

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет компьютерных технологий и прикладной математики



УТВЕРЖДАЮ:

Проректор по учебной работе,
качеству образования – первый
проректор

Хагуров Т.А.

«30» мая 2025 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
К.М.01.08 ТЕХНОЛОГИИ ОБРАБОТКИ ЯЗЫКА, ЗВУКОВЫХ
ДАННЫХ, ВКЛЮЧАЯ РАСПОЗНАВАНИЕ И СИНТЕЗ РЕЧИ**

Направление подготовки 09.03.03 Прикладная информатика

Направленность (профиль) Искусственный интеллект и машинное обучение

Форма обучения очная

Квалификация бакалавр

Краснодар 2025

Рабочая программа дисциплины Технологии обработки языка, звуковых данных, включая распознавание и синтез речи составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 09.03.03 Прикладная информатика, профиль Искусственный интеллект и машинное обучение.

Программу составил(и):

Е.В. Казаковцева, старший преподаватель кафедры анализа данных и искусственного интеллекта



Рабочая программа дисциплины Технологии обработки языка, звуковых данных, включая распознавание и синтез речи утверждена на заседании кафедры анализа данных и искусственного интеллекта протокол №13 от 20.05.2025

Заведующий кафедрой Коваленко А.В.



Утверждена на заседании учебно-методической комиссии факультета компьютерных технологий и прикладной математики протокол №4 от «23» мая 2025 г.

Председатель УМК факультета Коваленко А.В.



Рецензенты:

Трофимов Виктор Маратович, доктор физико-математических наук, старший научный сотрудник, профессор Кафедры информационных систем и программирования ФГБОУ ВО «Кубанский государственный технологический университет»

Попова Елена Витальевна, доктор экономических наук, кандидат физико-математических наук, профессор, Заведующий кафедрой информационных систем Федерального государственного бюджетного образовательного учреждения высшего образования «Кубанский государственный аграрный университет имени И.Т. Трубилина»

1 Цели и задачи изучения дисциплины (модуля)

1.1 Цель освоения дисциплины

Цели изучения дисциплины определены государственным образовательным стандартом высшего образования и соотнесены с общими целями ООП ВО по направлению подготовки 09.03.03 Прикладная информатика, в рамках которой преподается дисциплина.

Цели дисциплины Машинное обучение:

- познакомить студентов с основами обработки естественного языка (NLP);
- познакомить студентов с основами обработки звуковых данных;
- рассмотреть задачу распознавания и синтеза речи и способы её решения.

1.2 Задачи дисциплины:

- изучить основные архитектуры для решения задач NLP и работы с звуковыми данными;
- изучить большие языковые модели;
- научиться работать с библиотеками Python для обработки естественного языка и звуковых данных: gensim, NLTK, PyMorphy, а также с фреймворком PyTorch.

1.3 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Технологии обработки языка, звуковых данных, включая распознавание и синтез речи» относится к вариативной части Блока 1 «Дисциплины (модули)» учебного плана. Для успешного освоения данной дисциплины необходимы знания следующих дисциплин: Б1.О.05 Математический анализ, Б1.О.06 Векторная алгебра, Б1.О.08 Курс теории вероятностей, К.М.01.01 Математические модели нейронных сетей и К.М.01.03 Нейросетевые технологии.

1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Код и наименование индикатора*	Результаты обучения по дисциплине
ПК-4 Способность использовать знание основных методов искусственного интеллекта в последующей профессиональной деятельности в качестве научных сотрудников, преподавателей образовательных организаций высшего образования, инженеров, технологов	
ИПК-4.3 (40.011 А/02.5 Зн.2) Отечественный и международный опыт в области искусственного интеллекта в прикладных областях	Знает: основные достижения в области обработки естественного языка
	Умеет: применять отечественный и международный опыт по разработке систем искусственного интеллекта
	Владеет: необходимыми знаниями отечественного и международного опыта по анализу данных
ИПК-4.10 (40.011 А/02.5 Тд.2) Проведение наблюдений и измерений, составление их описаний и формулировка выводов при проведении исследований в области ИИ в прикладных областях	Знает: как проводить наблюдения и измерения в области обработки естественного языка
	Умеет: формулировать выводы по разработанным на основе ИИ моделям
	Владеет: знаниями о методах проведения измерений качества разработанных моделей
ПК-5 Способен выявить естественнонаучную сущность проблем, возникающих в ходе профессиональной деятельности в области моделирования и анализа сложных естественных и искусственных систем	
ИПК-5.3 (40.011 А/02.5 Зн.2) Отечественный и международный опыт в области	Знает: основные задачи в сфере обработки естественного языка

