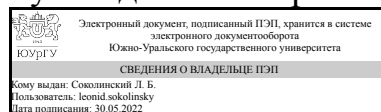


УТВЕРЖДАЮ:
Руководитель направления



Л. Б. Соколинский

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.02 Обработка естественного языка
для направления 02.04.02 Фундаментальная информатика и информационные
технологии

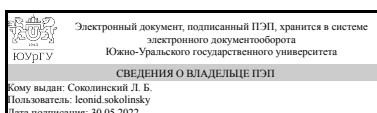
уровень Магистратура

форма обучения очная

кафедра-разработчик Системное программирование

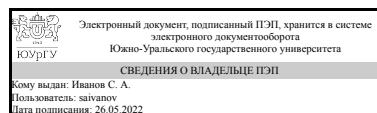
Рабочая программа составлена в соответствии с ФГОС ВО по направлению
подготовки 02.04.02 Фундаментальная информатика и информационные технологии,
утверждённым приказом Минобрнауки от 23.08.2017 № 811

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 обработке естественного языка. Основные задачи дисциплины: ознакомить студента с основными понятиями обработки текстов на естественном языке, дать описание базовых подходов и методов при решении задач обработки естественного языка, дать понимания работы различных типов искусственных нейронных сетей в рамках задач обработки текстов.

Краткое содержание дисциплины

Изложены наиболее важные понятия, определения и методы машинного обучения и искусственных нейронных сетей в задачах обработки естественного языка. В курс входят следующие разделы: введение в обработку естественного языка, машинное обучение для решения задач обработки естественного языка, искусственные нейронные сети для решения задач естественного языка. На практике студенты применяют навыки построения моделей машинного обучения и искусственных нейронных сетей на языке Python для решения задач обработки естественного языка.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-2 Способен разрабатывать системы хранения и обработки больших данных, в том числе на основе методов искусственного интеллекта	Знает: типовые решения, библиотеки программных модулей, шаблоны, классы объектов, используемые при разработке программного обеспечения для решения задач обработки естественного языка Умеет: применять типовые решения, библиотеки программных модулей, шаблоны, классы объектов при проектировании программного обеспечения Имеет практический опыт: проектирования и реализации приложений для решения задач обработки естественного языка с использованием методов машинного обучения и нейронных сетей

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
1.О.10 Нейронные сети	1.Ф.03 Анализ и прогнозирование временных рядов методами искусственного интеллекта

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

Дисциплина	Требования
------------	------------

1.О.10 Нейронные сети

Знает: технологии создания искусственных нейронных сетей с применением высокоуровневого языка программирования Python, математические основы и технологии создания и обучения искусственных нейронных сетей, математическую модель нейрона, основы линейной алгебры, технологии создания искусственных нейронных сетей, методы оптимизации, регуляризации и нормализации параметров нейронной сети и процесса ее обучения, методы оптимизации, регуляризации и нормализации параметров нейронной сети и процесса ее обучения. Умеет: создавать и обучать глубокие и сверточные искусственные нейронные сети на Python, применять современные методы машинного обучения на основе нейронных сетей, осуществлять формализацию задачи, построение математической модели, подготовку обучающего набора данных, подбор топологии и создание искусственной нейронной сети в соответствии с поставленной задачей, создавать глубокие и сверточные искусственные нейронные сети с применением специализированных библиотек на разных вычислительных платформах (CPU/GPU/TPU) Имеет практический опыт: анализа и оптимизации полученных решений на основе нейросетевого подхода, формулирования и решения задач в области машинного обучения с использованием нейросетевого подхода, решения задач распознавания образов на разных вычислительных платформах (CPU/GPU/TPU)

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 54,25 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		2
Общая трудоёмкость дисциплины	108	108
<i>Аудиторные занятия:</i>	48	48
Лекции (Л)	16	16
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	32	32
Лабораторные работы (ЛР)	0	0
<i>Самостоятельная работа (СРС)</i>	53,75	53,75
Изучение дополнительного материала по программированию на языке Python.	30	30
Изучение дополнительного материала по основам	18	18

линейной алгебры.		
Подготовка к зачету.	5,75	5.75
Консультации и промежуточная аттестация	6,25	6,25
Вид контроля (зачет, диф.зачет, экзамен)	-	зачет

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 для решения задач обработки естественного языка	12	4	8	0
3	Искусственные нейронные сети для решения задач естественного языка	30	10	20	0

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Введение в анализ естественного языка. Представления текстовых данных. Предобработка.	2
2-3	2	Методы машинного обучения для анализа естественного языка. Регрессионные модели. Деревья решений.	4
4-5	3	Искусственные нейронные сети для анализа текста. Слои LSTM. Одномерные свертки	4
6	3	Генерация символов и текста. Вопросно-ответный поиск.	2
7-8	3	Современные модели нейронных сетей для анализа естественного языка. Механизм внимания. Трансформеры. BERT, GPT.	4

