

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Соловьев Дмитрий Александрович

Должность: ректор ФГБОУ ВО Вавиловский университет

Дата подписания: 14.05.2024 13:28:28

Уникальный программный ключ:

528682d78e61e566ab037f01fe1a2172f735a12

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ



Федеральное государственное бюджетное образовательное учреждение высшего образования

«Саратовский государственный университет

генетики, биотехнологии и инженерии имени Н.И. Вавилова»

СОГЛАСОВАНО

И.о. заведующего кафедрой

/Ключиков А.В./

« 12 » апреля 2024 г.

УТВЕРЖДАЮ

Декан факультета

/Волощук Л.А./

« 12 » апреля 2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Дисциплина

**ОБРАБОТКА ЕСТЕСТВЕННОГО
ЯЗЫКА**

Направление
подготовки

38.04.05 Бизнес-информатика

Направленность
(профиль)

Управление бизнес анализом

Квалификация
выпускника

Магистр

Нормативный срок
обучения

2 года

Форма обучения

Очная

Разработчик: доцент, Шибайкин В.А.

Саратов 2024

1. Цель освоения дисциплины

Целью освоения дисциплины является формирование у студентов теоретических знаний об общих методах компьютерного анализа и синтеза текстов на естественных языках.

2. Место дисциплины в структуре ОПОП ВО

В соответствии с учебным планом по направлению подготовки 38.04.05 Бизнес-информатика дисциплина «Обработка естественного языка» относится к вариативной части Блока 1.

Для изучения данной дисциплины необходимы знания, умения и навыки, формируемые при изучении дисциплин «Искусственный интеллект в бизнес-среде», «Наука о данных для бизнеса», а также дисциплин бакалавриата.

Дисциплина «Обработка естественного языка» является базовой для изучения практики «Научно-исследовательская работа» и «Подготовка к процедуре защиты и защита выпускной квалификационной работы».

3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижениями компетенций

Изучение данной дисциплины направлено на формирование у обучающихся компетенций, представленных в табл. 1

Таблица 1

Требования к результатам освоения дисциплины

№ п/п	Код компетенции	Содержание компетенции (или ее части)	Индикаторы достижения компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
1	2	3	4	5	6	7
1.	ПК-3	Способен управлять контентом предприятия, процессами создания и использования информационных сервисов	ПК-3.3 Управляет работами по анализу требований в проектах малого и среднего уровня сложности в области ИТ	ключевые понятия, задачи и приложения обработки естественного языка, а также и методы обработки подготовки текстовых данных	применять методы подготовки данных для анализа текстов, создавать собственные приложения для обработки языковых задач и оценивать их производительность, интерпретировать полученные модели	инструментами для обработки текстов, методами и метриками для оценки языковых моделей, навыками исследования и изучения новых технологий и методов в области обработки естественного языка

4. Объём, структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 2 зачетные единицы, 72 часа.

Таблица 2 Объем дисциплины

	Количество часов				
	Всего	в т.ч. по семестрам			
		1	2	3	4
Контактная работа – всего, в т.ч.	32,1			32,1	
аудиторная работа:					
лекции	16			16	
лабораторные	16			16	
практические	-			-	
промежуточная аттестация	0,1			0,1	
контроль	-			-	
Самостоятельная работа	39,9			39,9	
Форма итогового контроля	зач.			зач.	
Курсовой проект (работа)	х			х	

Таблица 3

Структура и содержание дисциплины

№ п/п	Тема занятия Содержание	Неделя семестра	Контактная работа			Самостоятельная работа		Контроль	
			Вид занятия	Форма проведения	Количество часов	Количество часов	Вид	Форма	
1	2	3	4	5	6	7	8	9	
3 семестр									
1.	Введение в обработку естественного языка Обзор области обработки естественного языка, её задач и приложений. Рассмотрение истории развития обработки естественного языка, основные достижения и тенденции. Знакомство с основными терминами и понятиями.	1	Л	В	2	0	ТК	С	
2.	Работа с библиотеками обработки естественного языка Установка и настройка библиотек обработки естественного языка.	2	ЛЗ	М	2	4	ТК	ВК, ПО,ТЗ	
	Предобработка текстов Техники очищения текста. Устранение стоп-слов и нормализация текстов. Способы представления текстов для анализа. Основные инструменты и библиотеки.	3	Л	В	2	0	ТК	С	
3.	Реализация морфологического анализа Создание простого морфологического анализатора с использованием существующих инструментов.	4	ЛЗ	М	2	4	Д, ТК	ПО,ТЗ	
4.	Синтаксический анализ Обзор методов построения синтаксических деревьев. Принципы работы парсеров и их применение	5	Л	В	2	0	ТК	С	
5.	Построение синтаксического анализатора Практическая работа по созданию и тестированию синтаксического парсера	6	ЛЗ	М	2	4	ТК	ПО,ТЗ	
6.	Семантический анализ Изучение методов извлечения смысла из текста. Использование онтологий и семантических сетей	7	Л	В	2	0	ТК	С	
7.	Семантический анализ текста Использование WordNet и других инструментов для семантического анализа текстов.	8	ЛЗ	М	2	4	РК, Д	ПО, Т	

