

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по учебной работе

Дата подписания: 22.09.2023 11:02:08

Уникальный идентификатор:

d7a26b9e8ca85e98ec3de2eb454b4659d061f249

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное образовательное учреждение
высшего образования**

«Национальный исследовательский технологический университет «МИСИС»

Рабочая программа дисциплины (модуля)

Информационные технологии в лингвистике

Закреплена за подразделением

Кафедра иностранных языков и коммуникативных технологий

Направление подготовки

45.03.02 ЛИНГВИСТИКА

Профиль

Квалификация

Бакалавр

Форма обучения

очная

Общая трудоемкость

4 ЗЕТ

Часов по учебному плану

144

Формы контроля в семестрах:

в том числе:

экзамен 5

аудиторные занятия

51

самостоятельная работа

57

часов на контроль

36

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	5 (3.1)		Итого	
	Неделя 18			
Вид занятий	УП	РП	УП	РП
Лекции	17	17	17	17
Практические	34	34	34	34
Итого ауд.	51	51	51	51
Контактная работа	51	51	51	51
Сам. работа	57	57	57	57
Часы на контроль	36	36	36	36
Итого	144	144	144	144

Программу составил(и):
асс., Щекочихина А.В.

Рабочая программа

Информационные технологии в лингвистике

Разработана в соответствии с ОС ВО:

Самостоятельно устанавливаемый образовательный стандарт высшего образования - бакалавриат Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский технологический университет «МИСИС» по направлению подготовки 45.03.02 ЛИНГВИСТИКА (приказ от 02.04.2021 г. № 119 о.в.)

Составлена на основании учебного плана:

45.03.02 ЛИНГВИСТИКА, 45.03.02-БЛГ-23.plx , утвержденного Ученым советом НИТУ МИСИС в составе соответствующей ОПОП ВО 22.06.2023, протокол № 5-23

Утверждена в составе ОПОП ВО:

45.03.02 ЛИНГВИСТИКА, , утвержденной Ученым советом НИТУ МИСИС 22.06.2023, протокол № 5-23

Рабочая программа одобрена на заседании

Кафедра иностранных языков и коммуникативных технологий

Протокол от 16.06.2021 г., №10

Руководитель подразделения Бондарева Л.В.

1. ЦЕЛИ ОСВОЕНИЯ

1.1	Основной целью является формирование у студентов умений применять для решения практических задач комплекс знаний об автоматической обработке естественного языка.
-----	---

2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Блок ОП:		Б1.О
2.1	Требования к предварительной подготовке обучающегося:	
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Подготовка к процедуре защиты и защита выпускной квалификационной работы	

3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ

ОПК-5: Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности, работать с компьютером как средством получения, обработки и управления информацией для решения профессиональных задач, осуществлять моделирование, анализ и эксперименты в целях проведения детального исследования для решения задач в профессиональной области	
Знать:	
ОПК-5-31 Теоретические основы получения, обработки и управления информацией	
Уметь:	
ОПК-5-У1 Использовать современные методы компьютерной обработки информации для решения профессиональных задач	
Владеть:	
ОПК-5-В1 Приемами и методами моделирования и анализа информации в целях проведения детального исследования для решения практических задач	

4. СТРУКТУРА И СОДЕРЖАНИЕ

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Формируемые индикаторы компетенций	Литература и эл. ресурсы	Примечание	КМ	Выполняемые работы
	Раздел 1. Раздел 1. Компоненты прикладной лингвистики							
1.1	Основные понятия компьютерной лингвистики /Лек/	5	3	ОПК-5-31	Л1.1Л2.1			
1.2	Тест 1 /Ср/	5	1	ОПК-5-31	Л1.1Л2.1 Э1			
1.3	Автоматическая обработка текста: морфология /Лек/	5	2	ОПК-5-31	Л1.1Л2.1			
1.4	Тест 2 /Ср/	5	1	ОПК-5-31	Л1.1Л2.1 Э1			
1.5	Автоматическая обработка текста: синтаксис /Лек/	5	2	ОПК-5-31	Л1.1Л2.1			
1.6	Тест 3 /Ср/	5	1	ОПК-5-31	Л1.1Л2.1 Э1			
1.7	Корпусная лингвистика /Лек/	5	2	ОПК-5-31	Л1.1Л2.1			
1.8	Тест 4 /Ср/	5	2	ОПК-5-31	Л1.1Л2.1 Э1		КМ4	
1.9	Компьютерная семантика и извлечение информации /Лек/	5	2	ОПК-5-31	Л1.1Л2.1			
1.10	Тест 5 /Ср/	5	1	ОПК-5-31	Л1.1Л2.1 Э1			
	Раздел 2. Раздел 2. Направления прикладной лингвистики							

