

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Клочков Юрий Сергеевич
Должность: и.о. ректора
Дата подписания: 26.05.2024 16:44:48
Уникальный программный ключ: 4e7c4ea90328ec8e65c5d8058549a2538d7400d1

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ:

«_____» _____ 202__г.

РАБОЧАЯ ПРОГРАММА

дисциплины:	Автоматическая обработка естественного языка
направление подготовки:	01.04.02 Прикладная математика и информатика
направленность (профиль):	Машинное обучение и анализ данных
форма обучения:	очная

Рабочая программа рассмотрена
на заседании кафедры бизнес-информатики и математики

1. Цели и задачи освоения дисциплины

Цель дисциплины заключается в овладении фундаментальными знаниями в области автоматической обработки естественного языка и их использовании при решении научных и прикладных задач.

Задачи дисциплины:

- изучение базовых теоретических основ методологии автоматической обработки естественного языка;
- формирование умений применять пакеты программ и инструментальные средства для автоматической обработки естественного языка;
- формирование умений реализовать нейронную сеть с использованием пакетов программ для автоматической обработки естественного языка.

2. Место дисциплины в структуре ОПОП ВО

Дисциплина относится к дисциплинам части учебного плана, формируемой участниками образовательных отношений.

Необходимыми условиями для освоения дисциплины являются:

- знание основных понятий дискретной математики, прикладной алгебры, вычислительной математики;
- знание основ языка программирование Python, умение разрабатывать алгоритмы решения задач и записывать их на языке программирования;
- владение навыками использования компьютерных технологий и средств обработки информации.

Содержание дисциплины служит основой для освоения дисциплин Обучение с подкреплением, Глубинное обучение: дополнительные главы , для прохождения технологической (проектно-технологической) практики, научно-исследовательской работы.

3. Результаты обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих компетенций:

Таблица 3.1

Код и наименование компетенции	Код и наименование индикаторов достижения компетенций (ИДК)	Результаты обучения по дисциплине
ПКС-2 Способен разрабатывать (совершенствовать) и внедрять новые методы, модели,	ПКС-2.1. Способен разрабатывать методы, модели, алгоритмы, технологии и инструментальные средства работы с большими данными	(31) Знание этапов жизненного цикла инфраструктуры больших данных и показателей эффективности технологий больших данных

алгоритмы, технологии и инструментальные средства работы с большими данными		(У1) Проведение сравнительного анализ и выбора источников данных, средств хранения и обработки данных, оценка условий их приобретения и использования	
		(В1) Практический опыт в области управления этапами жизненного цикла методической и технологической инфраструктуры больших данных	
	ПКС-2.2. Способен оптимизировать и внедрять методы, модели, алгоритмы, технологии и инструментальные средства работы с большими данными в профессиональной деятельности		(З2) Знает основные задачи, решаемые с помощью автоматической обработки естественного языка, методы построения языковых моделей
			(У2) Умеет разрабатывать (совершенствовать) и использовать методы, модели, алгоритмы, технологии и инструментальные средства автоматической обработки естественного языка
		(В2) Владеет навыками автоматической обработки естественного языка в реальных практических задачах	

4. Объем дисциплины

Общий объем дисциплины составляет 4 зачетные единицы, 144 часа.

Таблица 4.1.

Форма обучения	Курс/семестр	Аудиторные занятия/контактная работа, час.			Самостоятельная работа, час.	Контроль	Форма промежуточной аттестации
		Лекции	Практические занятия	Лабораторные занятия			
очная	2/3	28	-	42	38	36	экзамен

5. Структура и содержание дисциплины

5.1. Структура дисциплины.

очная форма обучения (ОФО)

Таблица 5.1.1

№ п/п	Структура дисциплины		Аудиторные занятия, час.			СРС, час.	Всего, час.	Код ИДК	Оценочные средства
	Номер раздела	Наименование раздела	Л.	Пр.	Лаб.				
1	1	Основы автоматической обработки естественного языка	10	-	14	8	32	ПКС 2.1, ПКС 2.2	Задание на лабораторную работу № 1, Задание на лабораторную работу № 2.
2	2	Нейронные сети для задач обработки	9	-	13	15	37		Задание на лабораторн

		естественного языка							ую работу № 3, Задание на лабораторную работу № 4.
3	3	Применение нейронных сетей для решения задачи классификации текста	9	-	15	15	39		Задание на лабораторную работу № 5, Задание на лабораторную работу № 6.
4	Экзамен		-	-	-	36	36	ПКС 2.1, ПКС 2.2	Вопросы для экзамена в форме теста
Итого:			28	-	42	74	144		

5.2. Содержание дисциплины.

