

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по учебной работе

Дата подписания: 21.09.2023 10:08:55

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 текстовой информации**

Закреплена за подразделением

Кафедра АСУ

Направление подготовки

09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

Профиль

Квалификация

Бакалавр

Форма обучения

очная

Общая трудоемкость

3 ЗЕТ

Часов по учебному плану

108

Формы контроля в семестрах:

в том числе:

экзамен 7

аудиторные занятия

51

самостоятельная работа

21

часов на контроль

36

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	7 (4.1)		Итого	
	Неделя			
Неделя	18			
Вид занятий	УП	РП	УП	РП
Лекции	17	17	17	17
Лабораторные	34	34	34	34
Итого ауд.	51	51	51	51
Контактная работа	51	51	51	51
Сам. работа	21	21	21	21
Часы на контроль	36	36	36	36
Итого	108	108	108	108

Программу составил(и):

-, *ст.преп., Агабубаев Аслан Такабудинович*

Рабочая программа

Обработка текстовой информации

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 09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА (приказ от 02.04.2021 г. № 119 о.в.)

Составлена на основании учебного плана:

09.03.01 Информатика и вычислительная техника, 09.03.01-БИВТ-23.plx , утвержденного Ученым советом НИТУ МИСИС в составе соответствующей ОПОП ВО 22.06.2023, протокол № 5-23

Утверждена в составе ОПОП ВО:

09.03.01 Информатика и вычислительная техника, , 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 Ефимов Альберт Рувимович, к.филос.н.

1. ЦЕЛИ ОСВОЕНИЯ

1.1	Изучить методы и подходы из области искусственного интеллекта к анализу и синтезу речевых конструкций на естественном языке с формированием критериев оценивания возможности применения для решения практических задач в том числе после знакомства с современными сервисами и платформами интеллектуальной обработки естественного языка.
-----	--

2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Блок ОП:		Б1.В.ДВ.11
2.1	Требования к предварительной подготовке обучающегося:	
2.1.1	UX/UI - дизайн	
2.1.2	Автоматизация технологических процессов	
2.1.3	Архитектурирование	
2.1.4	Введение в IoT системы	
2.1.5	Математические модели социально-экономических систем	
2.1.6	Моделирование систем	
2.1.7	Мультиагентное моделирование систем	
2.1.8	Основы разработки цифровых платформ управления	
2.1.9	Современные инструментальные средства анализа данных	
2.1.10	Системная и программная инженерия	
2.1.11	Теория систем автоматического управления	
2.1.12	Введение в обработку больших данных	
2.1.13	Веб-аналитика	
2.1.14	Производственная практика	
2.1.15	Python для анализа данных	
2.1.16	Введение в прикладной ИИ	
2.1.17	Теория систем и системный анализ	
2.1.18	Основ теории информации	
2.1.19	Системы управления ресурсами предприятий	
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Аппаратные средства хранения и обработки данных	
2.2.2	Подготовка к процедуре защиты и защита выпускной квалификационной работы	

3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ

ПК-4: Способность разрабатывать математическую модель и проверить ее адекватность, провести анализ результатов моделирования, принять решение на основе полученных результатов	
Знать:	
ПК-4-31	Методы и подходы искусственного интеллекта применяемые для обработки естественного языка.
ПК-3: Готовность осуществлять и обосновывать выбор математического аппарата и программного обеспечения для решения поставленных задач; анализировать рынок программных и программно-технических средств, информационных продуктов и услуг для создания и модификации прикладных систем	
Знать:	
ПК-3-31	Современные направления развития Natural Language Processing.
ПК-4: Способность разрабатывать математическую модель и проверить ее адекватность, провести анализ результатов моделирования, принять решение на основе полученных результатов	
Уметь:	
ПК-4-У1	Оценивать задачи анализа и синтеза конструкций на естественном языке с целью применения методов и подходов ИИ для обработки естественного языка.
ПК-3: Готовность осуществлять и обосновывать выбор математического аппарата и программного обеспечения для решения поставленных задач; анализировать рынок программных и программно-технических средств, информационных продуктов и услуг для создания и модификации прикладных систем	
Уметь:	
ПК-3-У1	Определять области применения современных сервисов и библиотек ИИ для решения поставленной

практической задачи обработки естественного языка.