5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1-2	1	Введение в анализ естественного языка. Представления текстовых данных. Предобработка.	4
3-4	2	Методы машинного обучения для анализа естественного языка. Регрессионные модели. Деревья решений. Классификация документов.	4
5-6	2	Методы машинного обучения для анализа естественного языка. Моделирование тем и кластеризация документов.	4
7-8	3	Искусственные нейронные сети для анализа текста. Слои LSTM. Классификация документов.	4
9-10	3	Искусственные нейронные сети для анализа текста. Слои LSTM. Генерация символов и слов.	4
11-12	3	Трансферлернинг. Обучение модели BERT. Классификация документов.	4
13-14	3	Трансферлернинг. Обучение моделей BERT. Задача регрессии.	4
15-16	3	Генерация символов и текста.	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 дополнительного материала по программированию на языке Python.	Бизли, Д. Python. Книга рецептов / Д. Бизли, Б. К. Джонс ; перевод с английского Б. В. Уварова. — Москва : ДМК Пресс, 2019. — 646 с. — ISBN 978-5-97060-751-0. — Текст : электронный // Лань : электронно-библиотечная система.	2	30
Изучение дополнительного материала по основам линейной алгебры.	Практикум по линейной и тензорной алгебре : учебное пособие / О. Н. Казакова, Т. А. Фомина, С. В. Харитоновна, А. Р. Рустанов. — Оренбург : ОГУ, 2017. — 117 с. — Текст : электронный // Лань : электронно-библиотечная система.	2	18
Подготовка к зачету.	Основная литература 1-3. Дополнительная литература 1-3.	2	5,75

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	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.	зачет
2	2	Текущий контроль	Методы машинного обучения для анализа естественного языка. Классификация документов. Моделирование тем и кластеризация документов	5	5	5 баллов: полностью выполнено практическое задание, даны правильные ответы на контрольные вопросы. 4 балла: Не выполнено 1 задание практического задания или даны неверные ответы на контрольные вопросы. 3 балла: Не выполнено 2 задание практического задания или даны неверные ответы на контрольные	зачет

						<p>вопросы.</p> <p>2 балла: Не выполнено 1 или 2 задания практического задания и даны неверные ответы на контрольные вопросы.</p> <p>1 балл: Не выполнено 3 и более задания практического задания и даны неверные ответы на контрольные вопросы.</p> <p>0 баллов: задание не выполнено</p>	
3	2	Текущий контроль	<p>Искусственные нейронные сети для обработки текста.</p> <p>Слой LSTM.</p> <p>Классификация документов.</p> <p>Генерация символов и слов.</p>	5	5	<p>5 баллов: полностью выполнено практическое задание, даны правильные ответы на контрольные вопросы.</p> <p>4 балла: Не выполнено 1 задание практического задания или даны неверные ответы на контрольные вопросы.</p> <p>3 балла: Не выполнено 2 задание практического задания или даны неверные ответы на контрольные вопросы.</p> <p>2 балла: Не выполнено 1 или 2 задания практического задания и даны неверные ответы на контрольные вопросы.</p> <p>1 балл: Не выполнено 3 и более задания практического задания и даны неверные ответы на контрольные вопросы.</p> <p>0 баллов: задание не выполнено</p>	зачет
4	2	Текущий контроль	<p>Трансферлернинг.</p> <p>Обучение модели BERT.</p> <p>Классификация документов.</p>	5	5	<p>Компьютерный тест состоит из 5 вопросов, позволяющих оценить сформированность компетенций. На ответы отводится 7 мин. Стоимость одного вопроса - 1 балл.</p>	зачет
5	2	Текущий контроль	<p>Трансферлернинг.</p> <p>Обучение моделей BERT. Задача регрессии.</p>	5	5	<p>5 баллов: полностью выполнено практическое задание, даны правильные ответы на контрольные вопросы.</p> <p>4 балла: Не выполнено 1 задание практического задания или даны неверные ответы на контрольные вопросы.</p> <p>3 балла: Не выполнено 2 задание практического задания или даны неверные ответы на контрольные вопросы.</p> <p>2 балла: Не выполнено 1 или 2 задания практического задания и даны неверные ответы на контрольные вопросы.</p> <p>1 балл: Не выполнено 3 и более задания практического задания и даны неверные ответы на контрольные вопросы.</p>	зачет