Код и наименование индикатора*	Результаты обучения по дисциплине
моделирования и анализа сложных естественных и искусственных систем в прикладных областях	Умеет: использовать знания из области обработки естественного языка в практической деятельности
	Владеет: необходимыми знаниями библиотек (gensim, NLTK, PyMorphu и т.д.), с помощью которых можно решать задачи из области обработки естественного языка и звуковых данных
ИПК-5.10 (40.011 А/02.5 Тд.2) Проведение наблюдений и измерений, составление их описаний и формулировка выводов при проведении исследований в области моделирования и анализа сложных естественных и искусственных систем в прикладных областях	Знает: классические и нейросетевые подходы к решению задач обработки языка и звуковых данных
	Умеет: формулировать выводы по разработанным на основе ИИ моделям
	Владеет: знаниями о методах проведения измерений качества разработанных моделей

Результаты обучения по дисциплине достигаются в рамках осуществления всех видов контактной и самостоятельной работы обучающихся в соответствии с утвержденным учебным планом.

Индикаторы достижения компетенций считаются сформированными при достижении соответствующих им результатов обучения.

2. Структура и содержание дисциплины

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 2 зачетных единицы (72 часа), их распределение по видам работ представлено в таблице

Виды работ		Всего Часов	Форма обучения
			Очная
			7 семестр (часы)
Контактная работа, в том числе:		52,3	52,3
Аудиторные занятия (всего):		50	50
занятия лекционного типа		16	16
лабораторные занятия		34	34
практические занятия		-	-
семинарские занятия		-	-
Иная контактная работа:		2,3	2,3
Контроль самостоятельной работы (КСР)		2	2
Промежуточная аттестация (ИКР)		0,3	0,3
Самостоятельная работа, в том числе:		19,8	19,8
Проработка и повторение лекционного материала и материала учебников и учебных пособий		7,8	7,8
Подготовка к текущему контролю		12	12
Контроль:		-	-
Подготовка к экзамену		-	-
Общая трудоёмкость	час.	72	72
	в том числе контактная работа	52,3	52,3
	зач. Ед	2	2

2.2 Содержание дисциплины

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 7 семестре (очная форма обучения)

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	СРС
1	2	3	4	5	6	7
1.	Обработка естественного языка с помощью классических методов машинного обучения	16	4		10	6
2.	Архитектуры нейронных сетей для работы с речью и звуком	14	4		8	2
3.	Большие языковые модели	24	6		14	6
4.	Синтез речи нейросетевыми методами	11,8	2			5,8
ИТОГО по разделам дисциплины		69,8	16		34	19,8
Контроль самостоятельной работы (КСР)		2				
Промежуточная аттестация (ИКР)		0,3				
Подготовка к экзамену		-				
Общая трудоемкость по дисциплине		72				

Примечание: Л – лекции, ПЗ – практические занятия / семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

2.3 Содержание разделов (тем) дисциплины

2.3.1 Занятия лекционного типа

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1.	Обработка естественного языка с помощью классических методов машинного обучения	История развития области NLP Задача классификации текстов Задача классификации последовательности Задача sequence to sequence Предобработка текстовых данных	Р
2.	Архитектуры нейронных сетей для работы с речью и звуком	Популярные оптимизаторы в нейронных сетях (SGD, SGDM, Adam) Обнаружение недо- и переобучения нейронных сетей. Методы борьбы с переобучением Векторное представление слов (Word2Vec, Skip-gram, CBOW, FastText) Рекуррентные и двунаправленные нейронные сети LSTM-сети Популярные архитектуры нейросетей для классификации текстов Популярные архитектуры нейросетей для решения задачи NER Популярные архитектуры нейросетей для машинного перевода	Р
3.	Большие языковые модели	Механизм внимания. Архитектура трансформер Известные трансформеры: GPT и ruGPT, Bert и ruBert Техника Transfer Learning Нейросети для генерации описаний к изображению	Р

4.	Синтез речи нейросетевыми методами	Популярные архитектуры нейросетей для распознавания речи Популярные архитектуры нейросетей для синтеза речи	Р
----	------------------------------------	--	---

2.3.2 Занятия семинарского типа (практические / семинарские занятия/ лабораторные работы)

№	Наименование раздела (темы)	Тематика занятий/работ	Форма текущего контроля
1.	Обработка естественного языка с помощью классических методов машинного обучения	Библиотеки NLTK и PyMorphy для предобработки и анализа текстов Библиотека gensim	ЛР
2.	Архитектуры нейронных сетей для работы с речью и звуком	Возможности фреймворка PyTorch по анализу текстов Возможности фреймворка PyTorch по распознаванию и синтезу речи	ЛР
3.	Большие языковые модели	Возможности фреймворка PyTorch по генерации текстов	ЛР

Защита лабораторной работы (ЛР), выполнение курсового проекта (КП), курсовой работы (КР), расчетно-графического задания (РГЗ), написание реферата (Р), эссе (Э), коллоквиум (К), тестирование (Т) и т.д.

