ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Заведующий выпускающей кафедрой

Электронный документ, подписанный ПЭП, хранится в системе электронного документооборота Южно-Уральского государственного университета
СВЕДЕНИЯ О ВЛАДЕЛЬЦЕ ПЭП
Кому выдан: Соколинский Л. Б.
Пользователь: leonid.sokolinsky
Дата подписания: 15.05.2023

Л. Б. Соколинский

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.М0.03 Анализ естественного языка методами искусственного интеллекта
для направления 09.04.04 Программная инженерия
уровень Магистратура
магистерская программа Искусственный интеллект и инженерия данных
форма обучения очная
кафедра-разработчик Системное программирование

Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 09.04.04 Программная инженерия, утверждённым приказом Минобрнауки от 19.09.2017 № 932

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

Электронный документ, подписанный ПЭП, хранится в системе электронного документооборота Южно-Уральского государственного университета
СВЕДЕНИЯ О ВЛАДЕЛЬЦЕ ПЭП
Кому выдан: Соколинский Л. Б.
Пользователь: leonid.sokolinsky
Дата подписания: 15.05.2023

Л. Б. Соколинский

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

Электронный документ, подписанный ПЭП, хранится в системе электронного документооборота Южно-Уральского государственного университета
СВЕДЕНИЯ О ВЛАДЕЛЬЦЕ ПЭП
Кому выдан: Иванов С. А.
Пользователь: saivanov
Дата подписания: 13.05.2023

С. А. Иванов

Челябинск

1. Цели и задачи дисциплины

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

Краткое содержание дисциплины

Изложены наиболее важные понятия, определения и методы машинного обучения и искусственных нейронных сетей в задачах анализа естественного языка. В курс входят следующие разделы: введение в анализ естественного языка, машинное обучение и глубокие нейронные сети для решения задач анализа и обработки естественного языка, построение диалоговых систем. На практике студенты применят навыки построения моделей машинного обучения и искусственных нейронных сетей на языке Python для решения задач морфологического анализа, классификации и кластеризации текстовых документов, анализа тональности, определения семантической близости слов, машинного перевода, построения вопросно-ответных систем, автоматического реферирования текста, построения диалоговых систем.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-3 Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач	Знает: унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий
ПК-5 Способен руководить проектами по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов	Знает: функциональность современных инструментальных средств и систем программирования в области создания моделей искусственных нейронных сетей Умеет: проводить оценку и выбор моделей искусственных нейронных сетей и инструментальных средств для решения задач машинного обучения
ПК-6 Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных цифровых субтехнологий искусственного интеллекта в прикладных областях	Знает: принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем 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

Разработка интеллектуальных систем на языке R, Глубокие нейронные сети, Машинное обучение, Разработка систем искусственного интеллекта на языке Python	Нейробайесовские методы в машинном обучении, Нейросетевые технологии в задачах синтетических медиа, Управление проектами в сфере искусственного интеллекта, Обработка звуковых сигналов методами искусственного интеллекта

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

Дисциплина	Требования
Машинное обучение	Знает: классы методов и алгоритмов машинного обучения, фундаментальные научные принципы и методы исследований, функциональность современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения Умеет: ставить задачи и разрабатывать новые методы и алгоритмы машинного обучения, адаптировать с целью практического применения фундаментальные и новые научные принципы и методы исследований Имеет практический опыт: решать основные классы задач методами и алгоритмами машинного обучения, применения классических методов исследования для решения профессиональных задач
Разработка систем искусственного интеллекта на языке Python	Знает: основные инструменты языка Python для сбора данных, необходимых для разработки программного обеспечение с применением алгоритмов машинного обучения, основные веб-фреймворки на Python, подходы многопоточного и асинхронного программирования, принципы промышленной разработки интеллектуальных систем на языке Python Умеет: подбирать наиболее подходящие инструменты сбора, анализа, обработки и визуализации данных в Python, применять конкретные специализированные фреймворки языка Python для сбора, обработки и анализа данных для решения различных задач анализа данных, подбирать наиболее подходящие фреймворки и библиотеки для разработки веб-сервисов сбора, анализа и обработки данных Имеет практический опыт: сбора данных в различных форматах; предварительной обработки данных (приведение типов/форматов, заполнение пропусков фильтрация и т.п.); анализа и визуализации данных, анализа готовых информационных наборов данных; разработки и развертывания разработанного программного обеспечения для сбора и анализа данных в условиях решения реальных задач, разработки

