МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ
Декан факультета информационных систем и технологий
_____ К.В. Святов
«11» _____10_____ 2021 г.

РАБОЧАЯ ПРОГРАММА

Дисциплина (модуль) Методы интеллектуального анализа естественного языка
наименование дисциплины (модуля)

Уровень образования высшее образование - магистратура
(СПО/бакалавриат/магистратура/специалитет/подготовка кадров высшей квалификации)

Квалификация магистр
Техник/Бакалавр/Магистр/Инженер/ Исследователь. Преподаватель-исследователь

г. Ульяновск, 2021

Рабочая программа составлена

на кафедре	Информационные системы
факультета	информационных систем и технологий
в соответствии с учебным планом по направлению подготовки (специальности)	09.04.04 Программная инженерия
профиль (программа / специализация)	Искусственный интеллект и предиктивная аналитика

Составитель рабочей программы

доцент, доцент, к.т.н. *(должность, ученое звание, степень)*	*(подпись)*	Романов А.А. *(Фамилия И. О.)*
ст. преподаватель *(должность, ученое звание, степень)*	*(подпись)*	Алюнов Д.Ю. *(Фамилия И. О.)*

Рабочая программа рассмотрена на заседании кафедры

Заведующий кафедрой *(должность)*	*(подпись)*	Романов А.А. *(Фамилия И. О.)*

СОГЛАСОВАНО:

Руководитель ОПОП «11» октября 2021 г.	*(подпись)*	Романов А.А. *(Фамилия И. О.)*
Заведующий выпускающей кафедрой /научный руководитель ОПОП «11» октября 2021г.	*(подпись)*	Романов А.А *(Фамилия И. О.)*
Директор библиотеки «11» октября 2011 г.	*(подпись)*	Синдюкова Е.С. *(Фамилия И. О.)*

1 ОБЪЕМ ДИСЦИПЛИНЫ (МОДУЛЯ) В ЗАЧЕТНЫХ ЕДИНИЦАХ С УКАЗАНИЕМ АКАДЕМИЧЕСКИХ ЧАСОВ, ВЫДЕЛЕННЫХ НА КОНТАКТНУЮ РАБОТУ ОБУЧАЮЩИХСЯ С ПРЕПОДАВАТЕЛЕМ (ПО ВИДАМ УЧЕБНЫХ ЗАНЯТИЙ) И НА САМОСТОЯТЕЛЬНУЮ РАБОТУ ОБУЧАЮЩИХСЯ

Таблица 1

Бюджет времени с учетом формы обучения, семестра и видов занятий

Форма обучения	Очная				Очно-заочная				Заочная			
Семестр	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Контактная работа обучающихся с преподавателем (по видам учебных занятий), **всего часов**		**48**										
в том числе:												
- занятия лекционного типа (лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации педагогическими работниками), часов		**16**										
- занятия семинарского/ практического типа (семинары, практические занятия, практикумы, коллоквиумы и иные аналогичные занятия), часов												
- лабораторные занятия (включая работу обучающихся на реальных или виртуальных объектах профессиональной сферы), часов		**32**										
Самостоятельная работа обучающихся, часов		**87**										
в том числе:												
- групповые и индивидуальные консультации обучающихся с преподавателями												
- проработка теоретического курса		**40**										
- курсовая работа (проект)												
- расчетно-графическая работа												
- реферат												
- эссе												
- подготовка к занятиям семинарского/практического типа												
- подготовка к выполнению и защите лабораторных работ		**47**										
- взаимодействие в электронной информационно-образовательной среде вуза												
Промежуточная аттестация обучающихся, включая подготовку (Экзамен, Зачет, Зачет с оценкой, КП, КР)		**9**										
Итого, часов		**144**										
Трудоемкость, з.е.		**4**										

2 ЯЗЫК ПРЕПОДАВАНИЯ

Изучение дисциплины (модуля) осуществляется на русском языке.

3 ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ (МОДУЛЯ)

Целью освоения дисциплины (модуля) «Методы интеллектуального анализа естественного языка» является освоение студентами знаний в области обработки и анализа текстовой информации, математических, статистических методов, базовых методов глубокого обучения нейронных сетей для обработки текстовой информации. а также изучение основных проблем компьютерной обработки текстов и современных подходов к их решению.

