
УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«25» мая 2026 г.
Протокол № 2

**Рабочая программа дисциплины (модуля)
«Парсинг данных»**

Направление подготовки: 02.04.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Машинное обучение в индустрии

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Срок освоения программы: 2 года

Год набора: 2026

**Москва
2026**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения.....	5
3. Тематический план.....	8
4. Содержание дисциплины (модуля).....	8
5. Учебно-методическое обеспечение	10
6. Материально-техническое обеспечение	10
7. Методические и оценочные материалы	12

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Парсинг данных» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – магистратура по направлению 02.04.01 Математика и компьютерные науки, профиль Машинное обучение в промышленности, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 810 от 23.08.2017 года.

Изучение дисциплины (модуля) «Парсинг данных» позволяет студентам освоить актуальные методы работы с данными, что является необходимым в условиях современного информационного общества. Кроме того, навыки парсинга данных способствуют улучшению аналитических способностей и повышению конкурентоспособности на рынке труда.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки магистратуры по направлению 02.04.01 Математика и компьютерные науки, профиль Машинное обучение в промышленности и входит в Блок Факультативные дисциплины.

Дисциплина (модуль) изучается на 2 курсе в 3 семестре.

Цель изучения дисциплины (модуля): овладение студентами навыками извлечения, обработки и анализа данных из различных источников с использованием современных инструментов и технологий.

Задачи изучения дисциплины (модуля):

- формирование знаний основных понятий и принципов парсинга данных;
- формирование умений работать с HTML, CSS и другими форматами данных;
- формирование навыков использования библиотек Python для парсинга, таких как BeautifulSoup и lxml;
- формирование умений извлекать данные из API и обрабатывать полученную информацию;
- формирование навыков работы с регулярными выражениями для обработки текстовых данных;
- формирование умений хранить и обрабатывать данные в различных форматах (CSV, JSON, базы данных);
- формирование способности анализировать и интерпретировать собранные данные для принятия обоснованных решений.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- основные принципы работы с веб-страницами, структурой html и css;
- механизмы взаимодействия с api и форматами данных (json, xml, csv);
- концепцию регулярных выражений и их применение при обработке текста;
- подходы к хранению и последующей обработке извлечённых данных.

уметь:

- работать с библиотеками python для извлечения данных из html-документов; извлекать и преобразовывать данные из различных api; применять регулярные выражения для поиска и очистки текстовой информации; сохранять собранные данные в файлы и базы данных с учётом структуры и целостности.

владеть:

- навыком эффективно извлекать данные с динамических и статических веб-ресурсов;
- навыком обрабатывать и структурировать неоднородные данные;
- навыком автоматизировать процессы сбора информации;

— навыком анализировать и визуализировать результаты парсинга для принятия решений.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-6.	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1.	Знает основные методы самооценки и анализа своей деятельности, а также принципы управления временем и целеполагания
		УК-6.2.	Умеет ставить реалистичные и достижимые цели, определять приоритеты в своей деятельности, а также разрабатывать и внедрять планы по совершенствованию своих навыков и компетенций на основе полученной самооценки
		УК-6.3.	Имеет практический опыт применения методов самооценки в своей профессиональной деятельности, включая участие в тренингах, семинарах и проектах, направленных на развитие личной эффективности и профессионального роста
ОПК-1.	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики	ОПК-1.1.	Знает основные методы и подходы к решению задач прикладной и компьютерной математики, включая алгоритмы, математическое моделирование и теорию оптимизации, а также современные инструменты и технологии, используемые в этой области
		ОПК-1.2.	Умеет анализировать и формулировать математические задачи, применять соответствующие методы и алгоритмы для их решения, а также интерпретировать и

			представлять результаты в понятной и доступной форме
		ОПК-1.3.	Имеет практический опыт работы над проектами или исследованиями в области прикладной и компьютерной математики, включая участие в конкурсах, олимпиадах или научных публикациях, где были решены актуальные и значимые задачи
ПК-1.	Способен определять общие формы и закономерности области машинного обучения в индустрии	ПК-1.1.	Знает основные теоретические концепции и принципы, относящиеся к области машинного обучения в индустрии, а также ключевые закономерности и модели, которые помогают в анализе и интерпретации данных
		ПК-1.2.	Умеет проводить систематический анализ области разработки, выявлять и формулировать общие закономерности и тенденции, а также применять методы исследования для получения новых знаний и понимания
		ПК-1.3.	Имеет практический опыт работы в области машинного обучения в индустрии, включая участие в научных проектах, исследованиях или практических заданиях, где были выявлены и описаны общие формы и закономерности
ПК-3.	Способен решать задачи профессиональной деятельности, формулировать результат, увидеть следствия полученного результата	ПК-3.1.	Знает основные принципы и методы решения задач профессиональной деятельности, а также способы формулирования и представления результатов, включая анализ последствий и их значимость в контексте проекта
		ПК-3.2.	Умеет применять математические и компьютерные методы для решения конкретных задач,

