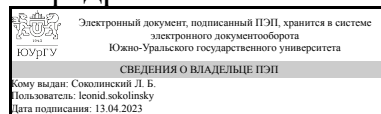


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Заведующий выпускающей
кафедрой



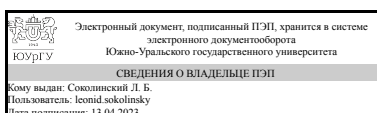
Л. Б. Соколинский

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.П0.06 Введение в обработку естественного языка
для направления 09.03.04 Программная инженерия
уровень Бакалавриат
профиль подготовки Инженерия информационных и интеллектуальных систем
форма обучения очная
кафедра-разработчик Системное программирование

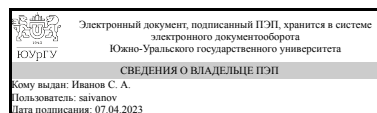
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.03.04 Программная инженерия, утверждённым приказом Минобрнауки от 19.09.2017 № 920

Зав.кафедрой разработчика,
д.физ.-мат.н., проф.



Л. Б. Соколинский

Разработчик программы,
к.физ.-мат.н., доцент



С. А. Иванов

1. Цели и задачи дисциплины

Целью дисциплины является формирование базовых представлений, знаний и умений в анализе естественного языка.

Краткое содержание дисциплины

Основные задачи дисциплины: ознакомить студента с основными понятиями анализа и обработки текстов на естественном языке, дать понимание базовых подходов и методов при решении задач анализа естественного языка, получить практический опыт работы с различными алгоритмами машинного обучения и архитектурами искусственных нейронных сетей в рамках задач обработки естественного языка.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-9 (ПК-6 модели) Способен создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов	Знает: архитектуры нейронных сетей для задач обработки естественного языка Умеет: ПК-6.1. У-2. Умеет применять современные инструментальные средства и системы программирования для разработки и обучения моделей искусственных нейронных сетей; Имеет практический опыт: создания и обучения нейросетевых моделей для решения задач в области обработки естественного языка
ПК-10 (ПК-9 модели) Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	Знает: ПК-9.2. 3-1. Знает принципы построения систем обработки естественного языка, методы и технологии искусственного интеллекта для анализа естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»; ПК-9.4. 3-1. Знает принципы построения систем распознавания и синтеза речи, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Распознавание и синтез речи»; Умеет: ПК-9.2. У-1. Умеет применять методы и подходы к планированию и реализации проектов по созданию и поддержке системы искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»; ПК-9.4. У-1. Умеет применять методы и подходы к планированию и реализации проектов по созданию и поддержке системы искусственного интеллекта на основе сквозной цифровой субтехнологии «Распознавание и синтез речи»;

	Имеет практический опыт: создания и обучения модели искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»
--	---

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Глубокое обучение, Подготовка данных для машинного обучения	Не предусмотрены

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Нет

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 4 з.е., 144 ч., 70,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		8
Общая трудоёмкость дисциплины	144	144
<i>Аудиторные занятия:</i>	60	60
Лекции (Л)	12	12
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	48	48
Лабораторные работы (ЛР)	0	0
<i>Самостоятельная работа (СРС)</i>	73,5	73,5
Подготовка к практическим занятиям	42,5	42,5
Подготовка к экзамену	31	31
Консультации и промежуточная аттестация	10,5	10,5
Вид контроля (зачет, диф.зачет, экзамен)	-	экзамен

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Введение в обработку естественного языка	8	2	6	0
2	Машинное обучение и глубокие нейронные сети для решения задач анализа и обработки естественного языка	52	10	42	0