ПК-4: Способность разрабатывать математическую модель и проверить ее адекватность, провести анализ результатов моделирования, принять решение на основе полученных результатов

Владеть:

ПК-4-В1 навыки использования библиотек Python для задач с лингвистикой

ПК-3: Готовность осуществлять и обосновывать выбор математического аппарата и программного обеспечения для решения поставленных задач; анализировать рынок программных и программно-технических средств, информационных продуктов и услуг для создания и модификации прикладных систем

Владеть:

ПК-3-В1 навыками математического описания сущности решения задачи

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. Современные тенденции развития и практические разработки в NLP							
1.1	Анализ рынка современных разработок для анализа\синтеза речи на русском языке (групповая работа) /Ср/	7	9		Л1.1Л2.3 Э3 Э4		КМ3	Р1
1.2	Компании разрабатывающие в внедряющие решения связанные с интеллектуальной разработкой ЕЯ /Лек/	7	2		Л1.1Л2.4 Э1 Э2		КМ3	Р1
	Раздел 2. Введение в NLP							
2.1	Основные термины и определения NLP /Лек/	7	4		Л1.1Л2.5			
	Раздел 3. Математический аппарат как базис для методов NLP							
3.1	Архитектуры нейросетей и особенности обучения для NLP /Лек/	7	4		Л2.4Л1.1			
3.2	Обработка звука и вероятностные модели обработки ЕЯ /Лек/	7	3		Л2.3Л1.1			
3.3	Решения задачи поиска фрагментов текста по поисковым запросам на русском языке /Лаб/	7	12		Л2.5Л2.4			
3.4	Моделирование структуры рекурсивной нейронной сети. (Лабраб1) /Ср/	7	2		Л1.1Л2.3		КМ2	Р3
3.5	Разработка программного кода для реализации самоорганизующиеся карты Кохонена. (Лабраб2) /Ср/	7	2		Л2.3Л1.1		КМ1	Р4
3.6	Подготовка корпуса текста для обучение на свободном корпусе текстов. (Лабраб3) /Ср/	7	2		Л2.5Л2.4		КМ2	Р5
	Раздел 4. Методы и подходы в NLP							

4.1	Байесовские методы машинного обучения. Вероятностные модели /Лек/	7	4		Л2.3Л2.4			
4.2	Решение задачи построения объектной структуры с семантическими связями их формализованного текста /Лаб/	7	8		Л2.3Л2.5			
4.3	Сравнительный анализ работы Байесовских и нейронных рекурсивных сетей на эталонном корпусе текстов в задачи категоризации текста. (групповая работа 2) /Ср/	7	6		Л2.4Л2.3		KM1	P2
Раздел 5. Современные сервисы и платформы для ИИ обработки ЕЯ								
5.1	Возможности ML Space /Лаб/	7	8		Л2.5Л1.1		KM1	P1
5.2	Категоризация свободного текста на русском языке /Лаб/	7	6		Л1.1Л2.3			

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
KM1	Контрольная работа2. Нейронные сети в интеллектуальной обработке текстов на ЕЯ	ПК-4-31;ПК-4-В1;ПК-4-У1	<ol style="list-style-type: none"> 1. Перечислите архитектуры нейронных сетей. 2. Что такое искусственная нейронная сеть 3. Что такое персептрон 4. Покажите архитектуру многослойного искусственного персептрона 5. Покажите архитектуру сверточной нейронной сети 6. Покажите архитектуру рекурсивной нейронной сети 7. Покажите архитектуру самообучающейся нейронной сети
KM2	Контрольная работа1. Байесовские сети в интеллектуальной обработке на ЕЯ	ПК-3-31;ПК-3-У1;ПК-3-В1	<ol style="list-style-type: none"> 1. Докажите теорема Байеса 2. Вероятностная модель в теории Байеса 3. Покажите на примере применений теоремы Байеса 4. покажите архитектуру Байесовской сети 5. Формализуйте задачу классификации 6. Покажите сходимость байесовского метода классификации 7. Пересилите методы байесовской классификации доступные в модулях Python

