

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования

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

ИНСТИТУТ ЛИНГВИСТИКИ

УНЦ компьютерной лингвистики

ТЕХНОЛОГИИ ОБРАБОТКИ ТЕКСТА И ЗВУЧАЩЕЙ РЕЧИ

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

45.03.03 Фундаментальная и прикладная лингвистика

Фундаментальная и прикладная лингвистика

Уровень квалификации выпускника (бакалавр)

Форма обучения (очная)

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2019

ТЕХНОЛОГИИ ОБРАБОТКИ ТЕКСТА И ЗВУЧАЩЕЙ РЕЧИ

Рабочая программа дисциплины

Составитель:

к. филол. н., доцент. УНЦ КЛ *А.Ч.Пиперски*

УТВЕРЖДЕНО

Протокол заседания УНЦ КЛ

№ 1 от 28.08.2019

ОГЛАВЛЕНИЕ

1. Пояснительная записка

1.1 Цель и задачи дисциплины

1.2. Формируемые компетенции, соотнесённые с планируемыми результатами обучения по дисциплине

1.3. Место дисциплины в структуре образовательной программы

2.. Структура дисциплины

3. Содержание дисциплины

4. Образовательные технологии

5. Оценка планируемых результатов обучения

5.1. Система оценивания

5.2. Критерии выставления оценок

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

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список источников и литературы

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

7. Материально-техническое обеспечение дисциплины

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

9. Методические материалы

9.1. Планы практических (семинарских, лабораторных) занятий

9.2. Методические рекомендации по подготовке письменных работ

9.3. Иные материалы

Приложения

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

I. Пояснительная записка

1.1. Цели и задачи курса

Курс “Технологии обработки текста и звучащей речи” – это курс, посвященный применению вычислительной техники и компьютеров для автоматического анализа и обработки лингвистических данных. Курс предполагает возможности лингвистических исследований с применением новейших доступных технологий, в частности, языка программирования Python и существующих библиотек по автоматической обработке естественного языка.

Предметом курса является язык программирования Python и написанные для него библиотеки для автоматической обработки естественного языка.

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

Задачи курса

Конкретные задачи курса – теоретические и практические.

Теоретические:

- ознакомить студентов с базовыми понятиями объектно-ориентированного программирования;
- ознакомить студентов с принципами работы объектно-ориентированного языка программирования высокого уровня интерпретаторного типа.

- Практические:
- развить навыки программирования в языке Python;
- научить автоматически обрабатывать лингвистические данные при помощи Python и существующих библиотек;
- развить навыки анализа полученных в результате автоматической обработки данных.

1.2. Формируемые компетенции и ожидаемые результаты обучения

Курс нацелен на **формирование** у студентов следующих профессиональных **компетенций**:

владение основными понятиями и категориями современной лингвистики (ОПК-1);

владением основными методами фонологического, морфологического, синтаксического, дискурсивного и семантического анализа с учетом языковых и экстралингвистических факторов (ПК-1);

владением основными методами инструментального анализа звучащей речи (ПК-2);
способностью пользоваться лингвистически ориентированными программными продуктами (ПК-9);

владением принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологии; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами (ПК-10);

способностью проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-12)

знать:

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

— принципы работы объектно-ориентированного языка высокого уровня интерпретаторного типа;

уметь:

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

— интерпретировать результаты автоматического анализа,

1.3. Место дисциплины в структуре образовательной программы

На этом курсе студенты получают сведения о языке программирования Python, наиболее распространенных его библиотеках (NLTK, Mystem и др.) и способах его использования для решения лингвистических задач.

В результате освоения дисциплины формируются знания, умения и навыки, необходимые для изучения следующих дисциплин и прохождения практик: практика по получению первичных профессиональных умений и навыков (научно-исследовательская).

Дисциплина «Технология обработки текста и звучащей речи» входит в базовую часть учебного плана по направлению подготовки 45.03.03 – «Фундаментальная и прикладная лингвистика». Дисциплина адресована бакалаврам, обучающимся по профилю «Фундаментальная и прикладная лингвистика». Читается на 4 курсе в 7 и 8 семестре.

2. СТРУКТУРА ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 5 зачетных единиц – 180 часов. Программа дисциплины включает лекции, практические аудиторные занятия и самостоятельную работу студента (92 часа), которая предусматривает выполнение домашних заданий и подготовку к зачету и экзамену.