	оригинальных программных сервисов сбора, анализа и обработки данных на Python
Разработка интеллектуальных систем на языке R	Знает: методы и критерии оценки качества моделей машинного обучения Умеет: выбирать, применять и интегрировать методы и инструментальные средства систем искусственного интеллекта, критерии их выбора и методы комплексирования в рамках создания интегрированных гибридных интеллектуальных систем различного назначения Имеет практический опыт:
Глубокие нейронные сети	Знает: математическую модель нейрона, технологии создания искусственных нейронных сетей, методы оптимизации, регуляризации и нормализации параметров нейронной сети и процесса ее обучения, принципы построения моделей глубоких нейронных сетей и глубокого машинного обучения (с подкреплением и без) Умеет: осуществлять формализацию задачи, построение математической модели, подготовку обучающего набора данных, подбор топологии и создание искусственной нейронной сети в соответствии с поставленной задачей, применять современные инструментальные средства и системы программирования для разработки и обучения моделей искусственных нейронных сетей Имеет практический опыт: формулирования и решения задач в области машинного обучения с использованием нейросетевого подхода

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 4 з.е., 144 ч., 58,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		2
Общая трудоёмкость дисциплины	144	144
Аудиторные занятия:	48	48
Лекции (Л)	16	16
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	32	32
Лабораторные работы (ЛР)	0	0
Самостоятельная работа (СРС)	85,5	85,5
Изучение основной и дополнительной литературы по анализу и обработке естественного языка	53,5	53.5
Подготовка к экзамену	32	32
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	6	2	4	0
2	Машинное обучение и глубокие нейронные сети для решения задач анализа и обработки естественного языка	36	12	24	0
3	Построение диалоговых систем	6	2	4	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.	2
3	2	Языковые модели. Word embeddings. Нейросетевые модели языка: word2vec, fasttext. Мера семантической близости. Классификация текстов на основе нейросетевых моделей языка.	2
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: машинной 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
8	3	Построение диалоговых систем. Архитектура диалоговых систем. Модули понимания естественного языка (NLU) и диалоговый менеджер (DM). Сложности построения диалоговых систем. Проектирование UX/UI диалоговых ассистентов в чатах и голосе. Обзор современных фреймворков для построения диалоговых систем: DeepPavlov, Rasa, Just AI Conversational Platform	2

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1-2	1	Построение диалоговых систем. Архитектура диалоговых систем. Модули понимания естественного языка (NLU) и диалоговый менеджер (DM). Сложности построения диалоговых систем. Проектирование UX/UI диалоговых ассистентов в чатах и голосе. Обзор современных фреймворков для построения диалоговых систем: DeepPavlov, Rasa, Just AI Conversational Platform	4
3-4	2	Классификация текстов методами машинного обучения. на основе частотных мер (TF-IDF). Деревья решений, наивный байесовский классификатор,	4

		логистическая регрессия в задаче классификации текстов.	
5-6	2	Нейросетевые модели языка: word2vec, fasttext. Задача определения семантической близости. Классификация текстов на основе нейросетевых моделей языка.	4
7-8	2	Задача кластеризации текстовой коллекции с применение методов pLSA и фреймворка BigARTM.	4
9-10	2	Классификация текстов с помощью различных архитектур глубоких нейронных сетей: CNN, LSTM.	4
11-12	2	Реализация вопросно-ответной системы на основе нейронных сетей encoder-decoder с механизмом внимания. Решение задачи автоматического реферирования (text summarization).	4
13-14	2	Архитектуры трансформеров: BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерации текстов.	4
15-16	3	Реализация модулей NLU и DM для чат-бота на основе одного из фреймворков диалоговых систем.	4

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
Изучение основной и дополнительной литературы по анализу и обработке естественного языка	Основная литература 1, 2. Дополнительная литература 1-3.	2	53,5
Подготовка к экзамену	Основная литература 1, 2. Дополнительная литература 1-3	2	32