КМ3	Тест. ИИ в обработки текстовой информации на ЕЯ	ПК-4-31;ПК-3-31	<ol style="list-style-type: none"> 1. Искусственный интеллект это 2. Кто создал основополагающие работы в области искусственного интеллекта -кибернетике 3. Назовите главное "мыслящее" устройство направления исследования в области искусственного интеллекта 4. Какие подходы к определению понятия «искусственный интеллект» существуют 5. Какие системы искусственного интеллекта (СИИ) входят в состав систем, основанных на языках 6. Какие системы искусственного интеллекта (СИИ) входят в состав систем эвристического поиска 7. Какие системы искусственного интеллекта (СИИ) входят в состав систем, основанных на языках 8. Какими характерными особенностями обладают системы искусственного интеллекта 9. Научное направление, связанное с попытками формализовать мышление человека называется ... 10. С каким объектом изучения тесно связаны термины "интеллект" и "информатика"
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	Групповая работа 1. Анализ рынка современных разработок для анализа и синтеза реси на ЕЯ	ПК-4-У1;ПК-4-В1	<ol style="list-style-type: none"> 1. Анализ современного рынка компаний, занимающихся разработкой ИС для обработки информации на русском языке 2. Формирование системы критериев оценки аналогичных разработок в области анализа информации на ЕЯ 3. Сравнительный анализ 3х произвольных разработок на эталонном корпусе тестовой информации на ЕЯ
P2	Групповая работа 2. Сравнительный анализ работы Байесовских и нейронных рекурсивных сетей на эталонном корпусе текстов в задачи категоризации текста.	ПК-3-У1;ПК-3-В1	<ol style="list-style-type: none"> 1. Анализ особенностей подходов для анализа информации на ЕЯ: на нейронных сетях и байесовских методах 2. Формирование системы критериев оценки аналогичных разработок в области анализа информации на ЕЯ основанных на нейро и Байесовских сетях 3. Сравнительный анализ разработок в области анализа информации на ЕЯ основанных на нейро и Байесовских сетях на эталонном корпусе текстовой информации на ЕЯ
P3	Лабораторная работа 1. Моделирование структуры рекурсивной нейронной сети.	ПК-4-У1;ПК-4-В1	<ol style="list-style-type: none"> 1. Исследование архитектур нейронных сетей 2. Анализ особенностей рекурсивной нейронной сети 3. Моделирование много слойной рекурсивной нейронной сети
P4	Лабораторная работа 2. Разработка программного кода для реализации самоорганизующиеся карты Кохонена	ПК-3-У1;ПК-3-В1	<ol style="list-style-type: none"> 1. Исследование архитектур самоорганизующийся карты Кохонена 2. Анализ особенностей самоорганизующийся карты Кохонена 3. Моделирование самоорганизующийся карты Кохонена
P5	Лабораторная работа 3. Подготовка корпуса текста для обучение на свободном корпусе текстов	ПК-4-У1;ПК-4-В1	<ol style="list-style-type: none"> 1. Постановка задачи классификации текстовой информации на русском языке 2. Сбор корпуса текстов на русском языке 3. Разметка корпуса текстов на русском языке 4. Формирование обучающей, тестовой и контрольной выборки из корпуса текстов на русском языке

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

1. Как называется область информационной технологии, изучающая методы превращения знаний в объект обработки на компьютере
2. В чем состоит главное назначение инженерии знаний ...
3. Как называются знания о конкретной ситуации в форме числовых, текстовых данных или простых утверждений ...
4. Как называются программы для ЭВМ, обладающие компетентностью, символическими рассуждениями, глубиной и самосознанием ...
5. Как называется искусственная система, имитирующая решение человеком сложных задач в процессе его жизнедеятельности ...
6. Укажите разрядность нейропроцессора
7. Укажите основные концепции развития СИИ
8. Что такое нечеткое множество
9. Состав базы знаний
10. Какие операции можно проводить с нечеткими знаниями
11. Укажите основные направления исследований СИИ
12. Безымянная функция для связывания формальных и фактических параметров на время
13. Сколько и какие точки зрения существуют на развитие СИИ
14. Укажите основные элементы нейропроцессора
15. Среди многих точек зрения, которые описывает научную область исследования СИИ перечислите доминирующие три...
16. Целью построения систем "Разработка интеллектуальных информационных систем или систем, основанных на знаниях" является...
17. Целью построения систем "Генерация и распознавание речи" является...
18. Целью построения систем "Разработка естественно-языковых интерфейсов и машинный перевод" является...
19. Целью построения систем "Обработка визуальной информации" является...
20. Целью построения систем "Обучение и самообучение" является...
21. Целью построения систем "Распознавание образов" является...
22. Целью построения систем "Игры и машинное творчество" является...

