

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Волков В.В.  
Должность: Ректор  
Дата подписания: 27.11.2023 18:33:01  
Уникальный программный ключ:  
ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a70e51517ff6d591

Автономная некоммерческая образовательная организация высшего образования  
«Европейский университет в Санкт-Петербурге»

УТВЕРЖДАЮ  
Ректор  /Волков В.В./  
« 31 » августа 2021 г.  
Протокол Ученого Совета  
№ 7 от 31 августа 2021 г.



Рабочая программа дисциплины  
**«Введение в обработку естественного языка (NLP)»**

дополнительная профессиональная программа  
«Прикладной анализ данных»

вид программы  
программа профессиональной переподготовки

язык обучения – русский  
форма обучения – очная

**Санкт-Петербург**

**Авторы:**

Браславский П.И, кандидат технических наук, доцент факультета социологии и философии АНООВО «ЕУСПб»

Рабочая программа дисциплины «Введение в обработку естественного языка (NLP)», входящая в состав дополнительной профессиональной программы профессиональной переподготовки «Прикладной анализ данных» утверждена на заседании Ученого совета университета.

**Содержание**

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ. ОБЪЕМ ДИСЦИПЛИНЫ.....	4
2. ТРЕБОВАНИЯ К УРОВНЮ ОСВОЕНИЯ СОДЕРЖАНИЯ ДИСЦИПЛИНЫ.....	4
3. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ.....	5
4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ .....	9
5. ОЦЕНОЧНЫЕ СРЕДСТВА.....	10
6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	15
7. ПРОГРАММНОЕ И МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПРОГРАММЫ.....	16

## 1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ. ОБЪЕМ ДИСЦИПЛИНЫ

**Цель обучения:** освоения дисциплины «Введение в обработку естественного языка (NLP)» — знакомство с основными методами и приложениями автоматической обработки естественного языка (ОЕЯ), получение практических навыков работы с инструментами ОЕЯ.

**Задачи обучения:**

- Знакомство с основными методами и подходами к ОЕЯ.
- Понимание принципов оценки качества методов ОЕЯ.
- Владение практическими навыками обработки больших коллекций текстов.

Изучение данной дисциплины способствует формированию профессиональных навыков по разработке теоретических моделей исследуемых процессов, явлений и объектов, относящихся к сфере профессиональной деятельности, их исследованию, оценке и интерпретации полученных результатов.

Отличительной особенностью реализуемого подхода к преподаванию дисциплины является разнообразных практических иллюстраций основных теоретических положений.

Общая трудоемкость дисциплины составляет 54 часа.

## 2. ТРЕБОВАНИЯ К УРОВНЮ ОСВОЕНИЯ СОДЕРЖАНИЯ ДИСЦИПЛИНЫ

В результате освоения дисциплины слушатель должен приобрести следующие знания и умения, необходимые для качественного изменения и (или) получения новых профессиональных компетенций:

**слушатель должен знать:**

- основные методы ОЕЯ;
- основные приложения ОЕЯ;
- основные инструменты ОЕЯ

**слушатель должен уметь:**

- правильно ставить задачи ОЕЯ;
- уметь применять алгоритмы ОЕЯ;
- правильно оформлять и представлять результаты работы.

**слушатель должен владеть:**

- практических навыков работы с инструментами ОЕЯ;
- способностью к самостоятельному обучению новым методам исследования, к изменению научного и научно-производственного профиля своей профессиональной деятельности.

В результате изучения дисциплины «Введение в обработку естественного языка (NLP)» слушатель приобретает следующие профессиональные компетенции (Таблица 1):

### Планируемые результаты обучения по дисциплине

Таблица 1

Код и название компетенции	Содержание компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапам формирования компетенций
ПК-2	способен организовать аналитическую работу в IT-проекте	<b>Знать:</b> <b>З (ПК-2)</b> – теоретические основы организации аналитической работы в IT-проекте
		<b>Уметь:</b> <b>У (ПК-2)</b> – анализировать данные необходимые для аналитической работы и распределять роли в IT-проекте
		<b>Владеть:</b> <b>В (ПК-2)</b> – навыками организации аналитической работы в IT-проекте

