



Основная профессиональная образовательная программа
45.04.01 Филология (Прикладная цифровая филология)

Министерство науки и высшего образования Российской Федерации

ИВАНОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Кафедра зарубежной филологии

ОДОБРЕНО:

Руководитель ОП

_____ С.А. Маник

(подпись)

«__29__» _____августа____2024__ г.

Рабочая программа дисциплины
Программирование и анализ данных

Уровень высшего образования:	магистратура
Квалификация выпускника:	магистр
Направление подготовки:	45.04.01 Филология
Направленность (профиль) образовательной программы:	Прикладная цифровая филология

Иваново 2024



Основная профессиональная образовательная программа

45.04.01 Филология (Прикладная цифровая филология)

1. Цели освоения дисциплины

Целью освоения дисциплины "Программирование и анализ данных" является формирование теоретических знаний об основных структурах данных, используемых при программировании, формах представления алгоритмов, методах тестирования программ, а также ключевых принципах, методах и средствах анализа данных, практических умений и навыков по применению языка программирования и современных методов анализа данных в различных сферах человеческой деятельности. Важно в ходе освоения дисциплины подготовить магистров к профессиональной деятельности в соответствии с квалификационной характеристикой магистра по направлению, в том числе формирование умений по выявлению необходимых усовершенствований и разработке новых, более эффективных средств анализа данных.

К основным задачам освоения дисциплины «Программирование и анализ данных» следует отнести:

- изучение существующих технологий подготовки данных к анализу;
- изучение основных методов поиска закономерностей, связей, правил в табулированных массивах данных большого объема, иллюстрированного их применения в различных областях деятельности;
- изучение методов статистического анализа данных;
- овладение практическими умениями и навыками реализации технологий анализа данных, формирования и проверки гипотез о их природе и структуре, варьирования применяемыми моделями;
- изучение принципов организации и проведения аналитического исследования, решение практических задач с целью прогнозирования и выработка рекомендаций;
- формирование умений и навыков применения универсальных программных пакетов и аналитических платформ для анализа данных.

Курс ориентирован на магистров ОП «Прикладная цифровая филология», примеры и задачи подбираются с учетом профессиональных интересов магистров. Особое внимание уделяется практическому освоению терминологии и определениям базовых терминов.

2. Место дисциплины в структуре ОП

Дисциплина относится к факультативной части образовательной программы, формируемой участниками образовательных отношений ФТД.01.

Студент, приступающий к изучению дисциплины, должен обладать знаниями, умениями, навыками/опытом практической деятельности, полученными ранее в ходе изучения дисциплин на уровне бакалавриата. Для освоения данной дисциплины обучающийся должен владеть следующими знаниями и компетенциями:

- Основы программирования
- Компьютерная лингвистика
- (Математическая) статистика

Успешное освоение данной дисциплины способствует готовности студентов к освоению дисциплины Основной иностранный язык: теория и практика перевода, Применение Big Data в гуманитарной сфере, а также опыта профессиональной деятельности производственной практики, научно-исследовательской работы и производственной практики и предвещает научно-исследовательскую работу по направлению «Филология», направленность «Прикладная цифровая филология».

Данная учебная дисциплина опирается на материал предшествующих теоретических курсов «Филология в системе гуманитарного знания», «Информационные технологии в науке и образовании», «Основы программирования для филологов», «Управление цифровыми проектами», «Терминоведение и базы данных», «Цифровые инструменты в переводе».