Задачами освоения дисциплины (модуля) являются формирование у обучающихся:

- Формирование базовых знаний в области компьютерной обработки текстовой информации как дисциплины, обеспечивающей технологические основы современных инновационных сфер деятельности.
- Обучение студентов принципам решения задач обработки естественного языка на основе методов машинного обучения.
- Формирование подходов к выполнению студентами исследований в области обработки естественного языка.

В результате изучения дисциплины (модуля) «Методы интеллектуального анализа естественного языка» обучающиеся на основе приобретенных знаний, умений и навыков достигают освоения компетенций на определенном уровне.

Аннотация дисциплины (модуля) представлена в Приложении А.

4 ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ), СООТНЕСЕННЫХ С ИНДИКАТОРАМИ ДОСТИЖЕНИЯ КОМПЕТЕНЦИЙ

Таблица 2

Планируемые результаты обучения по дисциплине (модулю), с указанием индикатора достижения компетенций

Код компетенции	**Формулировка компетенции**	**Код индикатора достижения компетенции (по данной дисциплине (модулю))**	**Индикаторы достижения компетенции (связанные с данной дисциплиной (модулем))**
Общепрофессиональные			
ОПК-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	ИД-1 $_{ОПК-2}$	Знать: современные информационно-коммуникационные и интеллектуальные технологии, инструментальные среды, программно-технические платформы для решения профессиональных задач
		ИД-2 $_{ОПК-2}$	Уметь: обосновывать выбор современных информационно-коммуникационных и интеллектуальных технологий, разрабатывать оригинальные программные средства для решения профессиональных задач

		ИД-3 $_{ОПК-2}$	Владеть: навыками разработки оригинальных программных средств, в том числе с использованием современных информационно-коммуникационных и интеллектуальных технологий, для решения профессиональных задач
ОПК-4	Способен применять на практике новые научные принципы и методы исследований;	ИД-1 $_{ОПК-4}$	Знать: новые научные принципы и методы исследований
		ИД-2 $_{ОПК-4}$	Умеет: применять на практике новые научные принципы и методы исследований
		ИД-3 $_{ОПК-4}$	Владеть: навыками применения новых научных принципов и методов исследования для решения профессиональных задач

5 МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина (модуль) относится к обязательной части блока Б 1 образовательной программы.

6 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ), СТРУКТУРИРОВАННОЕ ПО ТЕМАМ (РАЗДЕЛАМ) С УКАЗАНИЕМ ОТВЕДЕННОГО НА НИХ КОЛИЧЕСТВА АКАДЕМИЧЕСКИХ ЧАСОВ И ВИДОВ УЧЕБНЫХ ЗАНЯТИЙ

6.1 Тематический план изучения дисциплины (модуля)

Таблица 3

Тематический план с указанием выделенных академических часов на освоение каждого из разделов и проведение промежуточной аттестации

№	Наименование разделов (включая промежуточную аттестацию)	Очная (час)					Очно-заочная (час)					Заочная (час)				
		Лекции	Практические (сем.) занятия	Лабораторные работы	Самостоятельная работа	Всего	Лекции	Практические (сем.) занятия	Лабораторные работы	Самостоятельная работа	Всего	Лекции	Практические (сем.) занятия	Лабораторные работы	Самостоятельная работа	Всего
1	Раздел 1. Введение в компьютерную обработку текстов на естественном языке.	4		4	17	25										

2	Раздел 2. Основы теории нейронных сетей и алгоритмов машинного обучения.	4		12	30	46										
3	Раздел 3. Методы машинного обучения для обработки текста на естественном языке.	8		16	40	64										
4	Подготовка к промежуточной аттестации, консультации перед промежуточной аттестацией и сдача промежуточной аттестации				9	9										
	Итого часов	**16**		**32**	**96**	**144**										