5.2.1. Содержание разделов дисциплины (дидактические единицы).

Раздел 1. «Основы автоматической обработки естественного языка». Основные понятия задачи обработки естественного языка (Natural Language Processing – NLP). Задачи, решаемые при автоматической обработке естественного языка: распознавание речи (автоматическое преобразование речи в текстовые данные); реферирование и аннотирование текста; информационный поиск; классификация текста по темам (отнесение текста новостей к одной из N тем); анализ тональности текста (положительные/отрицательные отзывы); выделение именованных сущностей и фактов (извлечение из неструктурированного текста имен, или дат рождения, или марок автомобилей); вопросно-ответные системы; чат-боты (чат-бот Сбербанка для клиентской поддержки, чат-бот Тинькофф для поиска авиабилетов); машинный перевод; генерация текста; синтез речи (голосовые помощники, голосовые автоответчики). Предварительная обработка текста. Очистку текста от лишних символов. Токенизация текста. Лемматизация и стемминг текста. Векторизация текста.

Раздел 2. «Нейронные сети для задач обработки естественного языка». Архитектуры нейронных сетей, которые применяются при решении задач обработки естественного языка. Сверточные одномерные нейронные сети (CNN 1D). Рекуррентные нейронные сети (RNN). LSTM (Long short-term memory) – рекуррентные нейронные сети с долгой краткосрочной памятью. GRU (Gated Recurrent Units) – модификация рекуррентных нейронных сетей на основе механизма вентиляей. ELMO – нейронная сеть на основе LSTM. Transformer – архитектура, основанная на encoder и decoder частях. Архитектура BERT (Bidirectional Encoder Representations from Transformers) – как самая передовая на данный момент архитектура для решения задач обработки естественного языка. Варианты цифровых

представлений текста: частотный подход; тематическое моделирование; дистрибутивная семантика. Методы «мешок слов», One Hot Encoding (OHE), TF-IDF (term frequency, inverse document frequency). Тематическое моделирование. Дистрибутивная семантика. Вероятностный латентно семантический анализ (PLSA, Probabilistic latent semantic analysis). Латентное размещение Дирихле (LDA, Latent Dirichlet Allocation).

Раздел 3. «Применение нейронных сетей для решения задачи классификации текста». Инструментарий пакетов Keras и TensorFlow для построения модели нейронной сети, решающей задачу классификации текстов. Набор данных «Отзывы о фильмах от IMDb». Загрузка набора данных в онлайн сервис Google Colab. Решение задачи классификации (определения тональности). Решение задачи классификации текста на основе одномерной сверточной нейронной сети. Инструментарий пакетов Keras и TensorFlow для реализации сверточной нейронной сети. Решение задачи классификации на основе архитектур LSTM и GRU. Инструментарий Keras и TensorFlow для построения рекуррентных нейронных сетей и их модификаций.

5.2.2. Содержание дисциплины/модуля по видам учебных занятий.

Лекционные занятия

Таблица 5.2.1

№ п/п	Номер раздела дисциплины	Объем, час.			Тема лекции
		ОФО	ЗФО	ОЗФО	
1	1	4	-	-	Основные понятия задачи обработки естественного языка (Natural Language Processing – NLP)
2	1	3	-	-	Задачи, решаемые при автоматической обработке естественного языка
3	1	3	-	-	Предварительная обработка текста
4	2	3	-	-	Нейронные сети и их применение для задач автоматической обработки естественного языка
5	2	3	-	-	Архитектуры нейронных сетей, которые применяются при решении задач обработки естественного языка
6	2	3	-	-	Цифровое представление текста
7	3	3	-	-	Задача классификация текста и её решение с помощью нейронных сетей
8	3	3	-	-	Решение задачи классификации на основе сверточной нейронной сети
9	3	3	-	-	Решение задачи классификации на основе рекуррентной нейронной сети
Итого:		28	-	-	

Практические занятия

Практические занятия учебным планом не предусмотрены.