			формулировать четкие и обоснованные результаты, а также анализировать их последствия для дальнейших действий и решений
		ПК-3.3.	Имеет практический опыт в решении профессиональных задач, включая участие в проектах, где были получены результаты и проанализированы их следствия, что способствовало принятию обоснованных решений
ПК-6.	Способен разрабатывать программное обеспечение для решения прикладных задач в сфере машинного обучения в индустрии	ПК-6.1.	Знает основные языки программирования, методы разработки программного обеспечения, а также принципы проектирования и архитектуры программных систем, применяемых в машинном обучении в индустрии
		ПК-6.2.	Умеет анализировать прикладные задачи, разрабатывать алгоритмы и реализовывать их в виде программного обеспечения, используя современные инструменты и технологии, а также проводить тестирование и отладку созданных решений
		ПК-6.3.	Имеет практический опыт разработки программного обеспечения в рамках реальных проектов, включая участие в командах, где были успешно реализованы решения для конкретных прикладных задач в сфере профессиональной деятельности

3. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостояте льная работа	
		Лекции	Семинары (практичес кие занятия)			
1	Введение в парсинг данных	1	1		3	Домашнее задание
2	Основные концепции парсинга данных	1	1		3	Домашнее задание
3	Инструменты и библиотеки для парсинга	1	1		4	Домашнее задание
4	Разбор HTML и CSS для парсинга веб- страниц	1	1		4	Домашнее задание
5	Использование Python для парсинга данных	2	2		4	Домашнее задание
6	Работа с библиотеками Beautiful Soup и lxml	2	2		4	Домашнее задание
7	Парсинг данных с использованием API	2	2		4	Домашнее задание
8	Обработка и хранение данных после парсинга	2	2		4	Домашнее задание
9	Основы работы с регулярными выражениями	1	1		4	Домашнее задание
10	Практическое занятие: проект по парсингу данных	1	1	4	4	Домашнее задание
11	Этические и правовые аспекты парсинга данных	1	1		4	Домашнее задание
	Зачет					Проект
Итого:		15	15	4	42	
Объем дисциплины (модуля) (в ак. ч.)		76				
Объем дисциплины (модуля) (в зач. ед.)		2				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Введение в парсинг данных	Определение парсинга данных. Зачем нужен парсинг данных. Области применения
2	Основные концепции парсинга данных	Структурированные и неструктурированные данные. Форматы данных: HTML, XML, JSON. Принципы работы парсеров
3	Инструменты и библиотеки для парсинга	Обзор популярных инструментов. Установка и настройка окружения
4	Разбор HTML и CSS для парсинга веб-страниц	Структура HTML-документа. Основы CSS-селекторов

5	Использование Python для парсинга данных	Установка Python и необходимых библиотек. Основы работы с Python для парсинга
6	Работа с библиотеками BeautifulSoup и lxml	Установка и использование BeautifulSoup. Установка и использование lxml. Сравнение библиотек
7	Парсинг данных с использованием API	Что такое API. Принципы работы с RESTful API. Примеры парсинга данных из API
8	Обработка и хранение данных после парсинга	Форматы хранения данных: CSV, JSON, базы данных. Основы работы с SQLite
9	Основы работы с регулярными выражениями	Что такое регулярные выражения. Синтаксис и применение в парсинге
10	Практическое занятие: проект по парсингу данных	Выбор проекта. Реализация проекта. Презентация результатов
11	Этические и правовые аспекты парсинга данных	Законодательство о защите данных. Этические нормы парсинга

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Чернышев, С. А. Основы программирования на Python : учебник для вузов / С. А. Чернышев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 349 с. — (Высшее образование). — ISBN 978-5-534-17139-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/567821>.

