Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Гриб Владислав Валерьевич

Должность: Ректор

Дата подписания: 06.09.2023 08:30:49

Уникальный программный ключ:

637517d24e103c3db032acf37e**Образовательное** частное учреждение высшего образования

«МОСКОВСКИЙ УНИВЕРСИТЕТ ИМЕНИ А.С. ГРИБОЕДОВА»

(ИМПЭ им. А.С. Грибоедова)

Институт международной экономики, лидерства и менеджмента

УТВЕРЖДАЮ

Директор института международной экономики, лидерства и менеджмента
А.А. Панарин
«22» июня 2023 г.

Рабочая программа дисциплины ВВЕДЕНИЕ В ОБРАБОТКУ ЕСТЕСТВЕННОГО ЯЗЫКА

Направление подготовки 09.03.03 Прикладная информатика (уровень бакалавриат)

Направленность (профиль): «Анализ данных»

Форма обучения: очная

Москва

Рабочая программа дисциплины «Введение в обработку естественного языка». Направление подготовки 09.03.03 Прикладная информатика, направленность (профиль): «Анализ данных» / А.А. Шестемиров – М.: ИМПЭ им. А.С. Грибоедова. – 22 с.

Рабочая программа дисциплины составлена на основании федерального государственного образовательного стандарта высшего образования – бакалавриат по направлению подготовки 09.03.03 Прикладная информатика, утвержденного приказом Министерства образования и науки Российской Федерации от «19» сентября 2017 № 922 (с изменениями и дополнениями) и Профессионального стандарта «Программист», утвержденного приказом Министерства труда и социальной защиты Российской Федерации от 18 ноября 2013 г. № 679н (зарегистрирован Министерством юстиции Российской Федерации 18 декабря 2013 г., регистрационный № 30635), с изменением, внесенным приказом Министерства труда и социальной защиты Российской Федерации от 12 декабря 2016 г. № 727н (зарегистрирован Министерством юстиции Российской Федерации 13 января 2017 г., регистрационный № 45230), Профессионального стандарта «Специалист по ин-формационным системам», утвержденный приказом Министерства труда и социальной защиты Российской Федерации от 18 ноября 2014 г. № 896н (зарегистрирован Министерством юстиции Российской Федерации 24 декабря 2014 г., регистрационный № 35361), с изменением, внесенным приказом Министерства труда и социальной защиты Российской Федерации от 12 декабря 2016 г. № 727н (зарегистрирован Министерством юстиции Российской Федерации 13 января 2017 г., регистрационный № 45230)

Ответственный рецензент:	Назарова Н.А., к.э.н., доцент, заместитель руководителя департамента налогов и налогового администрирования Финансового университета при Правительстве Российской Федерации				
	(Ф.И.О., уч. сп	тепень, уч. звание, должность)			
Рабочая программа дисципл ционного менеджмента и предприні		добрена на заседании кафедры иннова 3г., протокол №6			
Заведующий кафедрой	/	/к.э.н. А.А. Шестемиров/			
	(подпись)				
Согласовано от Библиотеки		/О.Е. Степкина/			
	(подпись)				

К.э.н. А.А. Шестемиров

Разработчики:

Раздел 1. Цели и задачи освоения дисциплины

Целью изучения дисциплины «Введение в обработку естественного языка» является овладение студентами навыками разработки программных приложений, использующих алгоритмы обработки естественного языка.

Задачами изучения дисциплины являются:

- ознакомление студентов с понятиями и моделями машинного обучения и искусственного интеллекта, применяемыми для обработки естественного языка; возможностями библиотек Руthon по обработке естественного языка;
- получение навыков использования специализированного программного обеспечения при обработке естественного языка.

РАЗДЕЛ 2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

ПРОФЕССИОНАЛЬНЫЕ КОМПЕТЕНЦИИ Трудовые функции Индикаторы достижения компетенции (для планирования результатов Код, наименование (код, наименовапрофессиональ обучения по элементам образовательной программы и соответствуюние)/уровень (подурощих оценочных средств) ных компетенций вень) квалификации ОТФ (код, наименование) / Профессиональный стандарт (код, наименование) Тип(ы) задач(и) профессиональной деятельности D. Разработка требований и проектирование программного обеспечения/ 06.001 Программист Проектный ПК-1 Способ-D/03.6Проектиро-ИПК-1.1 Знать принципы построения архитектуры программпроективание программного ного обеспечения и виды архитектуры программного обеспеность обеспечения чения; типовые решения, библиотеки программных модулей, ровать прикладное программное шаблоны, классы объектов, используемые при разработке прообеспечение граммного обеспечения; методы и средства проектирования программного обеспечения; методы и средства проектирования баз данных; методы и средства проектирования программных интерфейсов ИПК-1.2 Уметь использовать существующие типовые решения и шаблоны проектирования программного обеспечения; применять методы и средства проектирования программного обеспечения, структур данных, баз данных, программных интерфейсов; осуществлять коммуникации с заинтересованными сторонами ИПК-1.3 Владеть разработкой, изменением и согласованием архитектуры программного обеспечения с системным аналитиком и архитектором программного обеспечения; системой проектирования структур данных; системой проектирования баз данных; системой проектирования программных интерфейсов; оценкой и согласованием сроков выполнения поставленных задач С Выполнение работ и управление работами по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы / 06.015 Профессиональный стандарт «Специалист по информационным системам»

производственно-технологический, научно-исследовательский

ПК-2. Способность осуществлять инженернотехнологическую поддержку планирования управления требованиями по созданию (модификации) и сопровождению ИС
ПК-3. Способность применять системный подход для составления технико-экономического обоснования проектных решений и технического задания на разработку информационной системы

С/10.6 Инженернотехнологическая поддержка планирования управления требованиями ИПК-2.1. Знать инструменты и методы управления требованиями; предметной областью автоматизации; возможностями ИС; источниками информации, необходимой для профессиональной деятельности; современный отечественный и зарубежный опыт в профессиональной деятельности; инструменты и методы моделирования бизнес-процессов организации; основы реинжиниринга бизнес-процессов организации; управление содержанием проекта: документирование требований, систему анализа продукта, модерируемых совещаний; управление качеством: контрольные списки, верификация, валидация (приемо-сдаточные испытания); управление коммуникациями в проекте: базовые навыки управления (в том числе проведение презентаций, проведение переговоров, публичные выступления); культуру речи; правила деловой переписки. ИПК-2.2. Умеет анализировать входные данные; планировать работы.

ИПК-2.3. Владеть системным подходом для выбора технологии управления требованиями; представлениями исходных данных для разработки плана управления требованиями; системой согласования в части инженерно-технологического обеспечения плана управления требованиями с заинтересованными сторонами

С/01.6 Определение первоначальных требований заказчика к ИС и возможности их реализации в ИС на этапе предконтрактных работ;

ИПКЗ.1. Знает возможности типовой ИС; предметную область автоматизации; методы выявления требований; технологии межличностной и групповой коммуникации в деловом взаимодействии, основы конфликтологии; технологии подготовки и проведения презентаций; архитектуру, устройство и функционирование вычислительных систем; коммуникационное оборудование; сетевые протоколы; основы современных операционных систем; основы современных систем управления базами данных; устройство и функционирование современных ИС; современные стандарты информационного взаимодействия систем; программные средства и платформы инфраструктуры информационных технологий организаций; современные подходы и стандарты автоматизации организации (например, CRM, MRP, ERP..., ITIL, ITSM); основы теории систем и системного анализа; методики описания и моделирования бизнес-процессов, средства моделирования бизнеспроцессов; системы классификации и кодирования информации, в том числе присвоение кодов документам и элементам справочников; отраслевая нормативная техническая документация; источники информации, необходимой для профессиональной деятельности; современный отечественный и зарубежный опыт в профессиональной деятельности; формирование и механизмы рыночных процессов организации; основы менеджмента, в том числе менеджмента качества; основы бухгалтерского учета и отчетности организаций; основы налогового законодательства Российской Федерации; основы управленческого учета; основы финансового учета и бюджетирования; основы международных стандартов финансовой отчетности (МСФО); основы управления торговлей, поставками и запасами; основы организации производства; основы управления взаимоотношениями с клиентами и заказчиками (CRM); основы управления персоналом, включая вопросы оплаты труда; основы теории управления; современные инструменты и методы управления организацией, в том числе методы планирования деятельности, распределения поручений, контроля исполнения, принятия решений; методология ведения документооборота в организациях; инструменты и методы определения финансовых и производственных показателей деятельности организаций; основы организационной диагностики; инструменты и методы моделирования бизнес-процессов организации; основы реинжиниринга бизнес-процессов организации; управление содержанием проекта документирование требований, анализ продукта, модерируемые совещания; культура речи; правила деловой переписки.

