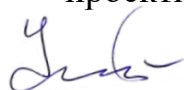


**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Утверждаю:
Зав. кафедрой компьютерных
интеллектуальных технологий
проектирования

 М.И. Чижов
«21» декабря 2021 г.

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**

«Технологии работы с естественным языком Nature Language Processing»

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Направленность (профиль): Искусственный интеллект

Квалификация выпускника магистр

Нормативный период обучения 2 года / 2 года и 5 м.

Форма обучения очная / заочная

Год начала подготовки 2022

Разработчик


подпись

В.В. Сокольников

Процесс изучения дисциплины «Технологии работы с естественным языком Nature Language Processing» направлен на формирование следующих компетенций:

УК-1 - Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий

УК-2 - Способен управлять проектом на всех этапах его жизненного цикла

ОПК-10 - Способен адаптировать и применять на практике классические и новые научные принципы и методы исследований для решения задач в области создания и применения технологий и систем искусственного интеллекта и методы исследований

ПК-2 - Способен участвовать в решении профессиональных проектных задач, выбирать и реализовывать командную роль в работе над проектом в соответствии с приоритетами собственной деятельности

Перечень планируемых результатов обучения и показателей оценивания сформированности компетенций на этапе промежуточной аттестации

№ п/п	Компетенция	Результаты обучения, характеризующие сформированность компетенции	Тип ОМ	Показатели оценивания
1	УК-1	Знает принципы разработки стратегии достижения поставленной цели как последовательность шагов, предвидя результат каждого шага и оценивая их влияние на внешнее окружение планируемой деятельности на взаимоотношения участников этой деятельности	Вопросы (тест) к зачету	Полнота знаний
		Знает как осуществляется поиск вариантов решения поставленной задачи на основе доступных источников информации; определяет в рамках выбранного алгоритма вопросы, подлежащие дальнейшей разработке	Стандартные задания	Наличие умений
		Владеет навыками анализа проблемной ситуации как системы, выявляя ее составляющие и связи между ними	Прикладные задания	Наличие навыков
2	УК-2	Знать принципы формирования результатов проекта (или отдельных его этапов) в форме отчета, статьи, выступления на научно-практической конференции	Вопросы (тест) к зачету	Полнота знаний
		Уметь организовать и координировать работу участников проекта, обеспечивает работу команды необходимыми ресурсами	Стандартные задания	Наличие умений
		Владеть навыками формирования план-графика реализации проекта в целом и контролирует его выполнение	Прикладные задания	Наличие навыков
3	ОПК-10	Знать методологии разработки программного обеспечения, методологии управления проектами разработки программного обеспечения, методы и средства организации проектных данных, лучшие практики управления разработкой программного обеспечения, нормативно-технические документы (стандарты и	Вопросы (тест) к зачету	Полнота знаний

		процессы управления инфраструктурой коллективной среды разработки		
		Уметь применять методологии разработки программного обеспечения, методологии управления проектами разработки программного обеспечения, методы и средства организации проектных данных, лучшие практики и отражать их в базе знаний, основные принципы и методы управления персоналом, нормативно-технические документы (стандарты и регламенты), описывающие процессы управления инфраструктурой коллективной среды разработки	Стандартные задания	Наличие умений
		Владеть навыками выбора инструментальных средств разработки, определения набора библиотек повторно используемых модулей, выбора средств создания и ведения репозитория, учета задач, сборки и непрерывной интеграции, базы знаний	Прикладные задания	Наличие навыков
4	ПК-2	Знать сквозные цифровые технологии искусственного интеллекта, включая «Компьютерное зрение», «Обработка естественного языка», «Рекомендательные системы и системы поддержки принятия решений», «Распознавание и синтез речи»; современный опыт применения систем искусственного интеллекта	Вопросы (тест) к зачету	Полнота знаний
		Уметь проводить сравнительный анализ методов и инструментальных средств для решения задач искусственного интеллекта; разрабатывать и оценивать модели искусственного интеллекта; программировать на языках высокого уровня, ориентированных на применение методов и алгоритмов искусственного интеллекта.	Стандартные задания	Наличие умений
		Владеть навыками выявления требований заказчика к результатам решения задач искусственного интеллекта; определения возможностей применения методов искусственного интеллекта в предметной области решаемой задачи; использования имеющейся методологической и технологической инфраструктуры анализа и обработки данных.	Прикладные задания	Наличие навыков