Лабораторные работы

Таблица 5.2.2

№ п/п	Номер раздела дисциплины	Объем, час.			Тема лабораторной работы
		ОФО	ЗФО	ОЗФО	

1	1	7	-	-	Решение задач обработки текстов на основе языка Python
2	1	7	-	-	Алгоритмы предварительной обработки текста и их реализация на Python
3	2	13	-	-	Построение нейронной сети с применением пакетов TensorFlow и Keras
4	3	5	-	-	Наборы данных для задачи классификации и их обработка средствами TensorFlow и Keras
5	3	5	-	-	Решение задачи классификации на основе сверточной нейронной сети в Google Colab
6	3	5	-	-	Решение задачи классификации на основе рекуррентной нейронной сети в Google Colab
Итого:		42	-	-	

Самостоятельная работа студента

Таблица 5.2.3

№ п/п	Номер раздела дисциплины	Объем, час.			Тема	Вид СРС
		ОФО	ЗФО	ОФО		
1	1	17	-	-	Основы автоматической обработки естественного языка	Подготовка к лабораторным работам, оформление отчетов по лабораторным работам
2	2	15	-	-	Нейронные сети для задач обработки естественного языка	Подготовка к лабораторным работам, оформление отчетов по лабораторным работам
3	3	15	-	-	Применение нейронных сетей для решения задачи классификации текста	Подготовка к лабораторным работам, оформление отчетов по лабораторным работам
4		27	-	-	1-3	Подготовка к экзамену (в форме теста)
Итого:		74	-	-		

5.2.3. Преподавание дисциплины ведется с применением следующих видов образовательных технологий:

- визуализация учебного материала в офисном пакете в диалоговом режиме (лекционные занятия);
- работа в малых группах (лабораторные работы);
- индивидуальные задания (лабораторные работы).

6. Тематика курсовых проектов

Курсовые проекты учебным планом не предусмотрены;

7. Контрольные работы

Контрольные работы учебным планом не предусмотрены.

8. Оценка результатов освоения дисциплины

8.1. Критерии оценивания степени полноты и качества освоения компетенций в соответствии с планируемыми результатами обучения приведены в Приложении 1.

8.2. Рейтинговая система оценивания степени полноты и качества освоения компетенций обучающихся очной формы обучения представлена в таблице 8.1.

Таблица 8.1

№ п/п	Виды мероприятий в рамках текущего контроля	Количество баллов
2 семестр		
1 текущая аттестация		
1	Лабораторная работа №1	15
2	Лабораторная работа №2	15
	ИТОГО за первую текущую аттестацию	30
2 текущая аттестация		
3	Лабораторная работа №3	15
4	Лабораторная работа №4	15
5		
	ИТОГО за вторую текущую аттестацию	30
3 текущая аттестация		
6	Лабораторная работа №5	20
7	Лабораторная работа №6	20
	ИТОГО за третью текущую аттестацию	40
	ВСЕГО	100

9. Учебно-методическое и информационное обеспечение дисциплины/модуля

9.1. Перечень рекомендуемой литературы представлен в Приложении 2.

9.2. Современные профессиональные базы данных и информационные справочные системы

- Электронный каталог/Электронная библиотека ТИУ <http://webirbis.tsogu.ru/>
- Цифровой образовательный ресурс – библиотечная система IPR SMART — <https://www.iprbookshop.ru/>
- Электронно-библиотечная система «Консультант студента» www.studentlibrary.ru
- Электронно-библиотечная система «Лань» <https://e.lanbook.com>
- Образовательная платформа ЮРАЙТ www.urait.ru
- Научная электронная библиотека ELIBRARY.RU <http://www.elibrary.ru>
- Национальная электронная библиотека (НЭБ)

9.3. Лицензионное и свободно распространяемое программное обеспечение, в т.ч. отечественного производства.

- Microsoft Windows;
- Microsoft Office Professional Plus;
- Python;

10. Материально-техническое обеспечение дисциплины/модуля

Помещения для проведения всех видов работы, предусмотренных учебным планом, укомплектованы необходимым оборудованием и техническими средствами обучения.

Таблица 10.1

Обеспеченность материально-технических условий реализации ОПОП ВО

п/п	Наименование учебных предметов, курсов, дисциплин (модулей), практики, иных видов учебной деятельности, предусмотренных учебным планом образовательной программы	Наименование помещений для проведения всех видов учебной деятельности, предусмотренной учебным планом, в том числе помещения для самостоятельной работы, с указанием перечня основного оборудования, учебно-наглядных пособий	Адрес (местоположение) помещений для проведения всех видов учебной деятельности, предусмотренной учебным планом (в случае реализации образовательной программы в сетевой форме дополнительно указывается наименование организации, с которой заключен договор)
	2	3	4
1.	Автоматическая обработка естественного языка	<p>Лекционные занятия: Учебная аудитория для проведения занятий лекционного типа; групповых и индивидуальных консультаций; текущего контроля и промежуточной аттестации. Оснащенность: Учебная мебель: столы, стулья. Моноблок - 1 шт., проектор - 1 шт., проекционный экран - 1 шт., акустическая система (колонки) - 4 шт., микрофон - 1 шт., документ-камера - 1 шт., телевизор - 2 шт.</p> <p>Лабораторные занятия: Учебная аудитория для проведения занятий семинарского типа (лабораторные занятия); групповых и индивидуальных консультаций; текущего контроля и промежуточной аттестации. Оснащенность: Учебная мебель: столы, стулья. Моноблок - 13 шт., проектор - 1 шт., интерактивная сенсорная доска - 1 шт., акустическая система (колонки) - 2 шт.</p>	<p>625039, г. Тюмень, ул. Мельникайте, д. 70.</p> <p>625039, г. Тюмень, ул. Мельникайте, д. 70.</p>

