

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

"БРАТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ"

УТВЕРЖДАЮ

Проректор по образовательной деятельности

_____ А.М. Патрусова

_____ 13 мая _____ 2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.01.05 Обработка естественного языка

Закреплена за кафедрой **Информатики, математики и физики**

Учебный план b010302_25_ИИиЗИ.plx

Направление: 01.03.02 Прикладная математика и информатика

Квалификация **Бакалавр**

Форма обучения **очная**

Общая трудоемкость **3 ЗЕТ**

Виды контроля в семестрах:

Экзамен 8

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	8 (4.2)		Итого	
Неделя	11			
Вид занятий	УП	РП	УП	РП
Лекции	11	11	11	11
Лабораторные	33	33	33	33
В том числе инт.	12	12	12	12
В том числе в форме практ. подготовки	33	33	33	33
Итого ауд.	44	44	44	44
Контактная работа	44	44	44	44
Сам. работа	28	28	28	28
Часы на контроль	36	36	36	36
Итого	108	108	108	108

Программу составил(и):
к.т.н., доцент, К.Н. Фигура _____

Рабочая программа дисциплины

Обработка естественного языка

разработана в соответствии с ФГОС:

Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 01.03.02 Прикладная математика и информатика (приказ Минобрнауки России от 10.01.2018 г. № 9)

составлена на основании учебного плана:

Направление: 01.03.02 Прикладная математика и информатика
утвержденного приказом ректора от 31.01.2025 № 61.

Рабочая программа одобрена на заседании кафедры

Информатики, математики и физики

Протокол от 16.04.2025 г. № 11

Срок действия программы: 4 года

Зав. кафедрой Горохов Д.Б.

Председатель МКФ

старший преподаватель Латушкина С.В. 28.04.2025г. протокол № 08

Ответственный за реализацию ОПОП _____ Горохов Д.Б.

Директор библиотеки _____ Сотник Т.Ф.

№ регистрации _____ 42 _____

Визирование РПД для исполнения в учебном году

Председатель МКФ

_____ 20__ г.

Рабочая программа пересмотрена, обсуждена и одобрена для
исполнения в 20__ -20__ учебном году на заседании кафедры**Информатики, математики и физики**

Внесены изменения/дополнения (Приложение _____)

Протокол от _____ 20__ г. № _____
Зав. кафедрой _____

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	Дисциплина «Обработка естественного языка» изучает передовые методы, модели, средства и технологии компьютерной обработки текстов на естественных языках, представляя в алгоритмическом виде процессы анализа и синтеза текста.
-----	---

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП

Цикл (раздел) ООП:		Б1.В.01.05
2.1	Требования к предварительной подготовке обучающегося:	
2.1.1	Введение в машинное обучение	
2.1.2	Комплексное обеспечение безопасности объекта информатизации	
2.1.3	Базы данных	
2.1.4	Основы информационной безопасности	
2.2	Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Производственная (преддипломная) практика	
2.2.2	Выполнение и защита выпускной квалификационной работы	

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)**ПК-2: Способен разрабатывать прототипы информационных систем****ПК-2.2: Осуществляет программную реализацию и тестирование прототипов информационных систем**

Знать: архитектуры нейронных сетей для задач обработки естественного языка;

Уметь: применять современные инструментальные средства и системы программирования для разработки и обучения моделей искусственных нейронных сетей;

Владеть: практическим опытом создания и обучения нейросетевых моделей для решения задач в области обработки естественного языка
;**ПК-3: Способен создавать программный код информационных систем****ПК-3.1: Владеет навыками создания программного кода с применением современных сред разработки**

Знать: принципы построения систем обработки естественного языка, методы и технологии искусственного интеллекта для анализа естественного языка;

Уметь: применять на практике методы анализа данных в задачах анализа текстов для решения задач в области обработки естественного языка;

Владеть: инструментами поиска новых технологий и методов в области искусственного интеллекта;

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Вид занятия	Наименование разделов и тем	Семестр / Курс	Часов	Индикаторы	Литература	Инте ракт.	Примечание
	Раздел	Раздел 1. Введение в обработку естественного языка						
1.1	Лек	Введение в обработку естественного языка (NLP). Основные задачи NLP.	8	2	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	1	Проблемная лекция
1.2	Лек	Представления текстовых данных. Предобработка текста, лемматизация, стемминг.	8	2	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	1	Лекция-визуализация
1.3	Лаб	Введение в анализ естественного языка. Представления текстовых данных. Предобработка текстовых данных. Лемматизация. Морфологический анализ. Реализация собственного POS-тэггера.	8	5	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	

