

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»

УТВЕРЖДАЮ  
Директор по образовательной  
деятельности

*М.И.И.* С.Т. Князев  
«10» *05* *2022*



**РАБОЧАЯ ПРОГРАММА МОДУЛЯ**

Код модуля	Модуль
1161160	Обработка естественного языка в прикладных задачах

Екатеринбург

2022

<b>Перечень сведений о рабочей программе модуля</b>	<b>Учетные данные</b>
<b>Образовательная программа</b> Алгоритмы искусственного интеллекта	<b>Код ОП</b> 09.03.01
<b>Направление подготовки</b> Информатика и вычислительная техника	<b>Код направления и уровня подготовки</b> 09.03.01

Области образования, в рамках которых реализуется модуль образовательной программы по СУОС УрФУ :

<b>№ п/п</b>	<b>Перечень областей образования, для которых разработан СУОС УрФУ</b>	<b>Уровень подготовки</b>
1.	Инженерное дело, технологии и технические науки	бакалавриат

Программа модуля составлена авторами:

<b>№ п/п</b>	<b>Фамилия Имя Отчество</b>	<b>Ученая степень, ученое звание</b>	<b>Должность</b>	<b>Подразделение</b>
1	Борисов Василий Ильич	кандидат технических наук, без ученого звания	Доцент	радиоэлектроники и телекоммуникаций
2	Созыкин Андрей Владимирович	кандидат технических наук, без ученого звания	Доцент	информационных технологий и систем управления

# 1. ОБЩАЯ ХАРАКТЕРИСТИКА МОДУЛЯ Обработка естественного языка в прикладных задачах

## 1.1. Аннотация содержания модуля

Модуль «Обработка естественного языка в прикладных задачах» состоит из одноименной дисциплины. Знакомит студентов с современными методами обработки естественного языка, основанными на глубоких нейронных сетях и машинном обучении. Содержание данного модуля позволяет студентам изучить, как использовать глубокие нейронные сети для классификации текстов, анализа настроений и автоматической генерации текста. В рамках дисциплины будут рассмотрены особые типы архитектуры нейронных сетей, подходящие для обработки текста: рекуррентные нейронные сети, включая LSTM и GRU, и одномерные сверточные сети.

## 1.2. Структура и объем модуля

Таблица 1

№ п/п	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах
1	Обработка естественного языка в прикладных задачах	3
ИТОГО по модулю:		3

## 1.3. Последовательность освоения модуля в образовательной программе

Пререквизиты модуля	Не предусмотрены
Постреквизиты и кореквизиты модуля	Не предусмотрены

## 1.4. Распределение компетенций по дисциплинам модуля, планируемые результаты обучения (индикаторы) по модулю

Таблица 2

Перечень дисциплин модуля	Код и наименование компетенции	Индикаторы достижения компетенции	Планируемые результаты обучения
1	2	3	4
Обработка естественного языка в прикладных задачах	ПК-1. Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать	ПК-1.2. Выбирает методы и инструментальные средства искусственного интеллекта для решения задач в зависимости от особенностей	ПК-1.2. З-1. Знает методы и инструментальные средства решения задач с использованием систем искусственного интеллекта в зависимости от особенностей проблемной области, критерии выбора методов и инструментальных средств решения интеллектуальных задач, подходы к выбору методов и

	<p>адекватные методы и инструментальные средства решения задач искусственного интеллекта</p>	<p>проблемной и предметной областей</p>	<p>инструментальных средств систем искусственного интеллекта, процесс, стадии и методологии разработки решений на основе искусственного интеллекта ПК-1.2. У-1. Умеет осуществлять оценку критериев выбора методов и инструментальных средств решения задач с помощью систем искусственного интеллекта и выбор методов и инструментальных средств в зависимости от особенностей проблемной и предметной областей</p>
	<p>ПК-7. Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта</p>	<p>ПК-7.2. Участвует в реализации проектов в области сквозной цифровой субтехнологии «Обработка естественного языка»</p>	<p>ПК-7.2. З-1. Знает принципы построения систем обработки естественного языка, методы и технологии искусственного интеллекта для анализа естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка» ПК-7.2. У-1. Умеет применять методы и подходы к планированию и реализации проектов по созданию и поддержке системы искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»</p>

### 1.5. Форма обучения

Обучение по дисциплинам модуля может осуществляться в очной формах.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**  
**Обработка естественного языка в**  
**прикладных задачах**