Код и название компетенции	Содержание компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапам формирования компетенций
ПК-4	способен составлять отчет об аналитических работах в IT-проекте	<b>Знать:</b> <b>З (ПК-4)</b> – основы составления отчетов об аналитических работах в IT-проекте
		<b>Уметь:</b> <b>У (ПК-4)</b> – собирать и систематизировать данные необходимые для составления отчетов об аналитических работах в IT-проектах
		<b>Владеть:</b> <b>В (ПК-4)</b> – навыками составления отчетов об аналитических работах в IT-проектах

### 3. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Содержание дисциплины соотносится с планируемыми результатами обучения по дисциплине через задачи, формируемые компетенции и их компоненты (знания, умения, навыки – далее ЗУВ) в соответствии с таблицей 2.

Таблица 2

#### Содержание дисциплины

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Коды ЗУВ (в соответствии с табл. 1)
1	Введение	Варианты названия дисциплины, определение, особенности, задачи, приложения, методы. Краткая история мировой и советской/российской ОЕЯ. Содержание и "идеология" курса. Ресурсы.	ПК-2 ПК-4	З (ПК-2) У (ПК-2) В (ПК-2) З (ПК-4) У (ПК-4) В (ПК-4)
2	Морфологический анализ	Чем занимается морфология, для чего нужна машинная морфология. Стемминг, лемматизация, частеречная (ЧР) разметка. Алгоритм Портера для английского. Лемматизация для русского языка: mystem, rymorphy2, АОТ. Словарь Зализняка как источник данных для лемматизации. Как строить гипотезы для незнакомых слов. Задача ЧР-разметки. Скрытые марковские модели для ЧР-разметки. Источники данных. Алгоритм Витерби.	ПК-2 ПК-4	З (ПК-2) У (ПК-2) В (ПК-2) З (ПК-4) У (ПК-4) В (ПК-4)
3	Классический сценарий информационного поиска, особенности задач информационного поиска	Базовые метрики качества поиска: точность, полнота, F1. Морфология в поиске. Частотные свойства терминов в коллекции: законы Хипса и Ципфа. Модель векторного пространства. Взвешивание терминов: подход tf.idf. Другие задачи ОЕЯ в поиске. Автоматическое реферирование. Постановка задачи, области применения, примеры. Типы рефератов. Методы реферирования текстовых документов. Постобработка рефератов. Оценка, инициативы по оценке методов автоматического реферирования. Веб-сниппеты (рефераты с учетом запроса): особенности задачи, методы. Оценка веб-сниппетов.	ПК-2 ПК-4	З (ПК-2) У (ПК-2) В (ПК-2) З (ПК-4) У (ПК-4) В (ПК-4)
4	Языковые модели (Language Models)	Области применения, ограничения. N-граммы, оценка вероятностей, доступные большие коллекции n-грамм. Оценка языковых	ПК-2 ПК-4	З (ПК-2) У (ПК-2) В (ПК-2)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Коды ЗУВ (в соответствии с табл. 1)
		моделей, перплексия. Сглаживание (smoothing), откат (backoff) и интерполяция (interpolation). Технические приемы при работе с языковыми моделями масштаба Веба, языковые модели с "памятью". Сглаживающие Лапласа, Гуда-Тьюринга, Кнезера-Нея.		З (ПК-4) У (ПК-4) В (ПК-4)
5	Синтаксический анализ (syntactic parsing)	Два формализма описания синтаксиса: система составляющих (constituency) и дерево зависимостей (dependency). Контекстно-свободные грамматики (КСГ): возможности и ограничения. Вероятностные КСГ. Алгоритм вероятностного синтаксического разбора. Оценка результатов разбора. Лексикализация ВКСГ.	ПК-2 ПК-4	З (ПК-2) У (ПК-2) В (ПК-2) З (ПК-4) У (ПК-4) В (ПК-4)
6	Извлечение информации из текстовых документов	Особенности задачи, источники данных. Именованные сущности (NEs), отношения. Основные подходы. Оценка. Машинный перевод: краткая история, сложности, подходы, приложения. Перевод, основанный на правилах (RBMТ) и статистический машинный перевод (СМП, STM). Источники данных для статистического машинного перевода. Выравнивание параллельного корпуса. IBM Models 1, 2. Фразовый СМП. Оценка систем МП: ручная, автоматическая (BLUE).	ПК-2 ПК-4	З (ПК-2) У (ПК-2) В (ПК-2) З (ПК-4) У (ПК-4) В (ПК-4)
7	Анализ тональности (sentiment analysis): источники данных и различные постановки задачи, примеры приложений.	Подход на основе классификации. Сложности анализа тональности: разнообразие форм выражения, ирония и сарказм, порядок и связность оценок. Использование словарей тонально окрашенных слов. Автоматическое пополнение/создание словарей. Обработка отрицаний. Фразы vs. отдельные слова. Аспекты/атрибуты: автоматическое выделение и списки, составленные вручную. Ресурсы для русского языка.	ПК-2 ПК-4	З (ПК-2) У (ПК-2) В (ПК-2) З (ПК-4) У (ПК-4) В (ПК-4)
8	Семантика, разные подходы и определения	Семантика, разные подходы и определения: логика высказываний, семантический веб, базы знаний, онтологии предметных областей, тезаурусы. Лексическая семантика: омонимы, многозначные слова, синонимы, антонимы, гипонимы/гиперонимы. Семантические словари -- тезаурусы. WordNet: синсеты и отношения. Методы определения семантической близости слов на основе тезауруса. Дистрибутивная семантика: семантическая близость на основе корпуса. Положительная поточечная взаимная информация (PPMI). Вычисления близости на основе структуры словосочетаний. Векторное представление семантики слов на основе на основе нейронных сетей.	ПК-2 ПК-4	З (ПК-2) У (ПК-2) В (ПК-2) З (ПК-4) У (ПК-4) В (ПК-4)

