

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по учебной работе

Дата подписания: 28.09.2023 12:48:13

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)

NLP-аналитика

Закреплена за подразделением

Кафедра АСУ

Направление подготовки

45.04.02 ЛИНГВИСТИКА

Профиль

Цифровая лингвистика и локализация

Квалификация

Магистр

Форма обучения

очная

Общая трудоемкость

3 ЗЕТ

Часов по учебному плану

108

Формы контроля в семестрах:

в том числе:

зачет с оценкой 4

аудиторные занятия

27

самостоятельная работа

81

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	4 (2.2)		Итого	
	10			
Неделя	УП	РП	УП	РП
Вид занятий				
Лекции	9	9	9	9
Практические	18	18	18	18
Итого ауд.	27	27	27	27
Контактная работа	27	27	27	27
Сам. работа	81	81	81	81
Итого	108	108	108	108

Программу составил(и):

дтн, Нач. каф., Темкин И.О.

Рабочая программа

NLP-аналитика

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, д.т.н., доцент

1. ЦЕЛИ ОСВОЕНИЯ

1.1	Целями освоения дисциплины (модуля) являются изучение основных нейросетевых парадигм и механизмов использования ИНС с различной архитектурой в задачах машинного обучения, а также формирование навыков использования этих инструментов для построения программных систем обработки и анализа данных
-----	--

2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Блок ОП:		Б1.В
2.1	Требования к предварительной подготовке обучающегося:	
2.1.1	Научно-исследовательская работа	
2.1.2	Основы машинного обучения	
2.1.3	Управление машинным переводом	
2.1.4	Учебная (консультационная) практика	
2.1.5	Квантитативная лингвистика	
2.1.6	Контроль качества переводческих проектов	
2.1.7	Написание научных статей для научных журналов	
2.1.8	Основы программирования	
2.1.9	Основы скорочтения	
2.1.10	Способы быстрого запоминания	
2.1.11	Нормативная база в области лингвистики	
2.1.12	Профильные интерфейсы лингвиста	
2.1.13	Современные подходы к управлению командами	
2.1.14	Форматирование и верстка	
2.1.15	Локализация программного обеспечения и игр	
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	

3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ

ПК-2: Способен автоматизировать лингвистические и локализационные проекты, а также управлять производственным процессом перевода
Знать:
ПК-2-31 принципы использования концепций машинного обучения при управлении лингвистическими системами
ПК-1: Способен разрабатывать цифровые лингвистические продукты
Знать:
ПК-1-31 принципы анализа данных в проблемной области и характеристики программных инструментов
ОПК-7: Способен работать с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации
Знать:
ОПК-7-31 основные характеристики современных аналитических и инструментальных платформ
ПК-2: Способен автоматизировать лингвистические и локализационные проекты, а также управлять производственным процессом перевода
Уметь:
ПК-2-У1 использовать нейросетевые программные при решении практик
ПК-1: Способен разрабатывать цифровые лингвистические продукты
Уметь:
ПК-1-У1 сформировать требования к разрабатываемой системе в соответствии с описанием проблемной области
ОПК-7: Способен работать с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации
Уметь:
ОПК-7-У1 проводить сравнительный анализ различных решений

ПК-2: Способен автоматизировать лингвистические и локализационные проекты, а также управлять производственным процессом перевода
Владеть:
ПК-2-В1 разработки и адаптации программных средств для решения практических задач на основе алгоритмов машинного обучения
ПК-1: Способен разрабатывать цифровые лингвистические продукты
Владеть:
ПК-1-В1 навыками разработки и адаптации программных инструментов для построения системы машинного обучения
ОПК-7: Способен работать с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации
Владеть:
ОПК-7-В1 проводить сравнительный анализ различных решений

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. Фундаментальные основы искусственных нейронных сетей. Архитектура и основные парадигмы обучения и области применения							
1.1	Введение в нейронные сети. Принципы использования ИНС в задачах распознавания, прогнозирования и управления /Лек/	4	2		Л1.2 Л1.3			
1.2	Базовые библиотеки Python для работы с лингвистическими данными. Сравнительный анализ различных функций активации при решении задач NLP. /Пр/	4	4					
	Раздел 2. Алгоритмы обучения ИНС. Обучение с учителем и самообучение							
2.1	Методы обучения. Алгоритмы обучения. Метрики обучения /Лек/	4	2					
2.2	Методы селекции наиболее адекватных ИНС. Применение ИНС для языковых систем. /Пр/	4	8		Л1.1			
2.3	Анализ различных методов машинного обучения для лингвистических моделей /Ср/	4	41					
	Раздел 3. Глубокое обучение. Рекуррентные и конволюционные сети							

3.1	Рекуррентные сети Джордана-Элмана. Концепция глубокого обучения. LSTM - сети. Архитектура и применение CNN - конволюционные сети. Архитектура и применение для языковых моделей /Лек/	4	5		Э1			
3.2	отработка практических навыков программирования в среде Python с подключением библиотек прогнозной аналитики для работы с языковыми моделями /Пр/	4	6					
3.3	Разработка NLP системы с использованием библиотек Python. /Ср/	4	40					

5. ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ

5.1. Вопросы для самостоятельной подготовки к экзамену (зачёту с оценкой)

Следующие вопросы для самостоятельной подготовки к экзамену, контролируют освоение следующих компетенций: УК-5-31, ОПК-8-31, ОПК-9-31, ПК-1-31

Искусственный интеллект: краткая история развития и области применения.

