

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ
Директор Таганрогского института
имени А. П. Чехова (филиала)
РГЭУ (РИНХ)
_____ С. А. Петрушенко
«20» мая 2025 г.

Рабочая программа дисциплины
Цифровые технологии и основы искусственного интеллекта

Направление подготовки
38.03.02 Менеджмент

Направленность (профиль) программы бакалавриата
38.03.02.04 Управление бизнес-процессами в цифровой экономике

Для набора 2025 года

Квалификация
бакалавр

КАФЕДРА экономики и гуманитарно-правовых дисциплин**Распределение часов дисциплины по семестрам / курсам**

Семестр (<Курс>. <Семестр на курсе>)	6 (3.2)		Итого	
	Неделя			
Неделя	17 2/6			
Вид занятий	уп	рп	уп	рп
Лекции	10	10	10	10
Лабораторные	10	10	10	10
Итого ауд.	20	20	20	20
Контактная работа	20	20	20	20
Сам. работа	88	88	88	88
Часы на контроль	36	36	36	36
Итого	144	144	144	144

ОСНОВАНИЕ

Учебный план утвержден учёным советом вуза от 28.02.2025 протокол № 9.

Программу составил(и): канд. техн. наук, Доц., внеш. совм. Веселая А.А.

Зав. кафедрой: Сердюкова Ю.А.

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1	сформировать у обучающихся представление о возможностях и перспективах развития цифровых технологий и искусственного интеллекта для обработки естественного языка
-----	---

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ОПК-2:	Способен осуществлять сбор, обработку и анализ данных, необходимых для решения поставленных управленческих задач, с использованием современного инструментария и интеллектуальных информационно-аналитических систем
ОПК-2.1:	Определяет источники информации и осуществляет их поиск на основе поставленных целей для решения профессиональных задач
ОПК-2.2:	Выбирает соответствующие содержанию профессиональных задач инструментарий обработки и анализа данных, современные информационные технологии и программное обеспечение
ОПК-2.3:	Составляет проекты распорядительных, организационных и информационно-справочных документов, осуществляет их информационную обработку и внедрение в управленческую деятельность с учетом заданных критериев качества документов
ОПК-5:	Способен использовать при решении профессиональных задач современные информационные технологии и программные средства, включая управление крупными массивами данных и их интеллектуальный анализ
ОПК-5.1:	Применяет общие или специализированные пакеты прикладных программ, предназначенных для анализа и обработки информации
ОПК-5.2:	Использует электронные библиотечные системы для поиска необходимой научной литературы и статистической информации
ОПК-5.3:	Работает с национальными и международными базами данных с целью поиска необходимой информации об экономических явлениях и процессах

В результате освоения дисциплины обучающийся должен:

Знать:
– основы интеллектуальных интернет-технологий (соотнесено с индикатором ОПК-2.1); – основы математической лингвистики и машинной обработки текстов на естественном языке (соотнесено с индикатором ОПК-5.1);;
Уметь:
– применять современные методы и инструменты обработки текстов на естественном языке(соотнесено с индикатором ОПК-2.2);
Владеть:
- создания онтологии(соотнесено с индикатором ОПК-2.3); - обработки естественного языка с помощью нейронных сетей(соотнесено с индикатором ОПК-5.3).

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ**Раздел 1. Обработка естественного языка**

№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
1.1	Введение в интеллектуальные интернет-технологии (Web Intelligence). Семантическая всемирная паутина (Semantic Web). URI. XML. DTD. RDF. OWL. Понятие онтологии. Модель онтологии. Простой словарь. Пассивный словарь. Активный словарь. Таксономия. Принципы проектирования и реализации онтологий. Protégé.	Лекционные занятия	6	2	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
1.2	Лабораторная работа 1. Создание онтологии в Protégé.	Лабораторные занятия	6	4	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
1.3	Формирование отчета по лабораторной работе 1	Самостоятельная работа	6	10	ОПК-2 ОПК-5 ОПК-5.1

					ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
1.4	Введение в обработку естественного языка (NLP). Обработка текста. Сегментация предложений. Токенизация. Нормализация текста. Удаление стоп-слов, специальных символов, цифр. Преобразование текста в общий регистр. Стемминг. Лемматизация.	Лекционные занятия	6	4	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
1.5	Лабораторная работа 2. Обработка естественного языка	Лабораторные занятия	6	2	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
1.6	Формирование отчета по лабораторной работе 2	Самостоятельная работа	6	10	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3

Раздел 2. Нейронные сети для обработки текста

№	Наименование темы, краткое содержание	Вид занятия / работы / форма ПА	Семестр / Курс	Количество часов	Компетенции
2.1	Введение в нейронные сети для обработки текста. Естественный язык и текст. Особенности обработки ЕЯ. Лингвистический анализ. Извлечение признаков. Прикладные задачи обработки текстов.	Самостоятельная работа	6	12	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
2.2	Векторная модель текста. Векторная модель текста и TF-IDF. Создание нейросети для работы с текстом	Самостоятельная работа	6	12	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
2.3	Классификация новостных текстов. Классификация сообщений электронной почты по темам с использованием библиотек scikit-learn, numpy, matplotlib и pytorch.	Лекционные занятия	6	2	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
2.4	Лабораторная работа 3. Классификация новостных текстов	Лабораторные занятия	6	2	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
2.5	Формирование отчета по лабораторной работе 3	Самостоятельная работа	6	10	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3

2.6	Базовые нейросетевые методы работы с текстами. Дистрибутивная семантика и векторные представления слов	Самостоятельная работа	6	12	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
2.7	Основные виды нейросетевых моделей для обработки текстов. Сверточные нейросети в обработке текста	Самостоятельная работа	6	12	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
2.8	Обучение модели Word2Vec на PyTorch. Обучение модели дистрибутивной семантики Word2Vec с использованием библиотек numpy, matplotlib и pytorch.	Лекционные занятия	6	2	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
2.9	Лабораторная работа 4. Обучение модели Word2Vec на PyTorch	Лабораторные занятия	6	2	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
2.10	Формирование отчета по лабораторной работе 4	Самостоятельная работа	6	10	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3
2.11	Подготовка к промежуточной аттестации	Экзамен	6	36	ОПК-2 ОПК-5 ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-2.1 ОПК-2.2 ОПК-2.3

4. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Структура и содержание фонда оценочных средств для проведения текущего контроля и промежуточной аттестации представлены в Приложении 1 к рабочей программе дисциплины.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Учебные, научные и методические издания

	Авторы, составители	Заглавие	Издательство, год	Библиотека / Количество
1	Рубашкин В. Ш.	Онтологическая семантика. Знания. Онтологии. Онтологически ориентированные методы информационного анализа текстов	Москва: Физматлит, 2012	http://biblioclub.ru/index.php?page=book&id=457687
2	Лейн Хобсон, Хапке Ханнес, Ховард Коул	Обработка естественного языка в действии	Санкт-Петербург: Питер, 2021	https://ibooks.ru/reading.php?short=1&productid=371695

5.1. Учебные, научные и методические издания

	Авторы, составители	Заглавие	Издательство, год	Библиотека / Количество
--	---------------------	----------	-------------------	-------------------------

	Авторы, составители	Заглавие	Издательство, год	Библиотека / Количество
1	Мартиросян К. В., Мишин В. В.	Интернет-технологии: учебное пособие	Ставрополь: Северо-Кавказский Федеральный университет (СКФУ), 2015	http://biblioclub.ru/index.php?page=book&id=457443
2	Потапов, А. С.	Технологии искусственного интеллекта	Санкт-Петербург: Университет ИТМО, 2010	http://www.iprbookshop.ru/68201.html

5.2. Профессиональные базы данных и информационные справочные системы

– Простое введение в обработку естественного языка (Simple Introduction to Natural Language Processing) – URL: <https://towardsdatascience.com/introduction-to-nlp-5bff2b2a7170>
– Онлайн-курс «Нейронные сети и обработка текста». <https://stepik.org/course/54098/>

5.3. Перечень программного обеспечения

Python
OpenOffice

5.4. Учебно-методические материалы для обучающихся с ограниченными возможностями здоровья