3. Планируемые результаты обучения по дисциплине



Основная профессиональная образовательная программа 45.04.01 Филология (Прикладная цифровая филология)

3.1. Компетенции, формированию которых способствует дисциплина

При освоении дисциплины формируются следующие компетенции в соответствии с ФГОС ВО по данному направлению подготовки:

ПК 4 - Способен самостоятельно осуществлять полный цикл создания, обработки, редактирования и продвижения цифрового контента разных жанров, дискурсов

ПК-5 - Способен использовать современные информационные технологии, управлять данными с использованием прикладных программ, баз данных и корпусов текстов, выполнять перевод текстов на иностранный язык с их помощью

3.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения формируемых компетенций

В результате освоения дисциплины обучающийся должен:

Знать:

- теоретические, методологические и практические основы использования компьютерных технологий в лингвистике, мета язык компьютерной и корпусной лингвистики (ПК-4, ПК-5);
- основные методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных (ПК-4, ПК-5);
- возможности отечественных и зарубежных универсальных программных средств и аналитических платформ, применяемых для анализа данных (ПК-4, ПК-5);

Уметь:

- создавать хранилища данных, выполнять их загрузку, извлекать данные из хранилищ (ПК-4, ПК-5);
- применять технологии интеллектуального анализа электронных массивов данных для решения конкретных практических проблем (ПК-4, ПК-5);
- использовать возможности и инструменты ИИ и компьютерных технологий при решении собственных задач в исследовательской и практической работе (ПК-4, ПК-5);
- применять полученные лингвистические знания и изученные вычислительные методы и алгоритмы к реальным лингвистическим данным (ПК-5);
- самостоятельно изучать, подключать и использовать компьютерные программы и модули, необходимые для решения конкретных задач компьютерной лингвистики (ПК-4, ПК-5);

Иметь практический опыт/Иметь навыки:

- внедрять наиболее популярные компьютерные лингвистические инструменты, библиотеки и приложения для работы с лингвистическими данными (ПК-4, ПК-5);
- методами работы с пакетом анализа данных, предназначенных для решения сложных статистических и социологических задач, подробно рассмотрев весь набор инструментов, входящих в пакет анализа данных (ПК-4, ПК-5).

4. Объем и содержание дисциплины

Объем дисциплины составляет 1 зачетных единиц (36 академических часов), в т.ч.: практическая подготовка (ПП) – 14 академических часов в очной форме

4.1. Содержание дисциплины по разделам (темам), соотнесенное с видами и трудоемкостью занятий лекционно-семинарского типа

Объем иной контактной работы и самостоятельной работы обучающегося по дисциплине указан в учебном плане образовательной программы.

№ п/п	Разделы (темы) дисциплины	Семестр	Виды занятий, их объем (в ак. часах, по очной форме обучения)	Формы текущего контроля успеваемости (по очной форме обучения)
-------	---------------------------	---------	---	--



Основная профессиональная образовательная программа 45.04.01 Филология (Прикладная цифровая филология)

			Занятия лекционного типа	Занятия семинарского типа	Формы промежуточной аттестации
1.	Вводный. Понятие больших данных и управление данными. Задачи анализа данных, понятие набора данных (dataset). Подготовительные операции для выполнения анализа данных: загрузка данных, трансформация данных, изучение данных, очистка данных, визуализация данных	3		2 практ. занятие	Входная диагностика: тест с последующим обсуждением результатов. Список вопросов, интересующих студента по содержанию дисциплины (сдается в письменном виде)
2.	Базовые технологии для анализа данных	3		4 практ. занятие	Решение учебных задач
3.	Анализ данных на Python	3		6 практ. занятие	Решение учебных задач
4.	Заключительный. Подведение и анализ промежуточных результатов освоения дисциплины	3		2 практ. занятие	Решение учебных задач
Итого за семестр:				14	Зачет

4.2. Развернутое описание содержания дисциплины по разделам (темам)

Раздел 1. Вводный. Понятие больших данных и управление данными. Задачи анализа данных, понятие набора данных (dataset). Подготовительные операции для выполнения анализа данных: загрузка данных, трансформация данных, изучение данных, очистка данных, визуализация данных

Задачи анализа данных, понятие набора данных (dataset). Подготовительные операции для выполнения анализа данных: загрузка данных, трансформация данных, изучение данных, очистка данных, визуализация данных. Технологический стек анализа данных, построенный на базе языка программирования Python. Язык программирования Python: основные характеристики, возможности языка для решения задач анализа данных и машинного обучения. Версии языка программирования Python, дистрибутивы 6 и библиотеки Python. Знакомство с дистрибутивом Anaconda и составом инструментов для задач анализа данных и машинного обучения, входящих в дистрибутив. Интерактивная оболочка IPython notebook: принципы работы и применение для решения задач анализа данных и машинного обучения