ИПК-3.2 Умеет проводить переговоры; проводить презентации; подготавливать протоколы мероприятий.

ИПК 3.3. Владеет основами выявления первоначальных требований заказчика к ИС; системой информирования заказчика о возможностях типовой ИС и вариантах ее модификации; определениями возможностей достижения соответствия ИС первоначальным требованиям заказчика; основами составления протоколов переговоров с заказчиком

РАЗДЕЛ 3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Введение в обработку естественного языка» изучается в восьмом семестре, относится к Б1.В. Части учебного плана, формируемой участниками образовательных отношений, Блока Б.1 «Дисциплины (модули)».

Общая трудоемкость дисциплины составляет 4 з.е.

Знания, умения, навыки, опыт практический деятельности, приобретенные при освоении настоящей дисциплины, необходимы для успешного освоения следующей дисциплины: «Компьютерные экспертные системы».

Раздел 4. Объем (трудоемкость) дисциплины (общая, по видам учебной работы, видам промежуточной аттестации)

Трудоемкость дисциплины и виды учебной работы (очная форма обучения)

				9 222W22 Pop.:2W 9	-					
			Контак		Иная СР					
3.e.		Всего Занятия Занятия семинарского часов лекцион-		инарского типа		Контакт- ная рабо- та по	Часы СР на подго- товку	Контроль		
	шсов	шсов	шсов	ного типа	Лаборатор- ные	Практические/ Семинарские	курсовой работе	кур.раб.		
				8 семестр						
1	144	28	24				56	36		
4	4 144	20	24	-	-	-	30	Экзамен		
	Всего по дисциплине									
4	144	28	24	-	-	-	56	36		

СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

№	Наименование разде-	Содержание темы			
п/п	лов и тем дисциплины				
Разд	Раздел №1 «Методы лингвистического анализа текстов»				
1	Тема 1.1 Этапы анализа	Изучаемые вопросы:			
	текста на естественном	1. Основные свойства текстов;			
	языке	2. Уровни рассмотрения текстов;			
		3. Графематический анализ;			
		4. Морфологический анализ.			
		5. Выделение словосочетаний;			
		6. Синтаксический анализ;			
		7. Семантический анализ.			
		Вопросы для самостоятельного изучения:			
		1. MyStem - морфологический анализ текста на русском языке.			

No	Наименование разде-	Содержание темы
п/п	лов и тем дисциплины	
		2. Контекстно-зависимый анализ текста
Разд	<u> </u> ел №2 «Модели компьют	ерного представления текстов»
2	Тема 2.1 Модели пред-	Изучаемые вопросы:
	ставления текстов	1. Модель «мешок слов»;
		2. Векторная модель представления текстов;
		3. Методы задания весов терминов.
		4. Качество модели представления текстов;
		5. Технология формирования векторного представления.
		Вопросы для самостоятельного изучения:
		1. Компьютерная модель языка;
		2. Лингвистические признаки.
		3. Контекстные признаки;
		4. Структурные признаки
	ел №3 «Методы классифі	:
3	Тема 3.1 Методы клас-	Изучаемые вопросы:
	сификации полнотек-	1. Общее описание методов классификации;
	стовых документов	2. Оценка качества классификации;
		3. Вероятностные методы классификации;
		4. Методы классификации на основе правил;
		 Комбинированные методы классификации Методы классификации на основе расстояний.
		Вопросы для самостоятельного изучения:
		1. Метрики классификации.
		 метрики классификации. Задачи классификации текстов
Разд	<u>।</u> ел №4 «Методы кластерн	*
4	Тема 4.1 Методы кла-	Изучаемые вопросы:
7	стерного анализа пол-	1. Общее описание методов кластерного анализа;
	нотекстовых докумен-	2. Оценка качества кластерного анализа;
	тов	3. Вероятностные методы кластерного анализа;
		4. Структурные методы кластерного анализа.
		5. Интерпретация результатов кластерного анализа;
		6. Сравнительный анализ методов кластерного анализа.
		Вопросы для самостоятельного изучения:
		1. Графовые методы анализа текста.
		2. Задачи кластеризации текстов.
	-	нологии обработки текстовых данных»
5	Тема 5.1 Модель	Изучаемые вопросы:
	word2vec	1. Предпосылки создания модели;
		2. Топология модели;
		3. Особенности применения;
		4. Метрики качества.
		Вопросы для самостоятельного изучения:
	T. 5216	1. История NLP и речевых технологий.
6	Тема 5.2 Модель	Изучаемые вопросы:
	doc2vec	1. Предпосылки создания модели;
		2. Топология модели;
		3. Особенности применения;
		4. Метрики качества.
		Вопросы для самостоятельного изучения 1. Модели glove, transformer
7	Тема 5.3 Генеративные	Изучаемые вопросы:
'	модели естественного	1. Предпосылки создания модели;
	языка: BERT	2. Топология модели;
	ASSERVA, EDITT	3. Особенности применения;
	I .	2. ceccemiceri ilpinicionin,

№	Наименование разде-	Содержание темы
п/п	лов и тем дисциплины	
		4. Метрики качества.
		Вопросы для самостоятельного изучения:
		1. Модели elmo, t-nlg.
8	Тема 5.4 Генеративные	Изучаемые вопросы:
	модели естественного	1. Предпосылки создания модели;
	языка: GPT-2, GPT-3	2. Топология модели;
		3. Особенности применения;
		4. Метрики качества.
		Вопросы для самостоятельного изучения:
		1. Исследование компании Gartner о перспективах развития АІ на
		ближайшие 2-5 лет.
Разд	ел №6 «Прикладные зада	чи анализа текстов»
9	Тема 6.1 Методы	Изучаемые вопросы:
	очистки текста	1. Методы очистки текста от элементов оформления.
		Вопросы для самостоятельного изучения:
		1. Методы выделения и классификации значимых фрагментов в
		текстах.
10	Тема 6.2 Комплексная	Изучаемые вопросы:
	технология классифи-	1. Технология классификации текстовых данных.
	кации текстовых дан-	Вопросы для самостоятельного изучения:
	ных	1. Методы выявления дубликатов и тематического упорядочения
		текстов.

Перечень разделов (модулей), тем дисциплины и распределение учебного времени по разделам/темам дисциплины, видам учебных занятий (в т.ч. контактной работы), видам текущего контроля

очная форма обучения

Разделы / Темы	Контактная работа			Часы СР на	Иная СР	Кон-	Всего часов	
	Заня- тия лек-	Заняті минар типа		Кон- такт- ная	под- готов- ку		Ь	
	цион-	Лаб.	Прак	работа	кур. р.			
	ного	p		ПО				
	типа	•	/сем.	кур.р				
		8 (семестр					
Раздел №1 «Методы лингви-	4	4				8		16
стического анализа текстов»								
Тема 1.1 Этапы анализа текста на	4	4				8		16
естественном языке								
Раздел №2 «Модели компьютер-	4	8				12		24
ного представления текстов»								
Тема 2.1 Модели представления	4	8				12		24
текстов								
Раздел №3 «Методы классифи-	4	4				8		16
кации»								
Тема 3.1 Методы классификации	4	4				8		16
полнотекстовых документов								
Раздел №4 «Методы кластерного	4	4				8		16
анализа»						_		
Тема 4.1 Методы кластерного ана-	4	4				8		16
лиза полнотекстовых документов								
Раздел №5 «Нейросетевые тех-	8	4				12		24
нологии обработки текстовых								
данных»	_							
Тема 5.1 Модель word2vec	2					2		4

Разделы / Темы	Контактная работа			Часы СР на	Иная СР	Кон- трол	Всего часов	
	Заня-	Заняті		Кон-	под-		Ь	
	ТИЯ	минар	ского	такт-	готов-			
	лек-	типа	1	ная	ку			
	цион-	Лаб.	Прак	работа	кур. р.			
	ного	p		по				
	типа		/сем.	кур.р				
Тема 5.2 Модель doc2vec	2					2		4
Тема 5.3 Генеративные модели естественного языка: BERT	2	2				4		8
Тема 5.4 Генеративные модели естественного языка: GPT-2, GPT-3	2	2				4		8
Раздел №6 «Прикладные задачи	4					8		12
анализа текстов»								
Тема 6.1 Методы очистки текста	2					4		6
Тема 6.2 Комплексная технология	2					4		6
классификации текстовых данных								
Экзамен							36	36
Итого за 8 семестр	28	24				56	36	144

ЗАНЯТИЯ СЕМИНАРСКОГО ТИПА для очной формы обучения

Семинарские занятия

Общие рекомендации по подготовке к семинарским занятиям. При подготовке к работе во время проведения занятий семинарского типа следует обратить внимание на следующие моменты: на процесс предварительной подготовки, на работу во время занятия, обработку полученных результатов, исправление полученных замечаний. Предварительная подготовка к учебному занятию семинарского типа заключается в изучении теоретического материала в отведенное для самостоятельной работы время, ознакомление с инструктивными материалами с целью осознания задач занятия. Работа во время проведения занятия семинарского типа включает несколько моментов: а) консультирование обучающихся преподавателями с целью предоставления исчерпывающей информации, необходимой для самостоятельного выполнения предложенных преподавателем задач, б) самостоятельное выполнение заданий согласно обозначенной учебной программой тематики.