При необходимости по заявлению обучающегося с ограниченными возможностями здоровья учебно-методические материалы предоставляются в формах, адаптированных к ограничениям здоровья и восприятия информации. Для лиц с нарушениями зрения: в форме аудиофайла; в печатной форме увеличенным шрифтом. Для лиц с нарушениями слуха: в форме электронного документа; в печатной форме. Для лиц с нарушениями опорно-двигательного аппарата: в форме электронного документа; в печатной форме.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Помещения для всех видов работ, предусмотренных учебным планом, укомплектованы необходимой специализированной учебной мебелью и техническими средствами обучения:

- столы, стулья;
- персональный компьютер / ноутбук (переносной);
- проектор;
- экран / интерактивная доска.

Лабораторные занятия проводятся в компьютерных классах, рабочие места в которых оборудованы необходимыми лицензионными и/или свободно распространяемыми программными средствами и выходом в Интернет, и/или в специализированных лабораториях, предусмотренных образовательной программой.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Методические указания по освоению дисциплины представлены в Приложении 2 к рабочей программе дисциплины.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1 Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

1.1 Показатели и критерии оценивания компетенций:

ЗУН, составляющие компетенцию	Показатели оценивания	Критерии оценивания	Средства оценивания
ОПК-2: Способен осуществлять сбор, обработку и анализ данных, необходимых для решения поставленных управленческих задач, с использованием современного инструментария и интеллектуальных информационно-аналитических систем			
З основы интеллектуальных интернет-технологий	знание основ интеллектуальных интернет-технологий	1 - продемонстрированы знание вопроса и самостоятельность мышления, ответ соответствует требованиям правильности, полноты и аргументированности, 0 - ответ на вопрос неконкретен, слабо аргументирован и неубедителен	С
У определять источники информации и осуществлять их поиск на основе поставленных целей для решения поставленных профессиональных задач	умение определять источники информации и осуществлять их поиск на основе поставленных целей для решения поставленных профессиональных задач	от 22 до 25 баллов выставляется студенту, если он выполнил без ошибок и недочетов 85-100% интерактивных упражнений по	ИУ

		<p>онлайн-курсу; от 18 до 21 баллов выставляется студенту, если он выполнил без ошибок и недочетов 71- 84%</p> <p>интерактивных упражнений по онлайн-курсу; от 15 до 17 баллов выставляется студенту, если он выполнил без ошибок и недочетов 60- 70%</p> <p>интерактивных упражнений по онлайн-курсу; от 12 до 14 баллов выставляется студенту, если он выполнил без ошибок и недочетов 50- 59%</p> <p>интерактивных упражнений по онлайн-курсу; от 0 до 11 баллов выставляется студенту при невыполнении (не полном выполнении) требований, предусмотренны х выше.</p>	
--	--	--	--

<p>В навыками составления проектов, их информационной обработки и внедрения в профессиональную деятельность</p>	<p>владение навыками составления проектов, их информационной обработки и внедрения в профессиональную деятельность</p>	<p>от 18 до 20 баллов выставляется студенту, если он своевременно выполнил все задания, предусмотренные в проектной работе и в процессе защиты продемонстрировал полноту теоретических знаний в объеме содержания учебной дисциплины, относящейся к проектной работе. Сумел ответить на дополнительные вопросы, связанные не только с процессом выполнения проектной работы, но и с пониманием совершенных действий и решенных задач; от 15 до 17 баллов выставляется студенту, если он выполнил все задачи, предусмотренные в проектной работе и в процессе защиты продемонстрировал наличие достаточных теоретических знаний в объеме содержания учебной дисциплины, относящейся к проектной работе. Сумел ответить на вопросы, связанные с процессом выполнения проектной работы; от 12 до 14 баллов выставляется студенту, если он более чем на половину выполнил</p>	<p><i>ПР</i></p>
---	--	---	------------------

