



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Магнитогорский государственный технический университет им. Г.И.
Носова»



УТВЕРЖДАЮ
Директор ИЭиАС
В.Р. Храмшин

26.01.2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

***ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА, РАСПОЗНАВАНИЕ И СИНТЕЗ
РЕЧИ***

Направление подготовки (специальность)
09.03.03 Прикладная информатика

Направленность (профиль/специализация) программы
Искусственный интеллект в цифровой экономике

Уровень высшего образования - бакалавриат

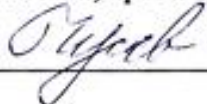
Форма обучения
очная

Институт/ факультет	Институт энергетики и автоматизированных систем
Кафедра	Бизнес-информатики и информационных технологий
Курс	4
Семестр	8

Магнитогорск
2022 год

Рабочая программа составлена на основе ФГОС ВО - бакалавриат по направлению подготовки 09.03.03 Прикладная информатика (приказ Минобрнауки России от 19.09.2017 г. № 922)


Рабочая программа рассмотрена и одобрена на заседании кафедры Бизнес-информатики и информационных технологий 25.01.2022, протокол № 5

Зав. кафедрой  Г.Н. Чусавитина


Рабочая программа одобрена методической комиссией ИЭиАС 26.01.2022 г. протокол № 5

Председатель  В.Р. Храмшин

Рабочая программа составлена:

доцент кафедры БИиИТ, канд. пед. наук  Е.В. Карманова

Рецензент:

Руководитель аналитических работ ЗАО «КонСОМ СКС» , канд. техн. наук
 В.А Ошурков

Лист актуализации рабочей программы

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2023 - 2024 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2024 - 2025 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2025 - 2026 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2026 - 2027 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

1 Цели освоения дисциплины (модуля)

Целью дисциплины является формирование у студентов прикладной информатики базовых представлений, знаний и умений в технологиях анализа естественного языка и синтеза речи.

2 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина Обработка естественного языка, распознавание и синтез речи входит в часть учебного плана формируемую участниками образовательных отношений образовательной программы.

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения дисциплин/ практик:

Информационные системы и технологии

Программирование

Программирование на Python

Интернет технологии

Основы искусственного интеллекта

Интеллектуальный анализ данных

Управление данными

Машинное обучение

Базы данных

Искусственные нейронные сети

Знания (умения, владения), полученные при изучении данной дисциплины будут необходимы для изучения дисциплин/практик:

Выполнение и защита выпускной квалификационной работы

Подготовка к сдаче и сдача государственного экзамена

Проектная деятельность

Производственная – преддипломная практика

Экспертные системы и системы поддержки принятия решений

3 Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля) и планируемые результаты обучения

В результате освоения дисциплины (модуля) «Обработка естественного языка, распознавание и синтез речи» обучающийся должен обладать следующими компетенциями:

Код индикатора	Индикатор достижения компетенции
ПК-1	Способен использовать, создавать и внедрять технологии искусственного интеллекта для решения задач в зависимости от особенностей проблемной и предметной областей
ПК-1.1	Осуществляет поиск данных в открытых источниках, специализированных библиотеках, репозиториях и архивах, выполняет подготовку, разметку, анализ, представление и визуализацию больших данных
ПК-1.2	Создает, поддерживает и использует системы искусственного интеллекта, методы и модели машинного обучения для решения задач искусственного интеллекта
ПК-1.3	Осуществляет оценку и выбор моделей искусственных нейронных сетей и инструментальных средств, использует и разрабатывает системы искусственного интеллекта на основе нейросетевых моделей и методов для решения поставленной задачи

4. Структура, объём и содержание дисциплины (модуля)

Общая трудоемкость дисциплины составляет 5 зачетных единиц 180 акад. часов, в том числе:

- контактная работа – 112,3 акад. часов;
- аудиторная – 110 акад. часов;
- внеаудиторная – 2,3 акад. часов;
- самостоятельная работа – 67,7 акад. часов;
- в форме практической подготовки – 0 акад. час;

Форма аттестации - зачет с оценкой

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа студента	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код компетенции
		Лек.	лаб. зан.	практ. зан.				
1. Введение в обработку естественного языка								
1.1 История NLP И речевых технологий. Области применения NPL	8	4	2		8	Подготовка к семинарскому занятию	Презентация по заданию семинарского занятия	ПК-1.2
1.2 Основы машинной обработки естественного языка. Представления текстовых данных.		2	4		4	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.2
1.3 Технологии предобработки текста, лемматизация, стемминг.		4	6		8	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.1
Итого по разделу		10	12		20			
2. Машинное обучение и глубокие нейронные сети для решения задач анализа и обработки естественного языка								
2.1 Методы машинного обучения для классификации текстовых документов на основе частотных мер (TF-IDF).	8	4	8		8	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.2
2.2 Классификация текстов на основе нейросетевых моделей языка.		4	8		4	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.3
2.3 Классификация текстов с помощью глубоких нейронных сетей: CNN, LSTM.		4	6		3,7	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.3
2.4 Кластеризация текстовых документов.		2	6		2	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.3, ПК-1.2

2.5 Задачи обработки последовательностей (машинный перевод, автоматическое реферирование, вопросно-ответные системы)	8	4	8		2	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.3, ПК-1.2
2.6 Трансформеры в задачах анализа текстов		4	8		4	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.3
Итого по разделу		22	44		23,7			
3. Построение диалоговых систем								
3.1 Архитектура диалоговых систем	8	4			8	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.2
3.2 Современные фреймворки для построения диалоговых систем		4	6		10	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.2, ПК-1.3
3.3 UX/UI диалоговых ассистентов		4	4		6	Выполнение лабораторной работы	Отчет по лабораторной работе	ПК-1.2
Итого по разделу		12	10		24			
Итого за семестр		44	66		67,7		зао	
Итого по дисциплине		44	66		67,7		зачет с оценкой	

5 Образовательные технологии

Основными образовательными технологиями, положенными в основу преподавания дисциплины «Обработка естественного языка, распознавание и синтез речи» являются:

- активные технологии обучения:

о метод ролевых игр - это разыгрывание участниками группы сценки с заранее распределенными ролями в интересах овладения определенной поведенческой или эмоциональной стороной жизненных ситуаций.