Дополнительная литература:

1. Федоров, Д. Ю. Программирование на python : учебное пособие для вузов / Д. Ю. Федоров. — 6-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 187 с. — (Высшее образование). — ISBN 978-5-534-19666-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/556864>.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое

Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Парсинг данных» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, проект, домашние задания, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар – это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Проект – исследовательская работа по дисциплине (модулю) и презентация результатов.

Для успешной подготовки к проекту: четко определите цели и задачи проекта, распределите роли и обязанности между участниками, а также установите сроки выполнения каждой части работы. Регулярно проводите встречи для обсуждения прогресса и решения возникающих вопросов.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу

с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Парсинг данных»

Оценивание уровня учебных достижений обучающихся по дисциплине (модулю) осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *зачета*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Зачтено	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Отлично	Зачтено	
8	Отлично	Зачтено	
7	Хорошо	Зачтено	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая
6	Хорошо	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	
3	Не сдан	Не зачтено	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

Дисциплина (модуль) «Парсинг данных» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашнее задание	40%	3	Набор задач по темам недели
Зачет	60%	1	Проект – исследовательская работа по дисциплине (модулю) и презентация результатов

Формула расчёта итоговой оценки по дисциплине (модулю) «Академия Backend»: $0,4 \times \text{Домашнее задание} + 0,6 \times \text{Зачет}$.

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание: Использование Python для парсинга данных

1. Установите Python на ваш компьютер и проверьте его корректную работу, выполнив простую программу, которая выводит "Hello, World!".
2. Установите необходимые библиотеки для парсинга данных, такие как requests и BeautifulSoup4, и создайте скрипт, который загружает веб-страницу и выводит её содержимое.
3. Напишите программу на Python, которая использует библиотеку requests для получения HTML-кода веб-страницы и выводит заголовки (теги <h1>, <h2>, и т.д.) в консоль.

4. Создайте скрипт, который парсит текст из определённого элемента на веб-странице (например, `<p>` или `<div>`) и сохраняет его в текстовый файл.
5. Разработайте простую программу, которая принимает URL веб-страницы в качестве входного параметра и выводит количество слов на этой странице.

Домашнее задание: Работа с библиотеками BeautifulSoup и lxml

1. Установите библиотеку BeautifulSoup и напишите скрипт, который загружает HTML-код веб-страницы и извлекает все ссылки (теги `<a>`).
2. Установите библиотеку lxml и создайте программу, которая загружает XML-файл и извлекает определённые данные (например, элементы с определённым тегом).
3. Сравните BeautifulSoup и lxml по производительности, написав один и тот же парсер для одной и той же веб-страницы с использованием обеих библиотек. Запишите время выполнения для каждой из реализаций.
4. Напишите скрипт, который использует BeautifulSoup для извлечения изображений (теги `<img>`) с веб-страницы и сохраняет их URL в CSV-файл.
5. Создайте программу, которая использует lxml для парсинга HTML-страницы и выводит все уникальные классы CSS, используемые на странице.

Домашнее задание: Парсинг данных с использованием API и обработка данных после парсинга

1. Изучите, что такое API, и напишите краткое описание, объясняющее его основные функции и назначение.
2. Создайте скрипт на Python, который использует библиотеку requests для получения данных из публичного RESTful API (например, API погоды или API для получения информации о фильмах) и выводит результаты в консоль.
3. Напишите программу, которая парсит данные из API и сохраняет их в формате JSON. Убедитесь, что структура данных корректна.
4. Разработайте скрипт, который извлекает данные из API и сохраняет их в формате CSV, включая заголовки столбцов.
5. Создайте базу данных SQLite и напишите программу, которая сохраняет данные, полученные из API, в таблицу базы данных, включая обработку возможных ошибок.

Примерное описание проекта

Проект по парсингу данных

Цель проекта: Создание приложения для парсинга данных из различных источников (веб-страниц и API) с последующей обработкой и хранением этих данных в структурированном виде.

Этапы реализации проекта:

1. Изучение основных концепций парсинга данных

- Исследование различий между структурированными и неструктурированными данными.
- Ознакомление с форматами данных: HTML, XML, JSON.
- Понимание принципов работы парсеров.

