

---

**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:** Программная инженерия в  
ритейле

**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.....                           | 7  |
| 4. Содержание дисциплины (модуля).....              | 7  |
| 5. Учебно-методическое обеспечение .....            | 9  |
| 6. Материально-техническое обеспечение .....        | 9  |
| 7. Методические и оценочные материалы .....         | 11 |

## **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 Программная инженерия в ретейле, 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 Программная инженерия в ретейле 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-2.      | Способен создавать и исследовать новые математические модели в естественных науках, совершенствовать и разрабатывать концепции, теории и методы | ОПК-2.1.              | Знает основные математические модели и методы, используемые в естественных науках, включая статистическое моделирование, дифференциальные уравнения и численные методы, а также современные подходы к исследованию и анализу данных в области программной инженерии в ретейле |
|             |                                                                                                                                                 | ОПК-2.2.              | Умеет разрабатывать и адаптировать математические модели для решения конкретных проблем в естественных науках, проводить их анализ и верификацию, а также интерпретировать полученные результаты в контексте научных исследований                                             |

|       |                                                                                                                                                        |          |                                                                                                                                                                                                                                                                            |
|-------|--------------------------------------------------------------------------------------------------------------------------------------------------------|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|       |                                                                                                                                                        | ОПК-2.3. | Имеет практический опыт создания и исследования математических моделей в рамках прикладных научных или инновационных проектов в сфере ретейла, включая участие в публикациях, конференциях или коллаборациях, где были разработаны и апробированы новые концепции и методы |
| ПК-3. | Способен применять методы анализа данных и математического моделирования для поддержки управленческих и операционных решений в информационных системах | ПК-3.1.  | Знает методы математического моделирования, анализа данных и оценки эффективности цифровых решений в рамках функционирования и развития информационных систем                                                                                                              |
|       |                                                                                                                                                        | ПК-3.2.  | Умеет анализировать практические ситуации, разрабатывать и применять математические и аналитические модели, а также интерпретировать полученные результаты с учётом показателей эффективности функционирования и результативности процессов                                |
|       |                                                                                                                                                        | ПК-3.3.  | Имеет практический опыт решения прикладных задач, связанных с оптимизацией процессов обработки данных, анализом поведения пользователей и повышением эффективности функционирования информационных систем                                                                  |

