



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования

КГЭУ

«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГЭУ»)

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
для проведения текущего контроля успеваемости и промежуточной
аттестации студентов по итогам освоения дисциплины

Обработка естественного языка

Направление подготовки 09.04.01 Информатика и вычислительная техника

Направленность (профиль) Инженерия искусственного интеллекта

Квалификация Магистр

Форма обучения Очная

Составлено авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

1. Цель и задачи текущего контроля и промежуточной аттестации студентов по дисциплине «Обработка естественного языка»

Цель текущего контроля - систематическая проверка степени освоения программы дисциплины «Обработка естественного языка», уровня сформированности знаний, умений, навыков, компетенций на текущих занятиях

Задачи текущего контроля:

1. определение индивидуального учебного рейтинга студентов;
2. своевременное выполнение корректирующих действий по содержанию и организации процесса обучения; обнаружение и устранение пробелов в усвоении учебной дисциплины;
3. подготовки к промежуточной аттестации.

В течение семестра при изучении дисциплины реализуется комплексная система поэтапного оценивания уровня освоения – балльно-рейтинговая система. За каждый вид учебных действий студенты получают определенное количество баллов. В течение семестра студент может набрать до 60-ти баллов.

Цель промежуточной аттестации - проверка степени усвоения студентами учебного материала за время изучения дисциплины, уровня сформированности компетенций после завершения изучения дисциплины. Аттестация проходит в форме зачета.

Задачи промежуточной аттестации:

1. определение уровня усвоения учебной дисциплины;
2. определение уровня сформированности компетенций.

2. Основное содержание текущего контроля и промежуточной аттестации студентов

В результате изучения дисциплины «Обработка естественного языка» формируются следующие компетенции или их составляющие:

2.1. Основное содержание текущего контроля

Коды компетенций	Совокупность ожидаемых результатов образования студентов в форме компетенций по завершении модуля / освоения дисциплины	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3
ПК-7	ПК-7.1 Руководит проектами в области сквозной цифровой субтехнологии «Компьютерное зрение» ПК-7.2 Руководит проектами в области сквозной цифровой субтехнологии «Обработка естественного языка» ПК-7.3 Исследует и анализирует развитие новых направлений и перспективных методов и технологий в области искусственного интеллекта, участвует в исследовательских проектах по развитию перспективных направлений в области искусственного интеллекта (алгоритмическая имитация биологических систем принятия решений, автономное самообучение и развитие адаптивности)	Контрольная работа; домашняя работа; практическая работа; зачёт

	алгоритмов к новым задачам, автономная декомпозиция сложных задач, поиск и синтез решений)	
ОПК-9	ОПК-9.1 Применяет инструментальные среды, программно-технические платформы для решения задач в области создания и применения искусственного интеллекта ОПК-9.2 Разрабатывает оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта	Контрольная работа; домашняя работа; практическая работа; зачёт

2.2 Основное содержание промежуточной аттестации студентов

Код и наименование компетенций, формируемые с участием дисциплины	Индикаторы достижения компетенции	Планируемые результаты обучения	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3	4
ПК-7	ПК-7.1 Руководит проектами в области сквозной цифровой субтехнологии «Компьютерное зрение» ПК-7.2 Руководит проектами в области сквозной цифровой субтехнологии «Обработка естественного языка» ПК-7.3 Исследует и анализирует развитие новых направлений и перспективных методов и технологий в области искусственного интеллекта, участвует в исследовательских проектах по развитию перспективных направлений в области искусственного интеллекта (алгоритмическая имитация биологических систем принятия решений, автономное	Знать: принципы построения систем компьютерного зрения, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение» Уметь: руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Компьютерное зрение» Знать: принципы построения систем обработки естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка» Уметь: руководить проектами по созданию, внедрению и	Контрольная работа; домашняя работа; практическая работа; зачёт

	самообучение и развитие адаптивности алгоритмов к новым задачам, автономная декомпозиция сложных задач, поиск и синтез решений)	поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка» Знать: современное состояние и перспективы развития новых направлений, методов и технологий в области искусственного интеллекта Уметь: проводить анализ новых направлений, методов и технологий в области искусственного интеллекта и определять наиболее перспективные для различных областей применения	
ОПК-9	ОПК-9.1 Применяет инструментальные среды, программно-технические платформы для решения задач в области создания и применения искусственного интеллекта ОПК-9.2 Разрабатывает оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта	Знать: инструментальные среды, программно-технические платформы для решения профессиональных задач Уметь: применять инструментальные среды, программно-технические платформы для решения профессиональных задач Знать: принципы разработки оригинальных программных средств для решения профессиональных задач Уметь: разрабатывать оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта	Контрольная работа; домашняя работа; практическая работа; зачёт

3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины

1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Контрольная работа	3 сем., 13 нед.	80
Самостоятельное изучение материала	3 сем., 1-15 нед.	20
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5		
Промежуточная аттестация по лекциям – <i>Зачет</i>		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0.5		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Выполнение и оформление практических работ	3 сем., 15 нед.	50
Домашняя работа №1	3 сем., 10 нед.	25
Домашняя работа №2	3 сем., 14 нед.	25
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – 1		
Промежуточная аттестация по практическим/семинарским занятиям – не предусмотрена		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – 0		
3. Лабораторные занятия: Не предусмотрены		
коэффициент значимости совокупных результатов лабораторных занятий – 0		

4. Практические занятия

№ п/п	Примерный перечень тем практических работ
1	Предварительная обработка текста для анализа.
2	Векторизация текста.
3	Классификация текста с использованием классических методов машинного обучения.
4	Классификация текста с использованием глубоких нейронных сетей.
5	Языковая модель. Обучение языковой модели.
6	Автоматическая генерация текста.
7	Поиск именованных сущностей в тексте.
8	Механизм внимания в нейронных сетях. Сети с архитектурой Transformer.
9	Перенос обучения в задачах обработки текстов.