Раздел 2. Базовые технологии для анализа данных

Знакомство с различными классами информационно-аналитических систем. Технологии Data Mining. Технологии анализа больших объемов данных (Big Data): причины возникновения, основные особенности функционирования и специфика создания приложений. Знакомство с библиотеками numpy и pandas и решением базовых задач подготовительных операций для выполнения анализа данных с помощью этих библиотек. Особенности построения информационно-аналитических систем с применением алгоритмов машинного обучения. Основные этапы создания информационно-аналитических систем с использованием алгоритмов машинного обучения. Постановки задач машинного обучения. Объекты и признаки. Типы признаков: бинарные, номинальные, порядковые, количественные. Типы задач машинного обучения: классификация, регрессия, прогнозирование, кластеризация.



Основная профессиональная образовательная программа
45.04.01 Филология (Прикладная цифровая филология)

Примеры задач, решаемых методами машинного обучения. Проблема недообучения / переобучения

Раздел 3. Анализ данных на Python

1. Работа со неструктурированными данными
2. Работа со структурированными данными (HTML, XML, JSON, CSV, H5, и другие)
3. Работа с библиотеками Numpy, Scipy, Pandas, Matplotlib, Plotly, Seaborn
4. Библиотеки машинного обучения Sklearn, Pytorch
5. Менеджер данных DVC

В теме рассматриваются основные библиотеки, используемые для научных вычислений и визуализации результатов. А также версионирование данных для моделей машинного обучения.



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

Раздел 4. Заключительный. Подведение итогов и анализ промежуточных результатов освоения дисциплины

5. Образовательные технологии

Рекомендуемые образовательные технологии: практические занятия с технологиями учебной дискуссии, проектная работа, представляемая в виде презентаций, технологии смешанного обучения, индивидуальные задания, самостоятельная работа студентов, контрольные работы. Студенты знакомятся с компьютерными и корпусными технологиями; применением машинного обучения в типичных случаях практики, разбирают практические кейсы, участвуют в лекциях приглашенных докладчиков.

При проведении практических занятий рекомендуется помимо обсуждения теоретических вопросов использование проектных методик, выполнение практических заданий на основе Интернет-ресурсов. Студенты представляют результаты своей самостоятельной и практической работы в виде презентаций с использованием программы PowerPoint и публикации материалов в ЭИОС

6. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся организована в виде изучения теоретического материала, решения задач и практического применения полученных навыков для машинного обучения и управления большими данными.

Студентам на ресурсе «Мой университет» выкладываются ссылки на записанные вебинары по разным темам, которые они могут посмотреть в удобное время; список площадок дистанционных курсов, на которые они должны подписаться и прислать отчет прохождения (скриншоты по неделям и успеваемость по курсу).

Полностью весь методический материал по обеспечению самостоятельной работы студентов приводится в Приложении 1 к РП.

7. Характеристика оценочных средств для текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине

Вопросы к обсуждению в ходе собеседования. Индивидуальные задания/работы с постановкой конкретной задачи и предписанным алгоритмом ее выполнения. Подготовка проектов.

Примерные вопросы к устному опросу.

1. Парадигмы программирования их суть и сильные стороны. Типичные представители различных парадигм, применение различных парадигм в Python.

2. Специфика типизации в языках программирования (различные аспекты типизации). Реализация типизации в Python.

3. Именованное переменных и других объектов в Python (правила и соглашения). Числовые типы: литералы, объявление и операции.

4. Присвоение по ссылке и по значению. Специфика создания объектов и присвоения в Python, особенности работы с объектами целочисленного типа.

5. Разница между копированием и присвоением. Проблема утечки динамической памяти, сборка мусора. Копирование, присвоение и стратегия управления динамической памятью в Python.