Рабочая программа дисциплины составлена авторами:

<b>№ п/п</b>	<b>Фамилия Имя Отчество</b>	<b>Ученая степень, ученое звание</b>	<b>Должность</b>	<b>Подразделение</b>
1	Борисов Василий Ильич	кандидат технических наук, без ученого звания	Доцент	Кафедра радиоэлектроники и телекоммуникаци й
2	Созыкин Андрей Владимирович	кандидат технических наук, без ученого звания	Доцент	Кафедра информационных технологий и систем управления

# 1. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ

## 1.1. Технологии реализации, используемые при изучении дисциплины модуля

- Традиционная (репродуктивная) технология
- Смешанная модель обучения с использованием онлайн-курса УрФУ;
- Исключительно электронного обучения с использованием внутреннего онлайн-курса УрФУ;
- С применением электронного обучения на основе электронных учебных курсов, размещенных на LMS-платформах УрФУ
- Разноуровневое (дифференцированное) обучение
  - Базовый уровень

*\*Базовый I уровень – сохраняет логику самой науки и позволяет получить упрощенное, но верное и полное представление о предмете дисциплины, требует знание системы понятий, умение решать проблемные ситуации. Освоение данного уровня результатов обучения должно обеспечить формирование запланированных компетенций и позволит обучающемуся на минимальном уровне самостоятельности и ответственности выполнять задания;*

*Продвинутый II уровень – углубляет и обогащает базовый уровень как по содержанию, так и по глубине проработки материала дисциплины. Это происходит за счет включения дополнительной информации. Данный уровень требует умения решать проблемы в рамках курса и смежных курсов посредством самостоятельной постановки цели и выбора программы действий. Освоение данного уровня результатов обучения позволит обучающемуся повысить уровень самостоятельности и ответственности до творческого применения знаний и умений.*

## 1.2. Содержание дисциплины

Таблица 1.1

Код раздела, темы	Раздел, тема дисциплины*	Содержание
1	Теоретические аспекты обработки естественного языка.	Синтаксический, морфологический, семантический и графематический анализ, омонимия, задачи лингвистического анализа
2	Предварительная обработка текста.	Очистка текста, токенизация, стемминг, лемматизация, удаление стоп-слов, фильтрация наиболее частотных и наименее частотных слов.
3	Векторизация текста.	Построение словаря, мешок слов, TF-IDF, word2vec, fasttext, LDA, LSI, GloVe.
4	Машинное обучение для обработки текстов.	Решение задач классификации и определения тональности методами классического машинного обучения на основе векторных моделей.
5	Нейронные сети в решении задач текстовой обработки.	Архитектуры нейронных сетей для обработки текстов: рекуррентные (LSTM, GRU), одномерные сверточные. Применение нейронных сетей для обработки текстов.
6	Языковая модель.	Языковая модель и дистрибутивная семантика. Обучение векторной модели. Задача генерации текста. Различные подходы к генерации текста.

7	Поиск именованных сущностей.	Задача поиска именованных сущностей в тексте. Применение нейронных сетей для поиска именованных сущностей.
8	Механизм внимания. Трансформер.	Механизм внимания в нейронных сетях. Применение механизма внимания для обработки текста. Нейронные сети с архитектурой Transformer. Нейронные сети BERT, GPT. Перенос обучения.

### 1.3. Направление, виды воспитательной деятельности и используемые технологии

Таблица 1.2

Направление воспитательной деятельности	Вид воспитательной деятельности	Технология воспитательной деятельности	Компетенция	Результаты обучения
Профессиональное воспитание	профориентационная деятельность	Технология формирования уверенности и готовности к самостоятельной профессиональной деятельности  Технология самостоятельной работы	ПК-7. Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	ПК-7.2. 3-1. Знает принципы построения систем обработки естественного языка, методы и технологии искусственного интеллекта для анализа естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка» ПК-7.2. У-1. Умеет применять методы и подходы к планированию и реализации проектов по созданию и поддержке системы искусственного интеллекта на основе сквозной

				цифровой субтехнологии «Обработка естественного языка»
--	--	--	--	--

1.4. Программа дисциплины реализуется полностью на иностранном языке.