Ролевая игра проводится в небольших группах (3-5 участников);

о технологии кейс-стади - техника обучения, использующая описание реальных ситуаций. Обучающиеся должны проанализировать ситуацию, разобраться в сути проблем, предложить возможные решения и выбрать лучшее из них. Кейсы базируются на реальном фактическом материале, или же приближены к реальной ситуации;

о разработка проекта - это способ достижения дидактической цели через детальную разработку проблемы (технологии), которая должна завершиться вполне реальным, осязаемым практическим результатом, оформленным тем или иным образом;

о работа в малых группах - это одна из самых популярных стратегий, так как она дает всем обучающимся (в том числе и стеснительным) возможность участвовать в работе, практиковать навыки сотрудничества, меж-личностного общения (в частности, умение активно слушать, вырабатывать общее мнение, разрешать возникающие разногласия);

- интерактивные лекции:

о лекций-дискуссий - преподаватель приводит отдельные примеры в виде ситуаций или кратко сформулированных проблем и предлагает студентам кратко обсудить, затем краткий анализ, выводы и лекция продолжается. Положительным в дискуссии является, то, что обучаемые соглашались с точкой зрения преподавателя с большой охотой, скорее в ходе дискуссии, нежели во время беседы, когда преподаватель лишь указывает на необходимость принять его позицию по обсуждаемому вопросу. Данный метод позволяет преподавателю видеть, насколько эффективно слушатели используют полученные знания в ходе дискуссии.

Активные технологии обучения преимущественно используются в рамках практических занятий, интерактивные лекции - в процессе изучения и закрепления нового учебного материала.

В качестве практико-ориентированного средства обучения выбран образовательный портал ФГБОУ ВО «МГТУ им. Г.И. Носова».

6 Учебно-методическое обеспечение самостоятельной работы обучающихся

Представлено в приложении 1.

7 Оценочные средства для проведения промежуточной аттестации

Представлены в приложении 2.

8 Учебно-методическое и информационное обеспечение дисциплины (модуля)

а) Основная литература:

Предобработка данных для нейросетевого управления : учебное пособие / Л. И. Воронова, В. Р. Брус, В. И. Воронов, А. Н. Баширов. — Москва : МГУСИ, 2021. — 49 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/215198> (дата обращения: 08.07.2022). — Режим доступа: для авториз. пользователей.

Мещерина, Е. В. Системы искусственного интеллекта : учебно-методическое пособие / Е. В. Мещерина. — Оренбург : ОГУ, 2019. — 96 с. — ISBN 978-5-7410-2315-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/160008> (дата обращения: 08.07.2022). — Режим доступа: для авториз. пользователей.

б) Дополнительная литература:

Волосова, А. В. Технологии искусственного интеллекта в ULS-системах : учебное пособие для вузов / А. В. Волосова. — Санкт-Петербург : Лань, 2022. — 308 с. — ISBN 978-5-8114-8839-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/208568> (дата обращения: 08.07.2022). — Режим доступа: для авториз. пользователей.

Толмачев, С. Г. Нейросетевые методы обработки информации : учебное пособие / С. Г. Толмачев. — Санкт-Петербург : БГТУ "Военмех" им. Д.Ф. Устинова, 2021. — 103 с. — ISBN 978-5-907324-33-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/220238> (дата обращения: 08.07.2022). — Режим доступа: для авториз. пользователей.

Филиппов, Ф. В. Моделирование нейронных сетей глубокого обучения : учебное пособие / Ф. В. Филиппов. — Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2019. — 79 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/180053> (дата обращения: 08.07.2022). — Режим доступа: для авториз. пользователей.

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: 08.07.2022). — Режим доступа: для авториз. пользователей.

в) Методические указания:

Представлено в приложении 1.

г) Программное обеспечение и Интернет-ресурсы:

Программное обеспечение

Наименование ПО	№ договора	Срок действия лицензии
7Zip	свободно	бессрочно
Anaconda	свободно	бессрочно
LibreOffice	свободно	бессрочно
JetBrains PyCharm Community Edition	свободно распространяемое ПО	бессрочно
Adobe Reader	свободно	бессрочно
Браузер Yandex	свободно	бессрочно
Браузер Mozilla Firefox	свободно распространяемое ПО	бессрочно

Профессиональные базы данных и информационные справочные системы

Название курса	Ссылка
Электронные ресурсы библиотеки МГТУ им. Г.И. Носова	https://magtu.informsystema.ru/Marc.html?locale=ru

Федеральное государственное бюджетное учреждение «Федеральный институт промышленной собственности»	URL: http://www1.fips.ru/
Поисковая система Академия Google (Google Scholar)	URL: https://scholar.google.ru/
Информационная система - Единое окно доступа к информационным ресурсам	URL: http://window.edu.ru/
Национальная информационно-аналитическая система – Российский индекс научного цитирования (РИНЦ)	URL: https://elibrary.ru/project_risc.asp

9 Материально-техническое обеспечение дисциплины (модуля)

Материально-техническое обеспечение дисциплины включает:

Материально-техническое обеспечение дисциплины включает:

Учебные аудитории для проведения занятий лекционного типа Персональный компьютер (или ноутбук) с пакетом Office, с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Доска, мультимедийный проектор, экран. Мультимедийные презентации к лекциям, учебно-наглядные пособия

Учебные аудитории для проведения лабораторных занятий Персональные компьютеры с пакетом Office; с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Браузер Yandex.