8 семестр

Раздел №1 «Методы лингвистического анализа текстов» Лабораторная работа 1. Инструменты анализа текстов (4 ч). Порядок выполнения работы:

1. Приобрести навыки работы с инструментальными средствами.

Литература:

Основная

Батура, Т. В. Математическая лингвистика и автоматическая обработка текстов на естественном языке : учебное пособие / Т. В. Батура. — Новосибирск : Новосибирский государственный университет, 2016. — 166 с. — ISBN 978-5-4437-0548-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.IPRsmart hop.ru/93489.html

Дополнительная

Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск : Ульяновский государственный технический университет, 2020. — 221 с. — ISBN 978-5-9795-2088-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.IPRsmart hop.ru/106136.html

Раздел №2 «Модели компьютерного представления текстов»

Лабораторная работа 2. Создание модели для заданного множества документов (веб-скрейпинг, предварительная обработка, модель «мешок слов») (4 ч). Порядок выполнения работы:

1. Используя технологию веб-скрейпинга собрать набор данных, провести предобработку и составить «мещок слов»

Литература: Основная

Батура, Т. В. Математическая лингвистика и автоматическая обработка текстов на естественном языке : учебное пособие / Т. В. Батура. — Новосибирск : Новосибирский государственный университет, 2016. — 166 с. — ISBN 978-5-4437-0548-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.IPRsmart hop.ru/93489.html

Дополнительная

Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск: Ульяновский государственный технический университет, 2020. — 221 с. — ISBN 978-5-9795-2088-9. — Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. — URL: https://www.IPRsmart hop.ru/106136.html

Лабораторная работа 3. Создание модели для заданного множества документов (веб-скрейпинг, предварительная обработка, векторное представление (4 ч). Порядок выполнения работы:

1. Используя технологию веб-скрейпинга собрать набор данных, провести предобработку и составить векторную модель.

Литература:

Основная

Батура, Т. В. Математическая лингвистика и автоматическая обработка текстов на естественном языке : учебное пособие / Т. В. Батура. — Новосибирск : Новосибирский государственный университет, 2016. — 166 с. — ISBN 978-5-4437-0548-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.IPRsmart hop.ru/93489.html

Дополнительная

Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск : Ульяновский государственный технический университет, 2020. — 221 с. — ISBN 978-5-9795-2088-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.IPRsmart hop.ru/106136.html

Раздел №3 «Методы классификации»

Лабораторная работа 4. Создание классификационных моделей для полнотекстовых документов (модель наивного Байеса, логистическая регрессия, бустинг, случайный лес) (4 ч). Порядок выполнения работы:

1. Для собранного набора реализовать классификационные модели.

Литература:

Основная

Батура, Т. В. Математическая лингвистика и автоматическая обработка текстов на естественном языке: учебное пособие / Т. В. Батура. — Новосибирск: Новосибирский государственный университет, 2016. — 166 с. — ISBN 978-5-4437-0548-4. — Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. — URL: https://www.IPRsmart hop.ru/93489.html

Дополнительная

Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск : Ульяновский государственный технический университет, 2020. — 221 с. — ISBN 978-5-9795-2088-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.IPRsmart hop.ru/106136.html

Раздел №4 «Методы кластерного анализа»

Лабораторная работа 5. Кластеризация полнотекстовых документов (4 ч). Порядок выполнения работы:

1. Для собранного набора реализовать модели кластеризации.

Литература:

Основная

Батура, Т. В. Математическая лингвистика и автоматическая обработка текстов на естественном языке: учебное пособие / Т. В. Батура. — Новосибирск: Новосибирский государственный университет, 2016. — 166 с. — ISBN 978-5-4437-0548-4. — Текст: электронный // Цифровой образо-вательный ресурс IPR SMART: [сайт]. — URL: https://www.IPRsmart hop.ru/93489.html

Дополнительная

Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск: Ульяновский государственный тех-

нический университет, 2020. — 221 с. — ISBN 978-5-9795-2088-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.IPRsmart hop.ru/106136.html

Раздел №5 «Нейросетевые технологии обработки текстовых данных» Лабораторная работа 6. Определение тональности сообщений из Твиттера (4 ч). Порядок выполнения работы:

1. Реализовать модели анализа эмоциональной окраски текста.

Литература: Основная

Батура, Т. В. Математическая лингвистика и автоматическая обработка текстов на естественном языке : учебное пособие / Т. В. Батура. — Новосибирск : Новосибирский государственный университет, 2016. — 166 с. — ISBN 978-5-4437-0548-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.IPRsmart hop.ru/93489.html

Дополнительная

Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск: Ульяновский государственный технический университет, 2020. — 221 с. — ISBN 978-5-9795-2088-9. — Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. — URL: https://www.IPRsmart hop.ru/106136.html

РАЗДЕЛ 5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Интерактивные образовательные технологии, используемые на аудиторных лабораторных занятиях

Очная форма обучения

Наименование разделов, тем	Используемые образовательные технологии	Часы
Раздел №1 «Методы лингвистического анализа текстов» Тема 1.1 Этапы анализа текста на естественном языке	Обсуждение решений профессио- нально-ориентированных заданий и задач; обсуждение и анализ решения кейсов	2
Раздел №2 «Модели компьютерного представления текстов» Тема 2.1 Модели представления текстов	Обсуждение решений профессио- нально-ориентированных заданий и задач; обсуждение и анализ решения кейсов	2
Раздел №3 «Методы классификации» Тема 3.1 Методы классификации полнотекстовых документов	Обсуждение решений профессионально-ориентированных заданий и задач; обсуждение и анализ решения кейсов	2
Раздел №4 «Методы кластерного анализа» Тема 4.1 Методы кластерного анализа полнотекстовых документов	Обсуждение решений профессио- нально-ориентированных заданий и задач; обсуждение и анализ решения кейсов	2
Раздел №5 «Нейросетевые технологии обра- ботки текстовых данных» Тема 5.1 Модель word2vec Тема 5.2 Модель doc2vec Тема 5.3 Генеративные модели естественного языка: ВЕЯТ Тема 5.4 Генеративные модели естественного языка: GPT-2, GPT-3	Обсуждение решений профессионально-ориентированных заданий и задач; обсуждение и анализ решения кейсов	2
Раздел №6 «Прикладные задачи анализа текстов» Тема 6.1 Методы очистки текста Тема 6.2 Комплексная технология классификации текстовых данных	Обсуждение решений профессионально-ориентированных заданий и задач; обсуждение и анализ решения кейсов	2

РАЗДЕЛ 6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

Самостоятельная работа

Наименование	Вопросы, выносимые
разделов/тем	на самостоятельное изучение
Раздел №1 «Методы лингвистического анализа тек-	1. MyStem - морфологический анализ тек-
стов»	ста на русском языке
Тема 1.1 Этапы анализа текста на естественном языке	2. Контекстно-зависимый анализ текста.
Раздел №2 «Модели компьютерного представления тек-	1. Компьютерная модель языка;
стов»	2. Лингвистические признаки.
Тема 2.1 Модели представления текстов	3. Контекстные признаки;
	4. Структурные признаки
Раздел №3 «Методы классификации»	1. Метрики классификации.
Тема 3.1 Методы классификации полнотекстовых докумен-	2. Задачи классификации текстов.
тов	
Раздел №4 «Методы кластерного анализа»	1. Графовые методы анализа текста
Тема 4.1 Методы кластерного анализа полнотекстовых доку-	2. Задачи кластеризации текстов
ментов	
Раздел №5 «Нейросетевые технологии обработки тексто-	1. История NLP и речевых технологий.
вых данных»	2. Модели glove, transformer.
Тема 5.1 Модель word2vec	3. Модели elmo, t-nlg.
Тема 5.2 Модель doc2vec	4. Исследование компании Gartner о пер-
Тема 5.3 Генеративные модели естественного языка: BERT	спективах развития АІ на ближайшие
Тема 5.4 Генеративные модели естественного языка: GPT-2,	2-5 лет.
GPT-3	
Раздел №6 «Прикладные задачи анализа текстов»	1. Методы выделения и классификации
Тема 6.1 Методы очистки текста	значимых фрагментов в текстах.
Тема 6.2 Комплексная технология классификации текстовых	2. Методы выявления дубликатов и тема-
данных	тического упорядочения текстов.