2.1	Анализ данных и статистика /Лек/	5	2	ОПК-5-31	Л1.1Л2.1			
2.2	Тест 6 /Ср/	5	1	ОПК-5-31	Л1.1Л2.1 Э1			
2.3	Машинный перевод /Лек/	5	2	ОПК-5-31	Л1.1Л2.1			
2.4	Тест 7 /Ср/	5	1	ОПК-5-31	Л1.1Л2.1 Э1			
2.5	Компьютерная лингводидактика /Лек/	5	2	ОПК-5-31	Л1.1Л2.1			
2.6	Тест 8 /Ср/	5	1	ОПК-5-31	Л1.1Л2.1 Э1			
	Раздел 3. Раздел 3. Основы Python для лингвистов							
3.1	Основные операторы, типы переменных /Пр/	5	10	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1			
3.2	Практическая работа 1 /Ср/	5	1	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1 Э1			
3.3	Словари и множества /Пр/	5	10	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1			
3.4	Практическая работа 2 /Ср/	5	1	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1 Э1			
3.5	Условия и циклы /Пр/	5	3	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1			
3.6	Практическая работа 3 /Ср/	5	6	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1 Э1		КМ11	Р1
	Раздел 4. Раздел 4. Библиотеки Python							
4.1	Мystem, Рymorphy /Пр/	5	2	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1		КМ12	
4.2	Практическая работа 4 /Ср/	5	8	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1 Э1		КМ11	
4.3	Pandas /Пр/	5	2	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1		КМ12	
4.4	Практическая работа 5 /Ср/	5	8	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1 Э1		КМ11	
4.5	Регулярные выражения /Пр/	5	2	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1		КМ12	
4.6	Практическая работа 6 /Ср/	5	8	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1 Э1		КМ11	
4.7	NLTK /Пр/	5	2	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1		КМ12	
4.8	Практическая работа 7 /Ср/	5	8	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1 Э1		КМ11	
	Раздел 5. Раздел 5. Подведение итогов							
5.1	Устное собеседование /Пр/	5	3	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1		КМ10	

5.2	Практическая работа 8 /Ср/	5	6	ОПК-5-31 ОПК-5-У1 ОПК-5-В1	Л1.1Л2.1 Э1		КМ11	
5.3	Итоговый тест /Ср/	5	2	ОПК-5-31	Л1.1Л2.1 Э1		КМ9	

5. ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ

5.1. Контрольные мероприятия (контрольная работа, тест, коллоквиум, экзамен и т.п), вопросы для самостоятельной подготовки

