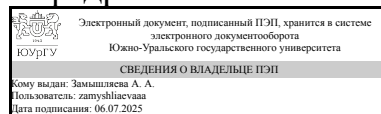


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Заведующий выпускающей
кафедрой



А. А. Замышляева

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.М0.04 Интеллектуальный анализ естественного языка
для направления 02.04.02 Фундаментальная информатика и информационные технологии

уровень Магистратура

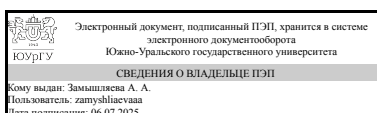
магистерская программа Разработка интеллектуальных систем

форма обучения очная

кафедра-разработчик Прикладная математика и программирование

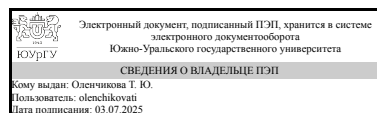
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 02.04.02 Фундаментальная информатика и информационные технологии, утверждённым приказом Минобрнауки от 23.08.2017 № 811

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



А. А. Замышляева

Разработчик программы,
к.техн.н., доцент



Т. Ю. Оленчикова

1. Цели и задачи дисциплины

Целью дисциплины является ознакомление студентов с методами интеллектуального анализа текстов, освоение принципов морфологического, синтаксического и семантического анализа, извлечения информации из текстов, классификации текстов, основами машинного перевода. Основные задачи дисциплины: - научить использовать известные программные инструменты обработки текстов - ставить и решать конкретные задачи анализа текстов - применять методы машинного обучения для анализа текстов - решать задачи извлечения из текстов именованных сущностей - выявлять фейковые новости - определять психологические особенности автора по его тексту

Краткое содержание дисциплины

Естественный язык. Типология естественных языков. Уровни анализа, Принципы морфологического, синтаксического и семантического анализа, извлечения информации из текстов. Извлечение информации из текстов, классификации. Методы машинного обучения для решения задач обработки текстов.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-2 Способен исследовать и разрабатывать архитектуры интеллектуальных информационных систем для различных предметных областей на основе комплексов методов и инструментальных средств систем искусственного интеллекта	Знает: принципы построения интеллектуальных систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе интеллектуального анализа текстов Имеет практический опыт: классификации и тематического моделирования текстов на основе интеллектуального анализа

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Компьютерное зрение, Интеллектуальные системы для автоматизации производственных процессов, Методы и технологии искусственного интеллекта в задачах синтетических медиа	Не предусмотрены

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

Дисциплина	Требования
Компьютерное зрение	Знает: Умеет: применять алгоритмы компьютерного зрения для распознавания образов, очистки изображений и других

	прикладных задач Имеет практический опыт: применения методов компьютерного зрения, позволяющих производить детектирование, отслеживание и классификацию объектов на изображениях и в видеопотоке
Методы и технологии искусственного интеллекта в задачах синтетических медиа	Знает: принципы построения систем распознавания и синтеза речи, методы и подходы к планированию и реализации проектов по созданию интеллектуальных систем на основе распознавания и синтеза речи Умеет: руководить проектами по созданию, внедрению и поддержке интеллектуальных систем на основе распознавания и синтеза речи Имеет практический опыт:
Интеллектуальные системы для автоматизации производственных процессов	Знает: Умеет: использовать системы искусственного интеллекта в задачах повышения эффективности производственных процессов, автоматизации выбора и оценки параметров производства, анализа требований к качеству продукции Имеет практический опыт: разработки интеллектуальных систем для решения задач логистики и автоматизации производственных процессов

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 4 з.е., 144 ч., 46,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		4
Общая трудоёмкость дисциплины	144	144
<i>Аудиторные занятия:</i>	36	36
Лекции (Л)	12	12
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	0	0
Лабораторные работы (ЛР)	24	24
<i>Самостоятельная работа (СРС)</i>	97,5	97,5
Подготовка к лекциям и лабораторным работам	79,5	79,5
Подготовка к экзамену	18	18
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 интеллектуальный анализ текстов	6	2	0	4

2	Методы лингвистического анализа текстов	18	6	0	12
3	Прикладные задачи интеллектуального анализа естественного языка	12	4	0	8

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Основные определения. Постановки задач обработки текстов и подходы к их решению. Особенности естественных языков. Типология естественных языков. Уровни анализа. Лингвистическое исследование.	2
2	2	Методы морфологического анализа. Проблема морфологической многозначности. Методы снятия морфологической многозначности.	2
3	2	Синтаксис естественных языков. Способы описания синтаксической структуры предложения. Методы синтаксического анализа.	2
4	2	Способы формализации семантики. Методы семантического анализа. Реляционно-ситуационный анализ текста. Дистрибутивная семантика.	2
5	3	Извлечение информации из текстов. Классификация и кластеризация текстов.	2
6	3	Методы машинного обучения для решения задач обработки естественного языка.	2