Искусственные нейронные сети (ИНС) и искусственный интеллект.

Аналитические платформы (АП). Общие сведения и типы задач, решаемых при помощи АП.

Предобработка данных. Основные процедуры.

Варианты классификации ИНС. Рассмотреть конкретный пример.

Статистические и ИНС модели сложных объектов в схеме «вход-выход».

Правило Хэбба. Классический персептрон. Функции активации.

Обучение с учителем. Обучение ИНС персептронного типа. Метод обратного распространения.

Классы задач, решаемых с помощью ИНС обратного распространения. Рассмотреть пример.

Обучение и оптимизация ИНС обратного распространения (feed forward back propagation - MLP). Используемые критерии.

Прогнозирование процессов с использованием ИНС MLP.

Градиентные методы обучения ИНС MLP

Обучение MLP методом Левенберга-Маркуардта.

Алгоритм выбора оптимальной ИНС в задаче распознавания образов.

Радиально-базисные функции (RBF). Структура RBF нейронных сетей

Алгоритм обучения RBF сетей. Сравнение RBF и MLP.

Обучение без учителя. Основные виды самообучающихся ИНС. Особенности datasets.

Самообучающиеся ИНС. Классы решаемых задач.

Сеть Кохонена. Рассмотреть пример работы сети

Нейроны Гроссберга. Архитектура сетей встречного распространения.

Рекуррентные сети. Классические сети с обратными связями.

Рекуррентные сети. Виды рекуррентных сетей.

Использование рекуррентных сетей для анализа временных рядов. Пример.

Рекуррентные сети и глубокое обучение. Рекуррентный нейрон.

LSTM сети. Архитектура сетей.

Алгоритм работы LSTM сети.

Неокогнитрон. Принципы работы сверточных сетей (CNN).

Архитектура CNN и основные области применения.

Обучение CNN.

Программные системы (платформы) построения ИНС с глубоким обучением.

5.2. Перечень работ, выполняемых по дисциплине (модулю, практике, НИР) - эссе, рефераты, практические и расчетно-графические работы, курсовые работы, проекты и др.

В рамках выполнения практических работ, контролируется освоение следующих компетенций: УК-5-У1, ОПК-8-У1, ОПК-9-У1, ПК-1-У1

УК-5-В1, ОПК-8-В1, ОПК-9-В1, ПК-1-В1

1. Подготовка данных, необходимых для корректного обучения нейронной сети. Первичная обработка. Структурирование. Использование процедур data mining.
2. Элементарные основы нейроматематики. Сравнительный анализ различных функций активации при решении задач распознавания образов.
3. Способы описания пространства образов.
4. Обучение с учителем и самообучение. Примеры и модели.
5. Интерпретация результатов моделирования, полученных с использованием ИНС
6. Алгоритмы классификации и кластеризации.
7. Рынок инструментов Data Mining.
8. Рынок инструментов Data Mining, в частности, его развитие, поставщики инструментов, классификация инструментов.
9. Критерии, по которым можно сравнивать и выбирать инструмент Data Mining.
10. Инструменты Data Mining.
11. Программные продукты Cognos и система STATISTICA Data Miner
12. Средства анализа и схема работы.

7. Сферы применения Data Mining

8. Основные сферы деятельности человека, где может успешно применяться технология Data Mining.

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

Экзаменационные билеты по дисциплине содержат два теоретических вопроса и хранятся на территории кафедры

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

Дисциплина считается освоенной при выполнении следующих условий:

- текущий лекционный контроль имеет положительные оценки ("удовлетворительно"; "хорошо"; "отлично");
- выполнены и защищены все практические и самостоятельные работы на положительную оценку ("удовлетворительно"; "хорошо"; "отлично").

Знания обучающихся оцениваются по пятибалльной шкале оценками «отлично», «хорошо», «удовлетворительно» и «неудовлетворительно».

Оценки «отлично» заслуживает обучающийся, обнаруживший всестороннее, систематическое и глубокое знание учебного материала, самостоятельно выполнивший все предусмотренные программой задания, хорошо усвоивший теоретический материал, активно работавший на лабораторных занятиях, разбирающийся в основных научных концепциях по изучаемой дисциплине, проявивший творческие способности и научный подход в понимании и изложении учебного материала, ответ отличается богатством и точностью использованных терминов, материал излагается последовательно и логично.