11. Методические указания по организации СРС

11.1. Методические указания по подготовке к лабораторным занятиям.

Лабораторные занятия способствуют углублённому изучению дисциплины и служат основной формой подведения итогов самостоятельной работы студентов. Основная цель лабораторных занятий заключается не только углубить и закрепить теоретические знания, но и сформировать практические компетенции, необходимые будущим специалистам.

На лабораторных занятиях приветствуется активное участие в обсуждении конкретных ситуаций, способность на основе полученных знаний находить наиболее

эффективные решения поставленных проблем, уметь находить полезный дополнительный материал по тематике занятий.

Магистранту рекомендуется следующая схема подготовки к занятию:

1. Проработать конспект лекций;
2. Изучить рекомендованную литературу;
3. При затруднениях сформулировать вопросы к преподавателю
4. После выполнения лабораторной работы оформит отчет и подготовиться к защите.

11.2. Методические указания по организации самостоятельной работы.

Самостоятельная работа является одной из важнейших форм изучения любой дисциплины. Она позволяет систематизировать и углубить теоретические знания, закрепить умения и навыки, способствует развитию умений пользоваться научной и учебно-методической литературой. Познавательная деятельность в процессе самостоятельной работы требует от магистранта высокого уровня активности и самоорганизованности.

В учебном процессе выделяют два вида самостоятельной работы: аудиторная и внеаудиторная.

Аудиторная самостоятельная работа по дисциплине выполняется на учебных занятиях под непосредственным руководством преподавателя и по его заданию.

Внеаудиторная самостоятельная работа магистрантов представляет собой логическое продолжение аудиторных занятий. Затраты времени на выполнение этой работы регламентируются рабочим учебным планом. Режим работы выбирает сам обучающийся в зависимости от своих способностей и конкретных условий.

Самостоятельная работа может осуществляться индивидуально или группами магистрантов в зависимости от цели, объема, конкретной тематики самостоятельной работы, уровня сложности, уровня умений магистрантов.

Самостоятельная работа включает в себя работу с конспектом лекций, изучение и конспектирование рекомендуемой литературы, подготовка мультимедиа-сообщений/докладов, подготовка реферата, тестирование, решение задач и упражнений по образцу, решение вариативных задач, выполнение чертежей, схем, расчетов (графических работ), решение ситуационных (профессиональных) задач, подготовка к деловым играм, проектирование и моделирование разных видов и компонентов профессиональной деятельности, научно-исследовательскую работу и др.

Контроль результатов внеаудиторной самостоятельной работы магистрантов может осуществляться в пределах времени, отведенного на обязательные учебные занятия по

дисциплине и внеаудиторную самостоятельную работу магистрантов по дисциплине, может проходить в письменной, устной или смешанной форме.

Планируемые результаты обучения для формирования компетенции и критерии их оценивания

Дисциплина: Автоматическая обработка естественного языка

Код, направление подготовки: 01.04.02 Прикладная математика и информатика

Направленность (профиль): Машинное обучение и анализ данных

Код компетенции	Код и наименование результата обучения по дисциплине	Критерии оценивания результатов обучения			
		1-2	3	4	5
ПКС-2	(З1) Знание этапов жизненного цикла инфраструктуры больших данных и показателей эффективности больших данных	Не знает основные задачи, жизненного цикла инфраструктуры больших данных и показателей эффективности больших данных	Демонстрирует знание основных задач, жизненного цикла инфраструктуры больших данных и показателей эффективности больших данных	Демонстрирует достаточные знания основных задач, жизненного цикла инфраструктуры больших данных и показателей эффективности больших данных	Демонстрирует исчерпывающие знания основных задач, жизненного цикла инфраструктуры больших данных и показателей эффективности больших данных
	(У1) Проведение сравнительного анализ и выбора источников данных, средств хранения и обработки данных, оценка условий их приобретения и использования	Не умеет проводить сравнительный анализ и выбора источников данных, средств хранения и обработки данных, оценка условий их приобретения и использования	Демонстрирует умение проводить сравнительный анализ и выбора источников данных, средств хранения и обработки данных, оценка условий их приобретения и использования	Демонстрирует достаточные умения проводить сравнительный анализ и выбора источников данных, средств хранения и обработки данных, оценка условий их приобретения и использования	Демонстрирует исчерпывающие умения проводить сравнительный анализ и выбора источников данных, средств хранения и обработки данных, оценка условий их приобретения и использования
	(В1) Владеет управлением этапами жизненного цикла методической и технологической инфраструктуры больших данных	Не владеет управлением этапами жизненного цикла методической и технологической инфраструктуры больших данных	Демонстрирует практические навыки управлением этапами жизненного цикла методической и технологической инфраструктуры больших данных	Демонстрирует достаточные практические навыки управлением этапами жизненного цикла методической и технологической инфраструктуры больших данных	Демонстрирует исчерпывающие навыки управлением этапами жизненного цикла методической и технологической инфраструктуры больших данных