2.3.3 Курсовые работы не предусмотрены

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Проработка и повторение лекционного материала, материала учебной и научной литературы, подготовка к семинарским занятиям	Методические указания для подготовки к лекционным и семинарским занятиям, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г. Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
2	Подготовка к лабораторным занятиям	Методические указания по выполнению лабораторных работ, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
3	Подготовка к решению задач и тестов	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
4	Подготовка к текущему контролю	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
5	Подготовка докладов	Методические указания для подготовки эссе, рефератов, курсовых работ, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. Образовательные технологии, применяемые при освоении дисциплины (модуля)

В соответствии с требованиями ФГОС в программа дисциплины предусматривает использование в учебном процессе следующих образовательные технологии: чтение лекций с использованием мультимедийных технологий; метод малых групп, разбор практических задач и кейсов.

При обучении используются следующие образовательные технологии:

– Технология коммуникативного обучения – направлена на формирование коммуникативной компетентности студентов, которая является базовой, необходимой для адаптации к современным условиям межкультурной коммуникации.

– Технология разноуровневого (дифференцированного) обучения – предполагает осуществление познавательной деятельности студентов с учётом их индивидуальных способностей, возможностей и интересов, поощряя их реализовывать свой творческий потенциал. Создание и использование диагностических тестов является неотъемлемой частью данной технологии.

– Технология модульного обучения – предусматривает деление содержания дисциплины на достаточно автономные разделы (модули), интегрированные в общий курс.

– Информационно-коммуникационные технологии (ИКТ) – расширяют рамки образовательного процесса, повышая его практическую направленность, способствуют интенсификации самостоятельной работы учащихся и повышению познавательной активности. В рамках ИКТ выделяются 2 вида технологий:

– Технология использования компьютерных программ – позволяет эффективно дополнить процесс обучения языку на всех уровнях.

– Интернет-технологии – предоставляют широкие возможности для поиска информации, разработки научных проектов, ведения научных исследований.

– Технология индивидуализации обучения – помогает реализовывать личностно-ориентированный подход, учитывая индивидуальные особенности и потребности учащихся.

– Проектная технология – ориентирована на моделирование социального взаимодействия учащихся с целью решения задачи, которая определяется в рамках профессиональной подготовки, выделяя ту или иную предметную область.

– Технология обучения в сотрудничестве – реализует идею взаимного обучения, осуществляя как индивидуальную, так и коллективную ответственность за решение учебных задач.

– Игровая технология – позволяет развивать навыки рассмотрения ряда возможных способов решения проблем, активизируя мышление студентов и раскрывая личностный потенциал каждого учащегося.

– Технология развития критического мышления – способствует формированию разносторонней личности, способной критически относиться к информации, умению отбирать информацию для решения поставленной задачи.

Комплексное использование в учебном процессе всех вышеназванных технологий стимулируют личностную, интеллектуальную активность, развивают познавательные процессы, способствуют формированию компетенций, которыми должен обладать будущий специалист.

Основные виды интерактивных образовательных технологий включают в себя:

– работа в малых группах (команде) - совместная деятельность студентов в группе под руководством лидера, направленная на решение общей задачи путём творческого сложения результатов индивидуальной работы членов команды с делением полномочий и ответственности;

– проектная технология - индивидуальная или коллективная деятельность по отбору, распределению и систематизации материала по определенной теме, в результате которой составляется проект;

– анализ конкретных ситуаций - анализ реальных проблемных ситуаций, имевших место в соответствующей области профессиональной деятельности, и поиск вариантов лучших решений;

– развитие критического мышления – образовательная деятельность, направленная на развитие у студентов разумного, рефлексивного мышления, способного выдвинуть новые идеи и увидеть новые возможности.

Подход разбора конкретных задач и ситуаций широко используется как преподавателем, так и студентами во время лекций, лабораторных занятий и анализа результатов самостоятельной работы. Это обусловлено тем, что при исследовании и решении каждой конкретной задачи имеется, как правило, несколько методов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

Темы, задания и вопросы для самостоятельной работы призваны сформировать навыки поиска информации, умения самостоятельно расширять и углублять знания, полученные в ходе лекционных и практических занятий.