1	2	3	4	5	6	7	8	9
8.	Обработка текстов на естественном языке Основные алгоритмы и подходы к обработке текстов.	9	Л	В	2	0	ТК	С
9.	Классификация текстов и извлечение сущностей Реализация алгоритмов классификации и извлечения сущностей из текстов	10	ЛЗ	В,М	2	4	ТК	ПО, ТЗ
10.	Нейронные сети для обработки естественного языка Введение в нейронные сети и их архитектуры, применяемые в обработке естественного языка. Обзор RNN, LSTM и GRU ¹ . Применение нейронных сетей в задачах языкового моделирования.	11	Л	В	2	0	ТК	С
11.	Разработка чат-бота Проектирование простого чат-бота, используя библиотеку для обработки естественного языка. Рассмотрение обработки запросов и генерации ответов.	12	ЛЗ	Т	2	4	ТК, ТР	Д, ПО,ТЗ
12.	Современные методы обработки естественного языка Обзор трансформеров и их влияние на обработке естественного языка. Инструменты.	13	Л	В	2	0	ТК	С
13.	Применение методов анализа тональности Практика анализа тональности текстовых данных с использованием предварительно обученных моделей и библиотек как TextBlob или VADER	14	ЛЗ		2	4	ТК	ПО,ТЗ
14.	Практические приложения обработки естественного языка и этические вопросы Обсуждение прикладных задач: анализ тональности, чат-боты, машинный перевод. Этические аспекты использования обработки естественного языка, включая предвзятость алгоритмов и конфиденциальность данных	15	Л	В	2	0	ТК	С
15.	Этические вопросы обработки языка и их обсуждение Групповое обсуждение и анализ кейсов из практики, связанных с предвзятостью моделей и конфиденциальностью данных. Проработка материалов и предложений по улучшению этических аспектов в обработке естественного языка	Н.н.	ЛЗ	КМ	2	11,9	РК, ТР	ПО,Т, Д
16.	Выходной контроль				0,1		Вых К	зач.
Итого:					32,1	39,9		

Примечание:

Условные обозначения:

Виды аудиторной работы: Л – лекция, ЛЗ – лабораторное занятие.

Формы проведения занятий: В – лекция-визуализация, Т – лекция/занятие, проводимое в традиционной форме, М – моделирование КМ-кейс метод.

Виды контроля: ВК – входной контроль; ТК – текущий контроль, ТР – творческая работа, РК – рубежный контроль. ВыхК – выходной контроль

Форма контроля: ПО- письменный опрос, С – собеседование, Д – доклад, ТЗ – типовое задание, Т – тестовое задание, зач. – зачет

5. Образовательные технологии

Организация занятий по дисциплине «Обработка естественного языка» проводится по видам учебной работы: лекции, практические занятия, текущий контроль.

Реализация компетентного подхода в рамках направления подготовки 38.04.05 Бизнес-информатика предусматривает использование в учебном процессе активных и интерактивных форм проведения занятий в сочетании с внеаудиторной работой для формирования и развития профессиональных навыков обучающихся.

Лекционные занятия проводятся в аудитории с применением мультимедийного проектора в виде учебной презентации. Основные моменты лекционных занятий конспектируются.

Целью лабораторных занятий является выработка практических навыков работы с программными инструментами для обработки текстов.

¹ Recurrent neural network, Long short-term memory, Gated Recurrent Unit.

Для достижения этих целей используются как традиционные формы работы – решение задач, так и интерактивные методы – моделирование, метод кейсов.

