

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Ястребов Олег Александрович

Должность: Ректор

Дата подписания: 29.05.2024 10:35:38

Уникальный программный ключ:

ca953a0120d891083f939673078ef1a989dae18a

Федеральное государственное автономное образовательное учреждение высшего образования

«Российский университет дружбы народов имени Патриса Лумумбы»

Факультет искусственного интеллекта

(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА И ВЕКТОРИЗАЦИЯ ТЕКСТА

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

02.03.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: РАЗРАБОТКА И ОБУЧЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

(наименование (профиль/специализация) ОП ВО)

2024 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Обработка естественного языка и векторизация текста» входит в программу бакалавриата «Искусственный интеллект: разработка и обучение интеллектуальных систем» по направлению 02.03.02 «Фундаментальная информатика и информационные технологии» и изучается в 6 семестре 3 курса. Дисциплину реализует NOT DEFINED. Дисциплина состоит из 7 разделов и 7 тем и направлена на изучение современных методов обработки естественного языка, основанными на глубоких нейронных сетях и машинном обучении.

Целью освоения дисциплины является получение теоретических знаний и практических навыков обработки естественно-языковых текстов; выявление сложностей, связанных с применением существующих методов обработки естественно-языковых текстов; использование полученных знаний по разработке, адаптации и использованию новейших средств информатики для обработки текстов на естественных языках.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Обработка естественного языка и векторизация текста» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
ПК-1	Способен создавать и оценивать различные модели машинного обучения и алгоритмы искусственного интеллекта с целью выбора наиболее эффективных решений для конкретных профессиональных задач	ПК-1.1 Умеет определять типы задач (классификация, регрессия, кластеризация и т.д.) и подходящие для них методы машинного обучения.; ПК-1.2 Владеет методами создания и обучения моделей с использованием различных алгоритмов и техник ML и AI; ПК-1.3 Применяет метрики для оценки производительности моделей, таких как точность, полнота, F1-мера и другие; ПК-1.4 Демонстрирует навыки обработки, представления и анализа данных для построения моделей машинного обучения;
ПК-2	Способен эффективно работать с большими объемами данных, включая их предварительную обработку, анализ и визуализацию, с целью извлечения полезной информации для обучения моделей искусственного интеллекта	ПК-2.1 Демонстрирует навыки анализа данных с использованием статистических методов и инструментов, таких как Python, R, SQL, NoSQL; ПК-2.2 Владеет методами работы с различными алгоритмами машинного обучения и глубинного обучения для решения различных задач;
ПК-3	Способен разрабатывать и оптимизировать алгоритмы обучения с подкреплением и системы автоматизированного принятия решений для создания интеллектуальных агентов и роботов	ПК-3.1 Демонстрирует умение разрабатывать и реализовывать алгоритмы обучения с подкреплением для обучения интеллектуальных агентов и роботов; ПК-3.2 Умеет интегрировать системы автоматизированного принятия решений в интеллектуальные агенты и роботы;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Обработка естественного языка и векторизация текста» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Обработка естественного языка и векторизация текста».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
ПК-1	Способен создавать и оценивать различные модели машинного обучения и алгоритмы искусственного интеллекта с целью выбора наиболее эффективных решений для конкретных профессиональных задач	Нейронные сети; Сбор и разметка данных для машинного обучения; Практикум по программированию; Методы машинного обучения; Глубинное обучение; <i>Типовые задачи и применение ИИ в лингвистике**;</i> <i>Типовые задачи и применение ИИ в финансах**;</i> <i>Практикум по программированию в лингвистике**;</i> <i>Практикум по программированию в финансах**;</i>	Преддипломная практика; Технологическая (проектно-технологическая) практика (производственная); Управление проектами; Практикум по программированию;
ПК-2	Способен эффективно работать с большими объемами данных, включая их предварительную обработку, анализ и визуализацию, с целью извлечения полезной информации для обучения моделей искусственного интеллекта	NoSQL базы данных; Нейронные сети; Искусственный интеллект и когнитивная психология; Программирование на языке Python; Сбор и разметка данных для машинного обучения; Инструменты обработки и визуализации данных; Онтология и графы знаний; Hadoop, SPARK; Обработка сигналов; <i>Введение в лингвистику**;</i> <i>Введение в финансы**;</i> <i>Типовые задачи и применение ИИ в лингвистике**;</i> <i>Типовые задачи и применение ИИ в финансах**;</i> <i>Практикум по программированию в лингвистике**;</i> <i>Практикум по программированию в финансах**;</i> Глубинное обучение; Технологическая (проектно-технологическая) практика (учебная);	Преддипломная практика; Технологическая (проектно-технологическая) практика (производственная);
ПК-3	Способен разрабатывать и оптимизировать алгоритмы обучения с	Нейронные сети; Инструменты обработки и визуализации данных;	Преддипломная практика; Технологическая (проектно-технологическая) практика

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	подкреплением и системы автоматизированного принятия решений для создания интеллектуальных агентов и роботов	Практикум по программированию; <i>Типовые задачи и применение ИИ в лингвистике**;</i> <i>Типовые задачи и применение ИИ в финансах**;</i>	(производственная); Практикум по программированию;

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Обработка естественного языка и векторизация текста» составляет «3» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			6
<i>Контактная работа, ак.ч.</i>	54		54
Лекции (ЛК)	18		18
Лабораторные работы (ЛР)	18		18
Практические/семинарские занятия (СЗ)	18		18
<i>Самостоятельная работа обучающихся, ак.ч.</i>	54		54
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	0		0
Общая трудоемкость дисциплины	ак.ч.	108	108
	зач.ед.	3	3