6. Булевский тип, сравнения и условные операторы в Python.

7. Циклы в Python, работа и устройство цикла for, типичное применение range и enumerate в цикле for.



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

8. Строки в Python. Принципы работы и основные операторы и функции.
9. Списки в Python. Различные способы создания и копирования списков в Python. Обход списка и поиск элементов в списке.
10. Списки в Python. Обращение к элементам списка и создание срезов. Стандартные агрегирующие функции, работающие со списками.
11. Списки в Python. Ключевые операции, проводящие к изменению списка и порождающие измененные списки.
12. Словари в Python. Основные способы создания, получения и изменения значений. Обход словарей.
13. Преобразование между словарями и списками в Python. Операции с представлениями словарей.
14. Операции со словарями, учитывающие возможное отсутствие ключа. Операции многоэлементного изменения словарей. Операции поэлементного извлечения из словаря и их использование.
15. Множества в Python. Основные способы создания, получения и изменения значений. Обход множеств.
16. Выполнение основных операций с парой множеств в Python.
17. Кортежи в Python. Отличия кортежей от списков. Распаковка и частичная распаковка кортежей.
18. Выражения генераторы и генераторы списков в Python. Использование условий в генераторах.
19. Генераторы множеств и словарей в Python. Использование условий в генераторах.
20. Функции стандартной библиотеки для работы с контейнерами.
21. Объявление и вызов функции в Python. Параметры функции со значением по умолчанию и комментирование функции. Получение информации о функции. Способы передачи параметров при вызове функции.
22. Передача переменного количества параметров (именованных и не именованных) в функции Python. Вызов функции с позиционными параметрами, находящимися в списке, и именованными параметрами, находящимися в словаре.
23. Анонимные функции в Python их возможности и ограничения. Типичные сценарии использования анонимных функций.
24. Синтаксис и семантика обработки исключительных ситуаций в Python.
25. Создание пользовательских исключений и инструкция assert.
26. Базовые операции для работы с файлами в Python.
27. Использование инструкции with ... as на примере работы с файлами.
28. Использование модулей pickle и shelve для сохранения объектов в файл и их восстановления.
29. Модули в Python и их отличие от скриптов Python. Варианты синтаксиса импорта модуля и объектов модуля. Применение импортированных объектов. Порядок поиска модулей и специфика их загрузки. Загрузка модулей из глобального репозитория.
30. Импорт кода из пакетов. Организация пакетов в Python.
31. Концепция класса и объекта. Принципы и механизмы ООП.
32. Объявление класса, конструктор, создание объектов и одиночное наследование в Python.
33. Управление доступом к атрибутам класса в Python.
34. Полиморфизм и утиная типизация и проверка принадлежности объекта к классу в языке Python.
35. Методы классов и статические переменные и методы в Python.
36. Интроспекция и динамические операции с объектами в Python.



