке уровня сформированности компетенций в рамках выборочного контроля при экзамене считается, что *полученная оценка за компонент проверяемой в билете компетенции обобщается на соответствующий компонент всех компетенций, формируемых в рамках данной учебной дисциплины.*

Типовые критерии и шкалы оценивания уровня сформированности компонентов компетенций приведены в общей части ФОС образовательной программы.

3.2. Оценка уровня сформированности компетенций

Общая оценка уровня сформированности всех компетенций проводится путем агрегирования оценок, полученных студентом за каждый компонент формируемых компетенций, с учетом результатов текущего и рубежного контроля в виде интегральной оценки по 4-х балльной шкале. Все результаты контроля заносятся в оценочный лист и заполняются преподавателем по итогам промежуточной аттестации.

Форма оценочного листа и требования к его заполнению приведены в общей части ФОС образовательной программы.

При формировании итоговой оценки промежуточной аттестации в виде экзамена используются типовые критерии, приведенные в общей части ФОС образовательной программы.