

Документ подписан простой электронной подписью  
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по учебной работе

Дата подписания: 28.09.2023 12:42:12

Уникальный идентификатор документа:

d7a26b9e8ca85e98ec3de2eb454b4659d061f249

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение  
высшего образования

«Национальный исследовательский технологический университет «МИСИС»

## Рабочая программа дисциплины (модуля)

# Обработка естественного языка

Закреплена за подразделением

Кафедра АСУ

Направление подготовки

45.04.02 ЛИНГВИСТИКА

Профиль

Цифровая лингвистика и локализация

Квалификация

Магистр

Форма обучения

очная

Общая трудоемкость

6 ЗЕТ

Часов по учебному плану

216

Формы контроля в семестрах:

в том числе:

экзамен 4

аудиторные занятия

66

зачет с оценкой 3

самостоятельная работа

114

часов на контроль

36

### Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	3 (2.1)		4 (2.2)		Итого	
	Неделя		10			
Вид занятий	УП	РП	УП	РП	УП	РП
Лекции	13	13	9	9	22	22
Практические	26	26	18	18	44	44
Итого ауд.	39	39	27	27	66	66
Контактная работа	39	39	27	27	66	66
Сам. работа	69	69	45	45	114	114
Часы на контроль			36	36	36	36
Итого	108	108	108	108	216	216

Программу составил(и):  
*ст.преп., Агабубаев Аслан*

Рабочая программа

**Обработка естественного языка**

Разработана в соответствии с ОС ВО:

Самостоятельно устанавливаемый образовательный стандарт высшего образования - магистратура Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский технологический университет «МИСИС» по направлению подготовки 45.04.02 ЛИНГВИСТИКА (приказ от 02.04.2021 г. № 119 о.в.)

Составлена на основании учебного плана:

45.04.02 ЛИНГВИСТИКА, 45.04.02-МЛГ-23-3.plx Цифровая лингвистика и локализация, утвержденного Ученым советом НИТУ МИСИС в составе соответствующей ОПОП ВО 22.06.2023, протокол № 5-23

Утверждена в составе ОПОП ВО:

45.04.02 ЛИНГВИСТИКА, Цифровая лингвистика и локализация, утвержденной Ученым советом НИТУ МИСИС 22.06.2023, протокол № 5-23

Рабочая программа одобрена на заседании

**Кафедра АСУ**

Протокол от 10.04.2023 г., №5

Руководитель подразделения Темкин Игорь Олегович

**1. ЦЕЛИ ОСВОЕНИЯ**

1.1	Целью освоения дисциплины «Интеллектуальный анализ данных» является формирование системного базового представления, первичных знаний, умений и навыков студентов по основам построения современных систем с элементами машинного обучения, познакомить студентов с ключевыми технологиями и инструментами интеллектуального анализа данных. Показать возможности программных сред и инструментов пред и пост обработки данных, рассмотреть работу с библиотеками, показать структуры данных (Series, Dataframe), агрегирование и визуализацию данных.
-----	---

**2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ**

Блок ОП:		Б1.В
<b>2.1</b>	<b>Требования к предварительной подготовке обучающегося:</b>	
2.1.1	Основы программирования	
2.1.2	Основы скорочтения	
2.1.3	Способы быстрого запоминания	
<b>2.2</b>	<b>Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:</b>	