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Введение в обработку естественного языка (NLP). Основные задачи NLP. Представления текстовых данных. Предобработка текста, лемматизация, стемминг.	2
2	2	Методы машинного обучения для классификации текстовых документов на основе частотных мер (TF-IDF). Деревья решений, наивный байесовский классификатор, логистическая регрессия в задаче классификации текстов.	1
3	2	Языковые модели. Word embeddings. Нейросетевые модели языка: word2vec, fasttext. Мера семантической близости. Классификация текстов на основе нейросетевых моделей языка.	1
4	2	Кластеризация текстовых документов. Тематическое моделирование Методы LSA, pLSA. Аддитивная регуляризация тематических моделей в BigARTM	2
5	2	Классификация текстов с помощью глубоких нейронных сетей: CNN, LSTM.	2
6	2	Задачи обработки последовательностей: машинный перевод, автоматическое реферирование (summarization), вопросно-ответные системы. Механизм внимания (attention). Архитектуры encoder- decoder-attention.	2
7	2	Transfer learning в задачах анализа текстов. Self-Attention. Архитектуры трансформеров: BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерации текстов. Fine- tuning трансформеров.	2

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1	1	Введение в анализ естественного языка. Представления текстовых данных. Предобработка текстовых данных. Лемматизация. Морфологический анализ. Реализация собственного POS-тэггера.	6
2	2	Классификация текстов методами машинного обучения. на основе частотных мер (TF-IDF). Деревья решений, наивный байесовский классификатор, логистическая регрессия в задаче классификации текстов.	6
3	2	Нейросетевые модели языка: word2vec, fasttext. Задача определения семантической близости.	6
4	2	Классификация текстов на основе нейросетевых моделей языка.	6
5	2	Задача кластеризации текстовой коллекции с применение методов pLSA и фреймворка BigARTM.	6
6	2	Классификация текстов с помощью различных архитектур глубоких нейронных сетей: CNN, LSTM.	6
7	2	Реализация вопросно-ответной системы на основе нейронных сетей encoder-decoder с механизмом внимания. Решение задачи автоматического реферирования (text summarization).	6
8	2	Архитектуры трансформеров: BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерации текстов.	6

5.3. Лабораторные работы

Не предусмотрены

5.4. Самостоятельная работа студента

Выполнение СРС

Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Подготовка к практическим занятиям	Гольдберг Й. Нейросетевые методы в обработке естественного языка (https://e.lanbook.com/book/131704) Москва : ДМК Пресс, 2019 Паттерсон Д., Гибсон А. Глубокое обучение с точки зрения практика (https://e.lanbook.com/book/116122) Москва : ДМК Пресс, 2018	8	42,5
Подготовка к экзамену	Гольдберг Й. Нейросетевые методы в обработке естественного языка (https://e.lanbook.com/book/131704) Москва : ДМК Пресс, 2019 Паттерсон Д., Гибсон А. Глубокое обучение с точки зрения практика (https://e.lanbook.com/book/116122) Москва : ДМК Пресс, 2018	8	31

6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учитывается в ПА
1	8	Текущий контроль	Проверка практического задания 1	1	1	Для прохождения собеседования по практической работе обучающийся предоставляет отчет. В процессе собеседования обучающийся должен свободно ориентироваться в отчете, сопоставлять теорию и полученные практические результаты, отвечать на дополнительные вопросы. При успешном прохождении собеседования обучающийся получает 1 балл	экзамен
2	8	Текущий контроль	Проверка практического задания 2	1	1	Для прохождения собеседования по практической работе обучающийся предоставляет отчет. В процессе собеседования обучающийся должен свободно ориентироваться в отчете, сопоставлять теорию и полученные практические результаты, отвечать на дополнительные вопросы. При успешном прохождении собеседования обучающийся получает 1 балл	экзамен
3	8	Текущий	Проверка	1	1	Для прохождения собеседования по	экзамен