Аудитории для самостоятельной работы: компьютерные классы; читальные залы библиотеки Персональные компьютеры с пакетом Office; с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Браузер Yandex.

Аудитории для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации Персональные компьютеры с пакетом Office; с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Браузер Yandex.

Аудитория для хранения и профилактического обслуживания учебного оборудования Мебель для хранения и обслуживания оборудования (шкафы, столы), учебно-методические материалы, компьютеры, ноутбуки, принтеры.

Учебно-методическое обеспечение самостоятельной работы обучающихся
Перечень тем для подготовки к занятиям:

№ занятия	Краткое содержание занятия, семинара
1	<p>История NLP И речевых технологий. Области применения NPL.</p> <p>Основные задачи NLP.</p> <p>Задание:</p> <p>Подготовить презентацию по теме технологии применения NLP в области:</p> <ul style="list-style-type: none"> – Финансы и страхование – Промышленность и логистика – Медицина IT и телекоммуникации – Юридическая практика – Медиа и реклама – Наука и образование – Государство и безопасность <p><i>Полезные источники:</i></p> <ul style="list-style-type: none"> • Альманах искусственный интеллект №2. Обработка естественного языка, распознавание и синтез речи. Аналитический сборник - https://aireport.ru/nlp
2	<p>Основы машинной обработки естественного языка. Представления текстовых данных.</p> <p>Статистическая модель. Библиотека spaCy.</p> <p>Задание:</p> <p>Установите библиотеку spaCy. Вывести статистическую модель предложенному тексту.</p> <p><i>Полезные источники:</i></p> <ul style="list-style-type: none"> – Беляева, Л. Н. Сетевые лингвистические технологии : монография / Л. Н. Беляева, О. Н. Камшилова, К. Р. Пиотровская. — Санкт-Петербург : РГПУ им. А. И. Герцена, 2019. — 111 с. — ISBN 978–5–8064–2701–5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/136751 (дата обращения: 08.07.2022). — Режим доступа: для авториз. пользователей. – Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. — М.: МИЭМ, 2011. — 272 с. - https://nsu.ru/xmlui/bitstream/handle/nsu/8982/17645.pdf?sequence=1&isAll

	owed=y
3	<p>Технологии предобработки текста, лемматизация, стемминг.</p> <p>Конвейер обработки текста.</p> <p>Задание:</p> <ol style="list-style-type: none"> 1. Реализуйте разбор текста на токены. 2. Проведите лемматизацию предложенного текста. 3. Опишите исключение (special case) для слова отдельного (одного) слова в предложенном тексте, заменив его лемму по умолчанию на иное. Добавьте это исключение в экземпляр Tokenizer. Для проверки выведите леммы всех слов в предложении. 4. Реализуйте задачу поиска соответствующих глаголов с помощью тегов частей речи. 5. Выделите именованные сущности в тексте. <p><i>Полезные источники:</i></p> <ul style="list-style-type: none"> – Васильев Юлий Обработка естественного языка. Python и spaCy на практике. — СПб.: Питер, 2021. — 256 с.: ил. — (Серия «Библиотека программиста»).
4	<p>Методы машинного обучения для классификации текстовых документов на основе частотных мер (TF-IDF).</p> <p>Деревья решений, наивный байесовский классификатор, логистическая регрессия в задаче классификации текстов.</p> <p>Задание:</p> <ol style="list-style-type: none"> 1. Реализуйте метод Bag-of-words (CountVectorizer) (мешок слов) для предложенного текста. Какие недостатки имеет данный метод? 2. Удалите стоп-слова из предложенного текста. 3. Удалите, все слова, которые встречаются менее, чем в 2 документах. 4. Реализуйте выделение биграмм и триграмм в тексте. 5. Выведите TF-IDF для предложенного текста <p><i>Полезные источники:</i></p> <p>Васильев Юлий Обработка естественного языка. Python и spaCy на практике. — СПб.: Питер, 2021. — 256 с.: ил. — (Серия «Библиотека программиста»).</p>
5	<p>Классификация текстов на основе нейросетевых моделей языка.</p> <p>Языковые модели. Word embeddings. Нейросетевые модели языка: word2vec, fasttext. Мера семантической близости.</p> <p>Задание:</p> <ol style="list-style-type: none"> 1. Изучите сервис https://rusvectors.org, проанализировать 2-3 статические модели отсюда. 2. Дан список слов words = ['день_NOUN', 'ночь_NOUN', 'человек_NOUN',

