

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«Российский государственный гуманитарный университет»
(ФГБОУ ВО «РГГУ»)**

ИНСТИТУТ ЛИНГВИСТИКИ
УНЦ компьютерной лингвистики

ПРОГРАММИРОВАНИЕ В ЛИНГВИСТИКЕ

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

45.03.03 Фундаментальная и прикладная лингвистика

Код и наименование направления подготовки/специальности

Фундаментальная и прикладная лингвистика

Наименование направленности (профиля)/ специализации

Уровень высшего образования: *бакалавриат*

Форма обучения: *Очная*

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2024

Программирование в лингвистике

Рабочая программа дисциплины

Составитель:

к. филол. н., доцент УНЦ компьютерной лингвистики А.Ч. Пиперски

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной лингвистики

№ 4 от 21.03.2024

ОГЛАВЛЕНИЕ

| | | |
|------|---|----|
| 1. | Пояснительная записка | 4 |
| 1.1. | Цель и задачи дисциплины | 4 |
| 1.2. | Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций | 4 |
| 1.3. | Место дисциплины в структуре образовательной программы | 5 |
| 2. | Структура дисциплины | 6 |
| 3. | Содержание дисциплины | 6 |
| 4. | Образовательные технологии | 8 |
| 5. | Оценка планируемых результатов обучения | 10 |
| 5.1 | Система оценивания | 10 |
| 5.2 | Критерии выставления оценки по дисциплине | 11 |
| 5.3 | Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине | 12 |
| 6. | Учебно-методическое и информационное обеспечение дисциплины | 13 |
| 6.1 | Список источников и литературы | 13 |
| 6.2 | Перечень ресурсов информационно-телекоммуникационной сети «Интернет». | 14 |
| 7. | Материально-техническое обеспечение дисциплины | 14 |
| 8. | Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов | 14 |
| 9. | Методические материалы | 15 |
| 9.1 | Планы практических занятий | 15 |
| 9.2 | Иные материалы | 17 |
| | Приложение 1. Аннотация рабочей программы дисциплины | 18 |

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Цель дисциплины — формирование у студентов знаний, умений и навыков в области программирования в лингвистике, владения языком программирования Python и библиотеками для работы с лингвистическими задачами.

Задачи:

Студент, освоивший дисциплину, должен знать современную научную и учебную литературу по компьютерной лингвистике и программированию в Python, уметь решать стандартные лингвистические задачи при помощи этого языка программирования, а также пользоваться модулями.

В результате освоения дисциплины (*модуля*) обучающийся должен:

знать

- методику сбора и анализа текстов, автоматической их обработки при помощи компьютера;

уметь:

- собирать и анализировать тексты, обрабатывать их при помощи языка программирования Python и его библиотек;

интерпретировать результаты автоматического анализа;

владеть:

- навыками программирования на языке Python.

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

| Компетенция (код и наименование) | Индикаторы компетенций (код и наименование) | Результаты обучения |
|--|--|---|
| ОПК-1. Способен применять в профессиональной деятельности основные понятия и категории современной лингвистики | ОПК-1.1 | Знает: понимает различие между языком как свойством homo sapiens и конкретными языками, между естественными и искусственными языками, между языком, речью и речевой деятельностью; знает функции языка; понимает отношение языка к мышлению, обществу и культуре и знаковую природу языка; знает основные компоненты и уровни языка как знаковой системы, типы отношений между единицами языка; систему лингвистических дисциплин, основные научные парадигмы в лингвистике и применяемые в ней |

| | | |
|--|---------|---|
| | | методы исследования; представляет себе связи лингвистики со смежными естественными и гуманитарными науками. |
| | ОПК-1.2 | Умеет: использовать основные термины лингвистики для характеристики устройства языковой системы; определить, в каком разделе лингвистики изучается то или иное явление языка, в каких аспектах изучается одно и то же языковое явление в разных лингвистических дисциплинах; охарактеризовать основные научные парадигмы в лингвистике по их целям, допущениям и применяемым в них методам. |
| | ОПК-1.3 | Владеет: основными понятиями и категориями современной лингвистики в объеме, позволяющем воспринимать научный лингвистический дискурс и грамотно формулировать высказывания о языке и лингвистике. |
| ОПК-2. Способен к ведению профессиональной деятельности с опорой на основы математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур | ОПК-2.3 | Владеет: основными методами решения типичных задач теории множеств, комбинаторики, математической логики, теории алгоритмов, теории вероятностей, математической статистики и теории информации; навыками планирования, написания и отладки простых программ для обработки языковых данных на изученном алгоритмическом языке, использования основных функций соответствующей среды программирования. |