Моделирование позволяет обучиться общим требованиям к использованиям цифровых технологий, поиску источников информации по цифровым технологиям практическому навыку описания объекта или процесса с использованием моделей.

Метод моделирования в наибольшей степени соответствует задачам высшего образования. Он способствует разделению сложного процесса моделирования на составные части, что позволяет лучше усваивать материал. Реализуется объяснительно-иллюстративный характер обучения

Лабораторные занятия проводятся в специальных аудиториях, оборудованных необходимым оборудованием.

Самостоятельная работа охватывает проработку обучающимися отдельных вопросов теоретического курса, выполнение домашних работ, включающих решение задач, анализ конкретных ситуаций и подготовку их презентаций, и т.п. Самостоятельная работа осуществляется в индивидуальном формате. Самостоятельная работа выполняется обучающимися на основе учебно-методических материалов дисциплины (приложение 2). Самостоятельно изучаемые вопросы курса включаются в вопросы выходного контроля.

6. Учебно-методическое и информационное обеспечение дисциплины

а) основная литература (библиотека Вавиловского университета)

№ п/п	Наименование, ссылка для электронного доступа или кол-во экземпляров в библиотеке	Автор(ы)	Место издания, издательство, год	Используется при изучении разделов (из п. 4, таб. 3)
1	2	3	4	5
1.	Обработка естественного языка с использованием языка программирования Python URL: https://e.lanbook.com/book/360923	А. Б. Мантусов	Элиста : КГУ, 2022	2-14
2.	Нейросетевые методы в обработке естественного языка. URL: https://www.iprbookshop.ru/124564.html	Гольдберг, Й.	Пресс, 2019	11-12
3.	Степанов, Ю. А. Системы искусственного интеллекта : учебное пособие URL: https://e.lanbook.com/book/427532	Ю. А. Степанов, А. В. Вылегжанина, Л. Н. Бурмин.	Кемерово : КемГУ, 2024.	Все разделы

б) дополнительная литература

№ п/п	Наименование, ссылка для электронного доступа или кол-во экземпляров в библиотеке	Автор(ы)	Место издания, издательство, год	Используется при изучении разделов (из п. 4.3)
1	2	3	4	5
1.	Гуманитарные проблемы искусственного интеллекта и его применения : монография URL: https://e.lanbook.com/book/426152	А. Б. Гехт, Р. В. Душкин, А. В. Неровный [и др.].	Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2024	15-16
2.	Лексические и грамматические аспекты разработки компьютерного сурдопереводчика русского языка : монография URL: https://znanium.ru/catalog/product/549352	М. Г. Гриф, О. О. Королькова, Л. Г. Панин [и др.].	- Новосибирск: Изд-во НГТУ, 2013	15-16

в) ресурсы информационно-телекоммуникационной сети «Интернет»

Для освоения дисциплины рекомендуются следующие сайты информационно-телекоммуникационной сети «Интернет»:

- официальный сайт университета: <https://vavilovsar.ru>;
- лексическая база слов <https://wordnet.princeton.edu>
- платформа для исследователей разных уровней
<https://jupyter.org/try>, <https://regex101.com/>
- Яндекс Диалоги (Алиса) <https://dialogs.yandex.ru/developer>
- коммуникационные сервисы <https://voximplant.ru> ,

г) периодические издания

- нет

д) информационные справочные системы и профессиональные базы данных:

Для пользования стандартами и нормативными документами рекомендуется применять информационные справочные системы и профессиональные базы данных, доступ к которым организован библиотекой университета через локальную вычислительную сеть.

Для пользования электронными изданиями рекомендуется использовать следующие информационные справочные системы и профессиональные базы данных:

1. Научная библиотека университета <https://www.vavilovsar.ru/biblioteka>

Базы данных содержат сведения о всех видах литературы, поступающей в фонд библиотеки. Более 1400 полнотекстовых документов (учебники, учебные пособия и т.п.) (доступ: с любого компьютера, подключенного к сети Internet).

2. Электронная библиотечная система "Лань" <https://e.lanbook.com>

Электронная библиотека издательства "Лань" - ресурс, включающий в себя как электронные версии книг издательства "Лань", так и коллекции полнотекстовых файлов других российских издательств (доступ: после регистрации с компьютера университета с любого компьютера, подключенного к сети Internet).