Код КМ	Контрольное мероприятие	Проверяемые индикаторы компетенций	Вопросы для подготовки
КМ1	Тест 1		Что такое лингвистика? Перечислите основные направления компьютерной лингвистики. Что такое релевантность и полнота поиска? Какие существуют факторы ранжирования? Как оценить релевантность поиска?
КМ2	Тест 2		Каким может быть анализ слов? Как анализировать слова на основе правил? Как работают анализаторы MyStem и Rymorphy? Какие существуют статистические методы анализа слов?
КМ3	Тест 3		Что такое грамматики зависимостей? Что такое грамматики непосредственных составляющих? Что такое парсинг? В чем заключается проблема комбинаторного взрыва? Какие существуют современные синтаксические анализаторы?
КМ4	Тест 4		Перечислите основные свойства корпуса. Какие бывают корпуса? Как размечаются корпуса? Для чего нужен корпусный менеджер? Как сделать корпус?
КМ5	Тест 5		Назовите виды семантических представлений. Раскройте понятия: сетевые модели, фреймы и сценарии, концептуальные графы. Что такое формальные онтологии? Какова структура формальных онтологий? Какие бывают компьютерные тезаурусы?
КМ6	Тест 6		Сформулируйте задачи машинного обучения. Раскройте понятия: метрические, статистические и линейные методы классификации, искусственные нейронные сети, кластеризация.
КМ7	Тест 7		Как осуществляется перевод на основе правил? Что такое модель языка и цепи Маркова? Какие существуют методы сглаживания? Как осуществляется оценка максимального правдоподобия? Что такое гибридный перевод? Как оценить качество перевода? Какие существуют современные системы машинного перевода?
КМ8	Тест 8		Раскройте понятия: дешифровка, типология, стилеметрия, глоттохронология. Для чего нужны частотные словари?

КМ9	Итоговый тест	<p>Что такое лингвистика? Перечислите основные направления компьютерной лингвистики. Что такое релевантность и полнота поиска? Какие существуют факторы ранжирования? Как оценить релевантность поиска?</p> <p>Каким может быть анализ слов? Как анализировать слова на основе правил? Как работают анализаторы MyStem и Rymorphy? Какие существуют статистические методы анализа слов?</p> <p>Что такое грамматики зависимостей? Что такое грамматики непосредственных составляющих? Что такое парсинг? В чем заключается проблема комбинаторного взрыва? Какие существуют современные синтаксические анализаторы?</p> <p>Перечислите основные свойства корпуса. Какие бывают корпуса? Как размечаются корпуса? Для чего нужен корпусный менеджер? Как сделать корпус?</p> <p>Назовите виды семантических представлений. Раскройте понятия: сетевые модели, фреймы и сценарии, концептуальные графы. Что такое формальные онтологии? Какова структура формальных онтологий? Какие бывают компьютерные тезаурусы?</p> <p>Сформулируйте задачи машинного обучения. Раскройте понятия: метрические, статистические и линейные методы классификации, искусственные нейронные сети, кластеризация.</p> <p>Как осуществляется перевод на основе правил? Что такое модель языка и цепи Маркова? Какие существуют методы сглаживания? Как осуществляется оценка максимального правдоподобия? Что такое гибридный перевод? Как оценить качество перевода? Какие существуют современные системы машинного перевода?</p> <p>Раскройте понятия: дешифровка, типология, стилеметрия, глоттохронология. Для чего нужны частотные словари?</p>
-----	---------------	--