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Программирование в лингвистике» относится к обязательной части блока дисциплин учебного плана.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин и прохождения практик: «Введение в теорию языка», «Понятийный аппарат математики», «Математическая логика», «Общая морфология», «Общий синтаксис», «Компьютерная и корпусная лингвистика».

В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: «Автоматический перевод», научно-исследовательская работа.

2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 9 з.е., 324 академических часа.

Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

| Семестр | Тип учебных занятий | Количество часов |
|---------|---------------------|------------------|
| | Лекции | 0 |
| | Семинары | 126 |
| | Всего: | 126 |

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 198 академических часов.

3. Содержание дисциплины

| № | Наименование раздела дисциплины | Содержание |
|----|--|--|
| 1. | Раздел 1. Работа с ЯП Python: классификация объектов | Язык программирования Python как объектно-ориентированный язык интерпретаторного типа. Переменные в Python. Объекты в Python и их классификация: изменяемые, неизменяемые, итерируемые. Обзор стандартных классов Python: integer, float, string. Арифметические операции. Понятие функции. Логические операции. |
| 2. | Раздел 2. Объекты типа string и их методы | Объекты класса string в Python и их основные характеристики. Применение объектов класса string. Основные методы объектов класса string: find, rfind, replace, split, strip, join. |
| 3. | Раздел 3. Объекты типа list и их методы | Объекты класса list в Python. Понятие итерации. Цикл while. Цикл for. Основные методы объектов класса list: append, pop, insert, index, reverse. |
| 4. | Раздел 4. Объекты типа set и их методы | Объекты класса set в Python. Разница между изменяемыми и неизменяемыми (хешируемыми) объектами. Логические операции на множествах. Основные методы объектов класса set: add, union, intersection. Операторы для работы с множествами. |

| | | |
|-----|--|---|
| 5. | Раздел 5. Функции. Рекурсия | Написание собственных функций в Python. Вызываемые объекты. Определение функции. Имя функции. Аргументы и параметры. Класс NoneType. Понятие рекурсии. |
| 6. | Раздел 6. Объекты типа dict и их методы | Объекты класса dict в Python. Понятие mapping objects. Ключи. Значения. Создание объектов класса dict. Основные методы объектов класса dict: pop, get, getitem. Понятие view objects: keys, values, items. Работа со стандартными библиотеками в Python. Библиотека collections. Класс defaultdict, Counter. Применение объектов класса defaultdict и Counter в реальных лингвистических задачах. Подсчет частотности слов в тексте. |
| 7. | Раздел 7. Работа с файлами в Python | Дескрипторы файлов в Python. Функция open и ее параметры. Кодировки: однобайтовые, Unicode. Абсолютный и относительный путь файла. Стандартные модули для работы с файлами os и pathlib. Обзор доступных в этих модулях функций для работы с путями файлов и директориями. |
| 8. | Раздел 8. Регулярные выражения в Python. Модуль regex | Понятие регулярных выражений, их использование для решения лингвистических задач по поиску в тексте. Базовые правила регулярных выражений: escape-последовательности, метасимволы, проверки, логические операции. Основные функции модуля regex. |
| 9. | Раздел 9. Типы файлов json, csv, модули pandas, json, csv. | Типы файлов, активно использующихся для работы с текстами и решениях лингвистических задач: json, csv. Основные функции модулей, предназначенных для работы с json и csv: чтение и запись таких файлов. Модуль pandas. Объект класса DataFrame; конвертация json и csv в объект класса DataFrame. Построение графиков с помощью метода scatter_matrix. |
| 10. | Раздел 10. Токенизация и сегментация в Python. Модули razdel, DeepPavlov. | Понятие токенизации. Токен. Автоматическая сегментация по предложениям. Основы работы с модулями razdel и DeepPavlov, автоматическая токенизация и сегментация текстов с их помощью. |
| 11. | Раздел 11. Автоматический морфологический анализ. Модули RNNMorph, Mystem. | Понятие автоматического морфологического анализа. Обзор существующих средств автоматического морфологического анализа. Основы работы с модулем Mystem: дескрипторы класса lemmatize, analyze и get_pos. Основы работы с библиотекой RNNMorph. |
| 12. | Раздел 12. Автоматический синтаксический анализ. Синтаксические парсеры. | Universal Dependencies: краткий обзор формата. Автоматический синтаксический анализ текста как задача. Библиотека UDPipe и основы работы с ней. Отображение синтаксических деревьев в графическом виде. |
| 13. | Раздел 13. Основы работы с модулем | Обзор возможностей модуля NLTK, базовые функции. Обработка сырого текста. Токенизация и сегментация |