6.1. Примерные задания для самостоятельной работы

- 1. Вы являетесь аналитиком в банке и в вашем распоряжении имеются клиентские истории, содержащие описание целей получения кредита. Необходимо:
 - выделить категории целей кредитования;
 - провести сегментацию клиентов по выделенным категориям кредитования.
- 2. Для фрагмента текста на естественном языке выполнить его морфологический разбор.
- 3. Для фрагмента текста на естественном языке выполнить его синтаксический разбор.
- 4. Для фрагмента текста на естественном языке построить семантическую сеть и карту понятий

РАЗДЕЛ 7. ОЦЕНОЧНЫЕ И МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЕ (ФОНД ОЦЕНОЧНЫХ СРЕДСТВ) ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

7.1. Форма промежуточной аттестации обучающегося по учебной дисциплине.

Контрольным мероприятием промежуточной аттестации обучающихся по учебной дисциплине «Введение в обработку естественного языка» в 8 семестре является экзамен, который проводится в устной форме.

Таблица 7.1

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ В СООТНОШЕНИИ С ОПЕНОЧНЫМИ СРЕДСТВАМИ

Планируемые результаты, характеризующие этапы формирования компетенции	Содержание учебного материала	Примеры контрольных вопросов и заданий для оценки знаний, умений, владений	Методы/ средства контроля
--	----------------------------------	--	---------------------------------

ИПК-1.1 Знать принципы поархитектуры строения программного обеспечения и виды архитектуры программного обеспечения; типовые решения, библиотеки программных модулей, шаблоны, классы объектов, используемые при разработке программного обеспечения; методы и средства проектирования программного обеспечения; методы и средства проектирования баз данных; методы и средства проектирования программных интерфейсов

Раздел №1 «Методы лингвистического анализа текстов» Тема 1.1 Этапы анализа текста на естественном языке Раздел №2 «Модели компьютерного представления текстов» Тема 2.1 Модели представле-

Тема 2.1 Модели представления текстов Раздел №3 «Методы класси-

фикации»
Тема 3.1 Методы классификации полнотекстовых документов

Раздел №4 «Методы кластерного анализа»

Тема 4.1 Методы кластерного анализа полнотекстовых документов

Раздел №5 «Нейросетевые технологии обработки текстовых данных»

Тема 5.1 Модель word2vec

Тема 5.2 Модель doc2vec

Тема 5.3 Генеративные модели естественного языка:

BERT

Тема 5.4 Генеративные модели естественного языка: GPT-2, GPT-3

Раздел №6 «Прикладные задачи анализа текстов» Тема 6.1 Методы очистки текста

Тема 6.2 Комплексная технология классификации текстовых данных

Раздел №1 «Методы лингвистического анализа текстов» Тема 1.1 Этапы анализа текста на естественном языке Раздел №2 «Модели компьютерного представления текстов»

Тема 2.1 Модели представления текстов

Раздел №3 «Методы классификации»

Тема 3.1 Методы классификации полнотекстовых документов

Раздел №4 «Методы кластерного анализа»

Тема 4.1 Методы кластерного анализа полнотекстовых документов

Раздел №5 «Нейросетевые технологии обработки текстовых данных»

Тема 5.1 Модель word2vec

Тема 5.2 Модель doc2vec

Тема 5.3 Генеративные модели естественного языка: BERT

Тема 5.4 Генеративные модели естественного языка: GPT-

2, GPT-3 Раздел №6 «Прикладные задачи анализа текстов» Тема 6.1 Методы очистки текста 1. Основные понятия и определения компьютерной лингвистики.

- 2. Задачи, направления компьютерной лингвистики.
- 3. Анализ текста. Сложности и особенности автоматического анализа текста на естественном языке.
- 4. Словарный и предиктивный морфологический анализ текста.
- 5. Словарный и бессловарный морфологический анализ текста.

Устный контроль/ опрос на семинарских занятиях, зачете, экзамене:

анализ докладов на семинарских занятиях; анализ защиты проектов; применение теоретических знаний при анализе (разборе) конкретных практикоориентированных ситуаций и профессионально-прикладных задач, анализ использования теоретических знаний в процессе решения кейсов, в ходе деловых игр;

письменный контроль, анализ содержания эссе; тестирование (выполнение тестовых заданий)

ИПК-1.2 Уметь использовать существующие типовые решения и шаблоны проектирования программного обеспечения; применять методы и средства проектирования программного обеспечения, структур данных, баз данных, программных интерфейсов; осуществлять коммуникации с заинтересованными сторонами

1. Инструментарий для морфологического анализа и синтеза текста.

- 2. Автоматизированное снятие омонимии. Постморфологический анализ.
- 3. Автоматизированное снятие омонимии. Синтаксическая сегмента-
- 4. Синтаксическая неоднозначность. Подходы к описанию синтаксиса в естественном языке.
- Понятие онтологии.
 Онтологические ресурсы и компьютерные тезаурусы.
- 6. Модели представления знаний в компьютерной семантике.

Анализ проявленных умений при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионально-ориентированных эссе; тести-

ориентированных эссе; тестирование (выполнение тестовых заданий); анализ защит профессионально-ориентированных проектов;

опрос на семинарских занятиях, зачете,

анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ решения конкретных практикоориентированных ситуаций и профессионально-прикладных задач, анализ выполнения контрольных работ

ИПК-1.3 Владеть разработкой, изменением и согласованием архитектуры программного обеспечения с системным аналитиком и архитектором программного обеспечения; системой проектирования структур данных; системой проектирования баз данных; системой проектирования программных интерфейсов; оценкой и согласованием сроков выполнения поставленных задач	Тема 6.2 Комплексная технология классификации текстовых данных Раздел №1 «Методы лингвистического анализа текстов» Тема 1.1 Этапы анализа текста на естественном языке Раздел №2 «Модели компьютерного представления текстов» Тема 2.1 Модели представления текстов» Тема 2.1 Модели представления текстов Раздел №3 «Методы классификации» Тема 3.1 Методы классификации полнотекстовых документов Раздел №4 «Методы кластерного анализа» Тема 4.1 Методы кластерного анализа полнотекстовых документов Раздел №5 «Нейросетевые технологии обработки текстовых данных» Тема 5.1 Модель word2vec Тема 5.2 Модель doc2vec Тема 5.3 Генеративные модели естественного языка: ВЕКТ Тема 5.4 Генеративные модели естественного языка: GPT-2, GPT-3 Раздел №6 «Прикладные задачи анализа текстов» Тема 6.1 Методы очистки текста Тема 6.2 Комплексная технология классификации текстовых данных	1. Классификация полнотекстовых документов. 2. Классификация полнотекстовых документов. Алгоритмы классификации с учителем. Алгоритм «наивной» байесовской классификации. 3. Классификация полнотекстовых документов. Алгоритмы классификации с учителем. Оценка результатов классификации. 4. Кластеризация и контент-анализа. Инструментальные средства кластеризации и контент-анализа. 5. Программные средства лингвистической обработки. 6. Системы автоматической обработки текстов на естественном языке. 7. Архитектура инструментальных систем для автоматической обработки и анализа текста на естественном языке.	Анализ проявленных навыков при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионально-ориентированных эссе; тестирование (выполнение тестовых заданий); анализ защит профессионально-ориентированных проектов; опрос на семинарских занятиях, зачете, экзамене; анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ решения конкретных практико-ориентированных ситуаций и профессиональноприкладных задач, анализ выполнения контрольных работ
ПК-2. Способность осуществлять и		і держку планирования управ	и вления требованиями по со-
зданию (модификации) и сопровож	кдению ИС		
ИПК-2.1. Знать инструменты и	Раздел №1 «Методы лингви-	1. Основные понятия и	Устный контроль/ опрос на
методы управления требования-	стического анализа текстов»	определения компью-	семинарских занятиях, зачете,
ми; предметной областью автоматизации; возможностями ИС;	Тема 1.1 Этапы анализа тек- ста на естественном языке	терной лингвистики. 2. Задачи, направления	экзамене;
источниками информации, необ-	Раздел №2 «Модели компью-	компьютерной лингви-	анализ докладов на семинар- ских занятиях; анализ защиты
ходимой для профессиональной	терного представления тек-	стики.	рефератов; анализ защиты про-
деятельности; современный оте-	стов»	3. Анализ текста. Слож-	ектов; применение теоретиче-
чественный и зарубежный опыт	Тема 2.1 Модели представле-	ности и особенности	ских знаний при анализе (раз-
в профессиональной деятельно-	ния текстов	автоматического анали-	боре) конкретных практико-
сти; инструменты и методы мо-	Раздел №3 «Методы класси-	за текста на естествен-	ориентированных ситуаций и

делирования бизнес-процессов организации; основы реинжиниринга бизнес-процессов организации; управление содержанием проекта: документирование требований, систему анализа продукта, мо-дерируемых совещаний; управление качеством: контрольные списки, вери-фикация, валидация (приемо-сдаточные испытания); управление коммуни-кациями в проекте: базовые навыки управления (в том числе проведение презентаций, проведение переговоров, публичные выступления); культуру речи; правила деловой переписки.