КМ10	Устное собеседование	<p>"Доклад по одной из тем: Сравнение систем автоматического анализа текстов: лингвистические задачи в сервисах Яндекс и Google, OpenCalais, RCO и др.</p> <p>Сравнительный анализ составления поисковых запросов в популярных русскоязычных поисковых системах (Google, Yandex, Rambler, Mail.ru, AltaVista, Yahoo, MSN, AOL)</p> <p>Основные типы компьютерных лингвистических ресурсов. Корпуса текстов, библиотеки, конкордансы. Словари, тезаурусы, онтологии. Лингвистические базы данных</p> <p>Обзор и сравнение сетевых ресурсов по корпусной лингвистике</p> <p>Специальные возможности программы MS Word для лингвистов (проверка правописания, рецензирование, автореферирование, использование шаблонов и т.д.)</p> <p>Особенности электронных переводческих словарей Lingvo и Multitran и их отличия от онлайн-переводчиков (Google, Yandex и т.п.)</p> <p>Технология подкастинга в обучении языкам</p> <p>Современные средства ввода текстовых данных в компьютер.</p> <p>Команда изучает сервисы, сравнивает принципы и алгоритмы их работы, выявляет преимущества и недостатки.</p> <p>Проверка орфографии на разных сервисах. Команда изучает сервисы, сравнивает принципы и алгоритмы их работы, выявляет преимущества и недостатки, слепые зоны.</p> <p>Проверка уникальности текстов. Команда сравнивает, как происходит анализ уникальности на разных ресурсах. Этапы, проблемы</p> <p>Статистические данные и лингвистика. SEO-статистика (search engine optimization). SEO-копирайтинг.</p> <p>Базы данных. Принципы и алгоритмы работы. Проблемы автоматического распознавания текстовых полей.</p> <p>Перевод рукописного текста в печатный.</p> <p>Автоматическая генерация текстов (например, научных статей). Принципы, этапы, проблемы.</p> <p>Исследование способов использования корпусов в лексикографии</p> <p>Русский WordNet</p> <p>FrameNet</p> <p>СУС-онтология</p> <p>Semantic Web</p> <p>Анализ референции в связном тексте</p> <p>Средства построения частотных словарей</p> <p>Логические модели анализа предложения и связного текста</p> <p>Автоматический анализ и синтез звучащей речи</p> <p>Компьютерная лексикография</p> <p>Компьютерная терминография</p> <p>Подходы к построению систем машинного перевода</p> <p>Языки представления знаний (семантики), используемые в информационных технологиях</p> <p>Программные средства обработки статистических данных</p> <p>Примеры задач по обработке статистических данных в среде табличного процессора Excel</p> <p>Средства обработки графической информации</p> <p>Фонетические информационные ресурсы</p> <p>Ошибки автоматического перевода</p> <p>Системы памяти переводов. Сравнение программ переводческой памяти (Trados, Deja Vu и т.п.)</p> <p>Программно-педагогические средства обучения ИЯ. Сравнение мультимедийных программ/ приложений по обучению иностранным языкам</p> <p>Сетевые формы коммуникации (мессенджеры, чаты, форумы) и их влияние на язык</p> <p>Средства обеспечения и поддержки локализации (Multilizer, Passolo и т.п.)"</p>
------	----------------------	---

КМ11	Практические работы		<p>Перечислите основные типы переменных и их особенности. Какие методы строк вы знаете? Что такое конкатенация? Какие операции можно выполнять с разными типами переменных?</p> <p>В чем отличие словаря от множества? Какие методы словарей и множеств вам известны? Для чего применяются словари и множества?</p> <p>В чем заключаются особенности применения циклов for и while? Как правильно задать условие? Что такое бесконечный цикл?</p> <p>Чем различаются анализаторы mystem и rymorphy? В чем преимущества каждого из них? Какие функции они выполняют? Что такое объект типа series? Какие операции с ним вам известны? Для чего нужна библиотека Pandas? Как импортировать файлы в Python? Методы работы с файлами.</p> <p>Что такое регулярные выражения? В каких сферах используются регулярные выражения? Какие возможности даёт знание регулярных выражений? Назовите некоторые регулярные выражения и для чего они применяются.</p> <p>Для чего применяется NLTK? Перечислите возможности NLTK. Используя полученные ранее знания, проведите анализ текста</p>
КМ12	Текущий контроль		<p>Работа на занятиях</p> <p>Перечислите основные типы переменных и их особенности. Какие методы строк вы знаете? Что такое конкатенация? Какие операции можно выполнять с разными типами переменных?</p> <p>В чем отличие словаря от множества? Какие методы словарей и множеств вам известны? Для чего применяются словари и множества?</p> <p>В чем заключаются особенности применения циклов for и while? Как правильно задать условие? Что такое бесконечный цикл?</p> <p>Чем различаются анализаторы mystem и rymorphy? В чем преимущества каждого из них? Какие функции они выполняют? Что такое объект типа series? Какие операции с ним вам известны? Для чего нужна библиотека Pandas? Как импортировать файлы в Python? Методы работы с файлами.</p> <p>Что такое регулярные выражения? В каких сферах используются регулярные выражения? Какие возможности даёт знание регулярных выражений? Назовите некоторые регулярные выражения и для чего они применяются.</p> <p>Для чего применяется NLTK? Перечислите возможности NLTK. Используя полученные ранее знания, проведите анализ текста</p>

5.2. Перечень работ, выполняемых по дисциплине (Курсовая работа, Курсовой проект, РГР, Реферат, ЛР, ПР и т.п.)