3. ЭБС IPR SMART <http://iprbookshop.ru>

ЭБС обеспечивает возможность работы с постоянно пополняемой базой лицензионных изданий (более 40000) по широкому спектру дисциплин - учебные, научные издания и периодика, представленные более 600 федеральными, региональными и вузовскими издательствами, научно-исследовательскими институтами и ведущими авторскими коллективами (доступ: после регистрации с компьютера университета с любого компьютера, подключенного к сети Internet).

4. ЭБС Znanium <https://znanium.ru>

Фонд ЭБС Znanium постоянно пополняется электронными версиями изданий, публикуемых Научно-издательским центром ИНФРА-М, коллекциями книг и журналов других российских издательств, а также произведениями отдельных авторов (доступ: с любого компьютера, подключенного к сети Internet; свободная регистрация).

5. Научная электронная библиотека eLIBRARY.RU <http://elibrary.ru>

Российский информационный портал в области науки, медицины, технологии и образования. На платформе аккумулируются полные тексты и

рефераты научных статей и публикаций (доступ: с любого компьютера, подключенного к сети Internet; свободная регистрация).

е) информационные технологии, используемые при осуществлении образовательного процесса:

К информационным технологиям, используемым при осуществлении образовательного процесса по дисциплине, относятся:

- персональные компьютеры, посредством которых осуществляется доступ к информационным ресурсам и оформляются результаты самостоятельной работы;
 - проекторы и экраны для демонстрации слайдов мультимедийных лекций;
 - активное использование средств коммуникаций (электронная почта, тематические сообщества в социальных сетях и т.п.).
- программное обеспечение:

№ п/п	Наименование раздела учебной дисциплины (модуля)	Наименование программы	Тип программы
1	2	3	4
1	Все разделы дисциплины	<i>Вспомогательное программное обеспечение:</i> «Р7-Офис» Предоставление неисключительных прав на программное обеспечение «Р7-Офис». Лицензиат – ООО «Солярис Технолоджис», г. Саратов. Договор № ЦЗ-1К-033 от 21.12.2022 г. Срок действия договора: с 01.01.2023 г. Лицензия на 3 года с правом последующего бессрочного использования, для образовательных учреждений.	Вспомогательная
2	Все разделы дисциплины	<i>Вспомогательное программное обеспечение:</i> Kaspersky Endpoint Security (антивирусное программное обеспечение). Лицензиат – ООО «Солярис Технолоджис», г. Саратов. Сублицензионный договор № 6-1128/2023/КСП-107 от 11.12.2023 г. Срок действия договора: 01.01.2024– 31.12.2024 г.	Вспомогательная

7. Материально-техническое обеспечение дисциплины (модуля)

Для проведения занятий лекционного и семинарского типов, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации необходимы аудитории с меловыми или маркерными досками, достаточным количеством посадочных мест и освещенностью. Для использования медиаресурсов необходимы проектор, экран, компьютер или ноутбук.

Для проведения лекционных, лабораторных занятий и самостоятельной работы по дисциплине кафедры «Цифровое управление процессами в АПК» имеются аудитории № 422, № 247, № 245, № 134а, № 134б, № 137.

Помещения для самостоятельной работы обучающихся аудитория читальные залы библиотеки оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.

8. Оценочные материалы

Оценочные материалы, сформированные для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине

«Обработка естественного языка» разработан на основании следующих документов:

- Федерального закона Российской Федерации от 29.12.2012 N 273-ФЗ «Об образовании в Российской Федерации» (с изменениями и дополнениями);

- приказа Минобрнауки РФ от 6 апреля 2021 г. N 245 Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры»;

Оценочные материалы представлены в приложении 1 к рабочей программе дисциплины и включают в себя:

- перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы;
- описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания;
- типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующие этапы формирования компетенций в процессе освоения образовательной программы;
- методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

9. Учебно-методическое обеспечение самостоятельной работы

Перечень учебно-методического обеспечения самостоятельной работы представлен в приложении 2 к рабочей программе по дисциплине «Обработка естественного языка».

10. Методические указания для обучающихся по изучению дисциплины «Обработка естественного языка»

Методические указания по изучению дисциплины «Обработка естественного языка» включают в себя:

1. Краткий курс лекций.

*Рассмотрено и утверждено на заседании кафедры
«Цифровое управление процессами в АПК»
«12» апреля 2024 г года (протокол № 12)*