фикации» Тема 3.1 Методы классификации полнотекстовых документов Раздел №4 «Методы кластерного анализа» Тема 4.1 Методы кластерного анализа полнотекстовых документов Раздел №5 «Нейросетевые технологии обработки текстовых данных» Тема 5.1 Модель word2vec Тема 5.2 Модель doc2vec Тема 5.3 Генеративные модели естественного языка: **BERT** Тема 5.4 Генеративные модели естественного языка: GPT-

2, GPT-3

- ном языке.
- 4. Словарный и предиктивный морфологический анализ текста.
- 5. Словарный и бессловарный морфологический анализ текста.

профессионально-прикладных задач, анализ использования теоретических знаний в процессе решения кейсов, в ходе деловых игр;

письменный контроль, анализ содержания эссе; тестирование (выполнение тестовых заданий)

	Раздел №6 «Прикладные задачи анализа текстов» Тема 6.1 Методы очистки текста Тема 6.2 Комплексная технология классификации текстовых данных		
ИПК-2.2. Умеет анализировать входные данные; планировать работы.	Раздел №1 «Методы лингвистического анализа текстов» Тема 1.1 Этапы анализа текстов» Тема 1.1 Этапы анализа текста на естественном языке Раздел №2 «Модели компьютерного представления текстов» Тема 2.1 Модели представления текстов Раздел №3 «Методы классификации» Тема 3.1 Методы классификации полнотекстовых документов Раздел №4 «Методы кластерного анализа» Тема 4.1 Методы кластерного анализа полнотекстовых документов Раздел №5 «Нейросетевые технологии обработки текстовых данных» Тема 5.1 Модель word2vec Тема 5.2 Модель doc2vec Тема 5.3 Генеративные модели естественного языка: ВЕКТ Тема 5.4 Генеративные модели естественного языка: GPT-2, GPT-3 Раздел №6 «Прикладные задачи анализа текстов» Тема 6.1 Методы очистки текста Тема 6.2 Комплексная технология классификации текстовых данных	1. Инструментарий для морфологического анализа и синтеза текста. 2. Автоматизированное снятие омонимии. Постморфологический анализ. 3. Автоматизированное снятие омонимии. Синтаксическая сегментация. 4. Синтаксическая неоднозначность. Подходы к описанию синтаксиса в естественном языке. 5. Понятие онтологии. Онтологические ресурсы и компьютерные тезаурусы. 6. Модели представления знаний в компьютерной семантике.	Анализ проявленных умений при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионально-ориентированных эссе; тестирование (выполнение тестовых заданий); анализ защит профессионально-ориентированных проектов; опрос на семинарских занятиях, зачете, анализ докладов на семинарских занятиях; анализ рефератов; анализ решения конкретных практико-ориентированных ситуаций и профессионально-прикладных задач, анализ выполнения контрольных работ
ИПК-2.3. Владеть системным подходом для выбора технологии управления требованиями; представлениями исходных данных для разработки пла-на управления требованиями; системой согласования в части инженерно-технологического обеспечения плана управления требованиями с заинтересованными сторонами	Раздел №1 «Методы лингвистического анализа текстов» Тема 1.1 Этапы анализа текстов анализа текстов ана естественном языке Раздел №2 «Модели компьютерного представления текстов» Тема 2.1 Модели представления текстов анализа «Методы классификации» Тема 3.1 Методы классификации полнотекстовых документов Раздел №4 «Методы кластерного анализа» Тема 4.1 Методы кластерного анализа полнотекстовых документов Раздел №5 «Нейросетевые технологии обработки текстовых данных» Тема 5.1 Модель word2vec Тема 5.2 Модель фос2vec Тема 5.3 Генеративные модели естественного языка: ВЕКТ Тема 5.4 Генеративные моде-	1. Классификация полнотекстовых документов. 2. Классификация полнотекстовых документов. Алгоритмы классификации с учителем. Алгоритм «наивной» байесовской классификации. 3. Классификация полнотекстовых документов. Алгоритмы классификации с учителем. Оценка результатов классификации. 4. Кластеризация и контент-анализ. Инструментальные средства кластеризации и контент-анализа. 5. Программные средства лингвистической обработки. 6. Системы автоматической обработки текстов на естественном языке.	Анализ проявленных навыков при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионально-ориентированных эссе; тестирование (выполнение тестовых заданий); анализ защит профессионально-ориентированных проектов; опрос на семинарских занятиях, зачете, экзамене; анализ докладов на семинарских занятиях; анализ защиты рефератов; анализ решения конкретных практико-ориентированных ситуаций и профессионально-прикладных задач, анализ выполнения контрольных работ

ли естественного языка: GPT-2, GPT-3
Раздел №6 «Прикладные задачи анализа текстов»
Тема 6.1 Методы очистки текста
Тема 6.2 Комплексная технология классификации тексто-

вых данных

7. Архитектура инструментальных систем для автоматической обработки и анализа текста на естественном языке.

ПК-3. Способность приме-ять системный подход для составления технико-экономического обоснования проектных решений и технического задания на разработку информационной системы

ИПК3.1. Знает возможности типовой ИС; предметную область автоматизации; методы выявления требований: технологии межличностной и групповой коммуникации в де-ловом взаимодействии, основы конфликтологии; технологии подготовки и проведения презентаций; архитектуру, устройство и функционирование вычислительных систем; коммуникационное оборудование; сетевые протоколы; основы современных операционных систем; основы современных систем управления базами данных; устройство и функционирование современных ИС; современные стандарты инфорвзаимодействия машионного систем; программные средства и платформы инфраструктуры информационных технологий организаций; современные подходы и стандарты автоматизации организации (например, CRM, MRP, ERP..., ITIL, ITSM); основы теории систем и системного анализа; методики описания и моделирования процессов, средства моделирования бизнес-процессов; системы классификации и кодирования информации, в том числе присвоение кодов документам и элементам справочников; отраслевая нормативная техническая документация; источники информации, необходимой для профессиональной деятельности; современный отечественный и зарубежный опыт в профессиональной деятельности; формирование и механизмы рыночных процессов организации; основы менеджмента, в том числе менеджмента качества; основы бухгалтерского учета и отчетности организаций; основы налогового законодательства Российской Федерации; основы управленческого учета; основы финансового учета и бюджетирования; основы международных стандартов финансовой отчетности (МСФО); основы управления торгов-лей, поставками и запасами; основы организации производства; основы управления взаимоотношениями с клиентами и заказчиками (CRM); основы

информационной системы
Раздел №1 «Методы лингвистического анализа текстов»
Тема 1.1 Этапы анализа текста на естественном языке
Раздел №2 «Модели компьютерного представления текстов»
Тема 2.1 Модели представления текстов
Раздел №3 «Методы классификации»
Тема 3.1 Методы классифи-

Тема 3.1 Методы классификации полнотекстовых документов

Раздел №4 «Методы кластерного анализа»

Тема 4.1 Методы кластерного анализа полнотекстовых документов

Раздел №5 «Нейросетевые технологии обработки текстовых данных»

Tема 5.1 Модель word2vec Tема 5.2 Модель doc2vec

Тема 5.3 Генеративные модели естественного языка: ВЕЯТ

Тема 5.4 Генеративные модели естественного языка: GPT-2. GPT-3

Раздел №6 «Прикладные задачи анализа текстов» Тема 6.1 Методы очистки текста

Тема 6.2 Комплексная технология классификации текстовых ланных

1. Основные понятия и определения компьютерной лингвистики.

- 2. Задачи, направления компьютерной лингвистики.
- 3. Анализ текста. Сложности и особенности автоматического анализа текста на естественном языке.
- 4. Словарный и предиктивный морфологический анализ текста.
- 5. Словарный и бессловарный морфологический анализ текста.