Подход разбора конкретных ситуаций широко используется как преподавателем, так и студентами при проведении анализа результатов самостоятельной работы.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

4. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Технологии обработки языка, звуковых данных, включая распознавание и синтез речи».

Оценочные средства включает контрольные материалы для проведения **текущего контроля** в форме тестовых заданий и лабораторных работ, и **промежуточной аттестации** в форме вопросов к зачету.

Текущий контроль успеваемости осуществляется в течение семестра, в ходе повседневной учебной работы и предполагает овладение материалами лекций, литературы, программы, работу студентов в ходе проведения лабораторных занятий, а также систематическое выполнение тестовых работ, решение практических задач и иных заданий для самостоятельной работы студентов. Данный вид контроля стимулирует у студентов стремление к систематической самостоятельной работе по изучению дисциплины. Он предназначен для оценки самостоятельной работы слушателей по решению задач, выполнению лабораторных работ, подведения итогов тестирования. Оценивается также активность и качество результатов практической работы на занятиях, участие в дискуссиях, обсуждениях и т.п. Индивидуальные и групповые самостоятельные, аудиторские работы по всем темам дисциплины организованы единообразным образом. Для контроля освоения содержания дисциплины используются оценочные средства. Они направлены на определение степени сформированности компетенций.

Промежуточная аттестация студентов осуществляется в рамках завершения изучения дисциплины и позволяет определить качество усвоения изученного материала, предполагает контроль и управление процессом приобретения студентами необходимых знаний, умения и навыков, определяемых по ФГОС ВО по соответствующему направлению подготовки в качестве результатов освоения учебной дисциплины.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на зачете;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Структура оценочных средств для текущей и промежуточной аттестации

	Результаты обучения	Наименование оценочного средства
--	---------------------	----------------------------------

№ п/п	Код и наименование индикатора		Текущий контроль	Промежуточная аттестация
1	ИПК-5.3 (40.011 А/02.5 Зн.1) Цели и задачи проводимых исследований и разработок в области моделирования и анализа сложных естественных и искусственных систем в прикладных областях	Знает: основные задачи в сфере обработки естественного языка Умеет: использовать знания из области обработки естественного языка в практической деятельности Владеет: необходимыми знаниями библиотек (gensim, NLTK, PyMorphy и т.д.), с помощью которых можно решать задачи из области обработки естественного языка и звуковых данных	Лабораторные работы №1, 2 по разделу «Обработка естественного языка с помощью классических методов машинного обучения», Реферат по темам 1-6	Вопрос на зачете 1-3, 7-10
2	ИПК-4.3 (40.011 А/02.5 Зн.2) Отечественный и международный опыт в области искусственного интеллекта в прикладных областях	Знает: основные достижения в области обработки естественного языка Умеет: применять отечественный и международный опыт по разработке систем искусственного интеллекта Владеет: необходимыми знаниями отечественного и международного опыта по анализу данных	Лабораторная работа 5 по разделу «Большие языковые модели», Реферат по темам 9-12, 13-15	Вопрос на зачете 13-15, 18-21, 23, 24
3	ИПК-4.10 (40.011 А/02.5 Тд.2) Проведение наблюдений и измерений, составление их описаний и формулировка выводов при проведении исследований области искусственного интеллекта в прикладных областях	Знает: как проводить наблюдения и измерения в области обработки естественного языка Умеет: формулировать выводы по разработанным на основе ИИ моделям Владеет: знаниями о методах проведения измерений качества разработанных моделей	Лабораторные работы №3, 4 по разделу «Архитектуры нейронных сетей для работы с речью и звуком», Реферат по темам 7, 8	Вопрос на зачете 4-6, 11, 12, 16, 17, 22
4	ИПК-5.10 (40.011 А/02.5 Тд.2) Проведение наблюдений и измерений, составление их описаний и формулировка	Знает: как проводить наблюдения и измерения в области нейросетевых технологий Умеет: формулировать выводы по	Лабораторная работа 5 по разделу «Большие языковые модели», Реферат по темам 7, 8	Вопрос на зачете 4-6, 11, 12, 16, 17, 22-24

выводов при проведении исследований в области моделирования и анализа сложных естественных и искусственных систем в прикладных областях	разработанным на основе ИИ моделям Владеет: знаниями о методах проведения измерений качества разработанных моделей		
---	--	--	--

Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Темы рефератов

1. История развития области NLP
2. Задача классификации текстов
3. Задача классификации последовательности
4. Задача sequence to sequence
5. Библиотеки NLTK и PyMorphu для предобработки и анализа текстов
6. Библиотека gensim
7. Популярные архитектуры нейросетей для классификации текстов
8. Популярные архитектуры нейросетей для решения задачи NER
9. Популярные архитектуры нейросетей для машинного перевода
10. Популярные архитектуры нейросетей для распознавания речи
11. Популярные архитектуры нейросетей для синтеза речи
12. Нейросети для генерации описаний к изображению
13. Возможности фреймворка PyTorch по анализу текстов
14. Возможности фреймворка PyTorch по генерации текстов
15. Возможности фреймворка PyTorch по распознаванию и синтезу речи

Пример лабораторной работы по разделу «Большие языковые модели»

1. Используйте одну из больших языковых моделей ruBert или ruGPT из фреймворка PyTorch для создания и обучения нейронной сети, позволяющей генерировать тексты на определенную тематику (тематику и датасет выбираете сами).

2. Сгенерируйте текст с помощью разработанной нейросети и проверьте его качество с помощью соответствующей метрики.

Зачетно-экзаменационные материалы для промежуточной аттестации (экзамен/зачет)

Вопросы к зачету

1. Популярные оптимизаторы в нейронных сетях (SGD, SGDM, Adam)
2. Обнаружение недо- и переобучения нейронных сетей. Методы борьбы с переобучением
3. История развития области NLP
4. Задача классификации текстов
5. Задача классификации последовательности
6. Задача sequence to sequence
7. Предобработка текстовых данных
8. Библиотеки NLTK и PyMorphu для предобработки и анализа текстов

9. Векторное представление слов (Word2Vec, Skip-gram, CBOW, FastText)
10. Библиотека gensim
11. Рекуррентные и двунаправленные нейронные сети
12. LSTM-сети
13. Механизм внимания. Архитектура трансформер
14. Известные трансформеры: GPT и ruGPT, Bert и ruBert
15. Техника Transfer Learning
16. Популярные архитектуры нейросетей для классификации текстов
17. Популярные архитектуры нейросетей для решения задачи NER
18. Популярные архитектуры нейросетей для машинного перевода
19. Популярные архитектуры нейросетей для распознавания речи
20. Популярные архитектуры нейросетей для синтеза речи
21. Нейросети для генерации описаний к изображению
22. Возможности фреймворка PyTorch по анализу текстов
23. Возможности фреймворка PyTorch по генерации текстов
24. Возможности фреймворка PyTorch по распознаванию и синтезу речи

Методические рекомендации, определяющие процедуры оценивания выполнения лабораторных работ:

Задание считается выполненным при выполнении следующих условий:

- предоставлен исходный код на языке Python с использованием соответствующих библиотек в среде Jupyter notebook / Google Colab;
- продемонстрирована работоспособность представленного кода;
- студент понимает исходный код и отвечает на вопросы по его организации.

Методические рекомендации, определяющие процедуры оценивания реферата (доклада, сообщения, презентации)

Неправильно оформленная работа не принимается.

Критерии оценки:

оценка «неудовлетворительно» выставляется при несоответствии заявленной темы реферата (доклада, сообщения, презентации) его содержанию, наличии грубых погрешностей в оформлении работы, использовании ненадлежащих нормативных и научных источников, приводящих к утрате научной значимости подготовленного реферата или реферат не подготовлен;

оценка «удовлетворительно» выставляется, если студентом допущены несущественные фактические ошибки в изложении материала и/или допущено использование ненадлежащих нормативных источников при сохранении актуальности темы реферата. Реферат (сообщение, доклад, презентация) представляет собой изложение результатов чужих исследований без самостоятельной обработки источников;

оценка «хорошо» выставляется в том случае, когда имеются отдельные погрешности в оформлении реферата. Реферат (сообщение, доклад, презентация) представляет собой самостоятельный анализ разнообразных научных исследований и теоретических данных, однако не в полной мере отражает требования, сформулированные к его и содержанию;

оценка «отлично» выставляется студенту, если студентом представлен реферат (сообщение, доклад, презентация), отвечающий требованиям по оформлению. Содержание реферата должно основываться на соответствующих литературных источниках. В реферате (сообщении, докладе, презентации) отражаются такие требования как актуальность содержания, высокий теоретический уровень, глубина и полнота факторов, явлений, проблем, относящихся к теме, информационная насыщенность, новизна, оригинальность изложения материала; структурная организованность, обоснованность предложения и выводов, сделанных в реферате (сообщении, докладе, презентации).