| | | |
|-----|--|---|
| | NLTK: базовые функции и возможности | модуля NLTK. Подсчет частотности лемм, TF/IDF документов. |
| 14. | Раздел 14. Работа с модулем NLTK: корпуса в составе модуля | Основы работы с готовыми корпусами. Корпуса в составе библиотеки NLTK: Gutenberg, инаугурационный корпус, языковые корпуса. |

4. Образовательные технологии

| Наименование раздела | Виды учебных занятий | Образовательные технологии |
|--|--|--|
| Раздел 1. Работа с ЯП Python: классификация объектов | Семинары Самостоятельная работа | Знакомство со средой разработки и ее возможностями, написание скриптов, решение типовых задач Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 2. Объекты типа string и их методы | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 3. Объекты типа list и их методы | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 4. Объекты типа set и их методы | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 5. Функции. Рекурсия | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 6. Объекты типа dict и их методы | Семинары | Написание скриптов с использованием пройденных материалов, решение типовых задач. |

| | | |
|--|---|--|
| | Самостоятельная работа | Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 7. Работа с файлами в Python | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 8. Регулярные выражения в Python. Модуль regex | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Представление возможных тем студенческих проектных работ. |
| Раздел 9. Типы файлов json, csv, модули pandas, json, csv. | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 10. Токенизация и сегментация в Python. Модули razdel, DeepPavlov. | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 11. Автоматический морфологический анализ. Модули RNNMorph, Mystem. | Лекция 2 часа. Семинары 6 часов. Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 12. Автоматический синтаксический анализ. Синтаксические парсеры. | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Раздел 13. Основы работы с модулем NLTK: базовые функции и возможности | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |

| | | |
|--|--|--|
| Раздел 14. Работа с модулем NLTK: корпуса в составе модуля | Семинары Самостоятельная работа | Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту. |
| Презентации студенческих проектов. | Семинар | Представление студентами презентаций и проектов. |

В период временного приостановления посещения обучающимися помещений и территории РГГУ для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

5. Оценка планируемых результатов обучения

5.1 Система оценивания

В каждом из трех семестров используется единая система оценивания:

| Форма контроля | Макс. количество баллов | |
|---|-------------------------|------------|
| | За одну работу | Всего |
| Текущий контроль: | | |
| - <i>опрос</i> | 5 баллов | 30 баллов |
| - <i>участие в дискуссии на семинаре</i> | 5 баллов | 10 баллов |
| - <i>домашняя работа</i> | 5 баллов | 20 баллов |
| Промежуточная аттестация – <i>зачет</i> (семестр 5) / <i>зачет с оценкой</i> (6) / <i>экзамен</i> (7) | | 40 баллов |
| Итого за семестр | | 100 баллов |

