

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Филологический факультет

УТВЕРЖДЕНО:
Декан
И.В. Тубалова

Рабочая программа дисциплины

Язык программирования Python

по направлению подготовки

45.04.03 Фундаментальная и прикладная лингвистика

Направленность (профиль) подготовки:
Компьютерная и когнитивная лингвистика

Форма обучения
Очная

Квалификация
Магистр

Год приема
2024

СОГЛАСОВАНО:
Руководитель ОП
З.И. Резанова

Председатель УМК
Ю.А. Тихомирова

Томск – 2025

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-6 Способен осуществлять эффективное управление разработкой программных средств информационных проектов в сфере своей профессиональной деятельности.

ПК-4 Способен разрабатывать проекты прикладной направленности в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-6.1 Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования

ИОПК-6.2 Разрабатывает алгоритмы и программы для решения лингвистических и междисциплинарных задач в том числе с применением высокопроизводительных вычислительных технологий

ИОПК-6.3 Разрабатывает и отлаживает программный код, направленный на решение лингвистических и междисциплинарных задач с применением современных языков программирования

ИПК-4.3 Обеспечивает выполнение проекта в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта, в соответствии с установленными целями, сроками и затратами

2. Задачи освоения дисциплины

– Формирование уверенных и системных знаний о принципах и особенностях создания программных решений, направленных на решение практических задач профессиональной деятельности.

– Свободное владение навыками написания части программного кода системы обработки естественного языка.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к Блоку 1 «Дисциплины (модули)».

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Первый семестр, экзамен

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования.

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 10 ч.

-практические занятия: 30 ч.

в том числе практическая подготовка: 30 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение в язык программирования Python.

Синтаксис, управляющие конструкции. Назначение переменных. Функция print.
Основные операторы.

Знакомство со средой выполнения программ на языке Python Jupyter Notebook.

Знакомство с сервисом Google Colab.

Тема 2. Строки.

Функции для работы со строками. Удаление, сложение и замены подстрок. Функции lower, upper. Многостроковые строки, f-строки.

Тема 3. Условия и цикл for.

Блок-схема алгоритма. Логические выражения, операторы «и», «или», «равно», «не равно». Ветвление алгоритмов. Множественные условия.

Тема 4. Цикл while и функции.

Рекурсивные алгоритмы. Написание функций с обязательными и необязательными аргументами. Вызов функций.

Тема 5. Работа с датафреймами.

Работа с файлами comma-separated value. Библиотека pandas, функции read_csv, head, shape, DataFrame и т. д. Выбор строк в датафрейме по условию.

Тема 6. Регулярные выражения.

Библиотека re. Синтаксис регулярных выражений.

Тема 7. Анонимные функции и генераторы списков.

Лямбда-функции. Сокращение алгоритмов. Автоматизированное наполнение списков.

Тема 8. Сбор данных с веб-сайтов.

Библиотеки BeautifulSoup, Selenium, scrapy, requests. Запрос по URL. Получение html-кода страницы. Поиск по html-коду страницы средствами библиотек BeautifulSoup и selenium. Систематизация и сбор полученных данных в датафрейм pandas.

Тема 9. Сбор данных из социальных сетей.

Библиотека snsscrape. Сбор микроблоговых сообщений, постов и получение ссылок на новости из них. Систематизация и сбор полученных данных в датафрейм pandas.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, выполнения домашних заданий и фиксируется в форме контрольной точки не менее одного раза в семестр.

Оценочные материалы текущего контроля размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в первом семестре принимается в форме проекта.

Проект сдается в виде презентации, а также прикрепляется в соответствующем элементе системы LMS iDo. При сдаче проекта оценивается выполнимость программного

кода, демонстрирование понимания основных функций и структур языка Python, умение оценивать и применять программные методы в соответствии с поставленной задачей. Проект реализуется по текстовому описанию следующего содержания:

«Итоговый проект по дисциплине «Язык программирования Python» включает разработку и предоставление программного кода скрейпера, а также текстового датасета, собранного с использованием этого скрейпера. Требования к представленным материалам приведены ниже:

Скрейпер:

- может быть прикреплен в курс «LMS iDo» в любом удобном вам виде (.ipynb, .py, ссылка на Colab Notebook, ссылка на облачное хранилище и т. д.);
- должен быть снабжен комментариями.

Датасет:

- должен быть прикреплен в формате .csv;
- должен содержать не менее 5 колонок: текст, заголовок, дата, автор, лемматизированный текст или текст с любой лингвистической разметкой, полученной при помощи библиотек для лингвистической обработки текста (Natasha, pymystem, pyMorphy и т. д.);
- может содержать иные атрибуты;
- может быть сопровожден комментариями.»

Проект проверяет компетенции ИОПК-6.1, ИОПК-6.2, ИОПК-6, ИПК-4.3.

Результаты экзамена определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценка «отлично» ставится при условиях: скрейпер полностью функционален, собирает данные в точном соответствии с требованиями, код структурирован, снабжен исчерпывающими комментариями, демонстрирующими логику работы; датасет содержит все обязательные колонки; ошибки отсутствуют или минимальны и не влияют на результат.

Оценка «хорошо» ставится при условиях: скрейпер выполняет задачу с незначительными отклонениями, код читаем, содержит достаточные комментарии; датасет включает все обязательные колонки.

Оценка «удовлетворительно» ставится при условиях: скрейпер работает с заметными ограничениями или ошибками, код частично структурирован, комментарии минимальны или недостаточно информативны; датасет содержит не все обязательные колонки.