Оценки «хорошо» заслуживает обучающийся, обнаруживший полное знание учебно-программного материала, не допускающий в ответе существенных неточностей, самостоятельно выполнивший все предусмотренные программой задания, активно работавший на практических занятиях, показавший систематический характер знаний по дисциплине, достаточный для дальнейшей учебы, а также способность к их самостоятельному пополнению.

Оценки «удовлетворительно» заслуживает обучающийся, обнаруживший знание основного учебного материала в объеме, необходимом для дальнейшей учебы и предстоящей работы по профессии, не отличавшийся активностью на практических занятиях, самостоятельно выполнивший основные предусмотренные программой задания, однако допустивший некоторые погрешности при их выполнении и в ответе на экзамене, обладающий необходимыми знаниями для устранения под руководством преподавателя допущенных погрешностей.

Оценки «неудовлетворительно» выставляется обучающемуся, обнаружившему пробелы в знаниях или отсутствие знаний по значительной части основного учебного материала, не выполнившего самостоятельно предусмотренные программой задания, допустившему принципиальные ошибки в выполнении предусмотренных программой заданий, не отработавшему основные практические занятия, допускающему существенные ошибки при ответе, и который не может продолжить обучение и приступить к профессиональной деятельности без дополнительных занятий по соответствующей дисциплине.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

6.1. Рекомендуемая литература

6.1.1. Основная литература

Авторы, составители	Заглавие	Библиотека	Издательство, год
---------------------	----------	------------	-------------------

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л1.1	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
Л1.2	Петровичев Е. И.	Введение в искусственные нейронные сети: учеб. пособие по дисц. "Нейротехнологии в управлении"	Библиотека МИСиС	М.: Изд-во МГГУ, 2008
Л1.3	Галушкин А. И.	Нейронные сети: основы теории: монография	Библиотека МИСиС	М.: Горячая линия - Телеком, 2010

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Э1	13) "Нейросети и нейрокомпьютеры" https://openedu.ru/course/urfu/PYDNN/ "Программирование глубоких нейронных сетей на Python" УрФУ	
Э2	https://openedu.ru/course/msu/PARPROG/ "Суперкомпьютеры и параллельная обработка данных" МГУ имени М.В.Ломоносова	

6.3 Перечень программного обеспечения

П.1	LMS Canvas
П.2	SAP (удаленный доступ предоставляет заказчик - IBS)
П.3	Python
П.4	R Studio
П.5	MATLAB
П.6	Statistica Neural Networks

6.4. Перечень информационных справочных систем и профессиональных баз данных

И.1	Полнотекстовые российские научные журналы и статьи:
И.2	— Научная электронная библиотека eLIBRARY https://elibrary.ru/
И.3	— Полнотекстовые деловые публикации информагентств и прессы по 53 отраслям https://polpred.com/news
И.4	— Российская Государственная библиотека https://www.rsl.ru
И.5	— Единое окно доступа к информационным ресурсам http://window.edu.ru
И.6	Иностранные базы данных (доступ с IP адресов МИСиС):
И.7	— аналитическая база (индексы цитирования) Web of Science https://apps.webofknowledge.com
И.8	— аналитическая база (индексы цитирования) Scopus https://www.scopus.com/
И.9	— наукометрическая система InCites https://apps.webofknowledge.com
И.10	— научные журналы издательства Elsevier https://www.sciencedirect.com/
И.11	— доступ к полным версиям книг издательства Springer на английском языке https://link.springer.com/

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, ПКс 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.

Подготовку к каждому лабораторному занятию Вы должны начать с ознакомления с планом лабораторного 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, выполнение дополнительных индивидуальных заданий на лабораторных работах.

Самостоятельная работа на лекции выполняется в конце каждой лекции и заключается в решении небольшой задачи, поставленной преподавателем по материалу прочитанной лекции.

Работа с лекцией включает в себя дополнение конспекта сведениями из рекомендованной литературы (с указанием использованного источника).

Возможны выступления обучающихся на лекции по отдельным вопросам обсуждаемой темы (проработанные самостоятельно под руководством преподавателя); сообщения занимают 7...10 мин. Такие выступления помогают четко выражать свои мысли, аргументировано излагать и отстаивать свою точку зрения при ответе на вопросы. Самостоятельное изучение практического материала планируется из расчета 0,3 ч на 1 ч лекции.

Работа с материалом лекции, выполненная через один-два дня после ее прослушивания, позволяет выделить неясные моменты, которые необходимо либо самостоятельно разобрать, пользуясь рекомендованными литературными источниками, либо обсудить с преподавателем на ближайшей консультации.

Внеаудиторную самостоятельную работу. Перечень лабораторных работ, а также список учебных и методических пособий для этих работ вывешивается в лаборатории и студенты имеют возможность подготовиться к выполнению этих работ.

Используется лабораторный практикум "Разработка автоматизированных экспертных систем". Внеаудиторная самостоятельная работа по лабораторным занятиям включает подготовку к выполнению работ, обработку полученных результатов, защиту работ.

Подготовка заключается в ознакомлении с названием, целью работы, основными теоретическими положениями и методическими указаниями по ее выполнению.