Методические рекомендации, определяющие процедуры оценивания на зачете:

«зачтено»: студент владеет теоретическими знаниями по всем разделам, однако допускает незначительные ошибки; студент умеет правильно объяснять материал по технологиям

обработки естественного языка, распознаванию и синтезу речи, иллюстрируя его примерами популярных архитектур нейронных сетей.

«не зачтено»: материал не усвоен или усвоен частично, студент затрудняется привести примеры по используемым архитектурам нейронных сетей для работы с речью и звуком, включая распознавание и синтез речи, довольно ограниченный объем знаний материала по используемым библиотекам для NLP.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на зачете;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине (модулю) предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

5.1 Основная литература:

1. Остроух, А. В. Системы искусственного интеллекта : монография / А. В. Остроух, Н. Е. Суркова. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 228 с. — ISBN 978-5-8114-8519-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/176662> (дата обращения: 18.08.2021). — Режим доступа: для авториз. пользователей.

2. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2024. — 243 с. — (Высшее образование). — ISBN 978-5-534-01042-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/537001> (дата обращения: 30.05.2024).

3. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — Москва : Издательство Юрайт, 2024. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/544780> (дата обращения: 30.05.2024).

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде в электронно-библиотечных системах «Лань».

5.2. Периодические издания:

1. Базы данных компании «Ист Вью» <http://dlib.eastview.com>
2. Электронная библиотека GREBENNIKON.RU <https://grebennikon.ru/>

5.3. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

Электронно-библиотечные системы (ЭБС):

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» <http://www.biblioclub.ru/>
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» www.znanium.com
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

Профессиональные базы данных

1. Scopus <http://www.scopus.com/>
2. ScienceDirect <https://www.sciencedirect.com/>
3. Журналы издательства Wiley <https://onlinelibrary.wiley.com/>
4. Научная электронная библиотека (НЭБ) <http://www.elibrary.ru/>
5. Полнотекстовые архивы ведущих западных научных журналов на Российской платформе научных журналов НЭИКОН <http://archive.neicon.ru>
6. Национальная электронная библиотека (доступ к Электронной библиотеке диссертаций Российской государственной библиотеки (РГБ)) <https://rusneb.ru/>
7. Президентская библиотека им. Б.Н. Ельцина <https://www.prilib.ru/>
8. База данных CSD Кембриджского центра кристаллографических данных (CCDC) <https://www.ccdc.cam.ac.uk/structures/>
9. Springer Journals: <https://link.springer.com/>
10. Springer Journals Archive: <https://link.springer.com/>
11. Nature Journals: <https://www.nature.com/>
12. Springer Nature Protocols and Methods: <https://experiments.springernature.com/sources/springer-protocols>
13. Springer Materials: <http://materials.springer.com/>
14. Nano Database: <https://nano.nature.com/>
15. Springer eBooks (i.e. 2020 eBook collections): <https://link.springer.com/>
16. "Лекториум ТВ" <http://www.lektorium.tv/>
17. Университетская информационная система РОССИЯ <http://uisrussia.msu.ru>

Информационные справочные системы

1. Консультант Плюс - справочная правовая система (доступ по локальной сети с компьютеров библиотеки)

Ресурсы свободного доступа

1. КиберЛенинка <http://cyberleninka.ru/>;
2. Американская патентная база данных <http://www.uspto.gov/patft/>
3. Министерство науки и высшего образования Российской Федерации <https://www.minobrnauki.gov.ru/>;
4. Федеральный портал "Российское образование" <http://www.edu.ru/>;
5. Информационная система "Единое окно доступа к образовательным ресурсам" <http://window.edu.ru/>;

6. Единая коллекция цифровых образовательных ресурсов <http://school-collection.edu.ru/> .
7. Проект Государственного института русского языка имени А.С. Пушкина "Образование на русском" <https://pushkininstitute.ru/>;
8. Справочно-информационный портал "Русский язык" <http://gramota.ru/>;
9. Служба тематических толковых словарей <http://www.glossary.ru/>;
10. Словари и энциклопедии <http://dic.academic.ru/>;
11. Образовательный портал "Учеба" <http://www.ucheba.com/>;
12. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы http://xn--273--84d1f.xn--p1ai/voprosy_i_otvety