### 3. Тематический план

| №<br>п/п                                  | Наименование<br>раздела дисциплины<br>(модуля)        | Трудоемкость, академические часы |                                           |          |                               | ТКУ<br>(текущий<br>контроль<br>успеваемости) |
|-------------------------------------------|-------------------------------------------------------|----------------------------------|-------------------------------------------|----------|-------------------------------|----------------------------------------------|
|                                           |                                                       | Очная форма                      |                                           |          |                               |                                              |
|                                           |                                                       | Аудиторная работа                |                                           | Контроль | Самостояте<br>льная<br>работа |                                              |
|                                           |                                                       | Лекции                           | Семинары<br>(практичес<br>кие<br>занятия) |          |                               |                                              |
| 1                                         | Введение в парсинг<br>данных                          | 1                                | 1                                         |          | 3                             | Домашнее задание                             |
| 2                                         | Основные концепции<br>парсинга данных                 | 1                                | 1                                         |          | 3                             | Домашнее задание                             |
| 3                                         | Инструменты и<br>библиотеки для<br>парсинга           | 1                                | 1                                         |          | 4                             | Домашнее задание                             |
| 4                                         | Разбор HTML и CSS<br>для парсинга веб-<br>страниц     | 1                                | 1                                         |          | 4                             | Домашнее задание                             |
| 5                                         | Использование Python<br>для парсинга данных           | 2                                | 2                                         |          | 4                             | Домашнее задание                             |
| 6                                         | Работа с<br>библиотеками<br>Beautiful Soup и lxml     | 2                                | 2                                         |          | 4                             | Домашнее задание                             |
| 7                                         | Парсинг данных с<br>использованием API                | 2                                | 2                                         |          | 4                             | Домашнее задание                             |
| 8                                         | Обработка и хранение<br>данных после<br>парсинга      | 2                                | 2                                         |          | 4                             | Домашнее задание                             |
| 9                                         | Основы работы с<br>регулярными<br>выражениями         | 1                                | 1                                         |          | 4                             | Домашнее задание                             |
| 10                                        | Практическое занятие:<br>проект по парсингу<br>данных | 1                                | 1                                         | 4        | 4                             | Домашнее задание                             |
| 11                                        | Этические и правовые<br>аспекты парсинга<br>данных    | 1                                | 1                                         |          | 4                             | Домашнее задание                             |
|                                           | Зачет                                                 |                                  |                                           |          |                               | Проект                                       |
| Итого:                                    |                                                       | 15                               | 15                                        | 4        | 42                            |                                              |
| Объем дисциплины (модуля)<br>(в ак. ч.)   |                                                       | 76                               |                                           |          |                               |                                              |
| Объем дисциплины (модуля)<br>(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<br>(издания, курса, документа)                            | Ссылка                                                                          |
|----|--------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| 1. | Научная электронная библиотека elibrary.ru библиотека                          | <a href="https://elibrary.ru/defaultx.asp">https://elibrary.ru/defaultx.asp</a> |
| 2. | База данных для IT-специалистов                                                | <a href="https://habr.com">https://habr.com</a>                                 |
| 3. | База данных ScienceDirect                                                      | <a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>       |
| 4. | Официальный сайт Министерства науки и высшего образования Российской Федерации | <a href="https://minobrnauki.gov.ru/">https://minobrnauki.gov.ru/</a>           |
| 5. | Федеральный портал «Российское образование»                                    | <a href="https://www.edu.ru/">https://www.edu.ru/</a>                           |
| 6. | Информационная система "Единое окно доступа к образовательным ресурсам"        | <a href="http://window.edu.ru/">http://window.edu.ru/</a>                       |
| 7. | Единая коллекция цифровых образовательных ресурсов                             | <a href="http://school-collection.edu.ru/">http://school-collection.edu.ru/</a> |
| 8. | Федеральный центр информационно - образовательных ресурсов                     | <a href="http://fcior.edu.ru/">http://fcior.edu.ru/</a>                         |

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 |
|---------------------------------------------------------------------|---------------|------------------------------------------|
| <b>Операционные системы:</b>                                        |               |                                          |
| Microsoft Imagine (Windows Client, Server)                          | зарубежное    | лицензионное                             |
| <b>Браузеры:</b>                                                    |               |                                          |
| Яндекс.Браузер                                                      | отечественное | свободно распространяемое                |
| Google Chrome                                                       | зарубежное    | свободно распространяемое                |
| <b>Офисные приложения:</b>                                          |               |                                          |
| Microsoft Imagine (Visio, OneNote)                                  | зарубежное    | лицензионное                             |
| TeXstudio                                                           | зарубежное    | свободно распространяемое                |
| Adobe Acrobat Reader                                                | зарубежное    | свободно распространяемое                |
| <b>Программное обеспечение для планирования и учета времени:</b>    |               |                                          |
| Toggle app                                                          | зарубежное    | свободно распространяемое                |
| <b>Системы управления проектами:</b>                                |               |                                          |
| Microsoft Imagine (Project)                                         | зарубежное    | лицензионное                             |
| <b>Системы управления базами данных:</b>                            |               |                                          |
| Microsoft Imagine (SQL Server)                                      | зарубежное    | лицензионное                             |
| <b>Системы резервного копирования (backup):</b>                     |               |                                          |
| Acronis Backup Advanced for HyperV                                  | зарубежное    | лицензионное                             |
| <b>Справочно-правовые системы:</b>                                  |               |                                          |
| КонсультантПлюс: справочно-правовая система                         | отечественное | лицензионное                             |
| <b>Средства антивирусной защиты:</b>                                |               |                                          |
| Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition | отечественное | лицензионное                             |
| <b>Среды разработки:</b>                                            |               |                                          |
| 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 |
| <b>Пакеты программных средств и библиотек:</b>           |            |                           |
| AutoPsy                                                  | зарубежное | свободно распространяемое |
| Interactive Disassembler (IDA)                           | зарубежное | свободно распространяемое |
| <b>Системы управления библиографической информацией:</b> |            |                           |
| Zotero                                                   | зарубежное | свободно распространяемое |
| <b>Сервисы и службы:</b>                                 |            |                           |
| 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:

