

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Галунин Сергей Александрович  
Должность: Директор департамента образования  
Дата подписания: 13.09.2021 19:57:32  
Уникальный программный ключ:  
1cb4f9edcd6d31e931c556ddefa3b376a443365a5419cb3e3965cc668ec8658b



**СПбГЭТУ «ЛЭТИ»**  
ПЕРВЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ

МИНОБРНАУКИ РОССИИ

федеральное государственное автономное образовательное учреждение высшего образования  
**«Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И.Ульянова (Ленина)»**  
(СПбГЭТУ «ЛЭТИ»)

**ТВЕРЖДАЮ**  
Директор департамента образования  
**С.А. Галунин**  
2021 г.

**РАБОЧАЯ ПРОГРАММА**

дисциплины

**«ОБРАБОТКА ЕСТЕСТВЕННЫХ ЯЗЫКОВ»**

для подготовки магистров

по направлению

01.04.02 «Прикладная математика и информатика»

по программе

**«Математическое и программное обеспечение вычислительных машин»**

Санкт-Петербург

2021

## ЛИСТ СОГЛАСОВАНИЯ

Разработчик

доцент, к.т.н.



И.А. Посов

Рабочая программа рассмотрена и одобрена на заседании кафедры АМ  
10.03.2021, протокол № 6

Заведующий кафедрой АМ  
заведующий кафедрой, д.пед.н., доцент



С.Н. Поздняков

Рабочая программа рассмотрена и одобрена учебно-методической комиссией  
ФКТИ, 25.03.2021, протокол № 2

Председатель УМК ФКТИ  
доцент, к.т.н.



В.С. Андреев

**Согласовано:**

Начальник ОМОЛА



О.В. Загороднюк

## 1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	АМ
Общая трудоемкость (ЗЕТ)	4
Курс	1
Семестр	1
<b>Виды занятий</b>	
Лекции (академ. часов)	17
Практические занятия (академ. часов)	17
Иная контактная работа (академ. часов)	1
Все контактные часы (академ. часов)	35
Самостоятельная работа, включая часы на контроль (академ. часов)	109
Всего (академ. часов)	144
<b>Вид промежуточной аттестации</b>	
Экзамен (курс)	1

## **2 АННОТАЦИЯ ДИСЦИПЛИНЫ**

### **«ОБРАБОТКА ЕСТЕСТВЕННЫХ ЯЗЫКОВ»**

В курсе рассматриваются задачи, которые требуют обработки текстов на естественных языках, в первую очередь русском и английском. Список задач включает в себя классификацию текстов, определение тональности, автоматическое реферирование, машинный перевод, многие другие задачи более низкого уровня. Из подходов к решению задач рассматриваются лингвистические подходы, статистические и подходы, использующие глубокое обучение. Курс предполагает решение практических заданий с помощью библиотек и ресурсов для обработки естественных языков для языка программирования python.

## **SUBJECT SUMMARY**

### **«NATURAL LANGUAGE PROCESSING»**

The course is devoted to the approaches to solve problems, that require processing of raw texts in natural languages, primarily Russian and English. The list of problems includes texts classification, sentiment analysis, automatic summarization, machine translation, a number of other low level tasks. The approaches being discussed include purely linguistic approaches, statistical approaches and deep learning approaches. The course supposes practice in coding in python using natural language processing libraries and resources.

## **3 ОБЩИЕ ПОЛОЖЕНИЯ**

### **3.1 Цели и задачи дисциплины**

1. Изучение задачи, которые возникают при обработке текстов на естественных языках и подходы к их решению; Умение связать решаемую задачу с задачами из области обработки естественных языков.
2. Формирование знаний о спектре подходов к решению разных задач обработки естественных языков. Умение выбрать подход к решению задачи.
3. Освоение математического аппарата, используемого в основе методов обработки естественных языков, и программных инструментов для обработки естественных языков. Формирование навыков использования математического аппарата и программных инструментов.

### **3.2 Место дисциплины в структуре ОПОП**

Дисциплина изучается на основе знаний, полученных при освоении программы бакалавриата или специалитета.