**1. 2. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ  
ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ  
Обработка естественного языка в  
прикладных задачах**

**Электронные ресурсы (издания)**

1. Карньюшина, В. В.; Информационные технологии в профессиональной деятельности учителя иностранных языков : учебное пособие.; Сургутский государственный педагогический университет, Сургут; 2019; <http://www.iprbookshop.ru/94283.html> (Электронное издание)
2. Борисова, Е. Г.; Язык общественно-политической коммуникации : учебное пособие для магистратуры по специальности «реклама и связи с общественностью».; Московский городской педагогический университет, Москва; 2012; <http://www.iprbookshop.ru/26673.html> (Электронное издание)
3. , Башмакова, Е. И.; Информационные технологии в лингвистике : учебное пособие для бакалавров.; Ай Пи Ар Медиа, Москва; 2021; <http://www.iprbookshop.ru/103344.html> (Электронное издание)
4. Леонидова, Г. Ф.; Прикладные программные средства библиотечно-информационной деятельности : практикум для обучающихся по направлению подготовки 51.03.06 «библиотечно-информационная деятельность», квалификация (степень) выпускника «бакалавр».; Кемеровский государственный институт культуры, Кемерово; 2019; <http://www.iprbookshop.ru/95569.html> (Электронное издание)

**Печатные издания**

1. Хорстманн, Кей С., К. С., Корнелл, Корнелл Г., Вейтман, В. В.; Java 2. Библиотека профессионала Т. 1. Основы; Вильямс, Москва ; СПб. ; Киев; 2006 (1 экз.)
2. Заболеева-Зотова, А. В., Камаев, В. А.; Лингвистическое обеспечение автоматизированных систем : учеб. пособие для студентов вузов, обучающихся по направлению "Информатика и вычисл. техника".; Высшая школа, Москва; 2008 (2 экз.)
3. , Mitkov, R.; The Oxford Handbook of Computational Linguistics; Oxford University Press, New York; 2005 (1 экз.)

**Профессиональные базы данных, информационно-справочные системы**

1. Смешанная модель обучения с использованием онлайн-курса УрФУ "Generating discrete sequences: language and music" <https://www.edx.org/course/generating-discrete-sequences-language-and-music>
2. Исключительно электронного обучения с использованием внутреннего онлайн-курса УрФУ [https://courses.openedu.urfu.ru/courses/course-v1:UrFU+NLP+fall\\_2020/course/](https://courses.openedu.urfu.ru/courses/course-v1:UrFU+NLP+fall_2020/course/)

3. С применением электронного обучения на основе электронных учебных курсов, размещенных на LMS-платформах УрФУ [https://courses.openedu.urfu.ru/courses/course-v1:UrFU+NLP+fall\\_2020/course/](https://courses.openedu.urfu.ru/courses/course-v1:UrFU+NLP+fall_2020/course/)
4. Цифровая библиотека научно-технических изданий Института инженеров по электротехнике и радиоэлектронике (Institute of Electrical and Electronic Engineers (IEEE)) на английском языке – <http://www.ieee.org/ieeexplore>
5. Oxford University Press – <http://www.oxfordjournals.org/en/>
6. Архив препринтов с открытым доступом – <https://arxiv.org/>

### **Материалы для лиц с ОВЗ**

Весь контент ЭБС представлен в виде файлов специального формата для воспроизведения синтезатором речи, а также в тестовом виде, пригодном для прочтения с использованием экранной лупы и настройкой контрастности.

### **Базы данных, информационно-справочные и поисковые системы**

1. Academic Search Ultimate EBSCO publishing – <http://search.ebscohost.com>
2. eBook Collections Springer Nature – <https://link.springer.com/>
3. Гугл Академия – <https://scholar.google.ru/>
4. Электронный научный архив УрФУ <https://elar.urfu.ru/>
5. Зональная научная библиотека (УрФУ) - <http://lib.urfu.ru/>
6. Портал информационно-образовательных ресурсов УрФУ <https://study.urfu.ru/>
7. Электронно-библиотечная система «Лань» – <https://e.lanbook.com/>
8. Университетская библиотека ONLINE – <https://biblioclub.ru/>
9. Электронно-библиотечная система "Библиокомплектатор" (IPRbooks) <http://www.bibliocomplectator.ru/available>
10. Электронные информационные ресурсы Российской государственной библиотеки <https://www.rsl.ru/>
11. Научная электронная библиотека «КиберЛенинка» <https://cyberleninka.ru/>

## **2. 3. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

### **Обработка естественного языка в прикладных задачах**

**Сведения об оснащённости дисциплины специализированным и лабораторным оборудованием и программным обеспечением**

Таблица 3.1

№ п/п	Виды занятий	Оснащённость специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения

1	Лекции	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Доска аудиторная</p> <p>Периферийное устройство</p> <p>Персональные компьютеры по количеству обучающихся</p> <p>Подключение к сети Интернет</p>	<p>Office 365 EDUA1 ShrdSvr ALNG SubsVL MVL PerUsr Faculty EES</p> <p>Используется бесплатно-распространяемое программное обеспечение:</p> <p>Python – <a href="https://www.python.org/">https://www.python.org/</a></p> <p>TensorFlow – <a href="https://www.tensorflow.org/">https://www.tensorflow.org/</a></p> <p>Hugging Face – <a href="https://huggingface.co/">https://huggingface.co/</a></p> <p>Веб - среда разработки для языка программирования Python: Google Colab – <a href="https://colab.research.google.com/">https://colab.research.google.com/</a></p>
2	Лабораторные занятия	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Доска аудиторная</p> <p>Персональные компьютеры по количеству обучающихся</p> <p>Подключение к сети Интернет</p>	<p>Office 365 EDUA5 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES</p> <p>Используется бесплатно-распространяемое программное обеспечение:</p> <p>Python – <a href="https://www.python.org/">https://www.python.org/</a></p> <p>TensorFlow – <a href="https://www.tensorflow.org/">https://www.tensorflow.org/</a></p> <p>Hugging Face – <a href="https://huggingface.co/">https://huggingface.co/</a></p> <p>Веб - среда разработки для языка программирования Python: Google Colab – <a href="https://colab.research.google.com/">https://colab.research.google.com/</a></p>

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ  
ПО ДИСЦИПЛИНЕ**

Обработка естественного языка в прикладных задачах

Оценочные материалы составлены автором(ами):

<b>№ п/п</b>	<b>Фамилия Имя Отчество</b>	<b>Ученая степень, ученое звание</b>	<b>Должность</b>	<b>Подразделение</b>
1	Борисов Василий Ильич	кандидат технических наук, без ученого звания	Доцент	радиоэлектроники и телекоммуникаций
2	Созыкин Андрей Владимирович	кандидат технических наук, без ученого звания	Доцент	информационных технологий и систем управления

## 1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ **Обработка естественного языка в прикладных задачах**

1.	• <b>Объем дисциплины в зачетных единицах</b>	• 3	
2.	• <b>Виды аудиторных занятий</b>	Лекции Лабораторные занятия	
3.	• <b>Промежуточная аттестация</b>	Зачёт	
4.	• <b>Текущая аттестация</b>	Домашняя работа	1

## 2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ **МОДУЛЯ Обработка естественного языка в прикладных задачах**

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	Индикаторы достижения компетенции	Планируемые результаты обучения	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3	4
ПК-1. Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта	ПК-1.2. Выбирает методы и инструментальные средства искусственного интеллекта для решения задач в зависимости от особенностей проблемной и предметной областей	ПК-1.2. 3-1. Знает методы и инструментальные средства решения задач с использованием систем искусственного интеллекта в зависимости от особенностей проблемной области, критерии выбора методов и инструментальных средств решения интеллектуальных задач, подходы к выбору методов и инструментальных средств систем искусственного интеллекта, процесс, стадии и методологии разработки решений на основе искусственного интеллекта ПК-1.2. У-1. Умеет осуществлять оценку критериев выбора методов	Лекции Лабораторные занятия Домашняя работа Зачёт

		и инструментальных средств решения задач с помощью систем искусственного интеллекта и выбор методов и инструментальных средств в зависимости от особенностей проблемной и предметной областей	
ПК-7. Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	ПК-7.2. Участвует в реализации проектов области сквозной цифровой субтехнологии «Обработка естественного языка»	ПК-7.2. З-1. Знает принципы построения систем обработки естественного языка, методы и технологии искусственного интеллекта для анализа естественного языка, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка» ПК-7.2. У-1. Умеет применять методы и подходы к планированию и реализации проектов по созданию и поддержке системы искусственного интеллекта на основе сквозной цифровой субтехнологии «Обработка естественного языка»	

### **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>Домашняя работа</i>	6, 8	100
<b>Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5</b>		
Промежуточная аттестация по лекциям – зачёт		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5		
<b>2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – не предусмотрено</b>		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<b>Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям– 0</b>		
Промежуточная аттестация по практическим/семинарским занятиям–нет		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям– не предусмотрено		
<b>3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий – 0.5</b>		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Защита лабораторных работ</i>	6, 1-16	100
<b>Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям - 1</b>		
Промежуточная аттестация по лабораторным занятиям – не предусмотрено		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям– 0		