Тематический календарный план курса

№ п/п	Раздел дисциплины/темы	Се ме ст р	Виды учебной работы (в часах)					Са мо сто ятель ная рабо та	Формы текущего контроля успеваемости, форма промежуточной аттестации (<i>по семестрам</i>)
			контактная						
			Ле кци и	Се ми нар	П ра кт ич ес кие за ня тия	Лабо ратор ные заня тия	Пром ежуто чная аттес тация		
	Раздел 1. Работа с ЯП Python: классификация объектов	7	1	1				6	
	Раздел 2. Объекты типа string и их методы	7	1	1				6	Решение задач
	Практикум по решению задач	7							
	Раздел 3. Объекты типа list и их методы	7	1	1				6	Решение задач
	Практикум по решению задач	7							
	Раздел 4. Объекты типа set и их методы	7	1	1				6	Решение задач
	Практикум по решению задач	7							
	Раздел 5. Функции. Рекурсия	7	1	1				6	
	Раздел 6. Объекты типа dict и их методы	7	1	1				6	Решение задач
	Практикум по решению задач	7							
	Раздел 7. Работа с	7						10	Решение задач

	<i>файлами в Python</i>								
	<i>Практикум по решению задач</i>	7							
	Зачет	7							Решение задач, ответ на контрольные вопросы
	<i>Раздел 8. Регулярные выражения в Python. Модуль regex</i>	8	1	1				6	Решение задач
	<i>Раздел 9. Типы файлов json, csv, модули pandas, json, csv.</i>	8	1	1				6	
	<i>Раздел 10. Токенизация и сегментация в Python. Модули razdel, DeepPavlov.</i>	8	2	2				6	Решение задач
	<i>Раздел 11. Автоматический морфологический анализ. Модули RNNMorph, Mystem.</i>	8	1	1				6	
	<i>Раздел 12. Автоматический синтаксический анализ. Синтаксические парсеры.</i>	8	1	1				6	Решение задач
	<i>Раздел 13. Основы работы с модулем NLTK: базовые функции и возможности</i>	8	1	1				6	
	<i>Раздел 14. Работа с модулем NLTK: корпуса в составе модуля</i>	8	1	1				10	Решение задач
	<i>Проектные работы студентов</i>	8							Защита проекта, презентация
	Итого		14	14				92	
	<i>экзамен</i>			<i>n</i>					Решение задач, ответ на вопросы, предоставление проектов
	ИТОГО:								

3. СОДЕРЖАНИЕ КУРСА

№	Наименование раздела дисциплины	Содержание
1.	<i>Раздел 1. Работа с ЯП Python: классификация объектов</i>	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.
	<i>Раздел 2. Объекты типа string и их методы</i>	Объекты класса string в Python и их основные характеристики. Применение объектов класса string. Основные методы объектов класса string: find, rfind, replace, split, strip, join.
	<i>Раздел 3. Объекты типа list и их методы</i>	Объекты класса list в Python. Понятие итерации. Цикл while. Цикл for. Основные методы объектов класса list: append, pop, insert, index, reverse.
	<i>Раздел 4. Объекты типа set и их методы</i>	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.
	<i>Раздел 5. Функции. Рекурсия</i>	Написание собственных функций в Python. Вызываемые объекты. Определение функции. Имя функции. Аргументы и параметры. Класс NoneType. Понятие рекурсии.
	<i>Раздел 6. Объекты типа dict и их методы</i>	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.
	<i>Раздел 7. Работа с файлами в Python</i>	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.
2.	<i>Раздел 8. Регулярные выражения в Python. Модуль regex</i>	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.

	<i>Раздел 9. Типы файлов json, csv, модули pandas, json, csv.</i>	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.
	<i>Раздел 10. Токенизация и сегментация в Python. Модули razdel, DeepPavlov.</i>	Понятие токенизации. Токен. Автоматическая сегментация по предложениям. Основы работы с модулями razdel и DeepPavlov, автоматическая токенизация и сегментация текстов с их помощью.
	<i>Раздел 11. Автоматический морфологический анализ. Модули RNNMorph, Mystem.</i>	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.
	<i>Раздел 12. Автоматический синтаксический анализ. Синтаксические парсеры.</i>	Universal Dependencies: краткий обзор формата. Автоматический синтаксический анализ текста как задача. Библиотека UDPipe и основы работы с ней. Отображение синтаксических деревьев в графическом виде.
	<i>Раздел 13. Основы работы с модулем NLTK: базовые функции и возможности</i>	Обзор возможностей модуля NLTK, базовые функции. Обработка сырого текста. Токенизация и сегментация модуля NLTK. Подсчет частотности лемм, TF/IDF документов.
	<i>Раздел 14. Работа с модулем NLTK: корпуса в составе модуля</i>	Основы работы с готовыми корпусами. Корпуса в составе библиотеки NLTK: Gutenberg, инаугурационный корпус, языковые корпуса.