## Структура дисциплины

Таблица 3.

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины, час.				Форма текущего контроля успеваемости, промежуточной аттестации
			Всего	Аудиторная работа по видам учебных занятий		СР <sup>1</sup>	
				Л	СЗ <sup>2</sup>		
1.	Введение	Варианты названия дисциплины, определение, особенности, задачи, приложения, методы. Краткая история мировой и советской/российской ОЕЯ. Сожержание и "идеология" курса. Ресурсы.	6	2	1	3	Практические задания, опросы, диспуты
2.	Морфологический анализ	Чем занимается морфология, для чего нужна машинная морфология. Стемминг, лемматизация, частеречная (ЧР) разметка. Алгоритм Портера для английского. Лемматизация для русского языка: mystem, rymorphy <sup>2</sup> , АОТ. Словарь Зализняка как источник данных для лемматизации. Как строить гипотезы для незнакомых слов. Задача ЧР-разметки. Скрытые марковские модели для ЧР-разметки. Источники данных. Алгоритм Витерби.	7	2	2	3	
3.	Классический сценарий информационного поиска, особенности задач информационного поиска	Базовые метрики качества поиска: точность, полнота, F1. Морфология в поиске. Частотные свойства терминов в коллекции: законы Хипса и Ципфа. Модель векторного пространства. Взвешивание терминов: подход tf.idf. Другие задачи ОЕЯ в поиске. Автоматическое реферирование. Постановка задачи, области применения, примеры. Типы рефератов. Методы реферирования текстовых документов. Постобработка рефератов. Оценка, инициативы по оценке методов автоматического реферирования. Веб-сниппеты (рефераты с учетом запроса): особенности задачи, методы. Оценка веб-сниппетов.	7	2	2	3	
4.	Языковые модели (Language Models)	Области применения, ограничения. N-граммы, оценка вероятностей, доступные большие коллекции n-грамм. Оценка языковых моделей, перплексия. Сглаживание (smoothing), откат (backoff) и интерполяция (interpolation). Технические приемы при работе с языковыми моделями масштаба	6	1	2	3	

<sup>1</sup> Самостоятельная работа, включает в себя часы на текущий контроль

<sup>2</sup> Могут включать в себя: лабораторные работы, круглые столы, мастер-классы, мастерские, деловые игры, ролевые игры, тренинги, семинары по обмену опытом, выездные занятия, консультации