и обеспечивает изучение последующих дисциплин:

1. «Разработка языковых процессоров»
2. «Анализ и интерпретация данных»

### 3.3 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения образовательной программы обучающийся должен достичь следующие результаты обучения по дисциплине:

<b>Код компетенции/ индикатора компетенции</b>	<b>Наименование компетенции/индикатора компетенции</b>
ОПК-2	Способен совершенствовать и реализовывать новые математические методы решения прикладных задач
<i>ОПК-2.1</i>	<i>Знает современные математические методы решения прикладных задач</i>
<i>ОПК-2.2</i>	<i>Умеет обосновывать выбор либо необходимость реализации новых математических методов решения прикладных задач</i>
<i>ОПК-2.3</i>	<i>Имеет навыки разработки новых математических методов решения прикладных задач</i>

## 4 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

### 4.1 Содержание разделов дисциплины

#### 4.1.1 Наименование тем и часы на все виды нагрузки

№ п/п	Наименование темы дисциплины	Лек, ач	Пр, ач	ИКР, ач	СР, ач
1	Введение	2	2		12
2	Задача классификации	3	3		18
3	Работа с последовательностями	3	3		21
4	Теория формальных языков в приложении к естественным	2	2		12
5	Семантика	3	3		18
6	Трансформеры последовательностей	3	3		22
7	Заключение	1	1	1	6
	Итого, ач	17	17	1	109
	Из них ач на контроль	0	0	0	35
	Общая трудоемкость освоения, ач/зе	144/4			

#### 4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	Введение	Задачи обработки естественного языка, этапы обработки текста.
2	Задача классификации	Наивный байесовый классификатор и другие традиционные методы классификации. Классификация с помощью нейронных сетей прямого распространения. Лингвистические приложения методов классификации.
3	Работа с последовательностями	Текст как последовательность символов или слов. N-грамм модели, модели основанные на рекуррентных нейронных сетях. Сглаживание, оценка модели. Классификация элементов последовательности.
4	Теория формальных языков в приложении к естественным	Контекстно-свободные грамматики в описании естественных языков, дерево разбора текста и граф зависимостей.
5	Семантика	Разные подходы к определению понятия смысла языковых единиц: исчисление предикатов, дистрибутивная семантика. Кластеризация. Нейросетевые подходы к векторному представлению слов (word embedding). Разрешение кореферентности.

№ п/п	Наименование темы дисциплины	Содержание
6	Трансформеры последовательностей	Понятие трансформера, энкодера, декодера, механизм внимания, современные нейросетевые модели. Машинный перевод, генерация текста.
7	Заключение	Обзор других современных направлений в обработке естественных языков.

#### 4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

#### 4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Определение авторства текста	1
2. N-грамм модели	2
3. Определение частей речи в тексте	2
4. Грамматический разбор текста	2
5. Поиск коллокаций	2
6. Word2Vec, кластеризация слов	2
7. BERT для анализа настроения	2
8. Выделение именованных сущностей	2
9. GPT2 для генерации текста	2
Итого	17

#### 4.4 Курсовое проектирование

Курсовая работа (проект) не предусмотрены.

#### 4.5 Реферат

Реферат не предусмотрен.

#### 4.6 Индивидуальное домашнее задание

Индивидуальное домашнее задание не предусмотрено.



#### 4.7 Доклад

Доклад не предусмотрен.

#### 4.8 Кейс

Кейс не предусмотрен.

#### 4.9 Организация и учебно-методическое обеспечение самостоятельной работы

Изучение дисциплины сопровождается самостоятельной работой студентов с рекомендованными преподавателем литературными источниками и информационными ресурсами сети Интернет.

Планирование времени для изучения дисциплины осуществляется на весь период обучения, предусматривая при этом регулярное повторение пройденного материала. Обучающимся, в рамках внеаудиторной самостоятельной работы, необходимо регулярно дополнять сведениями из литературных источников материал, законспектированный на лекциях. При этом на основе изучения рекомендованной литературы целесообразно составить конспект основных положений, терминов и определений, необходимых для освоения разделов учебной дисциплины.

Особое место уделяется консультированию, как одной из форм обучения и контроля самостоятельной работы. Консультирование предполагает особым образом организованное взаимодействие между преподавателем и студентами, при этом предполагается, что консультант либо знает готовое решение, которое он может предписать консультируемому, либо он владеет способами деятельности, которые указывают путь решения проблемы.

Текущая СРС	Примерная трудоемкость, ач
Работа с лекционным материалом, с учебной литературой	34

<b>Текущая СРС</b>	<b>Примерная трудоемкость, ач</b>
Опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	0
Самостоятельное изучение разделов дисциплины	0
Выполнение домашних заданий, домашних контрольных работ	75
Подготовка к лабораторным работам, к практическим и семинарским занятиям	0
Подготовка к контрольным работам, коллоквиумам	0
Выполнение расчетно-графических работ	0
Выполнение курсового проекта или курсовой работы	0
Поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	0
Работа над междисциплинарным проектом	0
Анализ данных по заданной теме, выполнение расчетов, составление схем и моделей, на основе собранных данных	0
Подготовка к зачету, дифференцированному зачету, экзамену	0
<b>ИТОГО СРС</b>	<b>109</b>