	<p>'семантика_NOUN', 'биткоин_NOUN']. Вывести с помощью модели 10 ближайших соседей для каждого слова и косинусные близости для каждого. Примечание. Используйте модель (ruwikiruscorpora_upros_skipgram_300_2_2018.vec.gz), размещенную на сервисе rusvectors.</p> <ol style="list-style-type: none"> 3. Найдите косинусную близость пары слов (функцией similarity()) преподаватель – студент. 4. Используя метод doesnt_match() найдите "лишнее слово" в группе слов: пицца пельмень хот-дог ананас <p><i>Полезные источники:</i></p> <p>https://rusvectors.org</p>
6	<p>Классификация текстов с помощью глубоких нейронных сетей: CNN, LSTM.</p> <p>Задание:</p> <ol style="list-style-type: none"> 1. Используя рекуррентную нейронную сеть с долгой краткосрочной памятью, выполните задачу классификации текстов. 2. Постройте CNN для задачи классификации текстов. <p><i>Полезные источники:</i></p> <ul style="list-style-type: none"> – Python для NLP: классификация нейронных сетей LSTM с несколькими метками текста с использованием Keras - https://russianblogs.com/article/56041415312/
7	<p>Кластеризация текстовых документов.</p> <p>Тематическое моделирование Методы LSA, pLSA. Аддитивная регуляризация тематических моделей в BigARTM</p> <p>Задание:</p> <ol style="list-style-type: none"> 1. Примените и оцените модуль извлечения именованных сущностей Natasha (https://github.com/natasha/natasha). Для тестирования используйте текст и соответствующую разметку. Тестовые данные содержат только разметку для людей (PER) и организаций (ORG). Рассчитайте F1 для каждого типа сущностей и общее значение F1. <p><i>Полезные источники:</i></p> <ul style="list-style-type: none"> – https://github.com/natasha/natasha
8	<p>Задачи обработки последовательностей (машинный перевод, автоматическое реферирование, вопросно-ответные системы).</p> <p>Механизм внимания (attention). Архитектуры encoder-decoder-attention.</p> <p>Задание:</p> <p>Реализуйте задачу выделения ключевых слов из деревьев синтаксических зависимостей для предложенного текста.</p>

	<p><i>Полезные источники:</i></p> <ul style="list-style-type: none"> – https://tpc.ispras.ru/
9	<p>Трансформеры в задачах анализа текстов.</p> <p>Transfer learning в задачах анализа текстов. Self-Attention. Архитектуры трансформеров: BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерации текстов. Fine-tuning трансформеров.</p> <p>Задание:</p> <p>Систематический анализ качества генерации текстов с помощью модели ruGPT-3 для различных сценариев (https://github.com/sberbank-ai/ru-gpts).</p> <p><i>Полезные источники:</i></p> <ul style="list-style-type: none"> – https://github.com/sberbank-ai/ru-gpts
10	<p>Архитектура диалоговых систем.</p> <p>Построение диалоговых систем. Архитектура диалоговых систем. Модули понимания естественного языка (NLU) и диалоговый менеджер (DM). Сложности построения диалоговых систем.</p> <p>Задание:</p> <p>Проведите сравнительный анализ двух моделей вопросно-ответного поиска для русского языка с помощью инструмента CheckList (https://github.com/marcotcr/checklist).</p> <p><i>Полезные источники:</i></p> <ul style="list-style-type: none"> – https://github.com/marcotcr/checklist
11	<p>Современные фреймворки для построения диалоговых систем</p> <p>DeepPavlov, Rasa, Just AI Conversational Platform</p> <p>Задание:</p> <p>Реализовать чат-бота с использованием одного из современных фреймворков на заданную тематику.</p> <p><i>Полезные источники:</i></p> <ul style="list-style-type: none"> – Трудно быть ботом: как сделать чатбота с помощью DeepPavlov - https://sysblok.ru/nlp/trudno-byt-botom-kak-sdelat-chatbota-s-pomoshhju-deepavlov/
12	<p>UX/UI диалоговых ассистентов.</p> <p>Проектирование UX/UI диалоговых ассистентов в чатах и голосовом помощнике.</p> <p>Задание:</p> <ol style="list-style-type: none"> 1. Изучите советы по проектированию UX/UI в следующих статьях: https://vc.ru/design/79928-gayd-po-golosovomu-ux-situacii-patterny-instrumenty

[-i-nemnogo-moralnoy-podderzhki](#)

<https://just-ai.com/blog/vse-cto-vy-hoteli-znat-pro-dialogovyj-ux-ui-v-proektirovanii-chat-botov>

2. Спроектируйте диалог, реализуя один навык чат –бота в любом он –лайн инструменте.

Полезные источники:

- <https://vc.ru/design/79928-gayd-po-golosovomu-ux-situacii-patterny-instrumenty-i-nemnogo-moralnoy-podderzhki>
- <https://just-ai.com/blog/vse-cto-vy-hoteli-znat-pro-dialogovyj-ux-ui-v-proektirovanii-chat-botov>

Оценочные средства для проведения промежуточной аттестации
Планируемые результаты обучения и оценочные средства для проведения промежуточной аттестации:

Код индикатора	Индикатор достижения компетенции	Оценочные средства
ПК-1 Способен использовать, создавать и внедрять технологии искусственного интеллекта для решения задач в зависимости от особенностей проблемной и предметной областей		
ПК – 1.1	Осуществляет поиск данных в открытых источниках, специализированных библиотеках, репозиториях и архивах, выполняет подготовку, разметку, анализ, представление и визуализацию больших данных	<p>Вопросы к зачету с оценкой:</p> <ol style="list-style-type: none"> 1. Представления текстовых данных. 2. Предобработка текста. Лемматизация, стемминг. <p>Задания:</p> <ol style="list-style-type: none"> 1. Вывести статистическую модель к предложенному тексту. 2. Реализуйте разбор текста на токены. 3. Проведите лемматизацию предложенного текста. 4. Опишите исключение (special case) для слова отдельного (одного) слова в предложенном тексте, заменив его лемму по умолчанию на иное. Добавьте это исключение в экземпляр Tokenizer. Для проверки выведите леммы всех слов в предложении. <p>Проектное задание: Используя API Вконтакте, реализуйте сбор, предобработку и визуализацию всех комментариев пользователей к постам отдельной группы социальной сети.</p>
ПК – 1.2	Создает, поддерживает и использует системы искусственного интеллекта, методы и модели машинного обучения для решения задач искусственного интеллекта	<p>Вопросы к зачету с оценкой:</p> <ol style="list-style-type: none"> 1. Методы машинного обучения для классификации текстовых документов на основе частотных мер (TF-IDF). 2. Деревья решений 3. Наивный байесовский классификатор 4. Логистическая регрессия в задаче классификации текстов. 5. Языковые модели. Word embeddings. 6. Кластеризация текстовых документов. 7. Задачи обработки последовательностей: машинной перевод, автоматическое реферирование (summarization), вопросно-ответные системы. 8. Механизм внимания (attention). 9. Построение диалоговых систем. Архитектура диалоговых систем.