4. Информационные и образовательные технологии

№ п/ п	Наименование раздела	Виды учебных занятий	Образовательные технологии
1	2	3	4
1.	<i>Раздел 1. Работа с ЯП Python: классификация объектов</i>	<i>Лекция 1. Семинар 1. Самостоятельная работа</i>	<i>Вводная лекция с использованием видеоматериалов и проекционной техники. Знакомство со средой разработки и ее возможностями, написание скриптов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту.</i>
	<i>Раздел 2. Объекты типа string и их методы</i>	<i>Лекция 2. Семинар 2. Самостоятельная работа</i>	<i>Лекция с использованием видеоматериалов и проекционной техники. Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту.</i>
	<i>Раздел 3. Объекты типа list и их методы</i>	<i>Лекция 3. Семинар 3. Самостоятельная работа</i>	<i>Лекция с использованием видеоматериалов и проекционной техники. Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту.</i>
	<i>Раздел 4. Объекты типа set и их методы</i>	<i>Лекция 4. Семинар 4. Самостоятельная работа</i>	<i>Лекция с использованием видеоматериалов и проекционной техники. Написание скриптов с использованием пройденных материалов, решение типовых задач. Консультирование и проверка домашних заданий через социальные сети и электронную почту.</i>
	<i>Раздел 5. Функции. Рекурсия</i>	<i>Лекция 5. Семинар 5. Самостоятельная</i>	<i>Лекция с использованием видеоматериалов и проекционной техники.</i>

		<i>я работа</i>	<p><i>Написание скриптов с использованием пройденных материалов, решение типовых задач.</i></p> <p><i>Консультирование и проверка домашних заданий через социальные сети и электронную почту.</i></p>
	<i>Раздел 6. Объекты типа dict и их методы</i>	<i>Лекция 6. Семинар 6. Самостоятельна я работа</i>	<p><i>Лекция с использованием видеоматериалов и проекционной техники.</i></p> <p><i>Написание скриптов с использованием пройденных материалов, решение типовых задач.</i></p> <p><i>Консультирование и проверка домашних заданий через социальные сети и электронную почту.</i></p>
	<i>Раздел 7. Работа с файлами в Python</i>	<i>Семинар 7. Самостоятельна я работа</i>	<p><i>Написание скриптов с использованием пройденных материалов, решение типовых задач.</i></p> <p><i>Консультирование и проверка домашних заданий через социальные сети и электронную почту.</i></p>
2.	<i>Раздел 8. Регулярные выражения в Python. Модуль regex</i>	<i>Лекция 1. Семинар 1. Самостоятельна я работа</i>	<p><i>Лекция с использованием видеоматериалов и проекционной техники.</i></p> <p><i>Написание скриптов с использованием пройденных материалов, решение типовых задач. Представление возможных тем студенческих проектных работ.</i></p>
	<i>Раздел 9. Типы файлов json, csv, модули pandas, json, csv.</i>	<i>Лекция 2. Семинар 2. Самостоятельна я работа</i>	<p><i>Лекция с использованием видеоматериалов и проекционной техники.</i></p> <p><i>Написание скриптов с использованием пройденных материалов, решение типовых задач.</i></p> <p><i>Консультирование и проверка домашних заданий через социальные сети и электронную почту.</i></p>
	<i>Раздел 10. Токенизация и сегментация в Python. Модули razdel, DeepPavlov.</i>	<i>Лекция 3. Семинар 3. Самостоятельна я работа</i>	<p><i>Лекция с использованием видеоматериалов и проекционной техники.</i></p> <p><i>Написание скриптов с использованием пройденных</i></p>