Собственные электронные образовательные и информационные ресурсы КубГУ

1. Электронный каталог Научной библиотеки КубГУ <http://megapro.kubsu.ru/MegaPro/Web>
2. Электронная библиотека трудов ученых КубГУ <http://megapro.kubsu.ru/MegaPro/UserEntry?Action=ToDb&idb=6>
3. Среда модульного динамического обучения <http://moodle.kubsu.ru>
4. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://infoneeds.kubsu.ru/>
5. Библиотека информационных ресурсов кафедры информационных образовательных технологий <http://mschool.kubsu.ru;>
6. Электронный архив документов КубГУ <http://docspace.kubsu.ru/>
7. Электронные образовательные ресурсы кафедры информационных систем и технологий в образовании КубГУ и научно-методического журнала "ШКОЛЬНЫЕ ГОДЫ" <http://icdau.kubsu.ru/>

6. Методические указания для обучающихся по освоению дисциплины (модуля)

Изучение курса «Технологии обработки языка, звуковых данных, включая распознавание и синтез речи» осуществляется в тесном взаимодействии с другими дисциплинами, связанными с анализом данных, искусственным интеллектом и программированием. Форма и способы изучения материала определяются с учетом специфики изучаемой темы. Однако во всех случаях необходимо обеспечить сочетание изучения теоретического материала, научного толкования того или иного понятия, даваемого в учебниках и лекциях, с самостоятельной работой студентов и выполнением практических заданий.

Лекционное занятие представляет собой систематическое, последовательное, монологическое изложение преподавателем-лектором учебного материала, как правило, теоретического характера. Такое занятие представляет собой элемент технологии представления учебного материала путем логически стройного, систематически последовательного и ясного изложения с использованием образовательных технологий.

Цель лекции – организация целенаправленной познавательной деятельности обучающихся по овладению программным материалом учебной дисциплины. Чтение курса лекций позволяет дать связанное, последовательное изложение материала в соответствии с новейшими данными науки, сообщить слушателям основное содержание предмета в целостном, систематизированном виде.

Задачи лекции заключаются в обеспечении формирования системы знаний по учебной дисциплине, в умении аргументировано излагать научный материал, в формировании профессионального кругозора и общей культуры, в отражении еще не получивших освещения в учебной литературе новых достижений науки, в оптимизации других форм организации учебного процесса.

Для подготовки к лекциям необходимо изучить основную литературу по заявленной теме и обратить внимание на те вопросы, которые предлагаются к рассмотрению в конце каждой темы. При изучении основной литературы, студент может в достаточном объеме

усвоить и успешно реализовать конкретные знания, умения, навыки и компетенции при выполнении следующих условий:

- 1) систематическая работа на учебных занятиях под руководством преподавателя и самостоятельная работа по закреплению полученных знаний и навыков;
- 2) добросовестное выполнение заданий преподавателя на практических занятиях;
- 3) выяснение и уточнение отдельных предпосылок, умозаключений и выводов, содержащихся в учебном курсе; взаимосвязей отдельных его разделов, используемых методов, характера их использования в практической деятельности;
- 4) сопоставление точек зрения различных авторов по затрагиваемым в учебном курсе проблемам; выявление неточностей и некорректного изложения материала в периодической и специальной литературе;
- 5) разработка предложений преподавателю в части доработки и совершенствования учебного курса.

Лабораторные занятия – являются формой учебной аудиторной работы, в рамках которой формируются, закрепляются и представляются студентами знания, умения и навыки, интегрирующие результаты освоения компетенций как в лекционном формате, так в различных формах самостоятельной работы. К каждому занятию преподавателем формулируются практические задания, требования и методические рекомендации к их выполнению, которые представляются в фонде оценочных средств учебной дисциплины.

В ходе самоподготовки к лабораторным занятиям студент осуществляет сбор и обработку материалов по тематике лабораторной работы, используя при этом открытые источники информации (публикации в научных изданиях, аналитические материалы, ресурсы сети Интернет и т.п.), а также практический опыт и доступные материалы объекта исследования.

Контроль за выполнением самостоятельной работы проводится при изучении каждой темы дисциплины на лабораторных занятиях.

Самостоятельная работа студентов по дисциплине «Технологии обработки языка, звуковых данных, включая распознавание и синтез речи» проводится с целью закрепления и систематизации теоретических знаний, формирования практических навыков по их применению при решении задач обработки естественного языка. Самостоятельная работа включает: изучение основной и литературы, проработку и повторение лекционного материала, материала учебной и научной литературы, подготовку к лабораторным занятиям, а также к контролируемой самостоятельной работе

Самостоятельная работа студентов по данному учебному курсу предполагает поэтапную подготовку по каждому разделу в рамках соответствующих заданий:

Первый этап самостоятельной работы студентов включает в себя тщательное изучение теоретического материала на основе лекционных материалов преподавателя, рекомендуемых разделов основной литературы, материалов периодических научных изданий, необходимых для овладения понятийно-категориальным аппаратом и формирования представлений о комплексе теоретического и аналитического инструментария, используемого в рамках данной отрасли знания.