		поставленные в проектной работе задачи, способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к проектной работе. от 0 до 11 баллов выставляется студенту, при невыполнении требований, предусмотренных выше.	
ОПК-5: Способен использовать при решении профессиональных задач современные информационные технологии и программные средства, включая управление крупными массивами данных и их интеллектуальный анализ			
З основы математической лингвистики и машинной обработки текстов на естественном языке	знание основ математической лингвистики и машинной обработки текстов на естественном языке	1 - продемонстрированы знание вопроса и самостоятельность мышления, ответ соответствует требованиям правильности, полноты и аргументированности, 0 - ответ на вопрос неконкретен, слабо аргументирован и неубедителен	С
У использовать электронные библиотечные системы для поиска необходимой научной литературы и статистической информации	умение использовать электронные библиотечные системы для поиска	от 22 до 25 баллов выставляется студенту, если он выполнил без ошибок и	ИУ

	<p>необходимой научной литературы статистической информации</p>	<p>и</p> <p>недочетов 85-100% интерактивных упражнений по онлайн-курсу; от 18 до 21 баллов выставляется студенту, если он выполнил без ошибок и недочетов 71-84% интерактивных упражнений по онлайн-курсу; от 15 до 17 баллов выставляется студенту, если он выполнил без ошибок и недочетов 60-70% интерактивных упражнений по онлайн-курсу; от 12 до 14 баллов выставляется студенту, если он выполнил без ошибок и недочетов 50-59% интерактивных упражнений по онлайн-курсу; от 0 до 11 баллов выставляется студенту при невыполнении (не полном выполнении) требований, предусмотренных выше.</p>	
--	---	--	--

<p>В навыками работы с национальными и международными базами данных с целью поиска необходимой информации об экономических явлениях и процессах</p>	<p>владение работы с национальными и международными базами данных с целью поиска необходимой информации об экономических явлениях и процессах</p>	<p>от 18 до 20 баллов выставляется студенту, если он своевременно выполнил все задания, предусмотренные в проектной работе и в процессе защиты продемонстрировал полноту теоретических знаний в объеме содержания учебной дисциплины, относящейся к проектной работе. Сумел ответить на дополнительные вопросы, связанные не только с процессом выполнения проектной работы, но и с пониманием совершенных действий и решенных задач; от 15 до 17 баллов выставляется студенту, если он выполнил все задачи, предусмотренные в проектной работе и в процессе защиты продемонстрировал наличие достаточных теоретических знаний в объеме содержания учебной дисциплины, относящейся к проектной работе. Сумел ответить на вопросы, связанные с процессом выполнения проектной работы; от 12 до 14 баллов выставляется студенту, если он более чем на половину выполнил</p>	<p><i>ПР</i></p>
---	---	---	------------------

		<p>поставленные в проектной работе задачи, способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к проектной работе. от 0 до 11 баллов выставляется студенту, при невыполнении требований, предусмотренных выше.</p>	
--	--	--	--

1.2 Шкалы оценивания:

Текущий контроль успеваемости и промежуточная аттестация осуществляется в рамках накопительной балльно-рейтинговой системы в 100-балльной шкале.

84-100 баллов (оценка «отлично»)

67-83 баллов (оценка «хорошо»)

50-66 баллов (оценка «удовлетворительно»)

0-49 баллов (оценка «неудовлетворительно»)

2 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Вопросы к экзамену:

1. Дайте определение Natural Language Processing (NLP). Какие основные задачи решаются в NLP? Приведите 3-5 примеров практического применения NLP.
2. Каково значение NLP в современном мире? Обсудите, как NLP влияет на технологии и повседневную жизнь.
3. Сравните библиотеки NLTK и SpaCy с точки зрения их возможностей, производительности и типичных сценариев использования.
4. Приведите примеры основных функций, которые предоставляют NLTK и SpaCy для обработки текста.
5. Опишите основные лингвистические уровни анализа текста: морфология, синтаксис, семантика и прагматика. Какова их роль в задачах NLP?
6. Какие этапы включает предварительная обработка текстовых данных? Объясните важность каждого этапа.