**3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ**

<b>ПК-6: Способен использовать потенциал смежных предметных областей для повышения уровня своей профессиональной экспертизы</b>	
<b>Знать:</b>	
ПК-6-31 современные программно-аналитические платформы и инструменты обработки и анализа данных, способы и механизмы управления данными подходы и математические методы формализации и решения отраслевых (бизнес) задач	
<b>ПК-1: Способен разрабатывать цифровые лингвистические продукты</b>	
<b>Знать:</b>	
ПК-1-31 Знать программные технологии сбора, обработки и анализа данных	
<b>ПК-6: Способен использовать потенциал смежных предметных областей для повышения уровня своей профессиональной экспертизы</b>	
<b>Уметь:</b>	
ПК-6-У1 применять современные технологии в разработке аналитических алгоритмов, программные прототипы прикладных данных применять системный подход в задачах управления	
<b>ПК-1: Способен разрабатывать цифровые лингвистические продукты</b>	
<b>Уметь:</b>	
ПК-1-У1 Уметь применять программные среды и инструменты в задачах обработки и анализа данных	
<b>ПК-6: Способен использовать потенциал смежных предметных областей для повышения уровня своей профессиональной экспертизы</b>	
<b>Владеть:</b>	
ПК-6-В1 навыками обработки и анализа данных на языках программирования C++ и Python навыками применения современных технологий сбора, обработки анализа данных в исследовании социально-экономических системах	
<b>ПК-1: Способен разрабатывать цифровые лингвистические продукты</b>	
<b>Владеть:</b>	
ПК-1-В1 навыками программной реализации математических моделей обработки данных	

**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. Глубокое обучение для обработки естественного языка							

1.1	Особенность применение подходов глубокого обучения к обработке естественного языка /Лек/	4	4					
1.2	Системы генерации текстов /Лек/	4	3	ПК-6-31				
1.3	Sentiment Analysis и системы Speech-to-Text и Text-to -Speech /Лек/	4	2	ПК-6-31				
1.4	Изучение NLTK библиотеки. /Пр/	4	4	ПК-1-31				
1.5	Применение BERT подхода для языковых моделей /Пр/	4	8					
1.6	Современные подходы применения глубокого обучения для языковых моделей /Пр/	4	6					
1.7	Разработка прототипа лингвистического приложения с использованием модели глубокого обучения для естественного языка. /Ср/	4	45					
	<b>Раздел 2. Введение в анализ данных</b>							
2.1	Введение в анализ данных /Лек/	3	1	ПК-1-31	Л1.2 Л1.5 Л1.7 Л1.8 Л1.11			
2.2	Основные методы машинного обучения. Ввиды обучения. Метрики. /Лек/	3	4	ПК-1-31	Л1.11			
2.3	Основные задачи машинного обучения. /Лек/	3	4	ПК-1-31	Л1.11			
2.4	Методы машинного обучения для лингвистических систем /Лек/	3	4					
	<b>Раздел 3. Машинное обучение для обработки естественного языка</b>							
3.1	Установка Python. Работа с Anaconda. /Пр/	3	2	ПК-1-У1 ПК-1 -В1	Л1.1 Л1.3 Л1.4 Л1.6 Л1.9 Л1.10 Э9			
3.2	Структуры данных. Создание. Свойства. Вывод значений. Срезы данных/копирование и ссылки. Удаление /фильтрация по условию. Добавление/выравнивание/ сортировка данных с использованием специализированных библиотек. /Пр/	3	4	ПК-1-У1 ПК-1 -В1	Л1.10 Э9			
3.3	Подготовка данных. Работа с пропусками и шумами /Пр/	3	4	ПК-1-У1 ПК-1 -В1	Л1.1 Э9			
3.4	Визуализация. Настройка графиков. Основные типы графиков. /Пр/	3	4		Л1.10 Э9			

3.5	Использование библиотеки sklearn для машинного обучения /Пр/	3	8	ПК-1-У1 ПК-1-В1	Л1.10 Э9			
3.6	Первичная обработка текста для использование системах ML и DL /Пр/	3	4					
3.7	Разработка прототипа лингвистического приложения с использованием модели машинного обучения для естественного языка. /Ср/	3	69					