## 5 Учебно-методическое обеспечение дисциплины

### 5.1 Перечень основной и дополнительной литературы, необходимой для освоения дисциплины

№ п/п	Название, библиографическое описание	К-во экз. в библ.
Основная литература		
1	Гольдберг Й. Нейросетевые методы в обработке естественного языка [Электронный ресурс] : руководство / Й. Гольдберг, 2019. -282 с.	неогр.
2	Лейн Хобсон Обработка естественного языка в действии [Электронный ресурс] / Хобсон Лейн, Ханнес Хапке, Коул Ховард, 2021. -576 с.	неогр.
Дополнительная литература		
1	Ганегедара Т. Обработка естественного языка с TensorFlow [Электронный ресурс] : руководство / Т. Ганегедара, 2020. -382 с.	неогр.
2	Риз Р. Обработка естественного языка на Java [Электронный ресурс], 2016. -264 с.	неогр.

### 5.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», используемых при освоении дисциплины

№ п/п	Электронный адрес
1	<a href="https://nlpub.ru/">https://nlpub.ru/</a>
2	<a href="http://opencorpora.org/">http://opencorpora.org/</a>
3	<a href="https://www.nltk.org/">https://www.nltk.org/</a>
4	<a href="https://raw.githubusercontent.com/jacobeisenstein/gt-nlp-class/master/notes/eisenstein-nlp-notes.pdf">https://raw.githubusercontent.com/jacobeisenstein/gt-nlp-class/master/notes/eisenstein-nlp-notes.pdf</a>
5	<a href="http://mmds.org/">http://mmds.org/</a>

### 5.3 Адрес сайта курса

Адрес сайта курса: <https://vec.etu.ru/moodle/course/view.php?id=7330>

## 6 Критерии оценивания и оценочные материалы

### 6.1 Критерии оценивания

Для дисциплины «Обработка естественных языков» формой промежуточной аттестации является экзамен.

#### Экзамен

<b>Оценка</b>	<b>Описание</b>
Неудовлетворительно	Курс не освоен. Студент испытывает серьезные трудности при ответе на ключевые вопросы дисциплины
Удовлетворительно	Студент в целом овладел курсом, но некоторые разделы освоены на уровне определений и формулировок теорем
Хорошо	Студент овладел курсом, но в отдельных вопросах испытывает затруднения. Умеет решать задачи
Отлично	Студент демонстрирует полное овладение курсом, способен применять полученные знания при решении конкретных задач.

## **Особенности допуска**

Для допуска к экзамену студенту необходимо решить задачи, выданные в течение семестра. Из каждой темы курса должна быть решена хотя бы одна задача. Также необходимы 80% посещаемости лекционных и практических занятий. На экзамене студенту необходимо устно ответить на два случайных вопроса по материалам лекций.

## **6.2 Оценочные материалы для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине**

Весь комплект контрольно-измерительных материалов для проверки сформированности компетенции (индикатора компетенции) размещен в закрытой части по адресу, указанному в п. 5.3

### 6.3 График текущего контроля успеваемости

Неделя	Темы занятий	Вид контроля
3	Задача классификации	
4		
5		Практическая работа
6	Работа с последовательностями	
7		
8		Практическая работа
9	Теория формальных языков в приложении к естественным	
10		Практическая работа
11	Семантика	
12		
13		Практическая работа
14	Трансформеры последовательностей	
15		
16		Практическая работа
17	Заключение	Практическая работа
1	Введение	
2		Практическая работа

### 6.4 Методика текущего контроля

Методические материалы; описание шкал оценивания; типовые контрольные задания или иные материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении к рабочей программе дисциплины.

Конкретные формы и процедуры текущего контроля знаний и промежуточной аттестации, а также методические указания для обучающихся по самостоятельной работе при освоении дисциплины доводятся до сведения обучающихся в течение первых недель обучения.

## 7 Описание информационных технологий и материально-технической базы

Тип занятий	Тип помещения	Требования к помещению	Требования к программному обеспечению
Лекция	Лекционная аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя, меловая или маркерная доска	
Практические занятия	Аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя, меловая или маркерная доска	
Самостоятельная работа	Помещение для самостоятельной работы	Оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.	1) Windows 7 и выше или дистрибутив Linux, основанный на Ubuntu или Fedora; 2) Microsoft Office 2007 и выше или Libre Office 6.0 и выше.

## **8 Адаптация рабочей программы для лиц с ОВЗ**

Адаптированная программа разрабатывается при наличии заявления со стороны обучающегося (родителей, законных представителей) и медицинских показаний (рекомендациями психолого-медико-педагогической комиссии). Для инвалидов адаптированная образовательная программа разрабатывается в соответствии с индивидуальной программой реабилитации.



## ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

<b>№ п/п</b>	<b>Дата</b>	<b>Изменение</b>	<b>Дата и номер протокола заседания УМК</b>	<b>Автор</b>	<b>Начальник ОМОЛА</b>