1.4	Ср	Подготовка к выполнению лабораторных работ	8	10	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	
1.5	Экзамен	Подготовка к экзамену	8	16	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	
	Раздел	Раздел 2.						
2.1	Лек	Методы машинного обучения для классификации текстовых документов на основе частотных мер (TF-IDF). Деревья решений, наивный байесовский классификатор, логистическая регрессия в задаче классификации текстов.	8	2	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	2	Лекция - визуализация
2.2	Лек	Языковые модели. Word embeddings. Нейросетевые модели языка: word2vec, fasttext. Мера семантической близости. Классификация текстов на основе нейросетевых моделей языка.	8	2	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	
2.3	Лек	Кластеризация текстовых документов. Тематическое моделирование Методы LSA, pLSA. Аддитивная регуляризация тематических моделей в BigARTM.	8	1	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	
2.4	Лек	Задачи обработки последовательностей: машинный перевод, автоматическое реферирование (summarization), вопросно-ответные системы. Механизм внимания (attention). Архитектуры encoder-decoder-attention.	8	2	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	2	Лекция - визуализация
2.5	Лаб	Классификация текстов методами машинного обучения. на основе частотных мер (TF-IDF). Деревья решений, наивный байесовский классификатор, логистическая регрессия в задаче классификации текстов.	8	4	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	
2.6	Лаб	Нейросетевые модели языка: word2vec, fasttext. Задача определения семантической близости. Классификация текстов на основе нейросетевых моделей языка.	8	4	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	
2.7	Лаб	Задача кластеризации текстовой коллекции с применением методов pLSA и фреймворка BigARTM.	8	4	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	

2.8	Лаб	Классификация текстов с помощью различных архитектур глубоких нейронных сетей: CNN, LSTM.	8	6	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	
2.9	Лаб	Реализация вопросно-ответной системы на основе нейронных сетей encoderdecoder с механизмом внимания. Решение задачи автоматического реферирования (text summarization).	8	6	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	4	(case-study (анализ конкретных ситуаций))
2.10	Лаб	Архитектуры трансформеров: BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерации текстов.	8	4	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	2	(case-study (ситуационный анализ))
2.11	Ср	Подготовка к выполнению лабораторных работ	8	18	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	
2.12	Экзамен	Подготовка к экзамену	8	20	ПК-3.1 ПК-2.2	Л1.1 Л1.2Л2.1 Л2.2 Э1	0	

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Образовательные технологии с использованием активных методов обучения (проблемная лекция)

Образовательные технологии с использованием активных методов обучения (лекция-визуализация)

Образовательные технологии с использованием интерактивных методов обучения (case-study (анализ конкретных ситуаций))

Образовательные технологии с использованием интерактивных методов обучения (case-study (ситуационный анализ))

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

6.1. Текущий контроль

Текущим контролем успеваемости обучающихся является межсессионная аттестация – единовременное подведение итогов текущей успеваемости не менее одного раза в семестр по всем дисциплинам/практикам.

Порядок проведения, содержание и особенности текущего контроля успеваемости представлены в разработанном Фонде оценочных средств для данной дисциплины.

6.2. Темы письменных работ

Не предусмотрено учебным планом.

6.3. Промежуточная аттестация

Промежуточная аттестация проводится в форме экзамена.

Порядок проведения, содержание и критерии оценивания итоговой промежуточной аттестации представлены в Фонде оценочных средств для данной дисциплины.

6.4. Перечень видов оценочных средств

Отчеты по лабораторным работам. Экзаменационные вопросы.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1. Рекомендуемая литература

7.1.1. Основная литература

Авторы,	Заглавие	Издательство,	Кол-во	Эл. адрес
---------	----------	---------------	--------	-----------