		<p>10. Модули понимания естественного языка (NLU) и диалоговый менеджер (DM).</p> <p>11. Сложности построения диалоговых систем.</p> <p>12. Проектирование UX/UI диалоговых ассистентов в чатах и голосе.</p> <p>13. Обзор современных фреймворков для построения диалоговых систем: DeepPavlov, Rasa, Just AI Conversational Platfor.</p> <p>Задания:</p> <ol style="list-style-type: none"> 1. Реализуйте задачу поиска соответствующих глаголов с помощью тегов частей речи. 2. Выделите именованные сущности в тексте. 3. Реализуйте метод Bag-of-words (CountVectorizer) (мешок слов) для предложенного текста. Какие недостатки имеет данный метод? 4. Удалите стоп-слова из предложенного текста. 5. Удалите, все слова, которые встречаются менее, чем в 2 документах. 6. Реализуйте выделение биграмм и триграмм в тексте. 7. Выведите TF-IDF для предложенного текста. <p>Проектное задание:</p> <p>Реализуйте задачу выделения ключевых слов из деревьев синтаксических зависимостей для предложенного текста.</p>
ПК-1.3	<p>Осуществляет оценку и выбор моделей искусственных нейронных сетей и инструментальных средств, использует и разрабатывает системы искусственного интеллекта на основе нейросетевых моделей и методов для решения поставленной задачи</p>	<p>Вопросы к зачету с оценкой:</p> <ol style="list-style-type: none"> 1. Нейросетевые модели языка: word2vec, fasttext. 2. Мера семантической близости. 3. Классификация текстов на основе нейросетевых моделей языка. 4. Кластеризация текстовых документов. 5. Тематическое моделирование Методы LSA, pLSA. 6. Аддитивная регуляризация тематических моделей в BigARTM 7. Классификация текстов с помощью глубоких нейронных сетей: CNN, LSTM. 8. Механизм внимания (attention). 9. Архитектуры encoder-decoder-attention. 10. Transfer learning в задачах анализа текстов. Self-Attention. 11. Архитектуры трансформеров: BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерации текстов. Fine-tuning трансформеров. <p>Задания:</p> <ol style="list-style-type: none"> 1. Используя рекуррентную нейронную сеть с долгой краткосрочной памятью, выполните

задачу классификации текстов.
2. Постройте CNN для задачи классификации текстов.

Проектное задание:

Реализовать чат-бота с использованием одного из современных фрейворков на заданную тематику.

б) Порядок проведения промежуточной аттестации, показатели и критерии оценивания:

Промежуточная аттестация по дисциплине «Обработка естественного языка, распознавание и синтез речи» включает теоретические вопросы, позволяющие оценить уровень усвоения обучающимися знаний, и практические задания, выявляющие степень сформированности умений и владений, проводится в форме зачета с оценкой.

Показатели и критерии оценивания зачета с оценкой:

– на оценку **«отлично»** (5 баллов) – обучающийся демонстрирует высокий уровень сформированности компетенций, всестороннее, систематическое и глубокое знание учебного материала, свободно выполняет практические задания, свободно оперирует знаниями, умениями, применяет их в ситуациях повышенной сложности.

– на оценку **«хорошо»** (4 балла) – обучающийся демонстрирует средний уровень сформированности компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.

– на оценку **«удовлетворительно»** (3 балла) – обучающийся демонстрирует пороговый уровень сформированности компетенций: в ходе контрольных мероприятий допускаются ошибки, проявляется отсутствие отдельных знаний, умений, навыков, обучающийся испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

– на оценку **«неудовлетворительно»** (2 балла) – обучающийся демонстрирует знания не более 20% теоретического материала, допускает существенные ошибки, не может показать интеллектуальные навыки решения простых задач.

– на оценку **«неудовлетворительно»** (1 балл) – обучающийся не может показать знания на уровне воспроизведения и объяснения информации, не может показать интеллектуальные навыки решения простых задач.

Методические рекомендации для самостоятельной работы студентов

В процессе выполнения самостоятельной работы студенты должны научиться воспринимать сведения на слух, фиксировать информацию в виде записей в тетрадях, работать с письменными текстами, самостоятельно извлекая из них полезные сведения и оформляя их в виде тезисов, конспектов, систематизировать информацию в виде заполнения таблиц, составления схем. Важно научиться выделять главные мысли в лекции преподавателя либо в письменном тексте; анализировать явления; определять свою позицию к полученным на занятиях сведениям, четко формулировать ее; аргументировать свою точку зрения: высказывать оценочные суждения; осуществлять самоанализ. Необходимо учиться владеть устной и письменной речью; вести диалог; участвовать в дискуссии; раскрывать содержание изучаемой проблемы в монологической речи; выступать с сообщениями и докладами.

Конспект лекции. Смысл присутствия студента на лекции заключается во включении его в активный процесс слушания, понимания и осмысления материала, подготовленного преподавателем. Этому способствует конспективная запись полученной информации, с помощью которой в дальнейшем можно восстановить основное содержание прослушанной лекции.