№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины, час.			Форма текущего контроля успеваемости, промежуточной аттестации	
			Всего	Аудиторная работа по видам учебных занятий			СР <sup>1</sup>
				Л	СЗ <sup>2</sup>		
		Веба, языковые модели с "памятью". Сглаживающие Лапласа, Гуда-Тьюринга, Кнезера-Нея.					
5.	Синтаксический анализ (syntactic parsing)	Два формализма описания синтаксиса: система составляющих (constituency) и дерево зависимостей (dependency). Контекстно-свободные грамматики (КСГ): возможности и ограничения. Вероятностные КСГ. Алгоритм вероятностного синтаксического разбора. Оценка результатов разбора. Лексикализация ВКСГ.	7	2	2	3	
6.	Извлечение информации из текстовых документов	Особенности задачи, источники данных. Именованные сущности (NEs), отношения. Основные подходы. Оценка. Машинный перевод: краткая история, сложности, подходы, приложения. Перевод, основанный на правилах (RBMТ) и статистический машинный перевод (СМП, STM). Источники данных для статистического машинного перевода. Выравнивание параллельного корпуса. IBM Models 1, 2. Фразовый СМП. Оценка систем МП: ручная, автоматическая (BLUE).	6	2	1	3	
7.	Анализ тональности (sentiment analysis): источники данных и различные постановки задачи, примеры приложений.	Подход на основе классификации. Сложности анализа тональности: разнообразие форм выражения, ирония и сарказм, порядок и связка оценок. Использование словарей тонально окрашенных слов. Автоматическое пополнение/создание словарей. Обработка отрицаний. Фразы vs. отдельные слова. Аспекты/атрибуты: автоматическое выделение и списки, составленные вручную. Ресурсы для русского языка.	6	1	2	3	
8.	Семантика, разные подходы и определения	Семантика, разные подходы и определения: логика высказываний, семантический веб, базы знаний, онтологии предметных областей, тезаурусы. Лексическая семантика: омонимы, многозначные слова, синонимы, антонимы, гипонимы/гиперонимы. Семантические словари -- тезаурусы. WordNet: синсеты и отношения. Методы определения семантической близости слов на основе тезауруса. Дистрибутивная	7	2	2	3	



№ п/п	Наименование и содержание тем	Основные понятия (категории) и проблемы, рассматриваемые в теме	Объем дисциплины, час.			Форма текущего контроля успеваемости, промежуточной аттестации	
			Всего	Аудиторная работа по видам учебных занятий			СР <sup>1</sup>
				Л	СЗ <sup>2</sup>		
		семантика: семантическая близость на основе корпуса. Положительная поточечная взаимная информация (PPMI). Вычисления близости на основе структуры словосочетаний. Векторное представление семантики слов на основе на основе нейронных сетей.					
9.	Промежуточная аттестация	Опрос по вопросам (устно)	2	-	-	2	
Всего:			54	14	14	26	

#### 4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

##### Общие положения.

Знания и навыки, полученные в результате лекций и семинарских занятий, закрепляются и развиваются в результате повторения материала, усвоенного в аудитории, путем чтения исследовательской литературы (из списков основной, дополнительной), статей по проблематики занятия и их анализа.

Самостоятельная работа обучающегося представляет самостоятельное изучение дополнительных материалов, Интернет-ресурсов и пр. Подготовка к семинарским занятиям, выполнение практических заданий, создание докладов, проектов и презентаций также является важной формой работы обучающихся. Самостоятельная работа может вестись как индивидуально, так и при содействии преподавателя. Вопросы и замечания, возникшие в ходе самостоятельного внеаудиторного чтения рекомендованной литературы, обсуждаются с преподавателем и другими обучающимися. Выносятся на обсуждение, как правило, актуальные проблемы и предлагается их рассмотреть с точки зрения того или иного теоретического подхода.

На занятиях материал излагается в проблемной форме. Основной упор в преподавании делается на изучение теоретических понятий и возможности их применения на конкретных примерах, в том числе в устных выступлениях обучающихся.

##### Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся.

- Исправление опечаток.
- Распознавание звучащей речи.
- Морфологический анализ: основные подходы.
- Нейронные языковые модели.
- Анализ тональности: постановка задачи, приложения, данные, методы решения.
- Синтаксический анализ: формализмы для представления синтаксической структуры, данные, подходы к решению, оценка качества.
- Методы извлечения информации на основе машинного обучения.
- Нейронные сети для автоматического реферирования.
- Дистрибутивные семантические модели.
- Машинный перевод: методы оценки качества.