			<p>материалов, решение типовых задач.</p> <p>Консультирование и проверка домашних заданий через социальные сети и электронную почту.</p>
	<p>Раздел 11.</p> <p>Автоматический морфологический анализ. Модули RNNMorph, Mystem.</p>	<p>Лекция 4.</p> <p>Семинар 4.</p> <p>Самостоятельная работа</p>	<p>Лекция с использованием видеоматериалов и проекционной техники.</p> <p>Написание скриптов с использованием пройденных материалов, решение типовых задач.</p> <p>Консультирование и проверка домашних заданий через социальные сети и электронную почту.</p>
	<p>Раздел 12.</p> <p>Автоматический синтаксический анализ. Синтаксические парсеры.</p>	<p>Лекция 5.</p> <p>Семинар 5.</p> <p>Самостоятельная работа</p>	<p>Лекция с использованием видеоматериалов и проекционной техники.</p> <p>Написание скриптов с использованием пройденных материалов, решение типовых задач.</p> <p>Консультирование и проверка домашних заданий через социальные сети и электронную почту.</p>
	<p>Раздел 13. Основы работы с модулем NLTK: базовые функции и возможности</p>	<p>Лекция 6.</p> <p>Семинар 6.</p> <p>Самостоятельная работа</p>	<p>Лекция с использованием видеоматериалов и проекционной техники.</p> <p>Написание скриптов с использованием пройденных материалов, решение типовых задач.</p> <p>Консультирование и проверка домашних заданий через социальные сети и электронную почту.</p>
	<p>Раздел 14. Работа с модулем NLTK: корпуса в составе модуля</p>	<p>Лекция 7.</p> <p>Семинар 7.</p> <p>Самостоятельная работа</p>	<p>Лекция с использованием видеоматериалов и проекционной техники.</p> <p>Написание скриптов с использованием пройденных материалов, решение типовых задач.</p> <p>Консультирование и проверка домашних заданий через социальные сети и электронную почту.</p>
	<p>Презентации студенческих проектов.</p>	<p>Семинар 8.</p>	<p>Представление студентами презентаций и проектов.</p>

5. Оценка планируемых результатов обучения

5.1. Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- домашние задания	5 баллов	5 баллов
- работа на семинарах, решение лингвистических задач	5 баллов	5 баллов
- самостоятельные работы по языкам мира (1-5)	2 балла	10 баллов
- проверочные работы по теории (1-2)	3 балла	6 баллов
- письменное домашнее задание по подсчёту индекса синтетизма	4 балла	4 балла
- коллоквиум №1	15 баллов	15 баллов
- коллоквиум №2	15 баллов	15 баллов
Промежуточная аттестация (экзамен по билетам)		40 баллов
Итого за семестр (дисциплину) экзамен		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/ А,В	«отлично»	<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/ С	«хорошо»	<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	«удовлетвори- тельно»	<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</p>

		<p>необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»	<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 выставляются 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 заданию

- 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)

1. Язык программирования Python и его характеристики. Среда разработки IDLE. Понятие сценарного и интерактивного режима.
2. Классификация объектов в Python. Стандартные классы: int, float, bool.
3. Арифметические и логические операции в Python.
4. Объекты класса string в Python. Их методы.
5. Понятие итерации. Разновидности циклов, особенности их применения.
6. Объекты класса list в Python. Их методы.
7. Изменяемые и неизменяемые объекты.
8. Объекты класса set в Python. Их методы. Логические операции с множествами.
9. Функции. Вызываемые объекты. Аргументы и параметры. Понятие рекурсии.
10. Объекты класса dict в Python. Их методы. Defaultdict и Counter.
11. Внешние библиотеки Python, способы их использования.
12. Дескрипторы файлов. Кодировки. Функция open.
13. Регулярные выражения и правила их составления.
14. Токенизация. Понятие токена. Необходимость токенизации для решения лингвистических задач.
15. Сегментация. Библиотека razdel, библиотека DeepPavlov.
16. Автоматический морфологический анализ: особенности. Используемые для него инструменты, возможности этих инструментов. Проблемы автоматического морфологического анализа.
17. Автоматический синтаксический анализ и доступные парсеры.
18. Формат Universal Dependencies.
19. Библиотека NLTK и существующие в ней функции.
20. Корпуса библиотеки NLTK и их характеристики.

6. Учебно-методическое и информационное обеспечение дисциплины

Изучение дисциплины обеспечивается авторскими учебными пособиями по курсу, а также рекомендованным списком научной литературы.

6.1. Список источников и литературы

Основные

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.

Дополнительные

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.

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/PlyaGusev/rnnmorph>

7. Материально-техническое обеспечение дисциплины (модуля)

Для реализации дисциплины «Технологии обработки текста и звучащей речи» необходим компьютер, проекционное оборудование и звуковые мониторы.