Примерная тематика контрольных работ:

Проектирование пайплайна для задач обработки естественного языка.

Примерные задания в составе контрольных работ:

Спроектировать последовательность действий для решения задачи анализа текста с помощью машинного обучения. Пайплайн должен включать:

1. Метод подготовки текста для обработки.
2. Подход к токенизации текста.
3. Подход к векторизации текста.
4. Используемую модель машинного обучения.
5. Метод обучения модели.
6. Метод оценки качества модели.
7. Использование обученной модели для решения задачи анализа текста.
8. Другие шаги, которые могут понадобиться при решении задачи.

Примеры задач обработки естественного языка, для которых нужно составлять пайплайны:

- Классификация текста.
- Определение эмоциональной окраски текста.
- Автоматическая генерация текста.
- Поиск именованных сущностей в тексте.

Примерная тематика домашних работ:

Домашняя работа №1:

Обучение языковой модели для текстов на русском языке.

Домашняя работа №2:

Дообучение предварительно обученной сети BERT

Примерные задания в составе домашних работ:

1. Обучите языковую модель для русского языка и используйте ее для генерации текста. Для этого:
 - Подготовьте набор данных с текстами на русском языке. Можно использовать готовые наборы данных или создать собственный.
 - Обучите языковую модель на подготовленном наборе данных.
 - Используя обученную языковую модель сгенерируйте пять примеров текстов на русском языке.
 - Выложите набор данных, код и обученную модель в открытый доступ на GitHub.
 - Оформите презентацию или технологическую статью о ходе работы, обосновании принятых решений и результатах работы.
 - (Не обязательное задание). Запишите видео с демонстрацией работы созданного решения.

2. Дообучите предварительно обученную сеть с архитектурой Transformer для классификации текстов на русском языке. Для этого:
 - Подготовьте набор данных с текстами на русском языке для классификации. Можно использовать готовые наборы данных или создать собственный.
 - Выберите предварительно обученную нейронную сеть с архитектурой Transformer, подходящую для задачи классификации текстов на русском языке.
 - Выполните дообучение выбранной нейронной сети на подготовленном наборе данных.
 - Проведите тестирование классификации текстов с помощью дообученной нейронной сети и оцените качество работы сети.
 - Выложите набор данных, код и дообученную модель в открытый доступ на GitHub.
 - Оформите презентацию или технологическую статью о ходе работы, обосновании принятых решений и результатах работы.
 - (Не обязательное задание). Запишите видео с демонстрацией работы созданного решения.

Пример дообучения нейронной сети BERT в TensorFlow – https://www.tensorflow.org/text/tutorials/fine_tune_bert

Ноутбук с примером кода решения – https://colab.research.google.com/github/tensorflow/text/blob/master/docs/tutorials/fine_tune_bert.ipynb

Пример дообучения нейронных сетей с архитектурой Transformer в Hugging Face – <https://huggingface.co/transformers/training.html>

Список примерных вопросов для зачета:

1. Теоретические аспекты обработки естественного языка.
2. Особенности обработки текста на английском языке.
3. Особенности обработки текста на русском языке.
4. Предварительная обработка текста. Очистка текста. Удаление стоп-слов/наиболее и наименее частотных слов.
5. Токенизация, стемминг, лемматизация текста.
6. Методы векторизации текста: построение словаря, мешок слов.

7. Методы векторизации текста: TF-IDF.
8. Методы векторизации текста: word2vec.
9. Методы векторизации текста: fasttext
10. Методы векторизации текста: GloVe.
11. Классические методы машинного обучения для решения задач классификации текста.
12. Классические методы машинного обучения для решения определения тональности текста.
13. Архитектуры нейронных сетей для обработки текстов: LSTM.
14. Архитектуры нейронных сетей для обработки текстов: GRU.
15. Архитектуры нейронных сетей для обработки текста: одномерные сверточные сети.
16. Классификация текста с помощью нейронных сетей.
17. Определение тональности текста с помощью нейронных сетей.
18. Языковая модель.
19. Обучение языковой модели.
20. Основные подходы к генерации текста.
21. Задача поиска именованных сущностей в тексте.
22. Применение нейронных сетей для поиска именованных сущностей.
23. Механизм внимания в нейронных сетях.
24. Применение механизма внимания для обработки текста.
25. Архитектура нейронных сетей Transformer.
26. Предварительно обученные нейронные сети для обработки текстов BERT.
27. Предварительно обученные нейронные сети для обработки текстов GPT.
28. Перенос обучения для задач обработки текстов.
29. Классификация текста с помощью сетей с архитектурой Transformer.
30. Генерация текста с помощью сетей с архитектурой Transformer.
31. Поиск именованных сущностей в тексте с помощью сетей с архитектурой Transformer.