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

Шкала оценивания дисциплины:

70 -80 баллов - удовлетворительно

80-90 баллов - хорошо

90- 100 баллов - отлично

Максимальные баллы по выполняемым мероприятиям.

1. Групповые работы:

Первая - 10 баллов

Вторая - 10 баллов

2. Лабораторные работы:

Первая - 10 баллов

Вторая - 10 баллов

Третья - 10 баллов

3. Контрольная работа:

Первая - 15 баллов

Вторая - 15 баллов

4. Общее тестирование: итого 10 баллов

5. Итоговый контроль: 10 баллов

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ**6.1. Рекомендуемая литература****6.1.2. Дополнительная литература**

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л2.1	Петровичев Е. И.	Введение в искусственные нейронные сети: учеб. пособие по дисц. "Нейротехнологии в управлении"	Библиотека МИСиС	М.: Изд-во МГТУ, 2008
Л2.2	Ясницкий Л. Н.	Введение в искусственный интеллект: учеб. пособие для студ. вузов, обуч. по спец. 010100 "Математика"	Библиотека МИСиС	М.: ACADEMIA, 2005
Л2.3	Алпайдин Э.	Машинное обучение: новый искусственный интеллект: пер. с англ.	Библиотека МИСиС	М.: Альпина Паблишер, 2017

	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л2.4	Галушкин А. И.	Нейронные сети: основы теории: монография	Библиотека МИСиС	М.: Горячая линия - Телеком, 2010
Л2.5	Смолин Д. В.	Введение в искусственный интеллект: конспект лекций: курс лекций	Электронная библиотека	Москва: Физматлит, 2007

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Э1	В продуктах компании используются технологии искусственного интеллекта для решения бизнес-задач.	https://www.abbyy.com/blog/libor-transition-why-ai-matters/
Э2	«ЦРТ-инновации» — научно-исследовательская компания, разработчик голосовых и бимодальных биометрических систем	https://www.speechpro.ru/
Э3	Алиса от Яндекс	https://yandex.ru/alice
Э4	Яндекс. Переводчик	https://translate.yandex.ru/collections

6.3 Перечень программного обеспечения

П.1	ТЕХНОЛОГИИ УПРАВЛЕНИЯ ПЕРЕВОДОМ ООО
П.2	Microsoft Visio 2016
П.3	Microsoft Visual Studio 2015
П.4	Microsoft SQL server 2016
П.5	Microsoft Office
П.6	LMS Canvas
П.7	MS Teams
П.8	R Studio
П.9	Python
П.10	PgAdmin III
П.11	Statistica Neural Networks

6.4. Перечень информационных справочных систем и профессиональных баз данных

И.1	ML Space
-----	----------

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

Ауд.	Назначение	Оснащение
Б-904а	Компьютерный класс	20 стационарных компьютеров (core i5-3470 8gb RAM), пакет лицензионных программ MS Office, демонстрационное оборудование: доска, проектор мультимедийный, экран, колонки, комплект учебной мебели
Б-907	Учебная аудитория	1 стационарный компьютер, пакет лицензионных программ MS Office, комплект учебной мебели на 42 посадочных места, демонстрационное оборудование: доска, проектор мультимедийный x 2, экран x 2, колонки
Б-934	Лекционная аудитория	4 кабины для синхронного перевода, мультимедийные экраны и проектор, ноутбук, пакет лицензионных программ MS Office, комплект учебной мебели на 130 посадочных мест
Б-902	Учебная аудитория	12 стационарных компьютеров (2 x core i5-3470 8gb RAM, 10 x ryzen5 2400g 32gb RAM), пакет лицензионных программ MS Office, демонстрационное оборудование: доска, проектор мультимедийный, комплект учебной мебели на 19 мест

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ

В дисциплине практические занятия проводятся на сервисах машинного обучения сторонних разработчиков. Необходима предварительная регистрация индивидуального аккаунта каждого ученика. А также уделить время перед началом практических занятий освоения графического интерфейса соответствующего сервиса указанного преподавателем.

Все лабораторные работы проводятся в самостоятельно м режиме с выбранной и согласованной с преподавателем задачей и на указанных характеристиках текстовой информации с выбранным естественным языкам