2. Выбор инструментов и библиотек для парсинга

- Обзор популярных инструментов для парсинга (Beautiful Soup, lxml, Scrapy и др.).
- Установка и настройка окружения (Python, необходимые библиотеки).
- 3. Разбор HTML и CSS для парсинга веб-страниц**
 - Изучение структуры HTML-документа, включая основные теги и атрибуты.
 - Ознакомление с основами CSS-селекторов для выборки элементов на странице.
- 4. Использование Python для парсинга данных**
 - Установка Python и необходимых библиотек (requests, BeautifulSoup, lxml).
 - Написание простых скриптов для получения и парсинга данных с веб-страниц.
- 5. Работа с библиотеками BeautifulSoup и lxml**
 - Установка и использование BeautifulSoup для парсинга HTML.
 - Установка и использование lxml для парсинга XML и HTML.
 - Сравнение возможностей и производительности обеих библиотек.
- 6. Парсинг данных с использованием API**
 - Изучение, что такое API и принципы работы с RESTful API.
 - Написание скриптов для получения данных из публичных API и их парсинга.
- 7. Обработка и хранение данных после парсинга**
 - Сохранение полученных данных в различных форматах (CSV, JSON).
 - Использование SQLite для хранения данных в базе данных.
 - Создание интерфейса для отображения и анализа собранных данных.

Критерии оценивания:

- 1. Техническая реализация (40%)**
 - Корректность работы парсеров.
 - Эффективность кода (время выполнения, использование памяти).
 - Чистота и читаемость кода (комментарии, структура).
- 2. Качество собранных данных (30%)**
 - Полнота и точность собранной информации.
 - Корректность форматов данных (JSON, CSV и т.д.).
- 3. Документация и презентация (20%)**
 - Наличие документации по проекту (инструкции по установке, использованию).
 - Презентация результатов работы (демонстрация приложения, примеры использования).
- 4. Креативность и оригинальность (10%)**
 - Уникальные решения и подходы к задаче.
 - Дополнительные функции или улучшения, реализованные в проекте.

Дополнительные рекомендации:

- Регулярно сохраняйте и коммитьте изменения в вашем коде с помощью систем контроля версий (например, Git).

Электронный документ

- Протестируйте ваш проект на различных данных, чтобы убедиться в его надежности и универсальности.
- Обсуждайте прогресс и возникающие проблемы с командой или наставником для получения обратной связи.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Как приоритизировать источники данных при парсинге?	По ценности данных, доступности и риску блокировки	УК-6
2.	Как планировать время на обработку больших объемов данных?	Оценка объема, выбор инструментов, буфер на ошибки	УК-6
3.	Как оценить эффективность написанного парсера?	По скорости, стабильности, полноте данных и затратам ресурсов	УК-6
4.	Как управлять рисками изменения структуры сайта?	Мониторинг, алерты, модульная архитектура парсера	УК-6
5.	Как оценить вычислительную сложность алгоритма парсинга?	Через анализ временной и пространственной сложности (O-нотация)	ОПК-1
6.	Какая структура данных лучше всего подходит для хранения DOM-дерева?	Древовидная структура (Tree)	ОПК-1
7.	Как математически моделировать частоту запросов для избежания бана?	Через распределение вероятностей и лимиты rate limiting	ОПК-1
8.	Что такое нормализация данных после парсинга?	Процесс приведения данных к единому формату и структуре	ОПК-1
9.	Как парсенные данные используются в ML-моделях?	Для обучения, валидации и тестирования моделей	ПК-1
10.	Что такое дата-дрифт источников данных?	Изменение распределения или структуры данных со	ПК-1

		временем	
11.	Какие этические стандарты сбора данных существуют в индустрии?	Соблюдение robots.txt, законов о персональных данных	ПК-1
12.	Как обеспечить воспроизводимость датасета для ML?	Версионирование данных, фиксация времени сбора, снапшоты	ПК-1
13.	Что такое парсинг данных?	Процесс извлечения и обработки данных из различных источников	ПК-3
14.	Какой протокол используется для взаимодействия с API?	HTTP/HTTPS	ПК-3
15.	Какой метод HTTP используется для получения данных?	GET	ПК-3
16.	Что означает аббревиатура API?	Интерфейс программирования приложений	ПК-3
17.	Какая библиотека Python чаще используется для парсинга HTML?	Beautiful Soup	ПК-6
18.	Какой модуль Python используется для отправки HTTP-запросов?	Requests	ПК-6
19.	Какая база данных встроена в Python для хранения результатов?	SQLite	ПК-6
20.	Какой статус-код HTTP означает успешный запрос?	200 OK	ПК-6