| <b>Десятибалльная оценка</b> | <b>Пятибалльная оценка</b> | <b>Оценка за зачет</b> | <b>Общая характеристика результата обучения по дисциплине (модулю)</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|------------------------------|----------------------------|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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.    | Что такое парсинг данных?<br>а) Процесс создания баз данных<br>б) Процесс извлечения и обработки данных из различных источников<br>в) Процесс написания веб-сайтов<br>г) Процесс сжатия данных | б              | ПК-3        |
| 2.    | Какой из перечисленных форматов данных является структурированным?<br>а) HTML<br>б) Не структурированный текст<br>в) JSON<br>г) Произвольный текст                                             | в              | ОПК-2       |
| 3.    | Какая библиотека Python чаще используется для парсинга HTML и XML?<br>а) NumPy<br>б) BeautifulSoup<br>в) Matplotlib<br>г) Pandas                                                               | б              | УК-6        |
| 4.    | Назовите один из основных форматов хранения данных после парсинга.                                                                                                                             | JSON           | ПК-3        |
| 5.    | Как называется инструмент, который позволяет получать данные с веб-страниц, используя Python?                                                                                                  | Beautiful Soup | ОПК-2       |
| 6.    | Какой протокол чаще всего используется для взаимодействия с API в парсинге?                                                                                                                    | HTTP           | ПК-3        |
| 7.    | Как называется язык разметки, который описывает структуру веб-страниц?                                                                                                                         | HTML           | ОПК-2       |
| 8.    | Какая база данных встроена в Python и часто используется для хранения результатов парсинга?                                                                                                    | SQLite         | УК-6        |
| 9.    | Какой метод HTTP-запроса используется для получения данных с сервера?<br>а) POST<br>б) PUT<br>в) GET<br>г) DELETE                                                                              | в              | ПК-3        |
| 10.   | Какой формат данных, помимо JSON, часто используется для обмена структурированной информацией в API?<br>а) CSV<br>б) XML<br>в) PDF<br>г) TXT                                                   | б              | ОПК-2       |

|     |                                                                                                                   |                                                |       |
|-----|-------------------------------------------------------------------------------------------------------------------|------------------------------------------------|-------|
| 11. | Какой модуль Python используется для отправки HTTP-запросов?<br>а) os<br>б) requests<br>в) sys<br>г) json         | б                                              | УК-6  |
| 12. | Что означает аббревиатура API?                                                                                    | Интерфейс программирования приложений          | ПК-3  |
| 13. | Какой тег в HTML используется для создания заголовков?                                                            | HL                                             | ОПК-2 |
| 14. | Как называется процесс проверки, не нарушает ли парсинг условий использования сайта?                              | Правовой аудит                                 | ПК-3  |
| 15. | Какой атрибут в HTML часто используется для поиска элементов с помощью парсера?                                   | class                                          | УК-6  |
| 16. | Какой метод в BeautifulSoup используется для поиска элемента по тегу и классу?                                    | find() или find_all()                          | ОПК-2 |
| 17. | Что такое User-Agent в HTTP-запросе?                                                                              | Строка, идентифицирующая клиентское приложение | ПК-3  |
| 18. | Какой статус-код HTTP означает успешный запрос?<br>а) 404<br>б) 500<br>в) 200<br>г) 301                           | в                                              | УК-6  |
| 19. | Как называется подход, при котором данные извлекаются с веб-сайта через официальное API, а не через парсинг HTML? | API-first                                      | ПК-3  |
| 20. | Какой инструмент позволяет автоматизировать парсинг динамических страниц, загружающих контент через JavaScript?   | Selenium                                       | ОПК-2 |