7. Что такое «шум» в текстовых данных и как его можно устранить? Опишите методы нормализации регистра, удаления пунктуации и чисел.
8. Объясните процесс токенизации. Какие существуют подходы к токенизации (например, по словам, по предложениям)?
9. Что такое стоп-слова и зачем их удаляют? Приведите примеры стоп-слов и объясните, когда их удаление может быть нежелательным.
10. Сравните стемминг и лемматизацию. В чем заключаются их основные отличия, преимущества и недостатки?
11. Опишите принципы работы алгоритмов стемминга Porter и Snowball. Какие существуют алгоритмы лемматизации и в чем их особенности (например, WordNetLemmatizer, SpaCy Lemmatizer)?
12. Как осуществляется обработка специальных символов и эмодзи на этапе предварительной обработки текста?
13. Объясните принципы работы моделей Bag of Words (BoW) и TF-IDF (Term Frequency-Inverse Document Frequency). В чем их различия и какие метрики они используют?
14. Какие ограничения имеют традиционные методы векторизации (BoW, TF-IDF)? Приведите примеры.
15. Что такое векторные представления слов (Word Embeddings)? Объясните концепцию Word2Vec и GloVe и в чем их преимущество перед традиционными методами векторизации.
16. Перечислите и кратко охарактеризуйте основные алгоритмы машинного обучения, используемые для классификации текста (Naive Bayes, SVM, Logistic Regression, Random Forest). В каких случаях предпочтительнее использовать каждый из них?
17. Что такое анализ тональности (Sentiment Analysis)? Опишите основные задачи и приведите примеры его применения.
18. Опишите пошаговый процесс построения простого классификатора текста (от предобработки до оценки).
19. Что такое Named Entity Recognition (NER)? Объясните принципы работы NER и перечислите основные типы сущностей, которые обычно извлекаются.
20. Приведите примеры применения NER в различных областях, таких как анализ новостей, медицинские тексты или юридические документы.
21. Какова роль аннотации данных (разметки) для обучения моделей NER? Опишите основные принципы разметки данных для NER.
22. Объясните основы рекуррентных нейронных сетей (RNN) и их применение в NLP. Какие проблемы присущи простым RNN при обработке длинных последовательностей?
23. Опишите архитектуры и преимущества моделей Long Short-Term Memory (LSTM) и Gated Recurrent Unit (GRU) по сравнению с базовыми RNN. Как они решают проблему «исчезающего/взрывающегося градиента»?
24. Приведите примеры применения рекуррентных сетей в задачах NLP, таких как классификация последовательностей или генерация текста.
25. Объясните архитектуру Трансформеров. Какова роль механизма внимания (Attention) в этой архитектуре и почему он так важен?
26. Опишите концепцию и принцип работы современных предобученных моделей на основе трансформеров: BERT, GPT, T5. В чем их основные различия и сферы применения?
27. Что такое «тонкая настройка» (fine-tuning) предобученных моделей BERT (или других трансформеров)? Опишите процесс и его преимущества для решения различных задач NLP.
28. Обсудите основные подходы к построению диалоговых систем (на основе правил и машинного обучения). В чем их преимущества и недостатки?
29. Объясните различия между управляемыми (rule-based) и неуправляемыми (generative) типами диалогов. Приведите примеры их использования.
30. Опишите жизненный цикл разработки чат-бота, начиная от определения требований до развертывания и поддержки.

31. Объясните концепцию Понимания Естественного Языка (NLU) в контексте диалоговых систем. Что такое определение намерений (intent detection) и извлечение слотов (slot filling)?
32. Что такое Генерация Естественного Языка (NLG)? Сравните подходы на основе шаблонов и генеративных моделей в NLG.
33. Какова роль Управления Диалогом (Dialog Management) в диалоговых системах? Опишите его основные функции.
34. Как NLP используется для улучшения взаимодействия с пользователем и персонализации в диалоговых системах?
35. Перечислите и кратко охарактеризуйте популярные платформы и фреймворки для создания чат-ботов (например, Rasa, Botpress, Google Dialogflow).
36. Обсудите текущие тренды и будущее диалоговых систем, включая голосовых помощников и мультимодальные диалоги.
37. Объясните основные метрики оценки качества для задач классификации текста: точность (accuracy), полнота (precision), F1-мера, ROC-AUC. В каких случаях предпочтительнее использовать каждую из них?
38. Какие метрики используются для оценки качества моделей в других задачах NLP, таких как машинный перевод (BLEU) и суммаризация текста (ROUGE)? Кратко опишите их.
39. Какова важность анализа ошибок моделей? Опишите стратегии улучшения моделей на основе анализа ошибок.
40. Объясните концепцию API (Application Programming Interface) для моделей машинного обучения. Зачем нужно создавать API для NLP-моделей?
41. Перечислите и кратко охарактеризуйте инструменты для создания API для Python-приложений (например, Flask, FastAPI). В чем их основные различия?
42. Опишите основные принципы деплоя моделей машинного обучения. Объясните роль докеризации и использования облачных платформ (например, AWS, GCP, Azure) в этом процессе.