5.2. Практические занятия, семинары

Не предусмотрены

5.3. Лабораторные работы

№ занятия	№ раздела	Наименование или краткое содержание лабораторной работы	Кол-во часов
1	1	Регулярные выражения	2
2	1	Токенизация	2
3, 4	2	Исправление опечаток	4
5, 6	2	Морфологический анализ	4
7, 8	2	Синтаксический анализ	4
9, 10	3	Извлечение именованных сущностей	4
11, 12	3	Анализ тональности	4

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 лекциям и лабораторным работам	ЭУМД 3: все разделы, 6 ЛР с метод. указ.; 2) ЭУМД 4: глава 1, С.21-34, глава 3, С. 74-102; 3) ЭУМД 5, глава 2, С. 30-52, глава 4, С. 76=156; 4) ЭУМД 7	4	79,5
Подготовка к экзамену	1) ЭУМД 4: глава 1, С.21-34, глава 3, С. 74-102; 2) ЭУМД 5, глава 2, С. 30-52, глава 4, С. 76=156; ЭУМД 3, все разделы, МУ к ЛР	4	18

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	4	Текущий контроль	ЛР1 Регулярные выражения	1	8	В задании 3 задачи. За решение каждой задачи можно получить до 2 баллов: задание выполнено полностью и правильно -2 балла; имеются незначительные ошибки -1 балл; приложение не работоспособно – 0 баллов; Ответ на контрольный вопрос должен продемонстрировать понимание механизмов NLP, за ответ можно получить оценку: ответ полный и правильный -2 балла; ответ не полный – 1 балл; ответ не удовлетворительный – 0 баллов. Максимальная оценка - 8 баллов	экзамен
2	4	Текущий контроль	ЛР2 Токенизация	1	10	В задании 4 задачи. За решение каждой задачи можно получить до 2 баллов: задание выполнено полностью и правильно -2 балла; имеются незначительные ошибки -1 балл; приложение не работоспособно – 0 баллов; Ответ на контрольный вопрос должен продемонстрировать понимание механизмов NLP, за ответ можно получить оценку: ответ полный и правильный -2 балла; ответ не полный – 1 балл; ответ не удовлетворительный – 0 баллов. Итого, максимальная оценка – 10 баллов	экзамен
3	4	Текущий контроль	ЛР3 Исправление опечаток	1	10	В задании 4 задачи. За решение каждой задачи можно получить до 2 баллов: задание выполнено полностью и правильно -2 балла; имеются незначительные ошибки -1 балл; приложение не работоспособно – 0 баллов; Ответ на контрольный вопрос должен продемонстрировать понимание механизмов NLP, за ответ можно получить оценку: ответ полный и	экзамен

						правильный -2 балла; ответ не полный – 1 балл; ответ не удовлетворительный – 0 баллов. Итого, максимальная оценка – 10 баллов	
4	4	Текущий контроль	ЛР4 Морфологический анализ	1	10	В задании 4 задачи. За решение каждой задачи можно получить до 2 баллов: задание выполнено полностью и правильно -2 балла; имеются незначительные ошибки -1 балл; приложение не работоспособно – 0 баллов; Ответ на контрольный вопрос должен продемонстрировать понимание механизмов NLP, за ответ можно получить оценку: ответ полный и правильный -2 балла; ответ не полный – 1 балл; ответ не удовлетворительный – 0 баллов. Итого, максимальная оценка – 10 баллов	экзамен
5	4	Текущий контроль	ЛР 5 Синтаксический анализ	1	10	В задании 4 задачи. За решение каждой задачи можно получить до 2 баллов: задание выполнено полностью и правильно -2 балла; имеются незначительные ошибки -1 балл; приложение не работоспособно – 0 баллов; Ответ на контрольный вопрос должен продемонстрировать понимание механизмов NLP, за ответ можно получить оценку: ответ полный и правильный -2 балла; ответ не полный – 1 балл; ответ не удовлетворительный – 0 баллов. Итого, максимальная оценка – 10 баллов	экзамен
6	4	Текущий контроль	ЛР 6 Извлечение именованных сущностей	1	10	В задании 4 задачи. За решение каждой задачи можно получить до 2 баллов: задание выполнено полностью и правильно -2 балла; имеются незначительные ошибки -1 балл; приложение не работоспособно – 0 баллов; Ответ на контрольный вопрос должен продемонстрировать понимание механизмов NLP, за ответ можно получить оценку: ответ полный и правильный -2 балла; ответ не полный – 1 балл; ответ не удовлетворительный – 0 баллов. Итого, максимальная оценка – 10 баллов	экзамен
8	4	Текущий контроль	ЛР 7 Анализ тональности текста	1	4	Оценка суммируется из оценок: 1) задание выполнено полностью и правильно -2 балла; имеются	экзамен

