УТВЕРЖДАЮ

Зав. кафедрой СП ЮУрГУ

\_\_\_\_\_\_\_\_\_Л.Б. Соколинский

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2022 г.

Фонд оценочных средств ООП «Искусственный интеллект и инженерия данных» по направлению 09.04.04 – Программная инженерия

Дисциплина «Анализ естественного языка методами искусственного интеллекта»

| **№ КМ** | **Вид КМ** | **Наименование КМ** | **Оценочные средства** |
| --- | --- | --- | --- |
|  | Текущий контроль | Реализация собственного POS-тэггера | Вопросы для подготовки к устному опросу:   1. Основные задачи NLP. 2. Представления текстовых данных. 3. Предобработка текста, лемматизация, стемминг |
|  | Текущий контроль | Классификация текстов методами машинного обучения | Вопросы для подготовки к устному опросу:  1. Классификации текстовых документов на  основе частотных мер  2. Деревья решений в задаче классификации текстов  3. Наивный байесовский классификатор в задаче классификации текстов  4. Логистическая регрессия в задаче классификации текстов |
|  | Текущий контроль | Классификация текстов на основе нейросетевых моделей языка | Вопросы для подготовки к устному опросу:  1. Задача классификации текстов  2. Языковые модели  3. Особенности нейросетевой модели языка word2vec  4. Особенности нейросетевой модели языка fasttext  5. Мера семантической близости. |
|  | Текущий контроль | Кластеризация текстовой коллекции методами тематического моделирования | Вопросы для подготовки к устному опросу:  1. Задача кластеризации текстовых документов  2. Суть метода тематического моделирования  3. Особенность метода LSA  4. Особенность метода pLSA  5. Аддитивная регуляризация тематических моделей |
|  | Текущий контроль | Классификация текстов с помощью различных архитектур глубоких нейронных сетей | Вопросы для подготовки к устному опросу:  1. Сверточная нейронная сеть (CNN)  2. Сеть с кратковременной памятью (LSTM)  3. Особенности применения нейронных сетей в задаче классификации текстов |
|  | Текущий контроль | Реализация вопросно-ответной системы. Решение задачи автоматического реферирования | Вопросы для подготовки к устному опросу:  1. Задача машинного перевода  2. Задача автоматического реферирования (summarization)  3. Задача вопросно-ответный системы  4. Механизм внимания (attention)  5. Архитектуры encoder-decoder-attention. |
|  | Текущий контроль | Архитектуры трансформеров: BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерации текстов | Вопросы для подготовки к устному опросу:  1. Transfer learning в задачах анализа текстов  2. Self-Attention  3. Модель BERT в задачах классификации текстов  4. Модель GPT в задачах классификации текстов  5. Cравнение BERT и GPT и влияние размера модели на качество  6. Методы предсказания пропущенных слов  7. Методы генерации текстов.  8. Fine-tuning трансформеров. |
|  | Текущий контроль | Разработка диалогового агента для чат-бота | Вопросы для подготовки к устному опросу:  1. Архитектура диалоговых систем.  2. Модули понимания естественного языка (NLU) и диалоговый менеджер (DM).  3. Проблемы при построения диалоговых систем.  4. Особенности проектирования UX/UI диалоговых ассистентов в чатах и голосе.  5. Современные фреймворки для построения диалоговых систем: DeepPavlov, Rasa, Just AI Conversational Platform |
|  | Промежуточный контроль | Итоговый тест | Вопросы для подготовки к экзамену по дисциплине:  1. Основные задачи NLP.  2. Представления текстовых данных.  3. Предобработка текста, лемматизация, стемминг.  4. Методы машинного обучения для классификации текстовых документов на основе частотных мер (TF-IDF).  5. Деревья решений  6. наивный байесовский классификатор  7. логистическая регрессия в задаче классификации текстов.  8. Языковые модели. Word embeddings.  9. Нейросетевые модели языка: word2vec, fasttext. Мера семантической близости.  10. Классификация текстов на основе нейросетевых моделей языка.  11. Кластеризация текстовых документов.  12. Тематическое моделирование Методы LSA, pLSA.  13. Аддитивная регуляризация тематических моделей в BigARTM  14. Классификация текстов с помощью глубоких нейронных сетей: CNN, LSTM.  15. Задачи обработки последовательностей: машинной перевод, автоматическое реферирование (summarization), вопросно-ответные системы.  16. Механизм внимания (attention).  17. Архитектуры encoder-decoder-attention.  18. Transfer learning в задачах анализа текстов. Self-Attention.  19. Архитектуры трансформеров: BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерации текстов. Fine-tuning трансформеров.  20. Построение диалоговых систем. Архитектура диалоговых систем.  21. Модули понимания естественного языка (NLU) и диалоговый менеджер (DM).  22. Сложности построения диалоговых систем.  23. Проектирование UX/UI диалоговых ассистентов в чатах и голосе.  24. Обзор современных фреймворков для построения диалоговых систем: DeepPavlov, Rasa, Just AI Conversational Platform. |

Паспорт фонда оценочных средств приведен в п. 6.3 РПД.

Разработчик С.А. Иванов