

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Ястребов Олег Александрович

Должность: Ректор

Дата подписания: 29.05.2024 11:28:12

Уникальный программный ключ:

ca953a01204891083f939673078ef1a989dae18a

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

ВВЕДЕНИЕ В ОБРАБОТКУ ЕСТЕСТВЕННОГО ЯЗЫКА

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

02.04.02 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):

УПРАВЛЕНИЕ ДАННЫМИ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

(наименование (профиль/специализация) ОП ВО)

2024 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Введение в обработку естественного языка» входит в программу магистратуры «Управление данными и искусственный интеллект» по направлению 02.04.02 «Фундаментальная информатика и информационные технологии» и изучается в 1 семестре 1 курса. Дисциплину реализует NOT DEFINED. Дисциплина состоит из 5 разделов и 7 тем и направлена на изучение

Целью освоения дисциплины является Целью освоения дисциплины «Введение в обработку естественного языка» является ознакомление слушателей с передовыми методами, моделями, средствами и технологиями компьютерной обработки текстов на естественных языках, дать умение представлять в алгоритмическом виде процессы анализа и синтеза текста. Рассматриваются задачи классификации текста, автоматической генерации текста с использованием рекуррентных нейронных сетей, включая LSTM и GRU, одномерных сверточных сетей, а также сетей с архитектурой Transformer.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Введение в обработку естественного языка» 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)
УК-1	Способен осуществлять поиск, критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.3 Имеет практический опыт работы с информационными источниками, опыт научного поиска, создания научных текстов.;
УК-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия	УК-5.3 Имеет практический опыт анализа философских и исторических фактов, опыт эстетической оценки явлений культуры;
ОПК-1	Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	ОПК-1.1 Обладает фундаментальными знаниями в области математических и естественных наук, теории коммуникаций; ОПК-1.2 Умеет осуществлять первичный сбор и анализ материала, интерпретировать различные математические объекты;
ОПК-2	Способен применять компьютерные / суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности	ОПК-2.2 Умеет анализировать типовые языки программирования, составлять программы; ОПК-2.3 Имеет практический опыт решения задач анализа, интеграции различных типов программного обеспечения, анализа типов коммуникации;
ПК-2	Способен к овладению методами математического моделирования объектов и процессов при разработке требований и проектировании программного обеспечения систем искусственного	ПК-2.1 Знает методы математического моделирования, используемые при разработке требований и проектировании программного обеспечения систем искусственного интеллекта;

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
	интеллекта	

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Введение в обработку естественного языка» 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 «Введение в обработку естественного языка».

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

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
УК-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия		Технологическая (проектно-технологическая) практика (учебная); Технологическая (проектно-технологическая) практика (производственная); История религий России; <i>Иностранный язык в профессиональной деятельности**</i> ; <i>Русский язык в профессиональной деятельности**</i> ; Машинный перевод;
УК-1	Способен осуществлять поиск, критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий		Алгоритмы и структуры данных; Методы к решению задач Information Extraction; Преддипломная практика;
ОПК-1	Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий		Обучение с подкреплением; Машинный перевод; Вопросно-ответные системы; Методы к решению задач Information Extraction; Алгоритмы и структуры данных;
ОПК-2	Способен применять компьютерные / суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной		Вопросно-ответные системы; Генеративные методы в NLP; Методы глубокого обучения для обработки естественного языка; Компьютерное зрение; Методы оптимизации; Практикум по

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	деятельности		программированию; Технологическая (проектно-технологическая) практика (производственная);
ПК-2	Способен к овладению методами математического моделирования объектов и процессов при разработке требований и проектировании программного обеспечения систем искусственного интеллекта		Технологическая (проектно-технологическая) практика (производственная); Преддипломная практика; Компьютерное зрение; Практикум по программированию; Обучение с подкреплением; Генеративный искусственный интеллект; <i>Искусственный интеллект в компьютерных играх**</i> ; <i>Искусственный интеллект в финансовых технологиях**</i> ; Генеративные методы в NLP;

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

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Введение в обработку естественного языка» составляет «4» 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)
			1
<i>Контактная работа, ак.ч.</i>	36		36
Лекции (ЛК)	18		18
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	18		18
<i>Самостоятельная работа обучающихся, ак.ч.</i>	81		81
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	27		27
Общая трудоемкость дисциплины	ак.ч.	144	144
	зач.ед.	4	4

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	Задачи обработки естественного языка	1.1	Синтаксический, морфологический, семантический и графематический анализ, омонимия, задачи лингвистического анализа	ЛК, СЗ
		1.2	Категоризация текстов, выделение тем, машинный перевод, ответы на вопросы, анализ тональности текста, автоматическое реферирование, использование машинного обучения в обработке естественного языка	ЛК, СЗ
Раздел 2	Предварительная обработка текста	2.1	Очистка текста, токенизация, стемминг, лемматизация, удаление стоп-слов, фильтрация наиболее частотных и наименее частотных слов	ЛК, СЗ
Раздел 3	Векторизация текста и машинное обучение для обработки текстов	3.1	Построение словаря, мешок слов, TF-IDF, word2vec, fasttext, LDA, LSI, GloVe	ЛК, СЗ
		3.2	Решение задач классификации и определения тональности методами классического машинного обучения на основе векторных моделей	ЛК, СЗ
Раздел 4	Нейронные сети в решении задач текстовой обработки	4.1	Архитектуры нейронных сетей для обработки текстов: рекуррентные (LSTM, GRU), одномерные сверточные. Применение нейронных сетей для обработки текстов	ЛК, СЗ
Раздел 5	Языковая модель	5.1	Языковая модель и дистрибутивная семантика. Обучение векторной модели. Задача генерации текста. Различные подходы к генерации текста	ЛК, 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.	
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.	
Для	Аудитория для самостоятельной работы	

самостоятельной работы	обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	
------------------------	---	--

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

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Хиценко В.П., Основы программирования : учебное пособие / Хиценко В.П. Новосибирск: Издательство Новосибирского государственного технического университета, 2015. Режим доступа:

<http://www.studentlibrary.ru/book/ISBN9785778227064.html>

2. Manning C.D., Schütze H. Foundations of Statistical Natural Language Processing. MIT Press, 1999.

3. Цитульский Антон Максимович, Иванников Александр Владимирович, Рогов Илья Сергеевич NLP - Обработка естественных языков // StudNet. 2020. №6.

<https://cyberleninka.ru/article/n/nlp-obrabotka-estestvennyh-yazykov>

Дополнительная литература:

1.

2.

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 «Введение в обработку естественного языка».

* - 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

«Введение в обработку естественного языка» представлены в Приложении к настоящей Рабочей программе дисциплины.

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

РАЗРАБОТЧИК:

_____	_____	Баум Валентина Владимировна
<i>Должность, БУП</i>	<i>Подпись</i>	<i>Фамилия И.О.</i>

РУКОВОДИТЕЛЬ БУП:

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

РУКОВОДИТЕЛЬ ОП ВО:

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