### 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	Экзамен		<ol style="list-style-type: none"> <li>1. Анализ данных, как область экспертной деятельности. Функции аналитика.</li> <li>2. Сферы применения технологий анализа данных. Задачи, составляющие основу анализа данных. Data Mining</li> <li>3. Обучение, как процесс построения модели. Роль учителя-аналитика.</li> <li>4. Основные процедуры на стадии подготовки датасета.</li> <li>5. Первичное преобразование данных.</li> <li>6. Шкалирование и кодирование данных</li> <li>7. Подготовка обучающих наборов. Корреляция и автокорреляция.</li> <li>8. Подготовка обучающих наборов. Ранговая корреляция.</li> <li>9. Подготовка обучающих наборов. Суть факторного анализа.</li> <li>10. Метод главных компонент.</li> <li>11. Методы классификации данных.</li> <li>12. Достоинства и недостатки метрических методов классификации.</li> <li>13. Классификация данных с использованием энтропийных критериев. ID3</li> <li>14. Классификация данных с использованием семейства деревьев - «случайный лес».</li> <li>15. Обучение и оптимизация классифицирующих ИНС. Используемые критерии.</li> <li>16. Регрессия. Виды регрессионных моделей</li> <li>17. Параметрическая и непараметрическая регрессия. Области применения в задачах анализа данных.</li> <li>18. Логистическая регрессии. Пример использования.</li> <li>19. Градиентный бустинг. Общий принцип работы.</li> <li>20. Кластеризация. Постановка задачи.</li> <li>21. Кластеризация. Общий обзор методов.</li> <li>22. Алгоритмы кластеризации. Иерархическая кластеризация.</li> <li>23. Алгоритмы кластеризации. Метод k-средних.</li> <li>24. Кластеризация и ИНС. Самообучающиеся сети.</li> <li>25. Самообучение и самоорганизация сетей Кохонена.</li> <li>26. Нечеткая кластеризация.</li> <li>27. Базы экспертных знаний, как нечеткие функции.</li> <li>28. Аппроксимация данных с помощью нечетких функций</li> </ol>

#### 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	Практическая работа №2		Настройка репозитория, установка библиотек и запуск Jupyter Notebook

P2	Практическая работа №3		Структуры данных. Создание. Свойства. Вывод значений. Срезы данных/копирование и ссылки. Удаление /фильтрация по условию. Добавление/выравнивание/сортировка данных. Работа с индексами. Мотивация/операции. Иерархическая индексация/типы.
P3	Практическая работа №4		Чистка данных. Пропущенные значения. Повторяющиеся значения/замена значений/применений функций. Математические операции. Сводка статистик / Арифметические операции. Одномерные статистики. Преобразование данных/оконные функции
P4	Практическая работа №5		Агрегирование данных. Механизм GroupBy / Типы группировок. Агрегирование, преобразование, исключение групп / Сводная таблица
P5	Практическая работа №6		Визуализация. Настройка графиков. Основные типы графиков

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

Экзаменационный билет содержит 2 вопроса по изученному материалу

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

Оценка практических работ:

Оценка "Отлично"- Студент имеет глубокие знания, умения, навыки, демонстрирует полное понимание проблемы, все задания выполнены.

Оценка "хорошо" -Студент имеет глубокие знания, умения, навыки, демонстрирует полное понимание проблемы, все задания выполнены. Законченный, полный ответ с минимальными недочетами.

Оценка "удовлетворительно" - Студент имеет низкий уровень знаний,умений, навыков, демонстрирует частичное понимание проблемы, большинство заданий выполнены,но ответ, содержит недочеты.

Оценка "неудовлетворительно" - Студент имеет пробелы в знаниях, умениях, навыках, демонстрирует непонимание проблемы, задания не выполнены, студенту требуются дополнительные занятия для освоения компетенций

Оценка за выполнение самостоятельных работ:

Оценка "отлично" - Выполнены все задания работы; работа выполнена в срок, оформление, структура и стиль работы образцовые; работа выполнена самостоятельно, присутствуют собственные обобщения, заключения и выводы.

Оценка "хорошо" - Выполнены все задания работы с незначительными замечаниями; работа выполнена в срок; в оформлении, структуре и стиле работы нет грубых ошибок; работа выполнена самостоятельно.