Для успешного выполнения этой работы советуем:

- подготовить отдельные тетради для каждого предмета. Запись в них лучше вести на одной стороне листа, чтобы позднее на чистой странице записать дополнения, уточнения, замечания, а также собственные мысли. С помощью разноцветных ручек или фломастеров можно будет выделить заголовки, разделы, термины и т.д.

- не записывать подряд все, что говорит лектор. Старайтесь вначале выслушать и понять материал, а затем уже зафиксировать его, не упуская основных положений и выводов. Сохраняйте логику изложения. Обратите внимание на необходимость точной записи определений и понятий.

- оставить место на странице свободным, если не успели осмыслить и записать часть информации.

- уделять внимание грамотному оформлению записей. Научитесь графически ясно и удобно располагать текст: вычленять абзацы, подчеркивать главные мысли, ключевые слова, помешать выводы в рамки и т.д. Немаловажное значение имеет и четкая структура лекции, в которую входит план, логически выстроенная конструкция освещения каждого пункта плана с аргументами и доказательствами, разъяснениями и примерами, а также список литературы по теме.

- научиться писать разборчиво и быстро. Чтобы в дальнейшем не тратить время на расшифровку собственных записей, следите за аккуратностью почерка, не экономьте бумагу за счет уплотнения текста. Конспектируя, пользуйтесь общепринятыми сокращениями слов и условными знаками, если есть необходимость, то придумайте собственные сокращения.

- уметь быстро и четко переносить в тетрадь графические рисунки и таблицы. Для этих целей приготовьте прозрачную линейку, карандаш и резинку. Старайтесь как можно точнее скопировать изображение с доски. Если наглядный материал трудно воспроизводим в условиях лекции, то сделайте его словесное описание с обобщающими выводами.

- просмотреть свои записи после окончания лекции. Подчеркните и отметьте разными цветами фломастера важные моменты в записях. Исправьте неточности, внесите необходимые дополнения. Не тратьте время на переписывание конспекта, если он оказался не совсем удачным. Совершенствуйтесь, записывая последующие лекции.

Доклад представляет собой устную форму сообщения информации. Он используется в вузе на семинарских занятиях и на научных студенческих конференциях.

Подготовка доклада осуществляется в два этапа: написание письменного текста на заданную тему и подготовка устного выступления перед аудиторией слушателей с освещением этой темы. Письменный доклад оформляется как реферат.

При работе над докладом следует учесть некоторые специфические особенности:

- Объем доклада должен согласовываться со временем, отведенным для выступления.
- При выборе темы нужно учитывать не только собственные интересы, но и интересы потенциальных слушателей. Ваше сообщение необходимо согласовывать с уровнем знаний и потребностей публики.

- Подготовленный текст доклада должен хорошо восприниматься на слух. Даже если отображенный вами материал сложен и неоднозначен, говорить желательно просто и ясно, не перегружая речь наукообразными оборотами и специфическими терминами.

Следует отметить, что иногда преподаватель не требует от студентов письменного варианта доклада и оценивает их работу исключительно по устному выступлению. Но значительно чаще письменный доклад проверяется и его качество также оценивается в баллах. Вне зависимости от того, нужно или не нужно будет сдавать на проверку текст будущего выступления, советуем не отказываться от письменной записи доклада. Это поможет избежать многих ошибок, которые случаются во время устной импровизации: отклонение от темы, нарушения логической последовательности, небрежное обращение с цитатами, злоупотребление деталями и т.д. Если вы хорошо владеете навыками свободной речи и обладаете высокой культурой мышления, то замените письменный доклад составлением тезисного плана. С его помощью зафиксируйте основные мысли и идеи, выстройте логику повествования, отберите яркие и точные примеры, сформулируйте выводы.

При подготовке к устному выступлению возьмите на вооружение некоторые советы:

- Лучший вариант выступления перед аудиторией – это свободная речь, не осложненная чтением текста. Но если у вас не выработано умение общаться с публикой без бумажки, то не пытайтесь сделать это сразу, без подготовки. Осваивать этот опыт нужно постепенно, от доклада к докладу увеличивая объем речи без заглядывания в текст.

- Если вы намерены считать доклад с заготовленных письменных записей, то постарайтесь, чтобы чтение было «художественным»: обозначайте паузой логические переходы от части к части, выделяйте интонационно особо важные мысли и аргументы, варьируйте темп речи.

- Читая доклад, не торопитесь, делайте это как можно спокойнее. Помните, что скорость произношения текста перед слушателями всегда должна быть более медленной, чем скорость вашей повседневной речи.

- Сверьте письменный текст с хронометром, для этого прочитайте его несколько раз с секундомером в руках. В случае, если доклад окажется слишком длинным или коротким, проведите его реконструкцию. Однако вместе с сокращениями или дополнениями не «потеряйте» тему. Не поддавайтесь искушению рассказать все, что знаете – полно и подробно.

- Обратите внимание на тембр и силу вашего голоса. Очень важно, чтобы вас было слышно в самых отдаленных частях аудитории, и при этом вы не «глушили» вблизи вас находящихся слушателей. Варьируйте тембр речи, он придаст ей выразительность и поможет избежать монотонности.

- Следите за своими жестами. Чрезмерная жестикуляция отвлекает от содержания доклада, а полное ее отсутствие снижает действенную силу выступления. Постарайтесь избавиться от жестов, демонстрирующих ваше волнение (когда крутятся ручки, теребятся пуговицы, заламываются пальцы). Используйте жесты – выразительные, описательные, подражательные, указующие – для полноты передачи ваших мыслей.