Критерии оценки

Один вопрос оценивается максимум в 50 баллов.

38-50 баллов, если: изученный материал изложен полно, определения даны верно; ответ показывает понимание материала; обучающийся может обосновать свои суждения, применить знания на практике, привести необходимые примеры, не только по учебнику и конспекту, но и самостоятельно составленные.

24-37 баллов, если: изученный материал изложен достаточно полно; при ответе допускаются ошибки, заминки, которые обучающийся в состоянии исправить самостоятельно при наводящих вопросах; обучающийся затрудняется с ответами на 1-2 дополнительных вопроса.

10-23 баллов, если: материал изложен неполно, с неточностями в определении понятий или формулировке определений; материал излагается непоследовательно; обучающийся не может достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; на 50% дополнительных вопросов даны неверные ответы.

0 баллов, если: при ответе обнаруживается полное незнание и непонимание изучаемого материала; материал излагается неуверенно, беспорядочно; даны неверные ответы более чем на 50% дополнительных вопросов.

Интерактивные упражнения

Выполнение интерактивных упражнений осуществляется на платформе Stepik по курсу «Нейронные сети для обработки текста».

Содержание курса:

1. Введение (в обработку естественного языка)
2. Векторная модель текста и классификация длинных текстов

3. Базовые нейросетевые методы работы с текстами
4. Языковые модели и генерация текста
5. Преобразование последовательностей: 1-к-1 и N-к-M
6. Transfer learning, адаптация моделей
7. Финальное соревнование на kaggle и заключение

Критерии оценки

- от 22 до 25 баллов выставляется студенту, если он выполнил без ошибок и недочетов 85-100% интерактивных упражнений по онлайн-курсу;
- от 18 до 21 баллов выставляется студенту, если он выполнил без ошибок и недочетов 71-84% интерактивных упражнений по онлайн-курсу;
- от 15 до 17 баллов выставляется студенту, если он выполнил без ошибок и недочетов 60-70% интерактивных упражнений по онлайн-курсу;
- от 12 до 14 баллов выставляется студенту, если он выполнил без ошибок и недочетов 50-59% интерактивных упражнений по онлайн-курсу;
- от 0 до 11 баллов выставляется студенту при невыполнении (не полном выполнении) требований, предусмотренных выше.

Проектные работы

Проектная работа №1 «Предварительная обработка текста на естественном языке»

Цель: Освоить инструменты предварительной обработки текста на естественном языке.

Инструментарий: Проектная работа № 1 (Часть 2) выполняется в среде Jupyter Notebook (Google Colab или локально) с использованием языка Python 3.x.

Обеспечение:

Пример №1.pdf

Задание №1.docx

Задание №2.ipynb

sample.csv

Задание:

Часть 1: на основе примера №1 выполните обработку собственного корпуса текста вручную, выполните задание в рабочей тетради «Задание №1.docx», загрузите результат в виде документа PDF.

Часть 2: на основе примера работы с датасетом Twitter выполните обработку собственного корпуса текста на языке Python, вставляя недостающие фрагменты кода, выполните задание в Блокноте Jupyter «Задание №2.ipynb», сравните размер словаря до и после обработки, загрузите результат работы в Блокноте Jupyter.