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

| | | | |
|--------------------|---------------------|------------|------------|
| 100-балльная шкала | Традиционная шкала | | Шкала ECTS |
| 95 – 100 | отлично | зачтено | A |
| 83 – 94 | | | B |
| 68 – 82 | хорошо | | C |
| 56 – 67 | удовлетворительно | | D |
| 50 – 55 | | | E |
| 20 – 49 | неудовлетворительно | не зачтено | FX |
| 0 – 19 | | | F |

5.2 Критерии выставления оценки по дисциплине

| Баллы/ Шкала ECTS | Оценка по дисциплине | Критерии оценки результатов обучения по дисциплине |
|-------------------------|-------------------------------|---|
| 100-83/ A,B | отлично/ зачтено | <p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p> |
| 82-68/ C | хорошо/ зачтено | <p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p> |
| 67-50/ D,E | удовлетворительно/ зачтено | <p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p> |

| Баллы/ Шкала ECTS | Оценка по дисциплине | Критерии оценки результатов обучения по дисциплине |
|-------------------------|---|---|
| 49-0/ F,FX | неудовлет- ворительно/ не зачтено | <p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p> |

5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

Пример заданий к семинарскому занятию

1) Вводится целое число n . Программа должна вывести " n коров", при этом слово "коровы" должно быть согласовано с числительным (1 корова, 10 коров).

2) Вводится число секунд n . Выведите, сколько это в днях, часах, минутах и секундах (например, 1:22:4:5)

3) Вводится три действительных числа. Выведите True, если третье число попадает в диапазон между первым и вторым, иначе False.

4) Посчитайте, может ли ладья попасть с одной клетки на другую за один ход. Вводится четыре целых числа - две координаты первой клетки и две координаты второй клетки. Выведите True, если ладья может пройти между клетками за один ход, иначе False.

5) Вводится две строки. Посчитайте, сколько процентов от длины первой строки составляет длина второй строки.

6) Вводится строка. Выведите строку, состоящую из первого символа исходной строки, повторенного 10 раз и последнего символа, тоже повторенного 10 раз. Например, из строки `abcd` должно получиться `aaaaaaaaaaddddd`.

7) Вводится строка с чётным количеством символов. Выведите 4 символа из середины строки.

8) С начала суток часовая стрелка повернулась на угол в n градусов. Определите сколько полных часов, минут и секунд прошло с начала суток. Запишите ответ в три переменные и выведите их на экран.

Экзаменационные вопросы (вопросы к зачету)

Язык программирования Python и его характеристики. Среда разработки IDLE. Понятие сценарного и интерактивного режима.

Классификация объектов в Python. Стандартные классы: `int`, `float`, `bool`.

Арифметические и логические операции в Python.

Объекты класса `string` в Python. Их методы.

Понятие итерации. Разновидности циклов, особенности их применения.

Объекты класса `list` в Python. Их методы.

Изменяемые и неизменяемые объекты.

Объекты класса set в Python. Их методы. Логические операции с множествами.

Функции. Вызываемые объекты. Аргументы и параметры. Понятие рекурсии.

Объекты класса dict в Python. Их методы. Defaultdict и Counter.

Внешние библиотеки Python, способы их использования.

Дескрипторы файлов. Кодировки. Функция open.

Регулярные выражения и правила их составления.

Токенизация. Понятие токена. Необходимость токенизации для решения лингвистических задач.

Сегментация. Библиотека razdel, библиотека DeepPavlov.

Автоматический морфологический анализ: особенности. Используемые для него инструменты, возможности этих инструментов. Проблемы автоматического морфологического анализа.

Автоматический синтаксический анализ и доступные парсеры.

Формат Universal Dependencies.

Библиотека NLTK и существующие в ней функции.