Устный контроль/ опрос на семинарских занятиях, зачете, экзамене;

анализ докладов на семинарских занятиях; анализ защиты проектов; применение теоретических знаний при анализе (разборе) конкретных практикоориентированных ситуаций и профессионально-прикладных задач, анализ использования теоретических знаний в процессе решения кейсов, в ходе деловых игр;

письменный контроль, анализ содержания эссе; тестирование (выполнение тестовых заданий)

управления срочным и метом управления сумунаминия сороженным выструменты и метом управления организациям, в том месте от произволяет в седения документоборога в организациям постать прогодствения проучения документоборога в организация управлениям постать прогодствения продукта, документоров предысающие управления продукта, документорования организация, управления содержанием простат, документорования продукта, документорования организация, управления организация орга				
ИПК 3.3. Владеет основами выявления первоначальных требованиям первоначальных требованиям возможностях типовой ИС и вариантах ее модификации; определениями возможностей достижения соответствия ИС первоначальным требованиям требованиям требованиям представления тектов первоначальным требованиям представления тектов первоначальным требованиям представления тектов первоначальным требованиям представления полнотекстовых докумен при решении кейсов, в ходе при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессионального байесовской классифи нально-ориентированных эссе;	вопросы оплаты труда; основы теории управлния; современные инструменты и методы управления организацией, в том числе методы планирования деятельности, распределения поручений, контроля исполнения, принятия решений; методология ведения документооборота в организациях; инструменты и методы определения финансовых и производственных показателей деятельности организаций; основы организационной диагностики; инструменты и методы моделирования бизнес-процессов организации; основы реинжиниринга бизнес-процессов организации; управление содержанием проекта документирование требований, анализ продукта, модерируемые совещания; культура речи; правила деловой переписки. ИПК-3.2 Умеет проводить переговоры; проводить презентации; подготавливать протоколы ме-	стического анализа текстов» Тема 1.1 Этапы анализа текста на естественном языке Раздел №2 «Модели компьютерного представления текстов» Тема 2.1 Модели представления текстов Раздел №3 «Методы классификации» Тема 3.1 Методы классификации полнотекстовых документов Раздел №4 «Методы кластерного анализа» Тема 4.1 Методы кластерного анализа полнотекстовых документов Раздел №5 «Нейросетевые технологии обработки текстовых данных» Тема 5.1 Модель word2vec Тема 5.2 Модель doc2vec Тема 5.3 Генеративные модели естественного языка: ВЕКТ Тема 5.4 Генеративные модели естественного языка: GPT-2, GPT-3 Раздел №6 «Прикладные задачи анализа текстов» Тема 6.1 Методы очистки текста Тема 6.2 Комплексная технология классификации тексто-	морфологического анализа и синтеза текста. 2. Автоматизированное снятие омонимии. Постморфологический анализ. 3. Автоматизированное снятие омонимии. Синтаксическая сегментация. 4. Синтаксическая неоднозначность. Подходы к описанию синтаксиса в естественном языке. 5. Понятие онтологии. Онтологические ресурсы и компьютерные тезаурусы. 6. Модели представления знаний в компью-	при решении кейсов, в ходе деловых игр; письменный контроль, анализ качества решений профессиональных задач в контрольных работах; анализ содержания профессиональноориентированных эссе; тестирование (выполнение тестовых заданий); анализ защит профессиональноориентированных проектов; опрос на семинарских занятиях, зачете, анализ докладов на семинарских занятиях; анализ рефератов; анализ решения конкретных практикоориентированных ситуаций и профессионально-прикладных задач, анализ выполнения кон-
ваний заказчика к ИС; системой информирования заказчика о возможностях типовой ИС и вариантах ее модификации; определениями возможностей достижения соответствия ИС первоначальным требованиям потекстов информирования префессионнальных задач в контрольных обайесовской классифинально-ориентированных эссе; тов. Тема 1.1 Этапы анализа тектов. Тема 1.1 Этапы анализа тектов. 2. Классификация полнотекстовых документов. Алгоритмы классификации с учителем. 4. Алгоритм «наивной» байесовской классифинально-ориентированных эссе;			~	•
информирования заказчика о возможностях типовой ИС и вариантах ее модификации; определениями возможностей достижения соответствия ИС первоначальным требованиям				
возможностях типовой ИС и Ваздел №2 «Модели компьювариантах ее модификации; терного представления текопределениями возможностей достижения соответствия ИС первоначальным требованиям ния текстов новых документов. Алгоритмы классификации с учителем. Алгоритм «наивной» байесовской классифинально-ориентированных эссе;				
вариантах ее модификации; терного представления текопределениями возможностей достижения соответствия ИС первоначальным требованиям ния текстов тов. Алгоритмы классифинальных задач в контрольных фикации с учителем. Алгоритм «наивной» анализ содержания профессиобайесовской классифинально-ориентированных эссе;			~	
определениями возможностей стов» фикации с учителем. достижения соответствия ИС тема 2.1 Модели представленервоначальным требованиям ния текстов фикации с учителем. Алгоритм «наивной» анализ содержания профессионально-ориентированных эссе;				
достижения соответствия ИС Тема 2.1 Модели представле- Алгоритм «наивной» анализ содержания профессио- первоначальным требованиям ния текстов байесовской классифи- нально-ориентированных эссе;			_	
первоначальным требованиям ния текстов байесовской классифи- нально-ориентированных эссе;	* ' '			~
	1 1 1	_		
I D 302 3 C			_	
заказчика; основами составления Раздел №3 «Методы класси- кации. тестирование (выполнение те-	заказчика; основами составления	Раздел №3 «Методы класси-	кации.	тестирование (выполнение те-

протоколов переговоров с заказфикании» 3. Классификация полстовых заданий); Тема 3.1 Методы классифинотекстовых докуменанализ защит профессиональнокации полнотекстовых докутов. Алгоритмы классиориентированных проектов; опрос на семинарских занятиях, ментов фикации с учителем. Раздел №4 «Методы кластер-Оценка результатов зачете, экзамене; ного анализа» классификации. анализ докладов на семинар-Тема 4.1 Методы кластерного 4. Кластеризация и конских занятиях: анализа полнотекстовых дотент-анализ. Инструанализ защиты рефератов; анакументов ментальные средства лиз решения конкретных прак-Раздел №5 «Нейросетевые кластеризации и контико-ориентированных ситуатехнологии обработки тектент-анализа. ций и профессиональностовых данных» 5. Программные средприкладных задач, анализ вы-Тема 5.1 Молель word2vec ства лингвистической полнения контрольных работ Тема 5.2 Модель doc2vec обработки. Тема 5.3 Генеративные моде-6. Системы автоматили естественного языка: ческой обработки тек-**BERT** стов на естественном Тема 5.4 Генеративные модеязыке. ли естественного языка: GPT-7. Архитектура инстру-2, GPT-3 ментальных систем для Раздел №6 «Прикладные заавтоматической обрадачи анализа текстов» ботки и анализа текста Тема 6.1 Методы очистки на естественном языке. Тема 6.2 Комплексная технология классификации текстовых данных

7.2.Перечень вопросов для подготовки к промежуточной аттестации (экзамен)

- 1. Основные понятия и определения компьютерной лингвистики.
- 2. Задачи, направления компьютерной лингвистики.
- 3. Анализ текста. Сложности и особенности автоматического анализа текста на естественном языке.
- 4. Словарный и предиктивный морфологический анализ текста.
- 5. Словарный и бессловарный морфологический анализ текста.
- 6. Инструментарий для морфологического анализа и синтеза текста.
- 7. Автоматизированное снятие омонимии. Постморфологический анализ.
- 8. Автоматизированное снятие омонимии. Синтаксическая сегментация.
- 9. Синтаксическая неоднозначность. Подходы к описанию синтаксиса в естественном языке.
- 10. Понятие онтологии. Онтологические ресурсы и компьютерные тезаурусы.
- 11. Модели представления знаний в компьютерной семантике.
- 12. Классификация полнотекстовых документов.
- 13. Классификация полнотекстовых документов. Алгоритмы классификации с учителем. Алгоритм «наивной» байесовской классификации.
- 14. Классификация полнотекстовых документов. Алгоритмы классификации с учителем. Оценка результатов классификации.
- 15. Кластеризация и контент-анализ. Инструментальные средства кластеризации и контент-анализа.
- 16. Программные средства лингвистической обработки.
- 17. Системы автоматической обработки текстов на естественном языке.
- 18. Архитектура инструментальных систем для автоматической обработки и анализа текста на естественном языке.