6.2 Теоретический курс

Таблица 4

Основные вопросы, освещаемые на лекциях

Раздел, тема учебной дисциплины (модуля), содержание темы
Раздел 1. Введение в компьютерную обработку текстовой информации.
1.1 Задачи обработки текстов. Понятие лингвистического процессора. Основные термины и определения. 1.2. Общие принципы построения систем автоматизированной обработки текстов. 1.3 Основные подходы к решению проблемы обработки и предобработки текстов, а также к построению генеративных систем. Обзор методов. 1.4 Парсинг текстов. Сбор базы. Базовая обработка текстовых данных. Регулярные выражения.
Раздел 2. Основы теории нейронных сетей и алгоритмов машинного обучения.
2.1 Понятие искусственного нейрона. Веса и связи. Метод обратного распространения ошибок. Функция Активации. 2.2 Принципы проектирования многослойной сети. Входной и выходной слой. Нормировка и предобработка данных данных. Скрытые слои нейросети - принципы их применения. Обучение нейросети. Обучающее и тестовое множество. Реализация нейросети с помощью библиотеки keras. 2.3 Метрики качества работы нейронной сети. Обучающая, валидационная и тестовая выборки. 2.4 Полносвязные нейронные сети. Bias и весовые коэффициенты. Слои Dropout, Batchnormalization. Схема обучения нейронной сети. 2.5. Сверточные нейронные сети. Одномерные сверточные нейронные сети. 2.6. Рекуррентные нейронные сети. LSTM нейронные сети.
Раздел 3. Методы машинного обучения для обработки текста на естественном языке.
3.1. Базовые подходы к преобразованию и обработке текстов. Численное представление текстов. Токенизация. Bag of words. 3.2. Семантический спектр текста. Косинусное сходство. 3.3. Задачи классификации текстов с применением полносвязных нейронных сетей. 3.4. Embedding представление слов. Обучение Embedding. Испльзование предобученных эмбеддингов. Gensim. 3.5. Использование рекуррентных нейронных сетей в задачах классификации текстов. 3.6. Задачи сегментации текстов. Морфологический анализ. Word2Wec эмбеддинги. Работа со словарями и корпусами. Разметка текстов для сегментации. 3.7. Подходы к кластеризации текстов. Методы k-means, «ближайшего соседа».

6.3 Практические (семинарские) занятия

Практические (семинарские) занятия учебным планом 09.04.04 «Программная инженерия» магистерская программа «Искусственный интеллект и предиктивная аналитика» не предусмотрены.

6.4 Лабораторный практикум

Таблица 6

Тематика лабораторных работ

Номер	Наименование лабораторной работы
1	Парсинг текстов. Базовая обработка текстов на Python. Регулярные выражения.
2	BOW представление текстов на естественном языке. Семантический спектр. Косинусное сходство.
3	Введение в нейронные сети. Обучение нейронной сети. Переобучение. Метрики качества
4	Полносвязные нейронные сети.
5	Классификация текстов с использованием полносвязных нейронных сетей
6	Классификация текстов с использованием эмбеддингов, рекурентных и сверточных нейронных сетей
7	Сегментация текста
8	Кластеризация текстов

6.5 Курсовой проект (работа), реферат, расчетно-графические работы

Учебным планом направления 09.04.04 «Программная инженерия» магистерская программа «Искусственный интеллект и предиктивная аналитика» курсовой проект (работа), реферат, расчетно-графические работы не предусмотрены.

6.6 Самостоятельная работа обучающихся

Виды самостоятельной работы распределяются в течение семестра. Подготовка к промежуточной аттестации ведется в установленные календарным учебным графиком сроки.

7 ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ (ОЦЕНОЧНЫХ МАТЕРИАЛОВ) ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Таблица 7

Наименование оценочных средств (оценочных материалов)

№ п/п	Код формируемой компетенции	Код индикатора достижения формируемой компетенции	Наименование оценочного средства
1.	ОПК-2	ИД-1 $_{ОПК-2}$	Выполнение лабораторных работ и собеседование по результатам их выполнения, зачет
		ИД-2 $_{ОПК-2}$	Выполнение лабораторных работ и собеседование по результатам их выполнения, зачет
		ИД-3 $_{ОПК-2}$	Выполнение лабораторных работ и собеседование по результатам их выполнения, зачет

2.	ОПК-4	ИД-1 $_{ОПК-4}$	Выполнение лабораторных работ и собеседование по результатам их выполнения, зачет
		ИД-2 $_{ОПК-4}$	Выполнение лабораторных работ и собеседование по результатам их выполнения, зачет
		ИД-3 $_{ОПК-4}$	Выполнение лабораторных работ и собеседование по результатам их выполнения, зачет