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	Учи-тыва-ется в ПА
1	2	Текущий контроль	ПЗ-1. Реализация собственного POS-тэггера.	4	3	3 балла: задание выполнено полностью, 2 балла: задание выполнено полностью, но допущены незначительные ошибки, или задание выполнено более, чем 50%, 1 балла: задание выполнено полностью, но допущены серьезные	экзамен

						ошибки, или задание выполнено менее, чем 50%, 0 баллов: задание не выполнено	
2	2	Текущий контроль	ПЗ-2. Классификация текстов методами машинного обучения	3	3	3 балла: задание выполнено полностью, 2 балла: задание выполнено полностью, но допущены незначительные ошибки, или задание выполнено более, чем 50%, 1 балла: задание выполнено полностью, но допущены серьезные ошибки, или задание выполнено менее, чем 50%, 0 баллов: задание не выполнено	экзамен
3	2	Текущий контроль	ПЗ-3. Классификация текстов на основе нейросетевых моделей языка	4	3	3 балла: задание выполнено полностью, 2 балла: задание выполнено полностью, но допущены незначительные ошибки, или задание выполнено более, чем 50%, 1 балла: задание выполнено полностью, но допущены серьезные ошибки, или задание выполнено менее, чем 50%, 0 баллов: задание не выполнено	экзамен
4	2	Текущий контроль	ПЗ-4. Кластеризация текстовой коллекции методами тематического моделирования	4	3	3 балла: задание выполнено полностью, 2 балла: задание выполнено полностью, но допущены незначительные ошибки, или задание выполнено более, чем 50%, 1 балла: задание выполнено полностью, но допущены серьезные ошибки, или задание выполнено менее, чем 50%, 0 баллов: задание не выполнено	экзамен
5	2	Текущий контроль	ПЗ-5. Классификация текстов с помощью различных архитектур глубоких нейронных сетей	5	3	3 балла: задание выполнено полностью, 2 балла: задание выполнено полностью, но допущены незначительные ошибки, или задание выполнено более, чем 50%, 1 балла: задание выполнено полностью, но допущены серьезные ошибки, или задание выполнено менее, чем 50%, 0 баллов: задание не выполнено	экзамен
6	2	Текущий контроль	ПЗ-6. Реализация вопросно-ответной системы. Решение задачи автоматического реферирования	6	3	3 балла: задание выполнено полностью, 2 балла: задание выполнено полностью, но допущены незначительные ошибки, или задание выполнено более, чем 50%, 1 балла: задание выполнено полностью, но допущены серьезные	экзамен

						ошибки, или задание выполнено менее, чем 50%, 0 баллов: задание не выполнено	
7	2	Текущий контроль	ПЗ-7. Архитектуры трансформеров: BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерации текстов.	5	3	3 балла: задание выполнено полностью, 2 балла: задание выполнено полностью, но допущены незначительные ошибки, или задание выполнено более, чем 50%, 1 балла: задание выполнено полностью, но допущены серьезные ошибки, или задание выполнено менее, чем 50%, 0 баллов: задание не выполнено	экзамен
8	2	Текущий контроль	ПЗ-8. Разработка диалогового агента для чат-бота	8	3	3 балла: задание выполнено полностью, 2 балла: задание выполнено полностью, но допущены незначительные ошибки, или задание выполнено более, чем 50%, 1 балла: задание выполнено полностью, но допущены серьезные ошибки, или задание выполнено менее, чем 50%, 0 баллов: задание не выполнено	экзамен
9	2	Текущий контроль	Введение в анализ естественного языка. Представления текстовых данных. Предобработка.	5	5	Компьютерный тест состоит из 5 вопросов, позволяющих оценить сформированность компетенций. На ответы отводится 7 мин. Стоимость одного вопроса - 1 балл.	экзамен
10	2	Текущий контроль	Трансферлернинг. Обучение модели BERT. Классификация документов.	5	5	Компьютерный тест состоит из 5 вопросов, позволяющих оценить сформированность компетенций. На ответы отводится 7 мин. Стоимость одного вопроса - 1 балл.	экзамен
11	2	Проме-жуточная аттестация	Итоговый тест	-	20	Компьютерный тест состоит из 20 вопросов, позволяющих оценить сформированность компетенций. На ответы отводится 1 час. 20 баллов: задание полностью выполнено без ошибок 1-19 баллов: задание выполнено частично или выполнено с ошибками 0 баллов: задание не выполнено	экзамен