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
Раздел 1	Морфологический и синтаксический анализ.	1.1	Особенности морфологии в разных языках, задачи машинной морфологии: лемматизация, стемминг, частеречная разметка, генерация.	ЛК, ЛР, СЗ
Раздел 2	Языковые модели.	2.1	Модели на основе n-грамм, оценка вероятностей, оценка качества, перплексия.	ЛК, ЛР, СЗ
Раздел 3	Лексическая семантика.	3.1	Вычислительное значение отдельных слов в контексте.	ЛК, ЛР, СЗ
Раздел 4	Информационный и вопросно-ответный поиск.	4.1	Частотные свойства слов, закон Ципфа, векторная модель информационного поиска, веса $tf*idf$.	ЛК, ЛР, СЗ
Раздел 5	Автоматическое реферирование.	5.1	Задача, приложения, типы рефератов.	ЛК, ЛР, СЗ
Раздел 6	Анализ тональности.	6.1	Анализ тональности.	ЛК, ЛР, СЗ
Раздел 7	Извлечение информации и машинный перевод.	7.1	Извлечение информации и машинный перевод.	ЛК, ЛР, СЗ

* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	Учебная аудитория для проведения занятий лекционного типа, практических занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Комплект специализированной мебели; технические средства: Моноблок Lenovo Idea Centre, выход в интернет, Проектор EPSON, Интерактивный дисплей SMART MX SERIES SBID-MX065-v2. Программное обеспечение: продукты Microsoft (ОС, пакет

		офисных приложений, в т.ч. MS Office/Office 365, Teams)
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве 11 шт.), доской (экраном) и техническими средствами мультимедиа презентаций.	Компьютерный класс для проведения занятий лекционного типа, практических занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Комплект специализированной мебели; технические средства: Моноблок MSI (no) - 13 Моноблок Lenovo c560 – 3, Моноблок hp pro one – 1, Интерактивная система Smart Unifi45 – 1, Проектор Notevision – 1, Проекционный экран Cactus – 1, Компьютерная гарнитура -17, Портативная камера для документов – 1. Имеется выход в интернет. Программное обеспечение: продукты Microsoft (ОС, пакет офисных приложений, в т.ч. MS Office/Office 365, Teams,).
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций.	Компьютерный класс для проведения занятий лекционного типа, практических занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Комплект специализированной мебели; технические средства: Моноблок MSI (no) - 13 Моноблок Lenovo c560 – 3, Моноблок hp pro one – 1, Интерактивная система Smart Unifi45 – 1,

		<p>Проектор Notevision – 1, Проекционный экран Cactus – 1, Компьютерная гарнитура -17, Портативная камера для документов – 1. Имеется выход в интернет. Программное обеспечение: продукты Microsoft (ОС, пакет офисных приложений, в т.ч. MS Office/Office 365, Teams,).</p>
<p>Для самостоятельной работы</p>	<p>Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.</p>	<p>Компьютерный класс для самостоятельной работы обучающихся. Комплект специализированной мебели (11 посадочных мест); технические средства: Моноблок Lenovo Idea Centre, проектор ViewSonic p9d6253, имеется выход в интернет, Проекционный экран Cactus, гарнитуры Plantronics Audio 655 Dsp Программное обеспечение: продукты Microsoft (ОС, пакет офисных приложений, в т.ч. MS Office/Office 365, Teams)</p>

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Гольдберг Й. Нейросетевые методы в обработке естественного языка [Электронный ресурс] : руководство / Й. Гольдберг, 2019. — 282 с.
2. Лейн Хобсон Обработка естественного языка в действии [Электронный ресурс] / Хобсон Лейн, Ханнес Хапке, Коул Ховард, 2021. — 576 с..

Дополнительная литература:

1. Ганегедара Т. Обработка естественного языка с TensorFlow [Электронный ресурс] : руководство / Т. Ганегедара, 2020..
2. Моделирование языковой деятельности в интеллектуальных системах (Под ред. А. Е. Кибрика и А. С. Нариньяни).

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров
 - Электронно-библиотечная система РУДН – ЭБС РУДН
<http://lib.rudn.ru/MegaPro/Web>
 - ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>

- ЭБС Юрайт <http://www.biblio-online.ru>
- ЭБС «Консультант студента» www.studentlibrary.ru
- ЭБС «Троицкий мост»

2. Базы данных и поисковые системы

- электронный фонд правовой и нормативно-технической документации

<http://docs.cntd.ru/>

- поисковая система Яндекс <https://www.yandex.ru/>
- поисковая система Google <https://www.google.ru/>
- реферативная база данных SCOPUS

<http://www.elsevierscience.ru/products/scopus/>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

1. Курс лекций по дисциплине «Обработка естественного языка и векторизация текста».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

8. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ И БАЛЛЬНО-РЕЙТИНГОВАЯ СИСТЕМА ОЦЕНИВАНИЯ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ ПО ДИСЦИПЛИНЕ

Оценочные материалы и балльно-рейтинговая система* оценивания уровня сформированности компетенций (части компетенций) по итогам освоения дисциплины «Обработка естественного языка и векторизация текста» представлены в Приложении к настоящей Рабочей программе дисциплины.

* - ОМ и БРС формируются на основании требований соответствующего локального нормативного акта РУДН.

РАЗРАБОТЧИК:

_____	_____	_____
<i>Должность, БУП</i>	<i>Подпись</i>	Андреева Марина Андреевна <i>Фамилия И.О.</i>

РУКОВОДИТЕЛЬ БУП:

_____	_____	_____
Заведующий кафедрой <i>Должность БУП</i>	<i>Подпись</i>	<i>Фамилия И.О.</i>

РУКОВОДИТЕЛЬ ОП ВО:

_____	_____	_____
<i>Должность, БУП</i>	<i>Подпись</i>	Воробьева Александра Андреевна <i>Фамилия И.О.</i>