8 ПЕРЕЧЕНЬ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1. Горожанина Е.И. Нейронные сети : учебное пособие / Горожанина Е.И.. — Самара : Поволжский государственный университет телекоммуникаций и информатики, 2017. — 84 c. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: https://www.iprbookshop.ru/75391.html (дата обращения: 08.10.2021). — Режим доступа: для авторизир. Пользователей
2. Павлова А.И. Информационные технологии: основные положения теории искусственных нейронных сетей : учебное пособие / Павлова А.И.. — Новосибирск : Новосибирский государственный университет экономики и управления «НИНХ», 2017. — 191 c. — ISBN 978-5-7014-0801-0. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: https://www.iprbookshop.ru/87110.html (дата обращения: 08.10.2021). — Режим доступа: для авторизир. пользователей. - DOI: https://doi.org/10.23682/87110
3. Гольдберг, Й. Нейросетевые методы в обработке естественного языка : руководство / Й. Гольдберг ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 282 с. — ISBN 978-5-97060-754-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/131704 (дата обращения: 08.10.2021). — Режим доступа: для авториз. пользователей.
4. Ганегедара, Т. Обработка естественного языка с TensorFlow : руководство / Т. Ганегедара ; перевод с английского В. С. Яценкова. — Москва : ДМК Пресс, 2020. — 382 с. — ISBN 978-5-97060-756-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/140584 (дата обращения: 08.10.2021). — Режим доступа: для авториз. пользователей.
5. Батура Т.В. Математическая лингвистика и автоматическая обработка текстов на естественном языке : учебное пособие / Батура Т.В.. — Новосибирск : Новосибирский государственный университет, 2016. — 166 c. — ISBN 978-5-4437-0548-4. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: https://www.iprbookshop.ru/93489.html (дата обращения: 08.10.2021). — Режим доступа: для авторизир. Пользователей
6. Daniel Jurafsky and James H. Martin. 2008. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Second Edition. Prentice Hall.
7. Christopher D. Manning and Hinrich Schütze. 1999. Foundations of Statistical Natural Language Processing. MIT Press.
8. Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python. O'Reilly Media, 2009 (http://www.nltk.org/book)

9 ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ КОНТАКТНОЙ И САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

1. Официальная документация по библиотеке машинного обучения Keras https://keras.io/
2. Официальная документация библиотеки машинного обучения TensorFlow https://www.tensorflow.org/
3. Бессмертный, И. А. Системы искусственного интеллекта : учебное пособие для вузов / И. А. Бессмертный. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2021. — 157 с. — (Высшее образование). — ISBN 978-5-534-07467-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/470638 (дата обращения: 08.10.2021).
4. Вакуленко С.А. Нейронные сети : учебное пособие / Вакуленко С.А., Жихарева А.А.. — Санкт-Петербург : Санкт-Петербургский государственный университет промышленных технологий и дизайна, 2019. — 110 с. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: https://www.iprbookshop.ru/102447.html (дата обращения: 08.10.2021). — Режим доступа: для авторизир. пользователей. - DOI: https://doi.org/10.23682/102447
5. Станкевич, Л. А. Интеллектуальные системы и технологии : учебник и практикум для среднего профессионального образования / Л. А. Станкевич. — Москва : Издательство Юрайт, 2021. — 397 с. — (Профессиональное образование). — ISBN 978-5-534-11659-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/476356 (дата обращения: 08.10.2021).

10. ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ РЕСУРСОВ

10.1 Справочные системы и современные профессиональные базы данных, к которым обучающимся обеспечен доступ (удаленный доступ), в том числе в случае применения электронного обучения, дистанционных образовательных технологий

1. Справочная система Гарант
2. База ГОСТы и СанПиНы https://standartgost.ru/
3. База СНИПы. Нормативно-техническая документация http://snipov.net/
4. Федеральный портал Единое окно доступа к образовательным ресурсам http://window.edu.ru/library
5. Научная электронная библиотека http://elibrary.ru/defaultx.asp
6. РГБ фонд диссертаций http://diss.rsl.ru/
7. Энциклопедия http://encyclopaedia.biga.ru