### **Источники для самостоятельной подготовки:**

- Ганегедара, Т. Обработка естественного языка с TensorFlow: монография / Т. Ганегедара ; пер. с англ. В. С. Яценкова. - Москва: ДМК Пресс, 2020. - 382 с. - ISBN 978-5-97060-756-5. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1094940>.
- Дуглас А., Л. Люк, Д.А. Анализ сетей (графов) в среде R. Руководство пользователя / Д.А. Люк ; пер. с англ. А.В. Груздева. - Москва: ДМК Пресс, 2017. - 250 с. - ISBN 978-5-97060-428-1. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1028090>
- Иванова, Н.Ю. Системное и прикладное программное обеспечение: учебное пособие / Н.Ю. Иванова, В.Г. Маняхина; Московский педагогический государственный университет. – Москва : Прометей, 2011. – 202 с.: ил.,табл., схем. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=105792>
- Кравцова, Е. Д. Логика и методология научных исследований [Электронный ресурс]: учеб. пособие / Е. Д. Кравцова, А. Н. Городищева. – Красноярск: Сиб. федер. ун-т, 2014. – 168 с. - ISBN 978-5-7638-2946-4 - Режим доступа: <http://znanium.com/catalog.php?bookinfo=507377>

## **5. ОЦЕНОЧНЫЕ СРЕДСТВА**

Проведение текущего контроля в рамках реализации данной дисциплины проходит в соответствии с Таблицей 3 данной рабочей программы дисциплины по основным понятиям (категориям) и проблемам, рассматриваемым в предложенных темах. Фиксация результатов текущего контроля в рамках реализации данной дисциплины не предусмотрена.

**Типовые задания к текущей аттестации (практические задания, опросы, диспуты).**

### **Тема 1. Введение.**

#### Опрос 1.

1. Три слова в списке имеют одинаковый тип словоизменения, одно -- отличается. Отметьте это слово. Выберите номер правильного ответа:

- 1) чело
- 2) табло
- 3) сверло
- 4) кайло.

2. Какие основные инструменты и приложения используются при обработке естественного языка?

3. История развития методов обработки естественного языка?

### **Тема 2. Морфологический анализ.**

#### Опрос 2.

1. Каковы основные подходы к морфологическому анализу?

2. Какова основная методика отбора данных для морфологического анализа?

3. Каковы основные инструменты морфологического анализа?

4. Какие термины относятся к морфологической обработке? Выберите номера

правильных ответов:

- 1) лемматизация;
- 2) лемминг;
- 3) стемминг;
- 4) прокрастинация;
- 5) парсинг.

### **Тема 3. Языковые модели.**

#### Практическое задание 1.

Отметьте правильный набор граммем (в нотации НКРЯ/mystem) для выделенного слова в предложении:

Лично я ловлю покемонов не одобряю.

- 1) S,жен,неод=вин,ед
- 2) V,несов,пе=непрош,ед,изъяв,1-л
- 3) V,несов,пе=ед,пов,2-л
- 4) S,муж,неод=им,ед.

#### Тема 4. Анализ тональности.

##### Практическое задание 2.

Замените слова в предложении на пару лемма|часть\_речи, используйте набор частеречных тегов mystem:

Под круглым прозрачным колпаком из неравномерно утолщающегося книзу стекла сияли торсионные весы

(Людмила Улицкая. Казус Кукоцкого)

Пример: Я иду в кино

-->

я|SPRO идти|V в|PR кино|S

#### Тема 5. Синтаксический анализ.

##### Практическое задание 3.

Постройте биграммную языковую модель на основе корпуса:

<s> Вася любит мороженое </s><sub>[SEP]</sub><s> Лена любит малину </s><sub>[SEP]</sub><s> Вася любит Лену </s><sub>[SEP]</sub><s> Георгий ест мороженое </s><sub>[SEP]</sub><s> Лена рисует яблоко </s><sub>[SEP]</sub><s> Георгий любит Катю </s><sub>[SEP]</sub><s> Георгий любит смотреть, как Лена ест мороженое </s>

Упорядочите предложения по убыванию оценок вероятностей на основе построенной языковой модели.

1. <s> Лена любит мороженое </s>

2. <s> Лена рисует малину </s>

3. <s> Вася любит Катю </s>.