7.3. Описание показателей и критериев оценивания сформированности компетенций на различных этапах их формирования; шкалы и процедуры оценивания

7.3.1. Оценивание ответов на вопросы и выполнения заданий для текущей и промежуточной аттестации

При оценке знаний учитывается уровень сформированности компетенций:

- 1. Уровень усвоения теоретических положений дисциплины, правильность формулировки основных понятий и закономерностей.
 - 2. Уровень знания фактического материала в объеме программы.
 - 3. Логика, структура и грамотность изложения вопроса.
 - 4. Умение связать теорию с практикой.

Шкала оценивания на экзамене

Оценка	Критерии выставления оценки		
Отлично	Обучающийся должен:		
	- продемонстрировать глубокое и прочное усвоение знаний программного мате		
	риала;		
	- исчерпывающе, последовательно, грамотно и логически стройно изложить тео-		
	ретический материал;		
	- правильно формулировать определения;		
	- продемонстрировать умения самостоятельной работы с литературой;		
	- уметь сделать выводы по излагаемому материалу.		
Хорошо	Обучающийся должен:		
	- продемонстрировать достаточно полное знание программного материала;		
	- продемонстрировать знание основных теоретических понятий;		
	достаточно последовательно, грамотно и логически стройно излагать материал;		
	- продемонстрировать умение ориентироваться в литературе;		
	- уметь сделать достаточно обоснованные выводы по излагаемому материалу.		
Удовлетворительно	Обучающийся должен:		
	- продемонстрировать общее знание изучаемого материала;		
	- показать общее владение понятийным аппаратом дисциплины;		
	- уметь строить ответ в соответствии со структурой излагаемого вопроса;		
	- знать основную рекомендуемую программой учебную литературу.		
Неудовлетворительно	Обучающийся демонстрирует:		
	- незнание значительной части программного материала;- не владение понятийным аппаратом дисциплины;- существенные ошибки при изложении учебного материала;		
	- неумение строить ответ в соответствии со структурой излагаемого вопроса		
	- неумение делать выводы по излагаемому материалу.		

7.4. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ОПРЕДЕЛЯЮЩИЕ ПРОЦЕДУРЫ ОЦЕНИ-ВАНИЯ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ, ХАРАК-ТЕРИЗУЮЩИХ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ.

Качество знаний характеризуется способностью обучающегося точно, структурированно и уместно воспроизводить информацию, полученную в процессе освоения дисциплины, в том виде, в котором она была изложена в учебном издании или преподавателем.

Умения, как правило, формируются на занятиях семинарского типа. Задания, направленные на оценку умений, в значительной степени требуют от обучающегося проявления стереотипности мышления, т.е. способности выполнить работу по образцам, с которыми он работал в процессе обучения. Преподаватель же оценивает своевременность и правильность выполнения задания.

Навыки можно трактовать как автоматизированные умения, развитые и закрепленные осознанным самостоятельным трудом. Навыки формируются при самостоятельном выполнении обучающимися практико-ориентированных заданий, моделирующих решение им производственных и социокультурных задач в соответствующей области профессиональной деятельности, как правило, при выполнении домашних заданий, курсовых проектов (работ), научно-исследовательских работ, прохождении практик, при работе индивидуально или в составе группы и т.д.

Устный опрос — это процедура, организованная как специальная беседа преподавателя с группой обучающихся (фронтальный опрос) или с отдельными обучающимися (индивидуальный опрос) с целью оценки сформированности у них основных понятий и усвоения учебного материала. Устный опрос может использоваться как вид контроля и метод оценивания формируемых компетенций (как и качества их формирования) в рамках самых разных форм контроля, таких как: собеседование, коллоквиум, зачет, экзамен по дисциплине. Устный опрос (УО) позволяет оценить знания и кругозор обучающегося, умение логически построить ответ, владение монологической речью и иные коммуникативные навыки. УО обладает большими возможностями воспитательного воздействия преподавателя. Воспитательная функция УО имеет ряд важных аспектов: профессионально-этический и нравственный аспекты, дидактический (систематизация материала при ответе, лучшее запоминание материала при интеллектуальной концен-

трации), эмоциональный (радость от успешного прохождения собеседования) и др. Обучающая функция УО состоит в выявлении деталей, которые по каким-то причинам оказались недостаточно осмысленными в ходе учебных занятий и при подготовке к зачёту или экзамену. УО обладает также мотивирующей функцией: правильно организованные собеседование, коллоквиум, зачёт и экзамен могут стимулировать учебную деятельность студента, его участие в научной работе.

Тесты являются простейшей формой контроля, направленной на проверку владения терминологическим аппаратом, современными информационными технологиями и конкретными знаниями в области фундаментальных и прикладных дисциплин. Тест может предоставлять возможность выбора из перечня ответов (один или несколько правильных ответов).

Семинарские занятия. Основное назначение семинарских занятий по дисциплине – обеспечить глубокое усвоение обучающимися материалов лекций, прививать навыки самостоятельной работы с литературой, воспитывать умение находить оптимальные решения в условиях изменяющихся отношений, формировать современное профессиональное мышление обучающихся. На семинарских занятиях преподаватель проверяет выполнение самостоятельных заданий и качество усвоения знаний, умений, определяет уровень сформированности компетенций.

Коллоквиум может служить формой не только проверки, но и повышения производительности труда студентов. На коллоквиумах обсуждаются отдельные части, разделы, темы, вопросы изучаемого курса, обычно не включаемые в тематику семинарских и других практических учебных занятий, а также рефераты, проекты и иные работы обучающихся.

Доклад, сообщение – продукт самостоятельной работы студента, представляющий собой публичное выступление по представлению полученных результатов решения определенной учебнопрактической, учебно-исследовательской или научной темы.

Контрольная работа - средство проверки умений применять полученные знания для решения задач определенного типа по теме или разделу.

Профессионально-ориентированное эссе — это средство, позволяющее оценить умение обучающегося письменно излагать суть поставленной проблемы, самостоятельно проводить анализ этой проблемы с использованием аналитического инструментария соответствующей дисциплины, делать выводы, обобщающие авторскую позицию по поставленной профессионально-ориентированной проблеме.

Реферат – продукт самостоятельной работы студента, представляющий собой краткое изложение в письменном виде полученных результатов теоретического анализа определенной научной (учебно-исследовательской) темы, где автор раскрывает суть исследуемой проблемы, приводит различные точки зрения, а также собственные взгляды на нее.

Ситуационный анализ — это комплексный анализ ситуации, имевший место в реальной практике профессиональной деятельности специалистов. Комплексный анализ включает в себя следующие составляющие: причинно-следственный анализ (установление причин, которые привели к возникновению данной ситуации, и следствий ее развертывания), системный анализ (определение сущностных предметно-содержательных характеристик, структуры ситуации, ее функций и др.), ценностно-мотивационный анализ (построение системы оценок ситуации, ее составляющих, выявление мотивов, установок, позиций действующих лиц); прогностический анализ (разработка перспектив развития событий по позитивному и негативному сценарию), рекомендательный анализ (выработка рекомендаций относительно поведения действующих лиц ситуации), программно-целевой анализ (разработка программ деятельности для разрешения данной ситуации).

Творческое задание — это частично регламентированное задание, имеющее нестандартное решение и позволяющее диагностировать умения интегрировать знания различных научных областей, аргументировать собственную точку зрения, доказывать правильность своей позиции. Может выполняться в индивидуальном порядке или группой обучающихся.

Деловая и/или ролевая игра — совместная деятельность группы обучающихся и преподавателя под управлением преподавателя с целью решения учебных и профессионально-ориентированных задач путем игрового моделирования реальной проблемной ситуации. Позволяет оценивать умение анализировать и решать типичные профессиональные задачи.

«Круглый стол», дискуссия — интерактивные оценочные средства, позволяющие включить обучающихся в процесс обсуждения спорного вопроса, проблемы и оценить их умение аргументировать собственную точку зрения. Занятие может проводить по традиционной (контактной) технологии, либо с использованием телекоммуникационных технологий.

Проект – конечный профессионально-ориентированный продукт, получаемый в результате планирования и выполнения комплекса учебных и исследовательских заданий. Позволяет оценить умения обучающихся самостоятельно конструировать свои знания в процессе решения практических задач и проблем, ориентироваться в информационном пространстве и уровень сформированности аналитиче-

ских, исследовательских навыков, навыков практического и творческого мышления. Может выполняться в индивидуальном порядке или группой обучающихся.

РАЗДЕЛ 8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Для успешного освоения дисциплины студенту необходимо посещать все контактные занятия и систематически в полном объеме выполнять все задания для самостоятельной работы.