- Установите зрительный контакт с аудиторией. Не стоит все время смотреть в окно, опускать глаза или сосредотачиваться на тексте. Старайтесь зрительно общаться со всеми слушателями, переводя взгляд от одних к другим. Не обращайтесь на опоздавших и не прерываете свой доклад замечаниями. Но вместе с тем следите за реакцией публики на ваше выступление (одобрение, усталость, интерес, скуку) и если сможете, вносите коррективы в речь с целью повышения интереса к его содержанию.

- Отвечать на вопросы в конце выступления надо кратко, четко и уверенно, без лишних подробностей и повторов. Постарайтесь предугадать возможные вопросы своих слушателей и подготовиться к ним заранее. Но если случится, что вы не знаете ответа на заданный вам вопрос, не бойтесь в этом признаться. Это значительно лучше, чем отвечать не по существу или отшучиваться.

- Проведите генеральную репетицию своего доклада перед друзьями или близкими. Это поможет заранее выявить некоторые недостатки – стилистически слабые места, труднопроизносимые слова и фразы, затянутые во времени части и т.д. Проанализируйте свою дикцию, интонации, жесты. Сделайте так, чтобы они помогали, а не мешали успешно представить публике подготовленный вами доклад.

Презентация – современный способ устного или письменного представления информации с использованием мультимедийных технологий.

Существует несколько вариантов презентаций.

- Презентация с выступлением докладчика
- Презентация с комментариями докладчика

- Презентация для самостоятельного просмотра, которая может демонстрироваться перед аудиторией без участия докладчика.

Подготовка презентации включает в себя несколько этапов:

1. Планирование презентации

От ответов на эти вопросы будет зависеть всё построение презентации:

- каково предназначение и смысл презентации (демонстрация результатов научной работы, защита дипломного проекта и т.д.);
- какую роль будет выполнять презентация в ходе выступления (сопровождение доклада или его иллюстрация);
- какова цель презентации (информирование, убеждение или анализ);
- на какое время рассчитана презентация (короткое - 5-10 минут или продолжительное - 15-20 минут);
- каков размер и состав зрительской аудитории (10-15 человек или 80-100; преподаватели, студенты или смешенная аудитория).

2. Структурирование информации

- в презентации не должна быть менее 10 слайдов, а общее их количество превышать 20 - 25.
- основными принципами при составлении презентации должны быть ясность, наглядность, логичность и запоминаемость;
- презентация должна иметь сценарий и четкую структуру, в которой будут отражены все причинно-следственные связи,
- работа над презентацией начинается после тщательного обдумывания и написания текста доклада, который необходимо разбить на фрагменты и обозначить связанные с каждым из них задачи и действия;
- первый шаг – это определение главной идеи, вокруг которой будет строиться презентация;
- часть информации можно перевести в два типа наглядных пособий: текстовые, которые помогут слушателям следить за ходом развертывания аргументов и графические, которые иллюстрируют главные пункты выступления и создают эмоциональные образы.
- сюжеты презентации могут разъяснять или иллюстрировать основные положения доклада в самых разнообразных вариантах.

Очень важно найти правильный баланс между речью докладчика и сопровождающими её мультимедийными элементами.

Для этого целесообразно:

- определить, что будет представлено на каждом слайде, что будет в это время говориться, как будет сделан переход к следующему слайду;
- самые важные идеи и мысли отразить и на слайдах и произнести словами, тогда как второстепенные – либо словами, либо на слайдах;
- информацию на слайдах представить в виде тезисов – они сопровождают подробное изложение мыслей выступающего, а не наоборот;
- для разъяснения положений доклада использовать разные виды слайдов: с текстом, с таблицами, с диаграммами;
- любая презентация должна иметь собственную драматургию, в которой есть:
«завязка» - пробуждение интереса аудитории к теме сообщения (яркий наглядный пример);
«развитие» - демонстрация основной информации в логической последовательности (чередование текстовых и графических слайдов);
«кульминация» - представление самого главного, нового, неожиданного (эмоциональный речевой или иллюстративный образ);
«развязка» - формулирование выводов или практических рекомендаций (видеоряд).

3. Оформление презентации

Оформление презентации включает в себя следующую обязательную информацию:

Титульный лист

- представляет тему доклада и имя автора (или авторов);
- на защите курсовой или дипломной работы указывает фамилию и инициалы научного руководителя или организации;
- на конференциях обозначает дату и название конференции.

План выступления

- формулирует основное содержание доклада (3-4 пункта);
- фиксирует порядок изложения информации;

Содержание презентации

- включает текстовую и графическую информацию;
- иллюстрирует основные пункты сообщения;
- может представлять самостоятельный вариант доклада;

Завершение

- обобщает, подводит итоги, суммирует информацию;
- может включать список литературы к докладу;
- содержит слова благодарности аудитории.

4. Дизайн презентации

Используйте брендинг вуза для оформления презентации, для этого на сайте МГТУ скачайте шаблон презентации.

Текстовое оформление

- Не стоит заполнять слайд слишком большим объемом информации - лучше всего запоминаются не более 3-х фактов, выводов, определений.
- Оптимальное число строк на слайде – 6 -11.
- Короткие фразы запоминаются визуально лучше. Пункты перечней не должны превышать двух строк на фразу.
- Цифровые материалы лучше представить в виде графиков и диаграмм.
- Необходимо обратить внимание на грамотность написания текста. Ошибки во весь экран производят неприятное впечатление

Шрифтовое оформление

- Для заголовка рекомендуемый размер шрифта 24-54 пункта, а для текста - 18-36 пунктов.
- Курсив, подчеркивание, жирный шрифт используются ограниченно, только для смыслового выделения фрагментов текста.
- Для основного текста не рекомендуются прописные буквы.