						0 баллов: задание не выполнено	
6	2	Текущий контроль	BERT. Генерация символов и текста.	5	5	5 баллов: полностью выполнено практическое задание, даны правильные ответы на контрольные вопросы. 4 балла: Не выполнено 1 задание практического задания или даны неверные ответы на контрольные вопросы. 3 балла: Не выполнено 2 задания практического задания или даны неверные ответы на контрольные вопросы. 2 балла: Не выполнено 1 или 2 задания практического задания и даны неверные ответы на контрольные вопросы. 1 балл: Не выполнено 3 и более задания практического задания и даны неверные ответы на контрольные вопросы. 0 баллов: задание не выполнено	зачет
7	2	Промежуточная аттестация	Итоговый тест	-	30	Компьютерный тест состоит из 30 вопросов, позволяющих оценить сформированность компетенций. На ответы отводится 1 час. Стоимость одного вопроса - 1 балл. 30 баллов: задание полностью выполнено без ошибок 1-29 баллов: задание выполнено частично или выполнено с ошибками 0 баллов: задание не выполнено	зачет

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
зачет	<p>При оценивании результатов учебной деятельности обучающегося по дисциплине используется балльно-рейтинговая система оценивания результатов учебной деятельности обучающихся (Положение о БРС утверждено приказом ректора от 24.05.2019 г. № 179, в редакции приказа ректора от 10.03.2022 г. № 25-13/09). Оценка за дисциплину формируется на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля. Зачтено: Величина рейтинга обучающегося по дисциплине 60...100 %. Незачтено: 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 30 вопросов. На выполнение теста дается 60 минут. В этом случае оценка за дисциплину рассчитывается на основе полученных оценок за контрольно-рейтинговые мероприятия текущего контроля и</p>	В соответствии с пп. 2.5, 2.6 Положения

	промежуточной аттестации. Фиксация результатов учебной деятельности по дисциплине проводится в день зачета при личном присутствии студента.	
--	---	--

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ						
		1	2	3	4	5	6	7
ПК-2	Знает: типовые решения, библиотеки программных модулей, шаблоны, классы объектов, используемые при разработке программного обеспечения для решения задач обработки естественного языка	+	+	+	+	+	+	+
ПК-2	Умеет: применять типовые решения, библиотеки программных модулей, шаблоны, классы объектов при проектировании программного обеспечения	+	+	+	+	+	+	+
ПК-2	Имеет практический опыт: проектирования и реализации приложений для решения задач обработки естественного языка с использованием методов машинного обучения и нейронных сетей	+	+	+	+	+	+	+

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. Методическое пособие

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	Паттерсон, Д. Глубокое обучение с точки зрения практика / Д. Паттерсон, А. Гибсон. — Москва : ДМК Пресс, 2018. — 418 с. — ISBN 978-5-97060-481-6. — Текст : электронный // Лань : электронно-библиотечная система. https://e.lanbook.com/book/116122
2	Основная литература	Электронно-библиотечная система издательства	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

		Лань	Пресс, 2018. — 294 с. — ISBN 978-5-97060-573-8. — Текст : электронный // Лань : электронно-библиотечная система. https://e.lanbook.com/book/111438
3	Основная литература	Электронно-библиотечная система издательства Лань	Козьмо, Л. П. Построение систем машинного обучения на языке Python / Л. П. Козьмо, В. Ричарт ; перевод с английского А. А. Слинкин. — 2-е изд. — Москва : ДМК Пресс, 2016. — 302 с. — ISBN 978-5-97060-330-7. — Текст : электронный // Лань : электронно-библиотечная система. https://e.lanbook.com/book/82818
4	Дополнительная литература	Электронно-библиотечная система издательства Лань	Джонс, М. Т. Программирование искусственного интеллекта в приложениях / М. Т. Джонс. — Москва : ДМК Пресс, 2011. — 312 с. — ISBN 978-5-94074-746-8. — Текст : электронный // Лань : электронно-библиотечная система. https://e.lanbook.com/book/1244
5	Дополнительная литература	Электронно-библиотечная система издательства Лань	Бизли, Д. Python. Книга рецептов / Д. Бизли, Б. К. Джонс ; перевод с английского Б. В. Уварова. — Москва : ДМК Пресс, 2019. — 646 с. — ISBN 978-5-97060-751-0. — Текст : электронный // Лань : электронно-библиотечная система. https://e.lanbook.com/book/131723
6	Дополнительная литература	Электронно-библиотечная система издательства Лань	Практикум по линейной и тензорной алгебре : учебное пособие / О. Н. Казакова, Т. А. Фомина, С. В. Харитоновна, А. Р. Рустанов. — Оренбург : ОГУ, 2017. — 117 с. — Текст : электронный // Лань : электронно-библиотечная система. https://e.lanbook.com/book/110639

Перечень используемого программного обеспечения:

Нет

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
Зачет, диф. зачет	110 (3г)	Проектор, компьютер
Лекции	110 (3г)	Проектор, компьютер
Практические занятия и семинары	110 (3г)	Проектор, компьютер