Во время лекций рекомендуется вести записи: выделять основные понятия, факты, выводы. Если какоето объяснение кажется непонятным, следует немедленно задать вопрос преподавателю. Для формирования необходимых компетенций рекомендуется принимать активное участие в обсуждении ставящихся перед аудиторией вопросов.

При подготовке к лабораторным занятиям необходимо тщательно изучить теоретический и методический материал, изложенный в лекциях.

Самостоятельная работа студентов является одной из основных форм внеаудиторной работы при реализации учебных планов и программ. Для реализации способностей и более глубокого освоения дисциплины предусмотрены следующие виды самостоятельной работы: текущая и проблемно - ориентированная.

Текущая работа по освоению дисциплины, направленная на углубление и закрепление знаний студента, развитие практических умений включает:

- работу с лекционным материалом, поиск и обзор литературы и электронных источников информации по индивидуальному заданию;
- опережающую самостоятельную работу;
- изучение вопросов, вынесенных на самостоятельную проработку;
- подготовку к лабораторным работам;
- подготовку к экзамену.

Начинать изучение дисциплины необходимо с ознакомления с целями и задачами дисциплины. В процессе учебы студенты используют ранее полученные и приобретенные знания и умения. Далее следует проработать отдельные вопросы по предложенным источникам литературы. Все неясные вопросы по дисциплине обучающийся может разрешить на консультациях, проводимых по расписанию. При подготовке к лабораторным занятиям обучающийся в обязательном порядке изучает теоретический материал в соответствии с перечнем основной учебной литературы и методическими указаниями.

Литература для самостоятельной работы:

- 1. Боярский, К. К. Введение в компьютерную лингвистику: учебное пособие / К. К. Боярский. Санкт-Петербург: Университет ИТМО, 2013. 73 с. Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. URL: https://www.IPRsmart hop.ru/71485.html
- 2. Лукашевич, Н. В. Тезаурусы в задачах информационного поиска: монография / Н. В. Лукашевич. Москва: Московский государственный университет имени М.В. Ломоносова, 2011. 512 с. ISBN 978-5-211-05926-9. Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. URL: https://www.IPRsmart hop.ru/13346.html

РАЗДЕЛ 9. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ *Основная литература*¹

- 1. Батура, Т. В. Математическая лингвистика и автоматическая обработка текстов на естественном языке : учебное пособие / Т. В. Батура. Новосибирск : Новосибирский государственный университет, 2016. 166 с. ISBN 978-5-4437-0548-4. Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. URL: https://www.IPRsmart hop.ru/93489.html
- 2. Боярский, К. К. Введение в компьютерную лингвистику: учебное пособие / К. К. Боярский. Санкт-Петербург: Университет ИТМО, 2013. 73 с. Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. URL: https://www.IPRsmart hop.ru/71485.html

Дополнительная литература¹

- 3. Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. Ульяновск : Ульяновский государственный технический университет, 2020. 221 с. ISBN 978-5-9795-2088-9. Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. URL: https://www.IPRsmart hop.ru/106136.html
- 4. Лукашевич, Н. В. Тезаурусы в задачах информационного поиска: монография / Н. В. Лукашевич. Москва: Московский государственный университет имени М.В. Ломоносова, 2011. 512 с. ISBN 978-5-211-05926-9. Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. URL: https://www.IPRsmart hop.ru/13346.html

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине: интернет-ресурсы, современные профессиональные базы данных, информационные справочные системы

Интернет-ресурсы

URL: https://www.IPRsmart hop.ru/ — электронно-библиотечная система IPRsmart .

Информационно-справочные и поисковые системы

Справочная правовая система «КонсультантПлюс»: http://www.con-sultant.ru

Современные профессиональные базы данных

URL:http://www.edu.ru/ – библиотека федерального портала «Российское образование»

URL:http://www.prlib.ru – Президентская библиотека

URL:http://www.rusneb.ru – Национальная электронная библиотека

URL:http://elibrary.rsl.ru/ – сайт Российской государственной библиотеки (раздел «Электронная библиотека»)

URL:http://elib.gnpbu.ru/ - сайт Научной педагогической электронной библиотеки им. К.Д. Ушинского

Комплект лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства

Комплект лицензионного программного обеспечения

Microsoft Open Value Subscription для решений Education Solutions № Tr000544893 от 21.10.2020 г. MDE Windows, Microsoft Office и Office Web Apps. (срок действия до 01.11.2023 г.)

Антивирусное программное обеспечение ESET NOD32 Antivirus Business Edition договор № ИС00-006348 от $14.10.2022 \, \Gamma$. (срок действия до $13.10.2025 \, \Gamma$.)

Программное обеспечение «Мираполис» система вебинаров - Лицензионный договор 244/09/16-к от 15.09.2016 (Спецификация к Лицензионному договору 244/09/16-к от 15.09.2016, от 11.05.2022 г.) (срок действия до 10.07.2023 г.)

Электронная информационно-образовательная среда «1С: Университет» договор от 10.09.2018 г. №ПРКТ-18281 (бессрочно)

Информационная система «ПервыйБит» сублицензионный оговор от 06.11.2015 г №009/061115/003 (бессрочно)

Система тестирования Indigo лицензионное соглашение (Договор) от 08.11.2018 г. №Д-54792 (бессрочно)

Информационно-поисковая система «Консультант Плюс» - договор об информационно поддержке от 26.12.2014, (бессрочно)

¹ Из ЭБС

Электронно-библиотечная система IPRsmart лицензионный договор от 01.09.2021 г. №8234/21С (срок действия до 31.08.2024 г.)

Научная электронная библиотека eLIBRARY лицензионный договор SCIENC INDEX № SIO - 3079/2022 от 12.01.2022 г. (срок действия до 27.01.2024 г.)

Свободно распространяемое программное обеспечение

Комплект онлайн сервисов GNU ImageManipulationProgram, свободно распространяемо программное обеспечение

Веб-браузер, Google Ghrome, свободное ПО, ежегодно обновляемое ПО.

Пакет офисных приложений, Office 2016, лицензионное соглашение - Договор №Тг000544893 от 21/10/2020 - 3 года

Пакет офисных приложений, OpenOffice, свободное ПО, ежегодно обновляемое ПО

Просмотр файлов в формате PDF, Adobe Reader, свободно распространяемое ΠO , ежегодно обновляемое ΠO

Просмотр файлов в формате DJV, WinDjView, свободное ПО, ежегодно обновляемое ПО

Файловый архиватор, 7 Zip, свободное ПО, ежегодно обновляемое ПО

Файловый менеджер, Far, свободно распространяемое ПО, ежегодно обновляемое ПО

Anaconda: дистрибутив языков программирования Python и R.

Программное обеспечение отечественного производства:

Программное обеспечение «Мираполис» система вебинаров - Лицензионный договор 244/09/16-к от 15.09.2016 (Спецификация к Лицензионному договору 244/09/16-к от 15.09.2016, от 11.05.2022 г.) (срок действия до 10.07.2023 г.)

Электронная информационно-образовательная среда «1С: Университет» договор от 10.09.2018 г. №ПРКТ-18281 (бессрочно)

Информационная система «ПервыйБит» сублицензионный договор от 06.11.2015 г. №009/061115/003 (бессрочно)

Система тестирования Indigo лицензионное соглашение (Договор) от 08.11.2018 г. №Д-54792 (бессрочно)

Информационно-поисковая система «Консультант Плюс» - договор МИ-ВИП-79717-56/2022 от 23.12.2021 (срок действия до 31.12.2022 г.)

Информационно-поисковая система «Консультант Плюс» - договор об информационно поддержке от 26.12.2014, (бессрочно)

Электронно-библиотечная система IPRsmart лицензионный договор от 01.09.2021 г. №8234/21С (срок действия до 31.08.2024 г.)

Научная электронная библиотека eLIBRARY лицензионный договор SCIENC INDEX № SIO -3079/2022 от $12.01.2022 \, \Gamma$. (срок действия до $27.01.2024 \, \Gamma$.)

РАЗДЕЛ 10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Учебная аудитория для проведения занятий	Оборудование: специализированная мебель (мебель
лекционного типа, занятий семинарского	аудиторная (9 столов, 9 стульев, доска аудиторная),
типа, групповых и индивидуальных кон-	стол преподавателя, стул преподавателя.
сультаций, текущего контроля и промежу-	Технические средства обучения: персональный ком-
точной аттестации	пьютер преподавателя; 9 компьютеров, мультиме-
	дийное оборудование (проектор, экран).
Помещение для самостоятельной работы	Специализированная мебель (10 столов, 10 стульев),
	персональные компьютеры с возможностью подклю-
	чения к сети «Интернет» и обеспечением доступа в
	электронную информационно-образовательную сре-
	ду Университета