Лицензионное программное обеспечение компьютера:

MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple.

Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

При необходимости рабочая программа дисциплины может быть адаптирована для обеспечения образовательного процесса лицам с ограниченными возможностями здоровья, в том числе для дистанционного обучения. Для этого от студента требуется представить заключение психолого-медико-педагогической комиссии (ПМПК) и личное заявление (заявление законного представителя).

В заключении ПМПК должно быть прописано:

- рекомендуемая учебная нагрузка на обучающегося (количество дней в неделю, часов в день);
- оборудование технических условий (при необходимости);
- сопровождение и (или) присутствие родителей (законных представителей) во время учебного процесса (при необходимости);
- организация психолого-педагогического сопровождения обучающегося с указанием специалистов и допустимой нагрузки (количества часов в неделю).

Для осуществления процедур текущего контроля успеваемости и промежуточной аттестации обучающихся, при необходимости могут быть созданы фонды оценочных средств, адаптированные для лиц с ограниченными возможностями здоровья и позволяющие оценить достижение ими запланированных в основной образовательной программе

результатов обучения и уровень сформированности всех компетенций, заявленных в образовательной программе.

Форма проведения текущей и итоговой аттестации для лиц с ограниченными возможностями здоровья устанавливается с учетом индивидуальных психофизических особенностей (устно, письменно (на бумаге, на компьютере), в форме тестирования и т.п.). При необходимости студенту предоставляется дополнительное время для подготовки ответа на зачете или экзамене.

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 семинарских/ практических/ лабораторных занятий

Семинар 1, 2. Введение. 8 часов.	<i>Вопросы для обсуждения.</i> 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. <i>Литература.</i> 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. —
----------------------------------	--

	<p>(Профессиональное образование). — ISBN 978-5-534-11961-9.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html Интерактивный учебник языка Python http://pythontutor.ru/</p> <p><i>Материально-техническое обеспечение</i> Компьютер, проекционное оборудование и звуковые мониторы. Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple. Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
<p><i>Семинар 3, 4. Объекты типа string. 8 часов</i></p>	<p><i>Вопросы для обсуждения.</i> Объекты класса list в Python. Понятие итерации. Цикл while. Цикл for. Основные методы объектов класса list: append, pop, insert, index, reverse.</p> <p><i>Литература.</i> 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. The Python Standard Library: https://docs.python.org/3/library/index.html Интерактивный учебник языка Python http://pythontutor.ru/</p> <p><i>Материально-техническое обеспечение</i> Компьютер, проекционное оборудование и звуковые мониторы. Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple. Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
<p><i>Семинар 5, 6. Объекты типа list. 8 часов.</i></p>	<p><i>Вопросы для обсуждения.</i> 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.</p> <p><i>Литература.</i> Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для среднего</p>

	<p>профессионального образования / Д. Ю. Федоров. — 2-е изд. — Москва : Издательство Юрайт, 2020. — 161 с. — (Профессиональное образование). — ISBN 978-5-534-11961-9.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html</p> <p>Интерактивный учебник языка Python http://pythontutor.ru/</p> <p><i>Материально-техническое обеспечение</i> Компьютер, проекционное оборудование и звуковые мониторы. Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple. Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
Семинар 7, 8. Объекты типа set. 8 часов.	<p><i>Вопросы для обсуждения.</i> Написание собственных функций в Python. Вызываемые объекты. Определение функции. Имя функции. Аргументы и параметры. Класс NoneType. Понятие рекурсии.</p> <p><i>Литература.</i> 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.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html</p> <p>Интерактивный учебник языка Python http://pythontutor.ru/</p> <p><i>Материально-техническое обеспечение</i> Компьютер, проекционное оборудование и звуковые мониторы. Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple. Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
Семинар 9, 10. Функции. Рекурсия. 8 часов.	<p><i>Вопросы для обсуждения.</i> 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.</p> <p>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.</p>

	<p><i>Литература.</i> 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.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html Интерактивный учебник языка Python http://pythontutor.ru/</p> <p><i>Материально-техническое обеспечение</i> Компьютер, проекционное оборудование и звуковые мониторы. Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple. Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
Семинар 11, 12. Объекты типа dict и их методы. 8 часов.	<p><i>Вопросы для обсуждения.</i> 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.</p> <p><i>Литература.</i> 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.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html Интерактивный учебник языка Python http://pythontutor.ru/</p> <p><i>Материально-техническое обеспечение</i> Компьютер, проекционное оборудование и звуковые мониторы. Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple. Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
Семинар 13, 14. Работа с файлами в Python. 8	<p><i>Вопросы для обсуждения.</i> Дескрипторы файлов в Python. Функция open и ее параметры.</p>