Критерии оценки:

от 18 до 20 баллов выставляется студенту, если он своевременно выполнил все задания, предусмотренные в проектной работе и в процессе защиты продемонстрировал полноту теоретических знаний в объеме содержания учебной дисциплины, относящейся к проектной работе. Сумел ответить на дополнительные вопросы, связанные не только с процессом выполнения проектной работы, но и с пониманием совершенных действий и решенных задач;

от 15 до 17 баллов выставляется студенту, если он выполнил все задачи, предусмотренные в проектной работе и в процессе защиты продемонстрировал наличие достаточных теоретических знаний в объеме содержания учебной дисциплины, относящейся к проектной работе. Сумел ответить на вопросы, связанные с процессом выполнения проектной работы;

от 12 до 14 баллов выставляется студенту, если он более чем на половину выполнил поставленные в проектной работе задачи, способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к проектной работе.

от 0 до 11 баллов выставляется студенту, при невыполнении требований, предусмотренных выше.

Проектная работа №2 «Классификация текста»

Цель: Научиться адаптировать и дорабатывать существующие архитектуры моделей для решения задач классификации текста, а также оптимизировать пайплайны обработки данных и обучения.

Инструментарий: Проектная работа № 2 выполняется в среде Jupyter Notebook (Google Colab или локально) с использованием языка Python 3.x.

Обеспечение:

Пример №2.pdf

review_polarity_.tar.gz

Задание:

на основе примера №2 выполните классификацию предложенного набора текстовых данных, примените другой алгоритм машинного обучения для улучшения производительности, загрузите результат работы в Блокноте Jupyter;

используя полученные навыки, выполните классификацию собственного набора текстовых данных, загрузите собственный набор текстовых данных и результат работы в Блокноте Jupyter.

Критерии оценки:

от 18 до 20 баллов выставляется студенту, если он своевременно выполнил все задания, предусмотренные в проектной работе и в процессе защиты продемонстрировал полноту теоретических знаний в объеме содержания учебной дисциплины, относящейся к проектной работе. Сумел ответить на дополнительные вопросы, связанные не только с процессом выполнения проектной работы, но и с пониманием совершенных действий и решенных задач;

от 15 до 17 баллов выставляется студенту, если он выполнил все задачи, предусмотренные в проектной работе и в процессе защиты продемонстрировал наличие достаточных теоретических знаний в объеме содержания учебной дисциплины, относящейся к проектной работе. Сумел ответить на вопросы, связанные с процессом выполнения проектной работы;

от 12 до 14 баллов выставляется студенту, если он более чем на половину выполнил поставленные в проектной работе задачи, способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к проектной работе.

от 0 до 11 баллов выставляется студенту, при невыполнении требований, предусмотренных выше.

Проектная работа №3 «Создание чат-бота»

Цель: Применить стек технологий, методов и алгоритмов для создания функционального прототипа чат-бота.

Инструментарий: Проектная работа № 3 выполняется в среде Jupyter Notebook (Google Colab или локально) с использованием языка Python 3.x.

Обеспечение:

Cosine simiiarity.ipynb

chatbot.txt

Задание:

создайте чат-бот, выполните задание в Блокноте Jupyter «Cosine simiiarity.ipynb», в поле «Ответ в виде текста» разместите фрагменты двух диалогов с чат-ботом: 1) на исходном корпусе текста «chatbot.txt» после добавления операции стемминга; 2) на собственном корпусе текста, в поле «Ответ в виде файла» загрузите итоговую рабочую тетрадь и текстовый файл с собственным корпусом текста (двумя файлами ipynb и txt, без архива).

Критерии оценки:

от 18 до 20 баллов выставляется студенту, если он своевременно выполнил все задания, предусмотренные в проектной работе и в процессе защиты продемонстрировал полноту теоретических знаний в объеме содержания учебной дисциплины, относящейся к проектной работе. Сумел ответить на дополнительные вопросы, связанные не только с процессом выполнения проектной работы, но и с пониманием совершенных действий и решенных задач;

от 15 до 17 баллов выставляется студенту, если он выполнил все задачи, предусмотренные в проектной работе и в процессе защиты продемонстрировал наличие достаточных теоретических знаний в объеме содержания учебной дисциплины, относящейся к проектной работе. Сумел ответить на вопросы, связанные с процессом выполнения проектной работы;

от 12 до 14 баллов выставляется студенту, если он более чем на половину выполнил поставленные в проектной работе задачи, способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к проектной работе.