Корпуса библиотеки NLTK и их характеристики.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы

Литература:

Основная:

Чернышев, С. А. Основы программирования на Python : учебное пособие для вузов / С. А. Чернышев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 349 с. — (Высшее образование). — ISBN 978-5-534-17139-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/544190>

Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для среднего профессионального образования / Д. Ю. Федоров. — 5-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 227 с. — (Профессиональное образование). — ISBN 978-5-534-17319-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/539652>

Дополнительная:

Рамальо, Л. Python. К вершинам мастерства / Лучано Рамальо ; пер. с англ. А.А. Слинкина. - Москва : ДМК Пресс, 2016. - 768 с. - ISBN 978-5-97060-384-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1028052>

6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

1. The Python Standard Library: <https://docs.python.org/3/library/index.html>
2. NLTK book <https://www.nltk.org/book/>
3. Интерактивный учебник языка Python <http://pythontutor.ru/>
4. Морфологический анализатор Pymorphy2 <https://pymorphy2.readthedocs.io/en/latest/>
5. Pymystem3: <https://github.com/nlpub/pymystem3>
6. RNNMorph: <https://github.com/IlyaGusev/rnnmorph>

7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения: учебные аудитории, оснащённые компьютером, проектором для демонстрации учебных материалов и обязательным подключением к Интернету.

Состав программного обеспечения:

1. Windows
2. Microsoft Office
3. Интерактивные среды разработки IDLE, Wing 101, Pycharm (распространяются свободно)

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением или могут быть заменены устным ответом; обеспечивается индивидуальное равномерное освещение не менее 300 люкс; для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств; письменные задания оформляются увеличенным шрифтом; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

- для глухих и слабослышащих: лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования; письменные задания выполняются на компьютере в письменной форме; экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих: в печатной форме увеличенным шрифтом, в форме электронного документа, в форме аудиофайла.
- для глухих и слабослышащих: в печатной форме, в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата: в печатной форме, в форме электронного документа, в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих: устройством для сканирования и чтения с камерой SARA CE; дисплеем Брайля PAC Mate 20; принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих: автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих; акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата: передвижными, регулируемые эргономическими партами СИ-1; компьютерной техникой со специальным программным обеспечением.

9. Методические материалы

9.1 Планы практических занятий

| Тема занятий | Вопросы для обсуждения |
|---------------------|--|
| Введение | Язык программирования Python как объектно-ориентированный язык интерпретаторного типа. Переменные в Python. Объекты в Python и их классификация: изменяемые, неизменяемые, итерируемые. Обзор стандартных классов Python: integer, float, string. Арифметические операции. Понятие функции. Логические операции. |
| Объекты типа string | Объекты класса string в Python. Типы кодировок. Доступ к элементам строк. Основные методы класса string: print, len, find, replace, lower. |

| Тема занятий | Вопросы для обсуждения |
|--|---|
| Объекты типа list | Объекты класса list в Python. Понятие итерации. Цикл while. Цикл for. Основные методы объектов класса list: append, pop, insert, index, reverse. |
| Объекты типа set | Объекты класса set в Python. Разница между изменяемыми и неизменяемыми (хешируемыми) объектами. Логические операции на множествах. Основные методы объектов класса set: add, union, intersection. Операторы для работы с множествами. |
| Функции. Рекурсия | Написание собственных функций в Python. Вызываемые объекты. Определение функции. Имя функции. Аргументы и параметры. Класс NoneType. Понятие рекурсии. |
| Объекты типа dict и их методы | Объекты класса dict в Python. Понятие mapping objects. Ключи. Значения. Создание объектов класса dict. Основные методы объектов класса dict: pop, get, getitem. Понятие view objects: keys, values, items. Работа со стандартными библиотеками в Python. Библиотека collections. Класс defaultdict, Counter. Применение объектов класса defaultdict и Counter в реальных лингвистических задачах. Подсчет частотности слов в тексте. |
| Работа с файлами в Python | Дескрипторы файлов в Python. Функция open и ее параметры. Кодировки: однобайтовые, Unicode. Абсолютный и относительный путь файла. Стандартные модули для работы с файлами os и pathlib. Обзор доступных в этих модулях функций для работы с путями файлов и директориями. |
| Регулярные выражения в Python. Модуль regex | Понятие регулярных выражений, их использование для решения лингвистических задач по поиску в тексте. Базовые правила регулярных выражений: escape-последовательности, метасимволы, проверки, логические операции. Основные функции модуля regex. |
| Типы файлов json, csv, модули pandas, json, csv | Типы файлов, активно используемых для работы с текстами и решениях лингвистических задач: json, csv. Основные функции модулей, предназначенных для работы с json и csv: чтение и запись таких файлов. Модуль pandas. Объект класса DataFrame; конвертация json и csv в объект класса DataFrame. Построение графиков с помощью метода scatter_matrix. |
| Токенизация и сегментация в Python. Модули razdel, DeepPavlov | Понятие токенизации. Токен. Автоматическая сегментация по предложениям. Основы работы с модулями razdel и DeepPavlov, автоматическая токенизация и сегментация текстов с их помощью. |
| Автоматический морфологический анализ. Модули RNNMorph, Mystem | Понятие автоматического морфологического анализа. Обзор существующих средств автоматического морфологического анализа. Основы работы с модулем Mystem: дескрипторы класса lemmatize, analyze и get_pos. Основы работы с библиотекой RNNMorph. |