		контроль	практического задания 3			практической работе обучающийся предоставляет отчет. В процессе собеседования обучающийся должен свободно ориентироваться в отчете, сопоставлять теорию и полученные практические результаты, отвечать на дополнительные вопросы. При успешном прохождении собеседования обучающийся получает 1 балл	
4	8	Текущий контроль	Проверка практического задания 4	1	1	Для прохождения собеседования по практической работе обучающийся предоставляет отчет. В процессе собеседования обучающийся должен свободно ориентироваться в отчете, сопоставлять теорию и полученные практические результаты, отвечать на дополнительные вопросы. При успешном прохождении собеседования обучающийся получает 1 балл	экзамен
5	8	Текущий контроль	Проверка практического задания 5	1	1	Для прохождения собеседования по практической работе обучающийся предоставляет разработанный код приложения. Представленный код должен выполнять поставленную задачу, его оформление должно соответствовать общепринятым соглашениям. В процессе собеседования обучающийся должен свободно ориентироваться в коде программы, сопоставлять теорию и полученные практические результаты, отвечать на дополнительные вопросы. При успешном прохождении собеседования обучающийся получает 1 балл	экзамен
6	8	Текущий контроль	Проверка практического задания 6	1	1	Для прохождения собеседования по практической работе обучающийся предоставляет разработанный код приложения. Представленный код должен выполнять поставленную задачу, его оформление должно соответствовать общепринятым соглашениям. В процессе собеседования обучающийся должен свободно ориентироваться в коде программы, сопоставлять теорию и полученные практические результаты, отвечать на дополнительные вопросы. При успешном прохождении собеседования обучающийся получает 1 балл	экзамен
7	8	Текущий контроль	Проверка практического задания 7	1	1	Для прохождения собеседования по практической работе обучающийся предоставляет разработанный код приложения. Представленный код должен выполнять поставленную задачу, его оформление должно соответствовать	экзамен

						общепринятым соглашениям. В процессе собеседования обучающийся должен свободно ориентироваться в коде программы, сопоставлять теорию и полученные практические результаты, отвечать на дополнительные вопросы. При успешном прохождении собеседования обучающийся получает 1 балл	
8	8	Промежуточная аттестация	Письменный экзамен	-	5	Задание для экзамена включает в себя ответ на два вопроса из списка. При оценивании результатов мероприятия используется балльнорейтинговая система оценивания результатов учебной деятельности обучающихся. Письменная работа позволяет оценить сформированность компетенции. Неправильный ответ на вопрос соответствует 0 баллов. Критерии оценивания контрольного мероприятия: наличие в ответе информации по теме — 0,5 баллов за вопрос; оформлено читаемо — 0,5 баллов за вопрос; присутствует логика изложения материала по теме — 0,5 баллов за вопрос; в целом верная информация — 0,5 баллов за вопрос; полнота раскрытия темы — 0,5 баллов за вопрос.	экзамен
9	8	Текущий контроль	Тест 1	1	3	Тест проводится на лекции. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 10 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	экзамен
10	8	Текущий контроль	Тест 2	1	3	Тест проводится на лекции. Каждый правильный ответ оценивается 1 баллом. Продолжительность тестирования 10 мин. 3 балла: даны верные ответы на все вопросы теста. 2 балла: даны верные ответы на 2 вопроса теста. 1 балл: дан верный ответ на 1 вопрос теста. 0 баллов: верные ответы отсутствуют	экзамен

6.2. Процедура проведения, критерии оценивания

Вид промежуточной	Процедура проведения	Критерии оценивания
-------------------	----------------------	---------------------

		система издательства Лань	2018
4	Дополнительная литература	Электронно- библиотечная система издательства Лань	Паттерсон Д., Гибсон А. Глубокое обучение с точки зрения практика (https://e.lanbook.com/book/116122) Москва : ДМК Пресс, 2018
5	Дополнительная литература	Электронно- библиотечная система издательства Лань	Козьмо Л. П., Ричарт В. Построение систем машинного обучения на языке Python (http://e.lanbook.com/books/element.php?pl1_id=82818) Москва : ДМК Пресс, 2016

Перечень используемого программного обеспечения:

1. -Python(бессрочно)

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Практические занятия и семинары	110 (3г)	Компьютерный класс, оборудованный персональными компьютерами с выходом в интернет, Платформа обработки данных на основе искусственного интеллекта https://aiplatform.ru/ ML Space: российская платформа для ML-разработки полного цикла https://sbercloud.ru/ru/aicloud/mlspace
Лекции	434 (3б)	Персональный компьютер, проектор
Экзамен	110 (3г)	Компьютерный класс, оборудованный персональными компьютерами с выходом в интернет