<p>часов.</p>	<p>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.</p> <p><i>Литература.</i></p> <p>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.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html</p> <p><i>Материально-техническое обеспечение</i></p> <p>Компьютер, проекционное оборудование и звуковые мониторы.</p> <p>Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple.</p> <p>Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
<p><i>Семинар 15. Регулярные выражения в Python. Модуль regex. 6 часов.</i></p>	<p><i>Вопросы для обсуждения.</i></p> <p>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.</p> <p><i>Литература.</i></p> <p>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.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html</p> <p><i>Материально-техническое обеспечение</i></p> <p>Компьютер, проекционное оборудование и звуковые мониторы.</p> <p>Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple.</p> <p>Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>

<p><i>Семинар 16. Типы файлов json, csv, модули pandas, json, csv. 4 часа.</i></p>	<p><i>Вопросы для обсуждения.</i> 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.</p> <p><i>Литература.</i> 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.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html</p> <p><i>Материально-техническое обеспечение</i> Компьютер, проекционное оборудование и звуковые мониторы. Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple. Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
<p><i>Семинар 17. Токенизация и сегментация в Python. Модули razdel, DeepPavlov. 6 часов.</i></p>	<p><i>Вопросы для обсуждения.</i> Понятие токенизации. Токен. Автоматическая сегментация по предложениям. Основы работы с модулями razdel и DeepPavlov, автоматическая токенизация и сегментация текстов с их помощью.</p> <p><i>Литература.</i> 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.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html</p> <p><i>Материально-техническое обеспечение</i> Компьютер, проекционное оборудование и звуковые мониторы. Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple. Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные</p>

	операционные системы.
<p><i>Семинар 18.</i> <i>Автоматический морфологический анализ.</i> <i>Модули RNNMorph, Mystem. 6 часов.</i></p>	<p><i>Вопросы для обсуждения.</i> 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. <i>Литература.</i> 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. The Python Standard Library: https://docs.python.org/3/library/index.html Морфологический анализатор Pymorphy2 https://pymorphy2.readthedocs.io/en/latest/ Pymystem3: https://github.com/nlpub/pymystem3 RNNMorph: https://github.com/IlyaGusev/rnnmorph</p> <p><i>Материально-техническое обеспечение</i> Компьютер, проекционное оборудование и звуковые мониторы. Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple. Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
<p><i>Семинар 19.</i> <i>Автоматический синтаксический анализ.</i> <i>Синтаксические парсеры. 6 часов.</i></p>	<p><i>Вопросы для обсуждения.</i> Universal Dependencies: краткий обзор формата. Автоматический синтаксический анализ текста как задача. Библиотека UDPipe и основы работы с ней. Отображение синтаксических деревьев в графическом виде. <i>Литература.</i> 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. The Python Standard Library: https://docs.python.org/3/library/index.html</p> <p><i>Материально-техническое обеспечение</i> Компьютер, проекционное оборудование и звуковые мониторы. Лицензионное программное обеспечение компьютера:</p>

	<p>MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple.</p> <p>Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
<p><i>Семинар 20. Основы работы с модулем NLTK: базовые функции и возможности. 4 часа</i></p>	<p><i>Вопросы для обсуждения.</i></p> <p>Обзор возможностей модуля NLTK, базовые функции. Обработка сырого текста. Токенизация и сегментация модуля NLTK. Подсчет частотности лемм, TF/IDF документов.</p> <p><i>Литература.</i></p> <p>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.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html NLTK book https://www.nltk.org/book/</p> <p><i>Материально-техническое обеспечение</i></p> <p>Компьютер, проекционное оборудование и звуковые мониторы.</p> <p>Лицензионное программное обеспечение компьютера: MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple.</p> <p>Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
<p><i>Семинар 21. Работа с модулем NLTK: корпуса в составе модуля. 4 часа</i></p>	<p><i>Вопросы для обсуждения.</i></p> <p>Основы работы с готовыми корпусами. Корпуса в составе библиотеки NLTK: Gutenberg, инаугурационный корпус, языковые корпуса.</p> <p><i>Литература.</i></p> <p>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.</p> <p>The Python Standard Library: https://docs.python.org/3/library/index.html NLTK book https://www.nltk.org/book/</p> <p><i>Материально-техническое обеспечение</i></p> <p>Компьютер, проекционное оборудование и звуковые мониторы.</p> <p>Лицензионное программное обеспечение компьютера:</p>