Оценка «неудовлетворительно» ставится при условиях: проект не представлен, скрейпер не функционирует или собирает данные, не соответствующие требованиям; код отсутствует или неработоспособен; датасет отсутствует, или в нем отсутствуют обязательные колонки.

Продолжительность экзамена 1,5 часа.

Оценочные материалы для проведения промежуточной аттестации размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «LMS iDo» - <https://lms.tsu.ru/course/view.php?id=29363>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

в) Методические указания по организации самостоятельной работы студентов - <https://lms.tsu.ru/mod/resource/view.php?id=752569>

В приведенной ниже таблице представлены методические указания по организации самостоятельной работы студентов по темам.

Тема	Самостоятельная работа
Тема 1. Введение в язык программирования Python.	<p>Студенты знакомятся с интерфейсом среды разработки, участвуя писать команды и запускать их, назначать переменные и подавать аргументы в функции.</p> <p>Самостоятельная работа: ознакомление с официальной документацией языка программирования Python (https://docs.python.org/3/)</p> <p>Hammond M. Python for Linguists / M. Hammond. – Cambridge University Press, 2020. – p. 1</p>
Тема 2. Строки.	<p>Каждый студент загружает и сохраняет данные в формате строки. Обсуждаются проблемы чтения и хранения строковых данных. Для первичной обработки данных изучает необходимые библиотеки и их функции.</p> <p>Hammond M. Python for Linguists / M. Hammond. – Cambridge University Press, 2020. – p. 14</p>
Тема 3. Условия и цикл for.	<p>Студенты учатся строить алгоритмы с ветвлением, для этого знакомятся с блок-схемами и логическими операторами. На основе полученных знаний о строках выполняют задания на написание кода по текстовым описаниям.</p> <p>Hammond M. Python for Linguists / M. Hammond. – Cambridge University Press, 2020. – p. 29</p>
Тема 4. Цикл while и функции.	<p>Студенты знакомятся с концепцией рекурсии и циклов с прекращением при условии. На основе полученных знаний о строках и логических операторах строят рекурсивные алгоритмы.</p> <p>Hammond M. Python for Linguists / M. Hammond. – Cambridge University Press, 2020. – p. 41</p>
Тема 5. Работа с датафреймами.	<p>Знакомство с библиотекой pandas. Ввод и вывод файла, чтение, запись, удаление данных внутри датафрейма.</p> <p>Самостоятельная работа: ознакомление с документацией библиотеки на официальном сайте (https://pandas.pydata.org/), написание кода по текстовому описанию алгоритма.</p>
Тема 6. Регулярные выражения.	<p>Знакомство с библиотекой re, знакомство с синтаксисом регулярных выражений.</p> <p>Самостоятельная работа: ознакомление с документацией библиотеки на официальном сайте (https://docs.python.org/3/library/re.html), написание кода по текстовому описанию алгоритма.</p>
Тема 7. Анонимные функции и генераторы списков.	<p>На основе написанного ранее кода студенты учатся сокращать код путем использования односторонних функций и автоматического наполнения списков.</p> <p>Hammond M. Python for Linguists / M. Hammond. – Cambridge University Press, 2020. – p. 275</p>
Тема 8. Сбор данных с веб-сайтов.	<p>Студенты знакомятся с библиотеками для сбора данных с веб-страниц. На основе полученных знаний о датафреймах систематизируют и</p>

	<p>сохраняют полученные данные, обрабатывают их.</p> <p>Самостоятельная работа: ознакомление с документацией библиотек на официальном сайте, написание кода по текстовому описанию алгоритма.</p>
Тема 9. Сбор данных из социальных сетей.	<p>Студенты знакомятся с библиотеками для сбора данных со страниц в социальных сетях. На основе полученных знаний о датафреймах систематизируют и сохраняют полученные данные, обрабатывают их.</p>

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

- Лутц М. Изучаем Python/ М. Лутц – СПб.: Символ-Плюс, 2011. - 1280 с.
- Hammond M. Python for Linguists / M. Hammond. – Cambridge University Press, 2020.
– 313 p.

б) дополнительная литература:

- Walters G. The Python Quick Syntax Reference / G. Walters – Apress, 2014. – 152 p.
- Мартелли А. Python. Справочник : полное описание языка / Алекс Мартелли, Анна Рейвенскрофт, Стив Холден ; перевод с английского А. Г. Гузикевича. - 3-е изд.. - Санкт-Петербург [и др.] : Диалектика, 2019. - 892 с.
- Browning J. B. Pro Python Second Edition // by J. Burton Browning, Marty Alchin. // Springer eBooks. URL: <http://dx.doi.org/10.1007/978-1-4842-0334-7>

в) ресурсы сети Интернет:

- открытые онлайн-курсы по языку программирования Python Codecademy (<https://codecademy.com>)
- документация языка программирования Python (<https://docs.python.org/3/>)
- документация библиотек pandas, BeautifulSoup и т. д.

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

- публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).
- интернет-браузер (Google Chrome, Yandex, и т. д.)
- публично доступный сервис для исполнения скриптов на языке Python Google Colab (<https://colab.research.google.com>)
- среда разработки для языка программирования Python Jupyter Notebook.

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

15. Информация о разработчиках

Аишева Динара Армановна, ассистент каф. общей, компьютерной и когнитивной лингвистики;
Шамигов Федор Федорович, специалист по учебно-методической работе.