10.2 Ресурсы информационно-телекоммуникационной сети «Интернет», необходимые для освоения дисциплины (модуля)

1. Электронно-библиотечная система издательства «Лань» http://e.lanbook.com/
2. Электронно-библиотечная система «Юрайт» https://urait.ru/
3. Электронно-библиотечная система издательства «IPR BOOKS» https://www.iprbookshop.ru/
4. Электронная библиотека полнотекстовых учебных и научных изданий УлГТУ http://venec.ulstu.ru/lib/faculty.php
5. Математический образовательный сайт http://old.exponenta.ru/default.asp

11 ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ И ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, ИСПОЛЬЗУЕМЫХ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ (МОДУЛЮ), ВКЛЮЧАЯ ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ

Таблица 8

Наименование и оснащенность помещений, используемых при осуществлении образовательного процесса по дисциплине (модулю)

№ п/п	Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения (подлежит ежегодному обновлению)
1	Учебные аудитории для проведения лекций	Учебная мебель: столы, стулья для обучающихся; стол, стул для преподавателя, доска. Аудитория, оснащенная комплексом технических средств обучения (проектор, экран, телевизор, компьютер.	Microsoft Windows; Adobe Reader; Microsoft Office (LibreOffice); Google Chrome;
2	Учебные аудитории для проведения лабораторных работ, практических работ, групповых и индивидуальных консультаций	Аудитория, оснащенная комплексом технических средств обучения (проектор, экран, телевизор, компьютер. Рабочие места, оборудованные компьютерами(ноутбуками) с выходом в интернет (Wi-Fi)	Проприетарные лицензии: Microsoft Windows; Microsoft Office, Microsoft Visual Studio Свободные и открытые лицензии: LibreOffice или OpenOffice, Mozilla Firefox, Adobe Reader, Архиватор 7-zip
3	Учебные аудитории для текущего контроля и промежуточной аттестации	Учебная мебель: столы, стулья для обучающихся; стол, стул для преподавателя, доска. Аудитория, оснащенная комплексом технических средств обучения (проектор, экран, телевизор, компьютер) (при наличии).	Microsoft Windows; Adobe Reader; Microsoft Office (LibreOffice);
4	Помещения для самостоятельной работы (читальный зал научной библиотеки)	Рабочие места, оборудованные компьютерами(ноутбуками) с выходом в интернет (Wi-Fi)	Проприетарные лицензии: Microsoft Windows; Microsoft Office, Microsoft Visual Studio Свободные и открытые лицензии: LibreOffice или OpenOffice, Mozilla Firefox, Adobe Reader, Архиватор 7-zip

Приложение А

Аннотация рабочей программы

Дисциплина (модуль)	Методы интеллектуального анализа естественного языка
Уровень образования	Магистратура
Квалификация	Магистр
Направление подготовки / специальность	09.04.04 / Программная инженерия
Профиль / программа / специализация	Искусственный интеллект и предиктивная аналитика
Дисциплина (модуль) нацелена на формирование компетенций	ОПК-2, ОПК-4
Цель освоения дисциплины (модуля)	Целью освоения дисциплины (модуля) «Методы интеллектуального анализа естественного языка» является освоение студентами знаний в области обработки и анализа текстовой информации, математических, статистических методов, базовых методов глубокого обучения нейронных сетей для обработки текстовой информации. а также изучение основных проблем компьютерной обработки текстов и современных подходов к их решению.
Перечень разделов дисциплины	1. Введение в компьютерную обработку текстов на естественном языке. 2. Основы теории нейронных сетей и алгоритмов машинного обучения. 3. Методы машинного обучения для обработки текста на естественном языке.
Общая трудоемкость дисциплины (модуля)	4 з.е.
Форма промежуточной аттестации	Зачет (2 семестр)

Лист дополнений и изменений

к рабочей программе дисциплины (модуля)
«Методы интеллектуального анализа естественного языка»

Учебный год: 20__/20__

Протокол заседания кафедры № ___ от «___» _____ 20___ г.

Принимаемые изменения:

_____
_____
_____
_____
_____
_____

Руководитель ОПОП _____ _____
личная подпись И.О. Фамилия

«____»_____20__ г.