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

37. Специальные методы для использования пользовательских классов со 21 стандартными операторами и функциями.
38. Основные возможности, поддерживаемые функциональными языками программирования. Поддержка функционального программирования в Python.
39. Концепция «функции – граждане первого класса» в языке программирования, поддержка этой концепции в Python. Специфика лямбда-функций в Python.
40. Глобальные и локальные переменные в функциях на примере Python. Побочные эффекты вызова функций и их последствия.
41. Вложенные функции и замыкания, специфика реализации в Python.
42. Функции высшего порядка и декораторы в Python.
43. Концепция map/filter/reduce. Работа map() в различных вариациях в Python.
44. Концепция map/filter/reduce. Работа filter() в Python. Использование модуля operator при работе с filter.
45. Концепция map/filter/reduce. Работа reduce() в Python. Использование reduce() с начальным значением, имеющим тип, отличный от возвращаемого редуцирующей функцией.
46. Итераторы в Python: встроенные итераторы, создание собственных итераторов, типичные способы обхода итераторов и принцип их работы.
47. Встроенные функции для работы с итераторами и возможности модуля itertools.
48. Функции генераторы и выражения генераторы: создание и применение в Python.
49. Специфика массивов, как структур данных. Динамические массивы – специфика работы, сложность операций. Специфика работы с array в Python.
50. Абстрактная структура данных стек: базовые и расширенные операции, их сложность. Реализация стека в Python.
51. Абстрактная структура данных очередь: базовые и расширенные операции, их сложность. Реализация очереди в Python.
52. Специфика реализации и скорости основных операций в очереди на базе массива и связанного списка.
53. Связанные списки: однонаправленные и двунаправленные. Сравнение скорости выполнения основных операций в связанных списках и в динамическом массиве.
54. Алгоритм сортировки выбором, сложность сортировки и возможности по ее улучшению.
55. Алгоритм сортировки вставками, его сложность. Алгоритм быстрого поиска в отсортированном массиве. Сложность поиска в отсортированном и не отсортированном массиве.
56. Алгоритм сортировки Шелла, сложность сортировки и возможности по ее улучшению.
57. Алгоритм быстрой сортировки, сложность сортировки и возможности по ее улучшению.
58. Алгоритм сортировки слиянием, сложность сортировки.

Критерии оценки опроса:

Оценка	Критерии
Отлично	Отвечает на все вопросы, продемонстрирована способность анализировать и обобщать информацию, сделаны обоснованные выводы, разъяснения
Хорошо	Отвечает не на все вопросы, но на большинство. Умеет в некоторой степени анализировать и обобщать информацию. Выводы умеет делать
Удовлетворительно	Отвечает не на все вопросы (меньше 50%). Демонстрирует некоторые способности анализировать и обобщать информацию. Выводы делать не может



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

Неудовлетворительно	Задание не выполнено
---------------------	----------------------

Также студенты выполняют индивидуальные задания по каждой теме, представляя результаты анализа и синтеза данных.

В конце магистранты сдают зачет. Если студент присутствовал на 60% занятий и выполнил не менее 5 заданий на практических заданиях, то он получает «зачет» автоматически как накопительную оценку. Если данное требование не выполнено, то проходит собеседование с преподавателем.

Типовые варианты тестовой работы представлены в фонде оценочных средств (Приложение 2).

8. Учебно-методическое и информационное обеспечение дисциплины

Основная литература:

1. Карякин, М. И. Технологии программирования и компьютерный практикум на языке Python : учебное пособие : [16+] / М. И. Карякин, К. А. Ватульян, Р. М. Мнухин ; Южный федеральный университет. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2022. – 244 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=698687> (дата обращения: 14.11.2024). – Библиогр. в кн. – ISBN 978-5-9275-4108-9. – Текст : электронный.
2. Баюк, О. А. Практикум по анализу данных на языках Python и R : учебное пособие для студентов, обучающихся по направлениям 38.03.01 «Экономика», 38.03.05 «Бизнес-информатика» : [16+] / О. А. Баюк, М. Р. Исаева, М. О. Самсонкин ; Финансовый университет при Правительстве Российской Федерации. – Москва : Прометей, 2023. – 100 с. : ил., табл., схем., граф. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=700938> (дата обращения: 14.11.2024). – Библиогр. в кн. – ISBN 978-5-00172-356-1. – Текст : электронный.
3. Программирование : учебное пособие / О. Ю. Горбадей, О. П. Рябычина, А. Н. Балаш, Н. И. Янович. – Минск : РИПО, 2023. – 300 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=717782> (дата обращения: 14.11.2024). – Библиогр. в кн. – ISBN 978-985-895-148-1. – Текст : электронный.
4. Златопольский, Д. М. Программирование : типовые задачи, алгоритмы, методы : учебное пособие : [12+] / Д. М. Златопольский. – 4-е изд. (эл.). – Москва : Лаборатория знаний, 2020. – 226 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=222873> (дата обращения: 14.11.2024). – Библиогр. в кн. – ISBN 978-5-00101-789-9. – Текст : электронный.