ОПИСАНИЕ ПОКАЗАТЕЛЕЙ, КРИТЕРИЕВ И ШКАЛ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ НА ЭТАПЕ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Показатели оценивания компетенций	Шкала и критерии оценки уровня сформированности компетенции			
	Неудовлетворительный	Минимально допустимый (пороговый)	Средний	Высокий
Полнота знаний	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущены не грубые ошибки.	Уровень знаний в объеме, соответствующем программе подготовки. Допущены некоторые погрешности.	Уровень знаний в объеме, соответствующем программе подготовки
Наличие умений	При выполнении стандартных заданий не продемонстрированы основные умения. Имели место грубые ошибки.	Продemonстрированы основные умения. Выполнены типовые задания с не грубыми ошибками. Выполнены все задания, но не в полном объеме (отсутствуют пояснения, неполные выводы)	Продemonстрированы все основные умения. Выполнены все основные задания с некоторыми погрешностями. Выполнены все задания в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения. Выполнены все основные и дополнительные задания без ошибок и погрешностей. Задания выполнены в полном объеме без недочетов.
Наличие навыков (владение опытом)	При выполнении стандартных заданий не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для выполнения стандартных заданий с некоторыми недочетами.	Продemonстрированы базовые навыки при выполнении стандартных заданий с некоторыми недочетами.	Продemonстрированы все основные умения. Выполнены все основные и дополнительные задания без ошибок и погрешностей. Продemonстрирован творческий подход к решению нестандартных задач.
Характеристика сформированности компетенции	Компетенция в полной мере не сформирована. Имеющихся знаний, умений, навыков недостаточно для решения практических (профессиональных) задач. Требуется повторное обучение.	Сформированность компетенции соответствует минимальным требованиям. Имеющихся знаний, умений, навыков в целом достаточно для решения практических (профессиональных) задач, но требуется дополнительная практика по большинству профессиональных задач.	Сформированность компетенций в целом соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в целом достаточно для решения стандартных профессиональных задач.	Сформированность компетенции полностью соответствует требованиям. Имеющихся знаний, умений, навыков и мотивации в полной мере достаточно для решения сложных профессиональных задач.

ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ МАТЕРИАЛОВ

Вопросы (тестовые задания) для оценки результатов обучения, характеризующих сформированность компетенций

<i>УК-1 - Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий</i>	
1.	Типы задач обработки естественного языка
2.	Традиционные методы и методы глубинного обучения
3.	Представления для слов естественного языка
4.	Векторное представление
5.	Целевая функция векторизации слов
6.	Построение векторизованной модели словаря
<i>УК-2 - Способен управлять проектом на всех этапах его жизненного цикла</i>	
1.	Сокращение размерности словаря
2.	Методы обучения
3.	Задачи классификации
4.	Учет структуры предложения
5.	Безконтекстная грамматика
6.	Проблема неопределенности модификатора
<i>ОПК-8 - Способен адаптировать и применять на практике классические и новые научные принципы и методы исследований для решения задач в области создания и применения технологий и систем искусственного интеллекта и методы исследований</i>	
1.	Типы зависимостей
2.	Статистическая модель языка
3.	N-граммы
4.	Рекуррентные нейронные сети
5.	Оценка качества перевода
6.	Отображение последовательностей (seq2seq)
<i>ПК-2 - Способен участвовать в решении профессиональных проектных задач, выбирать и реализовывать командную роль в работе над проектом в соответствии с приоритетами собственной деятельности</i>	
1.	Свёрточные нейронные сети
2.	Применение в системах анализа естественного языка
3.	Работа с частями слов
4.	Буквенные (character-based) модели
5.	Гибридные модели
6.	Контекстные представления
7.	Проблема неизвестных слов
8.	Механизмы внимания, их реализация, применение
9.	Самовнимание
10.	Трансформеры
11.	Перенесение обучения между языковыми моделями
12.	Обзор применения технологий работы с естественным языком

**Практические задания для оценки результатов обучения,
характеризующих сформированность компетенций**

УК-1 - Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	
1.	Какой из следующих методов можно использовать для нормализации ключевых слов в NLP, процесс преобразования ключевого слова в его базовую форму +а) Лемматизация б) Soundex с) Косинус сходства д) N-грамм
2.	N-граммы определяются как комбинация N ключевых слов вместе. Сколько би-грамм можно составить из данного предложения: “Analytics Vidhya is a great source to learn data science” а)7 б)8 +с)9 д)10
3.	Какой из следующих методов можно использовать для вычисления расстояния между двумя векторами слов в NLP? а) Лемматизация +б) Евклидово расстояние +с) Косинус сходства д) N-грамм
4.	Каковы возможные особенности корпуса текстов в NLP? а) Количество слов в документе б) Векторное обозначение слова с) Часть речевого тега д) Базовая грамматика зависимостей е) Все вышеперечисленное
5.	Вы создали матрицу терминов документа на основе входных данных 20 000 документов для модели машинного обучения. Что из перечисленного можно использовать для уменьшения размерности данных? 1 Нормализация ключевых слов. 2 Скрытое семантическое индексирование. 3 Скрытое распределение Дирихле. А) только 1 Б) 2, 3 В) 1, 3 +Г) 1, 2, 3
6.	Какой из методов синтаксического анализа текста можно использовать для обнаружения именной фразы, глагольной фразы, обнаружения субъекта и обнаружения объекта в NLP. а. Тегирование части речи б. Пропустить извлечение граммов и N-граммов в. Непрерывный мешок слов +г. Анализ зависимостей и анализ групп