Оценка "удовлетворительно" - Задания работы имеют значительные замечания, устраненные во время контактной работы с преподавателем; работа выполнена с нарушениями графика, в оформлении, структуре и стиле работы есть недостатки; работа выполнена самостоятельно.

Оценка "неудовлетворительно" - Часть работы или вся работа выполнена из фрагментов работ других авторов и носит несамостоятельный характер; задания в работе решены не полностью или решены неправильно; содержание работы не соответствует поставленной теме; при написании работы не были использованы литературные источники; оформление работы не соответствует требованиям.

Дисциплина считается освоенной при выполнении следующих условий:

- выполнены все лабораторные работы
- выполнены и защищены все самостоятельные работы;("удовлетворительно"; "хорошо"; "отлично");

## 6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

### 6.1. Рекомендуемая литература

#### 6.1.1. Основная литература

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л1.1	Сузи Р. А.	Язык программирования Python: учебное пособие	Электронная библиотека	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ)  Бином. Лаборатория знаний, 2007

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л1.2	Сотник С. Л.	Проектирование систем искусственного интеллекта: курс: учебное пособие	Электронная библиотека	Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), 2007
Л1.3	Буйначев С. К., Боклаг Н. Ю.	Основы программирования на языке Python: учебное пособие	Электронная библиотека	Екатеринбург: Издательство Уральского университета, 2014
Л1.4	Северенс Ч.	Введение в программирование на Python	Электронная библиотека	Москва: Национальный Открытый Университет «ИНТУИТ», 2016
Л1.5	Сергеев Н. Е.	Системы искусственного интеллекта: учебное пособие	Электронная библиотека	Таганрог: Южный федеральный университет, 2016
Л1.6	Балджи А. С., Хрипунова М. Б., Александрова И. А.	Математика на Python: учебно-методическое пособие	Электронная библиотека	Москва: Прометей, 2018
Л1.7	Мыльников Л. А., Краузе Б., Кютц М., Баде К., Шмидт И. А.	Интеллектуальный анализ данных в управлении производственными системами (подходы и методы): монография	Электронная библиотека	Москва: Библио-Глобус, 2017
Л1.8	Немтинов В. А., Карпушкин С. В., Мокрозуб В. Г., Малыгин Е. Н., Егоров С. Я.	Интеллектуальные системы проектирования и управления техническими объектами: учебное пособие	Электронная библиотека	Тамбов: Тамбовский государственный технический университет (ТГТУ), 2017
Л1.9	Шелудько В. М.	Основы программирования на языке высокого уровня Python: учебное пособие	Электронная библиотека	Ростов-на-Дону, Таганрог: Южный федеральный университет, 2017
Л1.10	Шелудько В. М.	Язык программирования высокого уровня Python: функции, структуры данных, дополнительные модули: учебное пособие	Электронная библиотека	Ростов-на-Дону, Таганрог: Южный федеральный университет, 2017
Л1.11	Алпайдин Э.	Машинное обучение: новый искусственный интеллект: пер. с англ.	Библиотека МИСиС	М.: Альпина Паблишер, 2017

#### 6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Э1	Лань [Электронный ресурс] : электронно-библиотечная система (ЭБС) / издательство Лань. – URL: <a href="http://e.lanbook.com/">http://e.lanbook.com/</a>	
Э2	Университетская библиотека онлайн [Электронный ресурс] : электронно-библиотечная система (ЭБС) / ООО Директмедиа Паблишинг. – URL: <a href="http://biblioclub.ru/">http://biblioclub.ru/</a>	
Э3	Юрайт [Электронный ресурс] : электронно-библиотечная система (ЭБС) / издательство Юрайт. – URL: <a href="https://biblio-online.ru">https://biblio-online.ru</a>	
Э4	Znanium.com [Электронный ресурс] : электронно-библиотечная система (ЭБС) / Научно-издательский центр ИНФРА-М. – URL: <a href="http://znanium.com/">http://znanium.com/</a>	
Э5	BOOK.ru [Электронный ресурс] : электронно-библиотечная система (ЭБС) / издательство КноРус. – URL: <a href="https://www.book.ru/">https://www.book.ru/</a>	
Э6	eLIBRARY.RU [Электронный ресурс] : электронная библиотека / Науч. электрон. б-ка. – URL: <a href="http://elibrary.ru/defaultx.asp">http://elibrary.ru/defaultx.asp</a>	