						незначительные ошибки -1 балл; приложение не работоспособно – 0 баллов; 2) ответ на контрольный вопрос должен продемонстрировать понимание механизмов NLP, за ответ можно получить оценку: ответ полный и правильный -2 балла; ответ не полный – 1 балл; ответ не удовлетворительный – 0 баллов. Итого, максимальная оценка – 4 балла	
10	4	Промежуточная аттестация	Экзамен	-	6	Экзаменационный билет содержит два теоретических вопроса. Суммируются оценки за каждый вопрос. Критерии оценки: полные и правильные ответы на вопрос билета и дополнительные вопросы - 3 балла; не полные или не совсем правильные ответы на вопрос билета и дополнительные вопросы - 2 балла; неправильный ответ на вопрос билета, правильные ответы на дополнительные вопросы - 1 балл; неудовлетворительные ответы на вопрос билета и дополнительные вопросы - 0 баллов Итого, максимальная оценка - 6 баллов	экзамен

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
экзамен	Экзамен проводится в очной форме по билетам. Процедура прохождения экзамена не является обязательной если по результатам текущего контроля БРС у студента положительная оценка и он с ней согласен. В каждом билете 2 теоретических вопроса. Студент устно отвечает на вопросы билета. Студент должен находиться в аудитории на протяжении всей процедуры экзамена. Число студентов, одновременно находящихся в аудитории, где сдается экзамен, не более 8 человек. На подготовку к ответу студенту отводится не более 30 мин. Когда обучающийся будет готов к ответу, ему задаются контрольные вопросы по содержанию билета. Студент должен УСТНО ответить на вопросы билета и дополнительные вопросы по теме билета в течение 5 мин. На этом основании преподаватель выставляет баллы за зачетную работу. Окончательная оценка за курс выставляется согласно БРС. Добор баллов осуществляется путем выполнения дополнительных заданий из KM1-KM8.	В соответствии с пп. 2.5, 2.6 Положения

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ KM									
		1	2	3	4	5	6	8	10		

ПК-2	Знает: принципы построения интеллектуальных систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе интеллектуального анализа текстов	+	+	+	+	+	+	+	+
ПК-2	Имеет практический опыт: классификации и тематического моделирования текстов на основе интеллектуального анализа	+	+	+	+	+	+	+	+

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. Методические указания по усвоению дисциплины для преподавателя и студента

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Методические указания по усвоению дисциплины для преподавателя и студента

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Дополнительная литература	Электронный каталог ЮУрГУ	Самбурский, М. С. Исследование применимости свёрточной и рекуррентной нейронных сетей для классификации русскоязычных поэтических текстов / М. С. Самбурский, Т. Ю. Оленчикова // Южно-Уральская молодежная школа по математическому моделированию: сб. тр. V Всеросс. студ. науч.-практ. конф. – Челябинск: Издательский центр ЮУрГУ, 2022. – С. 56–61 https://elib.susu.ru/cgi-bin/koha/opac-detail.pl?biblionumber=451386&query_desc=Уральская молодежная школа
2	Дополнительная литература	Электронный каталог ЮУрГУ	Чупин В.О., Метод выявления и анализа тональности наиболее обсуждаемых тем в социальных сетях /В.О.,Чупин, Т.Ю.Оленчикова // Южно-Уральская молодежная школа по математическому моделированию: сб. тр. VI Всеросс. студ. науч.-практ. конф. – Челябинск: Издательский центр ЮУрГУ, 2023., С.101-109 https://elib.susu.ru/cgi-bin/koha/opac-detail.pl?biblionumber=489422&query_desc=Уральская молодежная школа

3	Методические пособия для самостоятельной работы студента	Учебно-методические материалы кафедры	Оленчикова Т.Ю. Методические указания и задания к выполнению лабораторных работ по дисциплине «Интеллектуальный анализ текстов» https://prm.susu.ru/documents/dop/SRS_po_IANL-2025.zip
4	Основная литература	ЭБС 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
5	Основная литература	ЭБС 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
6	Дополнительная литература	ЭБС издательства Лань	Маккинни, У. Python и анализ данных. Первичная обработка данных с применением pandas, NumPy и Jupiter : справочник / У. Маккинни ; перевод с английского А. А. Слинкина. — 3-е изд. — Москва : ДМК Пресс, 2023. — 536 с. https://e.lanbook.com/book/348086

Перечень используемого программного обеспечения:

1. Microsoft-Office(бессрочно)
2. 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	336 (36)	Мультимедийная аудитория на 50 мест с предустановленным ПО MS Word, MS Power Poin, веб-браузер Chrome или Python
Лабораторные занятия	332 (36)	Компьютерный класс на 30 мест, проектор с предустановленным ПО MS Office, Python