	<p>MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple.</p> <p>Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>
<p><i>Семинар 22. Презентации студенческих проектов. 4 часа.</i></p>	<p>Представление студентами проектных работ.</p> <p><i>Материально-техническое обеспечение</i></p> <p>Компьютер, проекционное оборудование и звуковые мониторы.</p> <p>Лицензионное программное обеспечение компьютера:</p> <p>MsWindows и выше, MsOffice 2013 и выше или MacOS на компьютерах Apple.</p> <p>Интерактивные среды разработки IDLE, Wing 101, Pycharm распространяются свободно, устанавливаются на различные операционные системы.</p>

АННОТАЦИЯ

Курс “Технологии обработки текста и звучащей речи” – это курс, посвященный применению вычислительной техники и компьютеров для автоматического анализа и обработки лингвистических данных. Курс предполагает возможности лингвистических исследований с применением новейших доступных технологий, в частности, языка программирования Python и существующих библиотек по автоматической обработке естественного языка.

Предметом курса является язык программирования Python и написанные для него библиотеки для автоматической обработки естественного языка.

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

Задачи курса

Конкретные задачи курса – теоретические и практические.

Теоретические:

- ознакомить студентов с базовыми понятиями объектно-ориентированного программирования;
- ознакомить студентов с принципами работы объектно-ориентированного языка программирования высокого уровня интерпретаторного типа.
- Практические:
- развить навыки программирования в языке Python;
- научить автоматически обрабатывать лингвистические данные при помощи Python и существующих библиотек;
- развить навыки анализа полученных в результате автоматической обработки данных.

Курс нацелен на **формирование** у студентов следующих профессиональных **компетенций**:

- владение основными понятиями и категориями современной лингвистики (ОПК-1);
- владением основными методами фонологического, морфологического, синтаксического, дискурсивного и семантического анализа с учетом языковых и экстралингвистических факторов (ПК-1);
- владением основными методами инструментального анализа звучащей речи (ПК-2);

- способностью пользоваться лингвистически ориентированными программными продуктами (ПК-9);
- владением принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологии; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами (ПК-10);
- способностью проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-12)

В результате овладения дисциплиной студент должен:

- *знать:*
 - — методику сбора и анализа текстов, автоматической их обработки при помощи компьютера,
 - — принципы работы объектно-ориентированного языка высокого уровня интерпретаторного типа;
- *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.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости в форме проверки домашних заданий; итоговая аттестация по курсу — в форме экзамена.

Общая трудоемкость освоения дисциплины составляет 5 зачетных единиц, 180 часов.

Программой дисциплины предусмотрены лекции (14 часов) и практические занятия (112 часов), самостоятельная работа студента (92 часа) и аттестация в форме экзамена.

ЛИСТ ИЗМЕНЕНИЙ

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение №1	25.06.2020	4

1. Образовательные технологии (к п.4 на 2020 г.)

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.

2. Перечень БД и ИСС (к п. 6.2 на 2020 г.)

№п /п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru
4	Компьютерные справочные правовые системы Консультант Плюс, Гарант

3. Состав программного обеспечения (ПО) (к п. 7 на 2020 г.)

№п /п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное

4	AutoCAD 2010 Student	Autodesk	свободно распространяемое
5	Archicad 21 Rus Student	Graphisoft	свободно распространяемое
6	SPSS Statistics 22	IBM	лицензионное
7	Microsoft Share Point 2010	Microsoft	лицензионное
8	SPSS Statistics 25	IBM	лицензионное
9	Microsoft Office 2013	Microsoft	лицензионное
10	ОС «Альт Образование» 8	ООО «Базальт СПО	лицензионное
11	Microsoft Office 2013	Microsoft	лицензионное
12	Windows 10 Pro	Microsoft	лицензионное
13	Kaspersky Endpoint Security	Kaspersky	лицензионное
14	Microsoft Office 2016	Microsoft	лицензионное
15	Visual Studio 2019	Microsoft	лицензионное
16	Adobe Creative Cloud	Adobe	лицензионное
17	Zoom	Zoom	лицензионное