Э7	ИНФОРМИО [Электронный ресурс] : электронный справочник [обеспечение всех типов образовательных учреждений нормативными, методическими, научно-практическими материалами]. – URL: <a href="http://www.informio.ru/">http://www.informio.ru/</a> .	
Э8	<a href="http://www.loginom.ru">www.loginom.ru</a>	
Э9	<a href="https://www.anaconda.com/distribution/">https://www.anaconda.com/distribution/</a>	

### 6.3 Перечень программного обеспечения

П.1	LMS Canvas
П.2	Microsoft Office
П.3	MS Teams
П.4	Loginom
П.5	Python

### 6.4. Перечень информационных справочных систем и профессиональных баз данных

И.1	ЭБС "Лань" ( <a href="https://e.lanbook.com">https://e.lanbook.com</a> )
И.2	Единое окно доступа к образовательным ресурсам ( <a href="http://window.edu.ru">http://window.edu.ru</a> )
И.3	Scopus - единая реферативная база данных научных публикаций ( <a href="http://www.scopus.com">www.scopus.com</a> )
И.4	Открытый образовательный видеопортал UniverTV.ru. Образовательные фильмы на различные темы. Лекции в ведущих российских и зарубежных вузах. Научная конференция или научно-популярная лекция по интересующему вас вопросу. <a href="http://univertv.ru/video/matematika/">http://univertv.ru/video/matematika/</a>

## 7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

Ауд.	Назначение	Оснащение
Л-826	Лаборатория	доска и маркеры, персональные компьютеры ОС Windows с администраторскими правами доступа, с проводными сетевыми платами, с СОМ-портами количеством не менее 6, сетевое коммуникационное оборудование CISCO: 6 коммутаторов и 6 маршрутизаторов, обжатые кабели витая пара прямые и кроссовые количеством не менее 12 каждый, консольные кабели количеством не менее 6
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.
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

## 8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ

Подготовка к лекциям.

Подготовка к лекционному занятию включает выполнение всех видов заданий, рекомендованных к каждой лекции, т.е. задания выполняются еще до лекционного занятия по соответствующей теме.

В ходе лекционных занятий необходимо вести конспектирование учебного материала, обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации. При необходимости задавать преподавателю уточняющие вопросы.

Работая над конспектом лекций, Вам всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть теоретическим материалом.

Подготовка к практическим занятиям.

Подготовку к каждому практическому занятию Вы должны начать с ознакомления с планом практического занятия, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованной к данной теме. Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса.

В процессе подготовки к лабораторным занятиям Вам необходимо обратить особое внимание на самостоятельное



изучение рекомендованной литературы. При всей полноте конспектирования лекции в ней невозможно изложить весь материал из-за лимита аудиторных часов. Поэтому самостоятельная работа с учебниками, учебными пособиями, научной, справочной литературой, материалами периодических изданий и Интернета является наиболее эффективным методом получения дополнительных знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у Вас отношение к конкретной проблеме.

Ваша самостоятельная работа может осуществляться в аудиторной и внеаудиторной формах. Самостоятельная работа в аудиторное время включает:

Самостоятельную работу по теоретическому курсу: аудиторную самостоятельную работу на лекциях, работу с лекционным материалом после лекции, подготовка к коллоквиумам и контрольным работам. Самостоятельная работа по практической части ( решение задач, программирование) может быть осуществлена как в аудитории, так и вне ее.

Преподавателем ведется специальный журнал, куда заносятся все промежуточные результаты студентов. Задания выдаются, проверяются и оцениваются в LMS Canvas.