#### Тема 6. Извлечение информации.

##### Практическое задание 4.

Постройте биграммную языковую модель на основе корпуса:

<s> Вася любит мороженое </s>

<s> Лена любит малину </s>

<s> Вася любит Лену </s>

<s> Георгий ест мороженое </s>

<s> Лена рисует яблоко </s>

<s> Георгий любит Катю </s>

<s> Георгий любит смотреть, как Лена ест мороженое </s>

Вычислите перплексию модели на предложении:

<s> Георгий любит малину </s>

Обратите внимание, что для вычисления вероятности предложения (и, соответственно, перплексии) используются вероятности 4-х биграмм (N=4 в формуле вычисления перплексии).

Введите действительное число.

#### Тема 7. Автоматическое реферирование.

##### Практическое задание 5, Диспут 1.

Пусть у нас есть корпус, содержащий 10000 предложений, размер словаря -- 1500 уникальных слов (включая специальные "слова" -- маркеры начала и конца предложений).

Некоторые частоты униграмм:

ем 100

дуриан 1

и 5000

не 3000

морщусь 50

и биграмм:

<s> ем 20

ем дуриан 0

дуриан и 0

и не 300

не морщусь 15

морщусь </s> 5.

Примените сглаживание Лапласа ( $\alpha=1$ , сглаживание "+1") для оценки вероятностей биграмм и оцените на их основе вероятность предложения <s> ем дуриан и не морщусь </s>

В качестве ответа введите натуральный логарифм оценки вероятности предложения.

### **Тема 8. Вычислительная семантика.**

#### Практическое задание 6, Диспут 2.

Пусть у нас есть два «реферата-образца»:

1: карп лещ лещ щука сазан

2: лещ карп лещ сазан плотва

Упорядочите "рефераты" ниже по убыванию значения ROUGE-2.

1) лещ карп лещ сазан плотва

2) плотва сазан лещ карп сазан

3) щука сазан карп лещ сазан

4) карп карась лещ окунь лещ

5) щука сазан лещ карп сазан

### **Тема 9. Ответы на вопросы.**

#### Опрос 3, Диспут 3.

Выберите правильный разбор на составляющие предложения:

*Советник губернатора Чукотского автономного округа Роман Абрамович Роман Копин победил на повторных выборах главы администрации Чаунского района*

1. [[[[Советник [[губернатора [Чукотского [автономного округа]]] [Романа Абрамовича]]] [Роман Копин]] [победил [на [повторных [выборах [главы [администрации [Чаунского района]]]]]]]]]]

2. [[[[Советник [[губернатора [Чукотского [автономного округа]]] [Романа Абрамовича]]] [Роман Копин]] [[победил на] [повторных [выборах [главы [администрации [Чаунского района]]]]]]]]

3. [[[[[Советник губернатора] [Чукотского [автономного округа]]] [Романа Абрамовича]] [Роман Копин]] [победил [на [[повторных выборах] [главы [администрации [Чаунского района]]]]]]]]

4. [[[[Советник [[губернатора [Чукотского [автономного округа]]] [Романа Абрамовича]]] [[Роман Копин] победил]] [на [[повторных выборах] [главы [администрации [Чаунского района]]]]]]]]

[[[[Советник [[губернатора Чукотского] [[автономного округа] [Романа Абрамовича]]]]] [Роман Копин]] [победил [на [повторных [выборах [главы [администрации [Чаунского района]]]]]]]]]]

### **Тема 10. Машинный перевод (МП).**

#### Практическое задание 7, Диспут 4.

Выберите правильный разбор предложения в терминах зависимостей:

0 root

1 Активно

2 обсуждается

3 роль

4 нашей

5 страны

6 в

7 современном

8 быстро

9 меняющемся

10 мире,

11 проходящем

12 через

13 переломный

14 этап.

В приведенных ниже вариантах разбора каждая пара -- зависимость (хозяин, слуга), знаки препинания не учитываются, используется традиционный подход при установлении зависимостей с участием предлогов (не как в universal dependencies).