Код компетенции	Код и наименование результата обучения по дисциплине	Критерии оценивания результатов обучения			
		1-2	3	4	5
	(32) Знает основные задачи, решаемые с помощью автоматической обработки естественного языка, методы построения языковых моделей	Не знает основные задачи, решаемые с помощью автоматической обработки естественного языка, методы построения языковых моделей	Демонстрирует знание основных задач, решаемых с помощью автоматической обработки естественного языка, методы построения языковых моделей	Демонстрирует достаточные знания основных задач, решаемых с помощью автоматической обработки естественного языка, методы построения языковых моделей	Демонстрирует исчерпывающие знания основных задач, решаемых с помощью автоматической обработки естественного языка, методы построения языковых моделей
	(У2) Умеет разрабатывать (совершенствовать) и использовать методы, модели, алгоритмы, технологии и инструментальные средства автоматической обработки естественного языка	Не умеет разрабатывать (совершенствовать) и использовать методы, модели, алгоритмы, технологии и инструментальные средства автоматической обработки естественного языка	Демонстрирует умение разрабатывать (совершенствовать) и использовать методы, модели, алгоритмы, технологии и инструментальные средства автоматической обработки естественного языка	Демонстрирует достаточные умения разрабатывать (совершенствовать) и использовать методы, модели, алгоритмы, технологии и инструментальные средства автоматической обработки естественного языка	Демонстрирует исчерпывающие умения разрабатывать (совершенствовать) и использовать методы, модели, алгоритмы, технологии и инструментальные средства автоматической обработки естественного языка
	(В2) Владеет навыками автоматической обработки естественного языка в реальных практических задачах	Не владеет навыками автоматической обработки естественного языка в реальных практических задачах	Демонстрирует практические навыки автоматической обработки естественного языка в реальных практических задачах	Демонстрирует достаточные практические навыки автоматической обработки естественного языка в реальных практических задачах	Демонстрирует исчерпывающие навыки автоматической обработки естественного языка в реальных практических задачах

КАРТА
обеспеченности дисциплины учебной и учебно-методической литературой

Дисциплина: Автоматическая обработка естественного языка

Код, направление подготовки: 01.04.02 Прикладная математика и информатика

Направленность (профиль): Машинное обучение и анализ данных

№ п/п	Название учебного, учебно-методического издания, автор, издательство, вид издания, год издания	Количество экземпляров в БИК	Контингент обучающихся, использующих указанную литературу	Обеспеченность обучающихся литературой, %	Наличие электронного варианта в ЭБС (+/-)
1	Нейросетевые методы в обработке естественного языка / Й. Гольдберг. - Москва : ДМК Пресс, 2019. - 282 с. - ЭБС "Лань". : [сайт]. - URL: https://e.lanbook.com/book/131704	ЭР*	30	100	+
2	Обработка естественного языка с TensorFlow / Т. Ганегедара. - Москва : ДМК Пресс, 2020. - 382 с. - ЭБС "Лань". : [сайт]. - URL: https://e.lanbook.com/book/140584	ЭР*	30	100	+
3	Обработка естественного языка на Java / Р. Риз. - Москва : ДМК Пресс, 2016. - URL: https://e.lanbook.com/book/93272	ЭР*	30	100	+
4	Нейронные сети: основы теории / А. И. Галушкин. - [Б. м.] : Горячая линия-Телеком, 2017. - 496 с. - ЭБС Лань. : [сайт]. - URL: https://e.lanbook.com/book/111043	ЭР*	30	100	+

ЭР – электронный ресурс для автор. пользователей доступен через Электронный каталог/Электронную библиотеку ТИУ <http://webirbis.tsogu.ru/>