от 0 до 11 баллов выставляется студенту, при невыполнении требований, предусмотренных выше.

Вопросы для собеседования

1. Какие этапы включает предварительная обработка текстовых данных? Объясните важность каждого этапа.
2. Что такое векторные представления слов (Word Embeddings)? Объясните концепцию Word2Vec и GloVe и в чем их преимущество перед традиционными методами векторизации.
3. Перечислите и кратко охарактеризуйте основные алгоритмы машинного обучения, используемые для классификации текста (Naive Bayes, SVM, Logistic Regression, Random Forest). В каких случаях предпочтительнее использовать каждый из них?
4. Опишите пошаговый процесс построения простого классификатора текста (от предобработки до оценки).
5. Опишите жизненный цикл разработки чат-бота, начиная от определения требований до развертывания и поддержки.
6. Что такое «шум» в текстовых данных и как его можно устранить? Опишите методы нормализации регистра, удаления пунктуации и чисел.
7. Объясните процесс токенизации. Какие существуют подходы к токенизации (например, по словам, по предложениям)?
8. Что такое стоп-слова и зачем их удаляют? Приведите примеры стоп-слов и объясните, когда их удаление может быть нежелательным.
9. Сравните стемминг и лемматизацию. В чем заключаются их основные отличия, преимущества и недостатки?
10. Что такое анализ тональности (Sentiment Analysis)? Опишите основные задачи и приведите примеры его применения.
11. Объясните архитектуру Трансформеров. Какова роль механизма внимания (Attention) в этой архитектуре и почему он так важен?
12. Обсудите основные подходы к построению диалоговых систем (на основе правил и машинного обучения). В чем их преимущества и недостатки?
13. Объясните различия между управляемыми (rule-based) и неуправляемыми (generative) типами диалогов. Приведите примеры их использования.
14. Объясните основные метрики оценки качества для задач классификации текста: точность (accuracy), полнота (precision), F1-мера, ROC-AUC. В каких случаях предпочтительнее использовать каждую из них?
15. Какие метрики используются для оценки качества моделей в других задачах NLP, таких как машинный перевод (BLEU) и суммаризация текста (ROUGE)? Кратко опишите их.

Критерии оценки

Максимальное количество баллов – 15 (каждый вопрос оценивается в 1 балл).

1 - продемонстрированы знание вопроса и самостоятельность мышления, ответ соответствует требованиям правильности, полноты и аргументированности; 0 - ответ на вопрос неконкретен, слабо аргументирован и не убедителен

3 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Процедуры оценивания включают в себя текущий контроль и промежуточную аттестацию.

Текущий контроль успеваемости проводится с использованием оценочных средств, представленных в п. 2 данного приложения. Результаты текущего контроля доводятся до сведения студентов до промежуточной аттестации.

Промежуточная аттестация проводится в форме экзамена.

Экзамен проводится по расписанию промежуточной аттестации в устном виде. Количество вопросов в экзаменационном задании – 2. Объявление результатов производится в день экзамена. Результаты аттестации заносятся в экзаменационную ведомость и зачетную книжку студента. Студенты, не прошедшие промежуточную аттестацию по графику сессии, должны ликвидировать задолженность в установленном порядке.

Приложение 2
к рабочей программе

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Учебным планом предусмотрены только лабораторные занятия.

Лабораторные занятия направлены на получение навыков работы в программе «1С: Бухгалтерия предприятия». В процессе подготовки к выполнению лабораторных работ необходимо изучить рекомендованный преподавателем материал из основной и дополнительной литературы.

Основным методом приобретения знаний в высшей школе является самостоятельная работа студентов. Контроль самостоятельной работы студентов над учебной программой курса осуществляется в ходе занятий посредством тестирования. В ходе самостоятельной работы каждый студент обязан прочитать основную и по возможности дополнительную литературу.

Контроль знаний проводится во время итоговой аттестации (зачет) в виде устных ответов студентов на вопросы.

Для подготовки к занятиям, текущему контролю и промежуточной аттестации студенты могут воспользоваться электронно-библиотечными системами.