На втором этапе на основе сформированных знаний и представлений по данному разделу студенты выполняют лабораторные работы, нацеленные на формирование умений и навыков в рамках заявленных компетенций. На данном этапе студенты осуществляют самостоятельный поиск эмпирических материалов в рамках конкретного задания, обобщают и анализируют собранный материал по схеме, рекомендованной преподавателем, формулируют выводы.

На сегодняшний день *тестирование* – один из самых действенных и популярных способов проверить знания в изучаемой области. Тесты позволяют очень быстро проверить наличие знаний у студентов по выбранной теме. Кроме того, тесты не только проверяют знания, но и тренируют внимательность, усидчивость и умение быстро ориентироваться и соображать. При подготовке к решению тестов необходимо проработать основные категории и понятия дисциплины, обратить внимание на ключевые вопросы темы.

Под *контролируемой самостоятельной работой (КСР)* понимают совокупность заданий, которые студент должен выполнить, проработать, изучить по заданию под

руководством и контролем преподавателя. Т.е. КСР – это такой вид деятельности, наряду с лекциями, лабораторными и практическими занятиями, в ходе которых студент, руководствуясь специальными методическими указаниями преподавателя, а также методическими указаниями по выполнению типовых заданий, приобретает и совершенствует знания, умения и навыки, накапливает практический опыт.

Текущий контроль самостоятельной работы студентов осуществляется еженедельно в соответствии с программой занятий. Описание заданий для самостоятельной работы студентов и требований по их выполнению выдаются преподавателем в соответствии с разработанным фондом оценочных средств по дисциплине «Технологии обработки языка, звуковых данных, включая распознавание и синтез речи».

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

7. Материально-техническое обеспечение по дисциплине (модулю)

Наименование специальных помещений	Оснащенность специальных помещений	Перечень лицензионного программного обеспечения
Учебные аудитории для проведения занятий лекционного типа (ауд. 129, 131, А-305, А-307)	Мебель: учебная мебель Технические средства обучения: экран, проектор, компьютер	MS Office Word 2016 и выше Ms Power Point 2016 и выше
Учебные аудитории для проведения текущего контроля (Ауд. 101, 102, 105/1, 106 и 106а)	Мебель: учебная мебель Технические средства обучения: Экран, компьютер Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации	Браузер Google Chrome, Jupyter Notebook 6.3.0 и выше (язык Python с библиотеками Numpy, Pandas, gensim, NLTK, PyMorphology, фреймворком PyTorch)
Учебные аудитории для проведения промежуточной аттестации (Ауд. 129, 131, А-305, А-307)	Мебель: учебная мебель	-
Учебные аудитории для проведения лабораторных работ (Ауд. 101, 102, 105/1, 106 и 106а)	Мебель: учебная мебель Технические средства обучения: экран, компьютер Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации	Браузер Google Chrome, Jupyter Notebook 6.3.0 и выше (язык Python с библиотеками Numpy, Pandas, gensim, NLTK, PyMorphology, фреймворком PyTorch)

Для самостоятельной работы обучающихся предусмотрены помещения, укомплектованные специализированной мебелью, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.

Наименование помещений для самостоятельной работы обучающихся	Оснащенность помещений для самостоятельной работы обучающихся	Перечень лицензионного программного обеспечения
Помещение для самостоятельной работы обучающихся (читальный зал Научной библиотеки)	Мебель: учебная мебель Комплект специализированной мебели: компьютерные столы Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации, веб-камеры, коммуникационное оборудование, обеспечивающее доступ к сети интернет (проводное соединение и беспроводное соединение по технологии Wi-Fi)	MS Office Word 2016 и выше Ms Power Point 2016 и выше
Помещение для самостоятельной работы обучающихся (Ауд. 101, 102, 103, 105/1, 106 и 106а)	Мебель: учебная мебель Комплект специализированной мебели: компьютерные столы Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации, веб-камеры, коммуникационное оборудование, обеспечивающее доступ к сети интернет (проводное соединение и беспроводное соединение по технологии Wi-Fi)	Браузер Google Chrome, Jupyter Notebook 6.3.0 и выше (язык Python с библиотеками Numpy, Pandas, gensim, NLTK, PyMorphy, фреймворком PyTorch)

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