7.	Что из нижеперечисленного относится к методам нормализации ключевых слов в NLP? +а. Стемминг б. Часть речи в. Распознавание именованных объектов +г. Лемматизация
УК-2 - Способен управлять проектом на всех этапах его жизненного цикла	
1.	Что из нижеперечисленного относится к вариантам использования NLP? а. Обнаружение объектов на изображении б. Распознавание лиц в. Речь биометрическая +г. Обобщение текста
2.	В корпусе из N документов один случайно выбранный документ содержит в общей сложности T терминов, а термин «привет» встречается K раз. Каково правильное значение произведения TF (частота терминов) и IDF (обратная частота документа), если термин «привет» встречается примерно в одной трети всех документов? а. $KT * \text{Log}(3)$ б. $T * \text{Log}(3) / K$ +в. $K * \text{Log}(3) / T$ г. $\text{Log}(3) / KT$
3.	В NLP алгоритм уменьшает вес часто используемых слов и увеличивает вес слов, которые редко используются в наборе документов. а. Термин Частота (TF) +б. Обратная частота документа (IDF) в. Word2Vec г. Скрытое распределение Дирихле (LDA)
4.	В NLP процесс удаления таких слов, как «и», «является», «а», «ан», «то» из предложения называется а. Стемминг б. лемматизация +в. Стоп-слово г. Все вышеперечисленное
5.	В NLP процесс преобразования предложения или абзаца в токены называется стеммингом. а. Истинный +б. Ложь
6.	В NLP токены преобразуются в числа перед передачей в любую нейронную сеть. +а. Истинный б. Ложь
7.	определить лишнее а. nltk б. scikit learn в. SpaCy +г. BERT

ОПК-10 - Способен адаптировать и применять на практике классические и новые научные принципы и методы исследований для решения задач в области ния и применения технологий и систем искусственного интеллекта и методы исследований

1.	TF-IDF помогает вам установить? а. наиболее часто встречающееся слово в документе +б. самое важное слово в документе
2.	В NLP процесс идентификации людей, организации из заданного абзаца предложения называется а. Стемминг б. Лемматизация в. Удаление стоп-слов +д. Распознавание именованных объектов
3.	Что из нижеперечисленного не относится к методам предварительной обработки в NLP? а. Стемминг и лемматизация б. Преобразование в нижний регистр в. Удаление знаков препинания г. Удаление стоп-слов +д. Анализ настроений
4.	При анализе текста преобразование текста в токены, а затем преобразование их в целочисленные векторы или векторы с плавающей запятой можно выполнить с помощью +а. CountVectorizer б. TF-IDF в. Мешок слов г. NER
5.	В NLP слова, представленные в виде векторов, называются нейронными вложениями слов. +а. Истинный б. Ложь
6.	В NLP поддерживается контекстное моделирование, с помощью которого одно из следующих вложений слов а. Word2Vec б. GloVe +в. BERT г. Все вышеперечисленное
7.	В NLP двунаправленный контекст поддерживается каким из следующих вложений а. Word2Vec +б. BERT в. GloVe г. Все вышеперечисленное
8.	Какое из следующих вложений Word может быть специально обучено для определенного предмета в NLP а. Word2Vec +б. BERT в. GloVe

	г. Все вышеперечисленное
ПК-2 - Способен участвовать в решении профессиональных проектных задач, выбирать и реализовывать командную роль в работе над проектом в соответствии с приоритетами собственной деятельности	
1.	Вложения Word охватывают несколько измерений данных и представляются в виде векторов. +а. Истинный б. Ложь
2.	В NLP векторы встраивания Word помогают установить расстояние между двумя токенами. +а. Истинный б. Ложь
3.	Языковые предубеждения вводятся из-за исторических данных, используемых во время обучения встраиванию слов, которое из приведенных ниже не является примером предвзятости. +а. Нью-Дели — в Индию, Пекин — в Китай б. Мужчина для компьютера, женщина для домохозяйки
4.	Что из следующего будет лучшим выбором для решения таких случаев использования NLP, как семантическое сходство, понимание прочитанного и рассуждения здравого смысла а. GloVe б. BERT +в. Open AI's GPT г. ULMFit
5.	Какая из следующих архитектур может быть обучена быстрее и требует меньшего количества обучающих данных а. Языковое моделирование на основе LSTM +б. Трансформаторная архитектура
6.	Одно и то же слово может иметь несколько вложений слов с помощью _____? а. GloVe б. Word2Vec +в. ELMo г. nltk
7.	Для данного токена его входное представление представляет собой сумму встраивания из токена, сегмента и встраивания позиции. а. ELMo б. GPT +в. BERT г. ULMFit
8.	. Обучает две независимые языковые модели LSTM слева направо и справа налево и неглубоко объединяет их а. GPT б. BERT

	B. ULMFit +r. ELMo
--	-----------------------------------