	Авторы,	Заглавие	Издательство,	Кол-во	Эл. адрес
Л1.1	Баланов А. Н.	Машинное обучение и искусственный интеллект: учебное пособие для вузов	Санкт-Петербург: Лань, 2024	1	https://e.lanbook.com/book/414920
Л1.2	Баланов А. Н.	Машинное обучение и искусственный интеллект: учебное пособие для вузов	Санкт-Петербург: Лань, 2025	1	https://e.lanbook.com/book/462248
7.1.2. Дополнительная литература					
	Авторы,	Заглавие	Издательство,	Кол-во	Эл. адрес
Л2.1	Митяков Е. С., Шмелева А. Г., Ладынин А. И.	Искусственный интеллект и машинное обучение: учебное пособие для вузов	Санкт-Петербург: Лань, 2025	1	https://e.lanbook.com/book/450827
Л2.2	Бессмертный И. А.	Искусственный интеллект. Введение в многоагентные системы: учебник для вузов	Москва: Юрайт, 2025	1	https://urait.ru/bcode/569279
7.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"					
Э1	Платформа обработки данных на основе искусственного интеллекта ML Space: российская платформа для ML-разработки полного цикла		https://aiplatform.ru		
7.3.1 Перечень программного обеспечения					
7.3.1.1	Python IDLE				
7.3.1.2	ОС Linux				
7.3.1.3	Anaconda				
7.3.1.4	Kaspersky Endpoint Security для бизнеса - Расширенный Russian Edition. 1000-1499 Node 1 year Educational Renewal License				
7.3.2 Перечень информационных справочных систем					
7.3.2.1	Справочно-правовая система «Консультант Плюс»				
7.3.2.2	«Университетская библиотека online»				
7.3.2.3	Электронный каталог библиотеки БрГУ				
7.3.2.4	Электронная библиотека БрГУ				
7.3.2.5	Университетская информационная система РОССИЯ (УИС РОССИЯ)				
8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)					
Аудитория	Назначение	Оснащение аудитории			Вид занятия
1348	Учебная аудитория (дисплейный класс)	<p>Основное оборудование:</p> <ul style="list-style-type: none"> - 11 Персональных компьютеров i5-13500/DDR5 16 GB/SSD 1TB/GeForce RTX4060 (монитор Asus VA24EHF); - компьютер Fordel Pro Intel i7-12700, DDR5 16 GB, SSD 1 TB, ATX 800 W, монитор MSI Pro MP243X, Model: 3PB5, 23,8", FHD@100Hz; - принтер HP LaserJet 1150; <p>Дополнительно:</p> <ul style="list-style-type: none"> - маркерная доска – 1 шт. <p>Учебная мебель:</p> <ul style="list-style-type: none"> - комплект мебели (посадочных мест/АРМ) – 24/12 шт.; - комплект мебели (посадочных мест/АРМ) для преподавателя – 1 шт.; 			Лек
1348	Учебная аудитория (дисплейный класс)	<p>Основное оборудование:</p> <ul style="list-style-type: none"> - 11 Персональных компьютеров i5-13500/DDR5 16 GB/SSD 1TB/GeForce RTX4060 (монитор Asus VA24EHF); - компьютер Fordel Pro Intel i7-12700, DDR5 16 GB, SSD 1 TB, ATX 800 W, монитор MSI Pro MP243X, Model: 3PB5, 23,8", FHD@100Hz; - принтер HP LaserJet 1150; <p>Дополнительно:</p> <ul style="list-style-type: none"> - маркерная доска – 1 шт. <p>Учебная мебель:</p> <ul style="list-style-type: none"> - комплект мебели (посадочных мест/АРМ) – 24/12 шт.; - комплект мебели (посадочных мест/АРМ) для преподавателя – 1 шт.; 			Лаб

1348	Учебная аудитория (дисплейный класс)	Основное оборудование: - 11 Персональных компьютеров i5-13500/DDR5 16 GB/SSD 1TB/GeForce RTX4060 (монитор Asus VA24EHF); - компьютер Fordel Pro Intel i7-12700, DDR5 16 GB, SSD 1 TB, ATX 800 W, монитор MSI Pro MP243X, Model: 3PB5, 23,8", FHD@100Hz; - принтер HP LaserJet 1150; Дополнительно: - маркерная доска – 1 шт. Учебная мебель: - комплект мебели (посадочных мест/АРМ) – 24/12 шт.; - комплект мебели (посадочных мест/АРМ) для преподавателя – 1 шт.;	Экзамен
2201	читальный зал №1	Комплект мебели (посадочных мест) Стеллажи Комплект мебели (посадочных мест) для библиотекаря Выставочные шкафы ПК i5-2500/H67/4Gb (монитор TFT19 Samsung) (10шт.); принтер HP Laser Jet P2055D (1шт.)	Ср

9. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Организация самостоятельной работы обучающихся зависит от вида учебных занятий:

Лекции.

В процессе формирования конспекта лекций, обучающийся должен кратко, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометить важные мысли, выделять ключевые слова, термины. Самостоятельно осуществлять проверку терминов с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удается разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации, лабораторном или практическом занятии.

Лабораторные работы.

При подготовке к лабораторным работам обучающийся должен осуществлять работу с конспектом лекций (обобщение, систематизация, углубление и конкретизация полученных теоретических знаний), разработать план проведения работ и быть готовым к его реализации на практике.

Самостоятельная работа обучающихся.

Проработка основной и дополнительной литературы, терминов, сведений, требующихся для запоминания и являющихся основополагающими в теме/разделе. Конспектирование прочитанных литературных источников. Проработка материалов по изучаемому вопросу, с использованием рекомендуемых ресурсов информационно-телекоммуникационной сети «Интернет». Выполнение заданий преподавателя, необходимых для подготовки к участию в интерактивной, активной, инновационных формах обучения по изучаемой теме.

Подготовка к экзамену.

При подготовке к экзамену необходимо ориентироваться на конспекты лекций, рекомендуемую литературу, использовать рекомендуемые ресурсы информационно-телекоммуникационной сети «Интернет».