Дополнительная литература:

1. Программирование, тестирование, проектирование, нейросети, технологии аппаратно-программных средств (практические задания и способы их решения) : учебник : [16+] / С. В. Веретехина, В. Л. Симонов, К. С. Кармицкий [и др.]. – Москва : Директ-Медиа, 2022. – 144 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=694782> (дата обращения: 14.11.2024). – Библиогр. в кн. – ISBN 978-5-4499-3321-8. – DOI 10.23681/694782. – Текст : электронный.
2. Исаева, Г. Н. Языки программирования : практикум по курсу «Языки программирования» : учебное пособие : [16+] / Г. Н. Исаева, Н. В. Логачёва, Ю. В. Стреналюк ; Технологический университет. – Москва : Директ-Медиа, 2024. – 109 с. : ил., табл. – Режим доступа: по подписке. –



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

URL: <https://biblioclub.ru/index.php?page=book&id=713440> (дата обращения: 14.11.2024). – Библиогр.: с. 106-107. – ISBN 978-5-4499-4411-5. – DOI 10.23681/713440. – Текст : электронный.
3. [Электронный ресурс biblioclub: <http://biblioclub.ru/index.php?page=book&id=480500>] Жуковский О. И. Информационные технологии и анализ данных: учебное пособие / О.И. Жуковский - Томск: Эль Контент, 2014. - 130 с.
4. Электронный ресурс: <http://www.sfs.uni-tuebingen.de/~hbaayen/publications/baayenCUPstats.pdf> Harald Baayen (2008). Analyzing Linguistic Data. A Practical Introduction to Linguistics using R, Cambridge University Press: Cambridge

Ресурсы информационно-телекоммуникационной сети «Интернет»:

Система электронной поддержки образовательного процесса «Мой университет» <https://uni.ivanovo.ac.ru>

Профессиональные базы данных и информационно-справочные системы:

ЭБС «Университетская библиотека онлайн» www.biblioclub.ru;

<http://lib.ivanovo.ac.ru/index.php/polnotekstovye-resursy/ebs-universitetskaya-biblioteka>

Электронная библиотека ИвГУ <http://lib.ivanovo.ac.ru/index.php/polnotekstovye-resursy/elibnew>

Электронный каталог НБ ИвГУ <http://lib.ivanovo.ac.ru/index.php/ek>

Программное обеспечение: операционная система Microsoft Windows, пакет офисных программ Microsoft Office и(или) LibreOffice, интернет-браузер Microsoft Edge и(или) Yandex Browser.

9. Материально-техническое обеспечение дисциплины

Учебные аудитории:

- для проведения занятий семинарского типа, консультаций, текущего контроля и промежуточной аттестации с комплектом специализированной учебной мебели и техническими средствами обучения;

Лаборатория, оснащенная лабораторным оборудованием, комплектом специализированной учебной мебели и техническими средствами обучения.

Помещение для самостоятельной работы, оснащенное комплектом специализированной учебной мебели, компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в ЭИОС.

Демонстрационное оборудование и учебно-наглядные пособия для занятий лекционного типа, обеспечивающие тематические иллюстрации: презентации PowerPoint, электронные словари и справочно-информационные порталы, видеоматериалы и т.п

Автор(ы) рабочей программы дисциплины: доцент кафедры английской филологии, д.ф.н., доц. С.А. Маник

Программа рассмотрена и утверждена на заседании кафедры зарубежной филологии

«29» августа 2024 г., протокол № 13.

Программа обновлена

протокол заседания кафедры № _____ от «_____» _____ 20__ г.

Согласовано:



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

Руководитель ОП _____ / _____

(подпись)

Программа обновлена
протокол заседания кафедры № _____ от «_____» _____ 20__ г.

Согласовано:

Руководитель ОП _____ / _____

(подпись)

Программа обновлена
протокол заседания кафедры № _____ от «_____» _____ 20__ г.

Согласовано:

Руководитель ОП _____ / _____

(подпись)