1. (0, 3) (3, 2) (2, 1) (3, 5) (5, 4) (3, 6) (6, 10) (10, 7) (10, 9) (9, 8) (10, 11) (11, 12) (12, 14) (14, 13)
2. (0, 2) (2, 1) (2, 3) (3, 5) (5, 4) (3, 6) (6, 10) (10, 7) (10, 9) (9, 8) (10, 11) (11, 12) (12, 14) (14, 13)
3. (0, 2) (2, 1) (2, 3) (3, 5) (5, 4) (5, 6) (6, 10) (10, 7) (10, 9) (9, 8) (10, 11) (11, 12) (12, 14) (14, 13)
4. (0, 2) (2, 1) (2, 3) (3, 5) (5, 4) (3, 6) (6, 7) (7, 8) (8, 9) (9, 10) (10, 11) (11, 12) (12, 14) (14, 13)
5. (0, 2) (2, 3) (3, 5) (5, 4) (3, 6) (6, 10) (10, 7) (10, 9) (2, 8) (10, 11) (11, 1) (11, 12) (12, 14) (14, 13)

### Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
Опрос	Ответ отсутствует или является односложным, или содержит существенные ошибки – не зачтено Слушатель в ответах демонстрирует знание всех теоретических положений, (развернуто) отвечает на все поставленные вопросы, предлагает обоснования при ответе на все или большинство поставленных вопросов; несущественные ошибки не снижают качество ответа — зачтено
Диспут	Пассивность, участие без представления аргументов и обоснования точки зрения, несформированность навыков профессиональной коммуникации в группе — не зачтено Представление аргументированной научной позиции, обоснование точки зрения в диспуте, демонстрация навыков профессиональной коммуникации в группе — зачтено
Практическое задание	Слушатель выполняет задание частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, полное и правильное выполнение задания в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

Форма промежуточной аттестации – зачет, выставляемый на основе письменной работы (эссе).

При аттестации используются система «зачтено» и «не зачтено» в соответствии с критериями оценивания.

В результате промежуточного контроля знаний обучающиеся получают аттестацию по дисциплине.

### Показатели, критерии и оценивание компетенций по уровням их формирования в процессе промежуточной аттестации

Таблица 4

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Коды ЗУВ (в соответствии с Таблицей 1)	Критерии оценивания	Оценка
зачет / письменная работа (эссе)	ПК-2 ПК-4	З (ПК-2) У (ПК-2) В (ПК-2) З (ПК-4) У (ПК-4) В (ПК-4)	<p>Эссе соответствует следующим требованиям:</p> <p>сформулирован исследовательский вопрос, корректно выбраны методы и собраны данные, тема раскрыта, соблюдены структура и научный стиль, соблюдены структура и научный стиль, сформулированы выводы, аргументация убедительна, правильно оформлен библиографический аппарат и т.д.</p> <p>Слушатель демонстрирует:</p> <p>глубокое усвоение программного материала; изложение данного материала исчерпывающе, последовательно, четко;</p> <p>умение делать обоснованные выводы; соблюдение норм устной и письменной литературной речи.</p> <p>Эссе представлено на защите на высоком профессиональном уровне</p> <p>Эссе соответствует следующим требованиям:</p> <p>сформулирован исследовательский вопрос, корректно выбраны методы и собраны данные, тема раскрыта, соблюдены структура и научный стиль, сформулированы выводы, аргументация убедительна, правильно оформлен библиографический аппарат и т.д.</p> <p>Слушатель демонстрирует:</p> <p>глубокое усвоение программного материала; изложение данного материала исчерпывающе, последовательно, четко;</p> <p>умение делать обоснованные выводы; соблюдение норм устной и письменной литературной речи.</p> <p>Эссе представлено на защите на высоком профессиональном уровне</p>	зачтено
			<p>Представленное эссе не отвечает предъявляемым требованиям (либо не предоставлено эссе);</p> <p>Слушатель демонстрирует:</p> <p>незнание значительной части программного материала;</p> <p>наличие существенных ошибок в определениях, формулировках, понимании теоретических положений;</p> <p>бессистемность при ответе на поставленный вопрос;</p> <p>отсутствие в ответе логически корректного анализа, аргументации, классификации;</p> <p>наличие нарушений норм устной и письменной литературной речи</p>	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «зачтено», показывают уровень сформированности у обучающегося компетенций.

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «не зачтено», показывают не сформированность у обучающегося компетенций по дисциплине.

**Типовые темы письменной работы (эссе) к промежуточной аттестации.**

- Морфологический анализатор для русского языка.
- Система автоматического реферирования новостных документов на русском языке.
- Система машинного перевода (русский—английский).
- Система анализа тональности для русского языка.

