

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Обработка естественного языка

Код модуля
1154967(1)

Модуль
Приложения искусственного интеллекта

Екатеринбург

Оценочные материалы составлены автором(ами):

№ п/п	Фамилия, имя, отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созькин Андрей Владимирович	кандидат технических наук, без ученого звания	Доцент	информационных технологий и систем управления

Согласовано:

Управление образовательных программ

Т.Г. Комарова

Авторы:

- Созыкин Андрей Владимирович, Доцент, информационных технологий и систем управления

1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ **Обработка естественного языка**

1.	Объем дисциплины в зачетных единицах	3	
2.	Виды аудиторных занятий	Лекции Практические/семинарские занятия	
3.	Промежуточная аттестация	Зачет	
4.	Текущая аттестация	Контрольная работа	1
		Домашняя работа	1

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ **Обработка естественного языка**

Индикатор – это признак / сигнал/ маркер, который показывает, на каком уровне обучающийся должен освоить результаты обучения и их предъявление должно подтвердить факт освоения предметного содержания данной дисциплины, указанного в табл. 1.3 РПМ-РПД.

Таблица 1

Код и наименование компетенции	Планируемые результаты обучения (индикаторы)	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3
ПК-5 -Способен управлять программно-техническими, технологическими и человеческими ресурсами	З-2 - Различать основные принципы и методы управления персоналом П-1 - Использовать технологии гибкого подхода к управлению У-1 - Определять оптимальные методы и принципы управления человеческими ресурсами	Домашняя работа Зачет Контрольная работа Лекции Практические/семинарские занятия

3. ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)**3.1. Процедуры текущей и промежуточной аттестации по дисциплине**

1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>контрольная работа</i>	3,18	100
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5		
Промежуточная аттестация по лекциям – зачет		
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
<i>домашняя работа</i>	3,18	100
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – 1		
Промежуточная аттестация по практическим/семинарским занятиям – нет		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – не предусмотрено		
3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий – не предусмотрено		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям – не предусмотрено		
Промежуточная аттестация по лабораторным занятиям – нет		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – не предусмотрено		
4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий – не предусмотрено		
Текущая аттестация на онлайн-занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям – не предусмотрено		
Промежуточная аттестация по онлайн-занятиям – нет		
Весовой коэффициент значимости результатов промежуточной аттестации по онлайн-занятиям – не предусмотрено		

3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
---	---------------------------------	------------------------------

Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено		
Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено		

4. КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ

4.1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

Таблица 4

Критерии оценивания учебных достижений обучающихся

Результаты обучения	Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам
Знания	Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Умения	Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Опыт /владение	Студент демонстрирует опыт в области изучения на уровне указанных индикаторов.
Другие результаты	Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов. Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения. Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.

4.2 Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

Таблица 5

Шкала оценивания достижения результатов обучения (индикаторов) по уровням

Характеристика уровней достижения результатов обучения (индикаторов)			
№ п/п	Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное)	Шкала оценивания	
		Традиционная характеристика уровня	Качественная характеристика уровня

	задание)			
1.	Результаты обучения (индикаторы) достигнуты в полном объеме, замечаний нет	Отлично (80-100 баллов)	Зачтено	Высокий (В)
2.	Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения	Хорошо (60-79 баллов)		Средний (С)
3.	Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания	Удовлетворительно (40-59 баллов)		Пороговый (П)
4.	Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка	Неудовлетворительно (менее 40 баллов)	Не зачтено	Недостаточный (Н)
5.	Результат обучения не достигнут, задание не выполнено	Недостаточно свидетельств для оценивания		Нет результата

5. СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ

5.1. Описание аудиторных контрольно-оценочных мероприятий по дисциплине модуля

5.1.1. Лекции

Самостоятельное изучение теоретического материала по темам/разделам лекций в соответствии с содержанием дисциплины (п. 1.2. РПД)

5.1.2. Практические/семинарские занятия

Примерный перечень тем

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.

Примерные задания

Задачи и ограничения обработки текстов на естественном языке

Классификация задач обработки текстов на естественном языке

Программное обеспечение для обработки естественного языка

Понимание языка и генерация текста

LMS-платформа

1. <https://lms.skillfactory.ru/courses/course-v1:SkillFactory+URFUML22p1s+SEP2022/course/>

5.2. Описание внеаудиторных контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля

Разноуровневое (дифференцированное) обучение.

Базовый

5.2.1. Контрольная работа

Примерный перечень тем

1. Проектирование пайплайна для задач обработки естественного языка.

Примерные задания

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.

LMS-платформа

1. <https://lms.skillfactory.ru/courses/course-v1:SkillFactory+URFUML22p1s+SEP2022/course/>

5.2.2. Домашняя работа

Примерный перечень тем

1. Обучение языковой модели для текстов на русском языке.

2. Дообучение предварительно обученной сети BERT.

Примерные задания

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.

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>

LMS-платформа

1. <https://lms.skillfactory.ru/courses/course-v1:SkillFactory+URFUML22p1s+SEP2022/course/>

5.3. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля

5.3.1. Зачет

Список примерных вопросов

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.
- LMS-платформа – не предусмотрена

5.4 Содержание контрольно-оценочных мероприятий по направлениям воспитательной деятельности

Направления воспитательной деятельности сопрягаются со всеми результатами обучения компетенций по образовательной программе, их освоение обеспечивается содержанием всех дисциплин модулей.