| Тема занятий | Вопросы для обсуждения |
|--|---|
| Автоматический синтаксический анализ. Синтаксические парсеры | Universal Dependencies: краткий обзор формата. Автоматический синтаксический анализ текста как задача. Библиотека UDPipe и основы работы с ней. Отображение синтаксических деревьев в графическом виде. |
| Основы работы с модулем NLTK: базовые функции и возможности | Обзор возможностей модуля NLTK, базовые функции. Обработка сырого текста. Токенизация и сегментация модуля NLTK. Подсчет частотности лемм, TF/IDF документов. |
| Работа с модулем NLTK: корпуса в составе модуля | Основы работы с готовыми корпусами. Корпуса в составе библиотеки NLTK: Gutenberg, инаугурационный корпус, языковые корпуса. |
| Презентации студенческих проектов | Представление студентами проектных работ. |

9.2 Иные материалы

Рекомендуемая литература для более глубокого освоения программы.

Основная литература

Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для среднего профессионального образования / Д. Ю. Федоров. — 2-е изд. — Москва : Издательство Юрайт, 2020. — 161 с. — (Профессиональное образование). — ISBN 978-5-534-11961-9.

Гниденко, И. Г. Технологии и методы программирования : учебное пособие для вузов / И. Г. Гниденко, Ф. Ф. Павлов, Д. Ю. Федоров. — Москва : Издательство Юрайт, 2020. — 235 с. — (Высшее образование). — ISBN 978-5-534-02816-4.

Дополнительная литература

Зыков, С. В. Программирование. Объектно-ориентированный подход : учебник и практикум для вузов / С. В. Зыков. — Москва : Издательство Юрайт, 2020. — 155 с. — (Высшее образование). — ISBN 978-5-534-00850-0.

Зыков, С. В. Программирование. Функциональный подход : учебник и практикум для вузов / С. В. Зыков. — Москва : Издательство Юрайт, 2020. — 164 с. — (Высшее образование). — ISBN 978-5-534-00844-9.

Гниденко, И. Г. Технологии и методы программирования : учебное пособие для вузов / И. Г. Гниденко, Ф. Ф. Павлов, Д. Ю. Федоров. — Москва : Издательство Юрайт, 2020. — 235 с. — (Высшее образование). — ISBN 978-5-534-02816-4.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Цель дисциплины — формирование у студентов знаний, умений и навыков в области программирования в лингвистике, владения языком программирования Python и библиотеками для работы с лингвистическими задачами.

Задачи:

Студент, освоивший дисциплину, должен знать современную научную и учебную литературу по компьютерной лингвистике и программированию в Python, уметь решать стандартные лингвистические задачи при помощи этого языка программирования, а также пользоваться модулями.

В результате освоения дисциплины (*модуля*) обучающийся должен:

знать

- методику сбора и анализа текстов, автоматической их обработки при помощи компьютера;

уметь:

- собирать и анализировать тексты, обрабатывать их при помощи языка программирования Python и его библиотек;

интерпретировать результаты автоматического анализа;

владеть:

- навыками программирования на языке Python.