## **6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

**Основная литература:**

- Ганегедара, Т. Обработка естественного языка с TensorFlow: монография / Т. Ганегедара ; пер. с англ. В. С. Яценкова. - Москва: ДМК Пресс, 2020. - 382 с. - ISBN 978-5-97060-756-5. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1094940>.
- Дуглас А., Л. Люк, Д.А. Анализ сетей (графов) в среде R. Руководство пользователя / Д.А. Люк; пер. с англ. А.В. Груздева. - Москва: ДМК Пресс, 2017. - 250 с. - ISBN 978-5-97060-428-1. - Текст: электронный. - URL: <https://znanium.com/catalog/product/1028090>

**Дополнительная литература:**

- Иванова, Н.Ю. Системное и прикладное программное обеспечение: учебное пособие / Н.Ю. Иванова, В.Г. Маняхина; Московский педагогический государственный университет. – Москва: Прометей, 2011. – 202 с: ил.,табл., схем. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=105792>
- Кравцова, Е. Д. Логика и методология научных исследований [Электронный ресурс]: учеб. пособие / Е. Д. Кравцова, А. Н. Городищева. – Красноярск: Сиб. федер. ун-т, 2014. – 168 с. - ISBN 978-5-7638-2946-4 - Режим доступа: <http://znanium.com/catalog.php?bookinfo=507377>

Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения программы:

**Информационно-справочные системы:**

- Гарант.Ру. Информационно-правовой портал: <http://www.garant.ru>
- Информационная система «Единое окно доступа к образовательным ресурсам»: <http://window.edu.ru/>
- Открытое образование. Ассоциация «Национальная платформа открытого образования»: <http://npoed.ru>
- Официальная Россия. Сервер органов государственной власти Российской Федерации: <http://www.gov.ru>
- Официальный интернет-портал правовой информации. Государственная система правовой информации: <http://pravo.gov.ru>
- Правовой сайт КонсультантПлюс: <http://www.consultant.ru/sys>
- Российское образование. Федеральный портал: <http://www.edu.ru>

**Тематические системы:**

- Google. Книги: <https://books.google.com>
- Internet Archive: <https://archive.org>
- Koob.ru. Электронная библиотека «Куб»: <http://www.koob.ru/philosophy/>
- Библиотека Ихтика [ihtik.lib.ru]: <http://ihtik.lib.ru/>
- Докусфера — Российская национальная библиотека: <http://leb.nl.ru>

- ЕНИП — Электронная библиотека «Научное наследие России»: <http://e-heritage.ru/>
- Интелрос. Интеллектуальная Россия: <http://www.intelros.ru/>
- Национальная электронная библиотека НЭБ: <http://www.rusneb.ru>
- Неприкосновенный запас: <http://magazines.russ.ru/nz/>
- Президентская библиотека: <http://www.prlib.ru>
- Российская государственная библиотека: <http://www.rsl.ru/>
- Российская национальная библиотека: <http://www.nlr.ru/poisk/>

## **7. ПРОГРАММНОЕ И МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ПРОГРАММЫ**

В ходе реализации образовательного процесса используются многофункциональные аудитории для проведения занятий лекционного типа, занятий семинарского типа, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Проведение занятий лекционного типа и семинарского типа обеспечивается демонстрационным оборудованием.

Для обучающихся с ограниченными возможностями здоровья и инвалидов (в случае необходимости) могут быть созданы специальные условия для получения образования.

### **Программное обеспечение**

При осуществлении образовательного процесса в рамках Университета слушателям рекомендовано использовать следующее лицензионное программное обеспечение:

- OS Microsoft Windows (OVS OS Platform)
- MS Office (OVS Office Platform)
- Adobe Acrobat Professional 11.0 MLP AOO License RU
- Adobe CS5.5 Design Standart Win IE EDU CLP
- ABBYY FineReader 11 Corporate Edition
- ABBYY Lingvo x5
- Adobe Photoshop Extended CS6 13.0 MLP AOO License RU
- Adobe Acrobat Reader DC /Pro – бесплатно
- Google Chrome – бесплатно
- Opera – бесплатно
- Mozilla – бесплатно
- VLC – бесплатно