### 3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<b>Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено</b>		
<b>Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено</b>		

## 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).

## 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).

## 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 задание)	Шкала оценивания		
		Традиционная характеристика уровня		Качественная характеристика уровня
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,	Неудовлетворительно	Не зачтено	Недостаточный (Н)

	имеются существенные ошибки и замечания, требуется доработка	(менее 40 баллов)		
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. Лабораторные занятия

Примерный перечень тем:

2. Предварительная обработка текста для анализа.
3. Векторизация текста.
4. Классификация текста с использованием классических методов машинного обучения.
5. Классификация текста с использованием глубоких нейронных сетей.
6. Языковая модель. Преподавание языковой модели.
7. Автоматическая генерация текста.
8. Поиск именованных сущностей в тексте.
9. Механизм внимания в нейронных сетях. Сети с архитектурой Transformer.
10. Передача обучения в задачах обработки текстов.

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

Разноуровневое (дифференцированное) обучение.

## Базовый

#### 5.2.1. Домашняя работа

##### Примерные задания домашних работ:

1. Обучить языковую модель для русского языка и использовать ее для генерации текста. Для выполнения задания необходимо сделать следующее:

- Подготовьте набор данных с текстами на русском языке. Вы можете использовать готовые наборы данных или создать свой собственный.
- Обучить языковую модель на подготовленном наборе данных.
- Используя обученную языковую модель, сгенерируйте пять примеров текстов на русском языке.
- Разместите набор данных, код и обученную модель в открытом доступе на GitHub.
- Сделать презентацию или технологическую статью о ходе работ, обоснование принятых решений и результатов работы.

\* (Необязательное задание). Запишите видео, показывающее, как работает созданное решение.

2. Обучить предварительно обученную сеть с архитектурой Transformer для классификации текстов в Русский. Для выполнения задания необходимо сделать следующее:

- Подготовить набор данных с текстами на русском языке для классификации. Вы можете использовать готовые наборы данных или создать свой собственный.

- Выберите предварительно обученную нейронную сеть с архитектурой Transformer, подходящую для задачи классификации текстов на русском языке.
  - Провести дообучение выбранной нейронной сети на подготовленных данных. установлен.
  - Выполните тестирование классификации текста с помощью обученной нейронной сети и оцените качество сети.
  - Разместите набор данных, код и готовую модель в открытом доступе на GitHub.
  - Сделать презентацию или технологическую статью о ходе работы, обоснование принятых решений и результатов работы.
- \* (Необязательное задание). Запишите видео, показывающее, как работает созданное решение.

Пример обучения нейросети BERT в TensorFlow — [https://www.tensorflow.org/text/tutorials/fine\\_tune\\_bert](https://www.tensorflow.org/text/tutorials/fine_tune_bert)

Ноутбук с образцом кода решения — [https://colab.research.google.com/github/tensorflow/text/blob/master/docs/tutorials/fine\\_tune\\_bert.ipynb](https://colab.research.google.com/github/tensorflow/text/blob/master/docs/tutorials/fine_tune_bert.ipynb)

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

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

#### 5.3.1. Зачёт

##### Перечень примерных вопросов для зачёта

1. Теоретические аспекты обработки естественного языка.
2. Особенности обработки текстов на английском языке.
3. Особенности обработки текстов на русском языке.
4. Предварительная обработка текста. Очистка текста. Удаление стоп-слов/наиболее и наименее частых слов.
5. Токенизация, стемминг, лемматизация текста.
6. Методы векторизации текста: построение словаря, набора 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 текста: LSTM.
14. Архитектуры нейронных сетей для обработки текста: ГРУ.
15. Нейросетевые архитектуры 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 текстов.
21. Задание на поиск именованных объектов в тексте.
22. Использование нейронных сетей для поиска именованных сущностей.
23. Механизм внимания в нейронных сетях.
24. Применение механизма внимания для обработки текста.
25. Архитектура нейронной сети преобразователя.
26. Предварительно обученные нейронные сети для обработки текста BERT.
27. Предварительно обученные нейронные сети для обработки текста GPT.
28. Перенос обучения задачам обработки текста.

29. Классификация текста с использованием сетей с архитектурой Transformer. 30. Генерация текста с использованием сетей с архитектурой Transformer.
31. Поиск именованных сущностей в тексте с использованием сетей с архитектурой Transformer.