Код работы	Название работы	Проверяемые индикаторы компетенций	Содержание работы
P1	Чтение дополнительной литературы		По согласованию с преподавателем

5.3. Оценочные материалы, используемые для экзамена (описание билетов, тестов и т.п.)

Экзамен не предусмотрен

5.4. Методика оценки освоения дисциплины (модуля, практики. НИР)

Оценочные средства	Баллы
Тесты по лекциям x8	32
Практические работы x8	32
Устное собеседование x2	20
Итоговый тест	16

В течение семестра студент получает баллы за выполненные задания. Итоговая оценка по дисциплине осуществляется посредством конвертации итогового балла (процента) в оценку по следующей схеме:

51 - 69 % - удовлетворительно
70 - 84% - хорошо
85 - 100% - отлично

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л1.1	Гусякова А. В.	Информационные технологии и лингвистика XXI века: учебное пособие	Электронная библиотека	Москва: Московский педагогический государственный университет (МПГУ), 2016

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л2.1	Моисеева И. Ю.	Квантитативная лингвистика и новые информационные технологии: учебное пособие	Электронная библиотека	Оренбург: Оренбургский государственный университет, 2017

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Э1	Информационные технологии в лингвистике (LMS Canvas)	https://lms.misis.ru/
----	--	---

6.3 Перечень программного обеспечения

П.1	LMS Canvas
П.2	Python

6.4. Перечень информационных справочных систем и профессиональных баз данных

И.1	Журналы Кембриджского университета (Cambridge University Press) URL: https://www.cambridge.org/core
-----	--

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

Ауд.	Назначение	Оснащение
Любой корпус Мультимедийная	Учебная аудитория для проведения занятий лекционного типа и/или для проведения практических занятий:	комплект учебной мебели до 36 мест для обучающихся, мультимедийное оборудование, магнитно-маркерная доска, рабочее место преподавателя, ПК с доступом к ИТС «Интернет», ЭИОС университета через личный кабинет на платформе LMS Canvas, лицензионные программы MS Office, MS Teams, ESET Antivirus
Любой корпус Мультимедийная	Учебная аудитория для проведения занятий лекционного типа и/или для проведения практических занятий:	комплект учебной мебели до 36 мест для обучающихся, мультимедийное оборудование, магнитно-маркерная доска, рабочее место преподавателя, ПК с доступом к ИТС «Интернет», ЭИОС университета через личный кабинет на платформе LMS Canvas, лицензионные программы MS Office, MS Teams, ESET Antivirus
Читальный зал №3 (Б)		комплект учебной мебели на 44 места для обучающихся, МФУ Xerox VersaLink B7025 с функцией масштабирования текстов и изображений, 8 ПК с доступом к ИТС «Интернет», ЭИОС университета через личный кабинет на платформе LMS Canvas, лицензионные программы MS Office, MS Teams, ESET Antivirus.
Читальный зал №4 (Б)		комплект учебной мебели на 20 рабочих мест, компьютеры с подключением к сети «Интернет» и доступом в электронную информационно-образовательную среду университета
Читальный зал электронных ресурсов		комплект учебной мебели на 55 мест для обучающихся, 50 ПК с доступом к ИТС «Интернет», ЭИОС университета через личный кабинет на платформе LMS Canvas, лицензионные программы MS Office, MS Teams, ESET Antivirus.

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ

1	Обучающиеся должны собрать электронное портфолио своих работ
2	Обучающиеся должны предоставить конспекты лекций
3	Обучающиеся должны освоить материалы лекционного курса
4	Обучающиеся должны принимать активное участие в обсуждениях по темам лекций и результатам практических работ
5	Обучающиеся должны соблюдать сроки выполнения заданий и тестов