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
экзамен	При оценивании результатов учебной деятельности обучающегося по дисциплине используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (Положение о БРС утверждено	В соответствии с пп. 2.5, 2.6 Положения

	приказом ректора от 24.05.2019 г. № 179, в редакции приказа ректора от 10.03.2022 г. No 25-13/09). Оценка за дисциплину формируется на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля. Отлично: Величина рейтинга обучающегося по дисциплине 85…100 %. Хорошо: Величина рейтинга обучающегося по дисциплине 75…84 %.Удовлетворительно: Величина рейтинга обучающегося по дисциплине 60…74 %. Неудовлетворительно: Величина рейтинга обучающегося по дисциплине 0…59 %. Если студент не согласен с оценкой, полученной по результатам текущего контроля, студент проходит мероприятие промежуточной аттестации в виде тестирования. Тестирование проводится в системе edu.susu.ru. Тест содержит 20 вопросов. На выполнение теста дается 20 минут. В этом случае оценка за дисциплину рассчитывается на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и промежуточной аттестации. Фиксация результатов учебной деятельности по дисциплине проводится в день экзамена при личном присутствии студента.	

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ									
		1	2	3	4	5	6	7	8	9	10	11
ПК-3	Знает: унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий	+	+	+	+	+	+	+	+	+	+	+
ПК-5	Знает: функциональность современных инструментальных средств и систем программирования в области создания моделей искусственных нейронных сетей	+	+	+	+	+	+	+	+	+	+	+
ПК-5	Умеет: проводить оценку и выбор моделей искусственных нейронных сетей и инструментальных средств для решения задач машинного обучения	+	+	+	+	+	+	+	+	+	+	+
ПК-6	Знает: принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»	+	+	+	+	+	+	+	+	+	+	+

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

7. Учебно-методическое и информационное обеспечение дисциплины

Печатная учебно-методическая документация

а) основная литература:

Не предусмотрена

б) дополнительная литература:

Не предусмотрена

в) отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:

Не предусмотрены

г) методические указания для студентов по освоению дисциплины:

1. Гольдберг Й. Нейросетевые методы в обработке естественного языка (https://e.lanbook.com/book/131704) Москва : ДМК Пресс, 2019

из них: учебно-методическое обеспечение самостоятельной работы студента:

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Основная литература	Электронно-библиотечная система издательства Лань	Ганегедара, Т. Обработка естественного языка с TensorFlow : руководство / Т. Ганегедара ; перевод с английского В. С. Яценкова. — Москва : ДМК Пресс, 2020. — 382 с. https://e.lanbook.com/book/140584
2	Основная литература	Электронно-библиотечная система издательства Лань	Гольдберг, Й. Нейросетевые методы в обработке естественного языка : руководство / Й. Гольдберг ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 282 с. https://e.lanbook.com/book/131704
3	Основная литература	Электронно-библиотечная система издательства Лань	Антонио, Д. Библиотека Keras – инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотек Theano и TensorFlow / Д. Антонио, П. Суджит ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2018. — 294 с. https://e.lanbook.com/book/111438
4	Дополнительная литература	Электронно-библиотечная система издательства Лань	Паттерсон, Д. Глубокое обучение с точки зрения практика / Д. Паттерсон, А. Гибсон. — Москва : ДМК Пресс, 2018. — 418 с. https://e.lanbook.com/book/116122
5	Дополнительная литература	Электронно-библиотечная система издательства Лань	Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль ; перевод с английского А. А. Слинкина. — 2-е изд. — Москва : ДМК Пресс, 2018. — 652 с. https://e.lanbook.com/book/107901
6	Дополнительная литература	Электронно-библиотечная система издательства Лань	Коэльо, Л. П. Построение систем машинного обучения на языке Python / Л. П. Коэльо, В. Ричарт ; перевод с английского А. А. Слинкин. — 2-е изд. — Москва : ДМК Пресс, 2016. — 302 с. https://e.lanbook.com/book/82818

Перечень используемого программного обеспечения:

1. Python Software Foundation-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

Лекции	434 (3б)	Компьютер, проектор
Практические занятия и семинары	114-1 (2)	Компьютерный класс, имеется выход в Интернет