Цветовое оформление

- На одном слайде не используется более трех цветов: фон, заголовок, текст.
- Цвет шрифта и цвет фона должны контрастировать – текст должен хорошо читаться, но не резать глаза.
- Если презентация большая, то есть смысл разделить её на части с помощью цвета – разный цвет способен создавать разный эмоциональный настрой.
- Нельзя выбирать фон, который содержит активный рисунок.

Композиционное оформление

- Следует соблюдать единый стиль оформления. Он может включать определенный шрифт (гарнитура и цвет), фон цвета или фоновый рисунок, декоративный элемент небольшого размера и т.д.
- Не приемлемы стили, которые будут отвлекать от презентации.
- Крупные объекты в композиции смотрятся неважно.
- Вспомогательная информация (управляющие кнопки) не должна преобладать над основной (текстом и иллюстрацией).

Анимационное оформление

- Основная роль анимации – дозирования информации. Аудитория, как правило, лучше воспринимает информацию порциями, небольшими зрительными фрагментами.
- Анимация используется для привлечения внимания или демонстрации развития какого-либо процесса
- Не стоит злоупотреблять анимационными эффектами, которые отвлекают от содержания или утомляют глаза читающего.
- Особенно нежелательно частое использование таких анимационных эффектов как вылет, вращение, волна, побуквенное появление текста.

Звуковое оформление

- Музыкальное сопровождение призвано отразить суть или подчеркнуть особенности темы слайда или всей презентации, создать определенный эмоциональный настрой.

- Музыка целесообразно включать тогда, когда презентация идет без словесного сопровождения.
- Звуковое сопровождение используется только по необходимости, поскольку даже фоновая тихая музыка создает излишний шум и мешает восприятию содержания.
- Необходимо выбрать оптимальную громкость, чтобы звук был слышан всем слушателем, но не был оглушительным.

Графическое оформление

- Рисунки, фотографии, диаграммы призваны дополнить текстовую информацию или передать её в более наглядном виде.
- Нельзя представлять рисунки и фото плохого качества или с искаженными пропорциями.
- Желательно, чтобы изображение было не столько фоном, сколько иллюстрацией, равной по смыслу самому тексту, чтобы помочь по-новому понять и раскрыть его.
- Следует избегать некорректных иллюстраций, которые неправильно или двусмысленно отражают смысл информации.
- Необходимо позаботиться о равномерном и рациональном использовании пространства на слайде: если текст первичен, то текстовый фрагмент размещается в левом верхнем углу, а графический рисунок внизу справа и наоборот.
- Иллюстрации рекомендуется сопровождать пояснительным текстом. Подписи к картинкам лучше выполнять сбоку или снизу, если это только не название самого слайда.
- Если графическое изображение используется в качестве фона, то текст на этом фоне должен быть хорошо читаем.

Таблицы и схемы

- Не стоит вставлять в презентацию большие таблицы – они трудны для восприятия. Лучше заменить их графиками, построенными на основе этих таблиц.
- Если все же таблицу показать надо, то следует оставить как можно меньше строк и столбцов, отобразив и разместив только самые важные данные.
- При использовании схем на слайдах необходимо выровнять ряды блоков схемы, расстояние между блоками, добавить соединительные линии при помощи инструментов Автофигур,
- При создании схем нужно учитывать связь между составными частями схемы: если они равнозначны, то заполняются одним шрифтом, фоном и текстом, если есть первостепенная информация, то она выделяется особым способом с помощью организационных диаграмм.

Аудио и видео оформление

- Видео, кино и теле материалы могут быть использованы полностью или фрагментарно в зависимости от целей, которые преследуются.
- Продолжительность фильма не должна превышать 15-25 минут, а фрагмента – 4-6 минут.
- Нельзя использовать два фильма на одном мероприятии, но показать фрагменты из двух фильмов вполне возможно.

Подготовка к зачёту/экзамену. Готовиться к зачёту/экзамену нужно заранее и в несколько этапов. Для этого:

- Просматривайте конспекты лекций сразу после занятий. Это поможет разобраться с непонятными моментами лекции и возникшими вопросами, пока еще лекция свежа в памяти.
- Бегло просматривайте конспекты до начала следующего занятия. Это позволит «освежить» предыдущую лекцию и подготовиться к восприятию нового материала.
- Каждую неделю отводите время для повторения пройденного материала. Непосредственно при подготовке:
- Упорядочьте свои конспекты, записи, задания.
- Прикиньте время, необходимое вам для повторения каждой части (блока) материала, выносимого на зачет.
- Составьте расписание с учетом скорости повторения материала, для чего
- Разделите вопросы для зачёта/экзамену на знакомые (по лекционному курсу, семинарам, конспектированию), которые потребуют лишь повторения и новые, которые придется осваивать самостоятельно. Начните с тем хорошо вам известных и закрепите их с помощью

конспекта и учебника. Затем пополните свой теоретический багаж новыми знаниями, обязательно воспользовавшись рекомендованной литературой.

- Правильно используйте консультации, которые проводит преподаватель. Приходите на них с заранее проработанными самостоятельно вопросами. Вы можете получить разъяснение по поводу сложных, не до конца понятых тем, но не рассчитывайте во время консультации на исчерпывающую информацию по содержанию всего курса.

Для успешного выполнения лабораторных работ рекомендуем использовать предложенную литературу в работе.

При прохождении МООС (при условии его заявления в изучаемом предмете) рекомендуется следить за сроками выполнения заданий курса, не выполнение заданий вовремя влечет понижение баллов в прогрессе курса.

