



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
«Дальневосточный федеральный университет»  
(ДВФУ)

**ВОСТОЧНЫЙ ИНСТИТУТ – ШКОЛА РЕГИОНАЛЬНЫХ И  
МЕЖДУНАРОДНЫХ ИССЛЕДОВАНИЙ**

«СОГЛАСОВАНО»

«УТВЕРЖДАЮ»

Руководитель ОП

Заведующий кафедрой  
лингвистики и межкультурной коммуникации

Николаева О.В.

(подпись)

(ФИО)

Г.Н. Ловцевич

(подпись) (ФИО.)

«25» ноября 2021 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
Обработка естественного языка и текстовая аналитика  
Направление подготовки 45.04.01 Филология  
(Европейские языки: коммуникация и лингвоаналитика)  
Форма подготовки очная

курс 2 семестр 3  
лекции 0 час.  
практические занятия 10 час.  
лабораторные работы 0 час.  
в том числе с использованием МАО 0 час.  
всего часов аудиторной нагрузки 10 час.  
самостоятельная работа 134 час.  
в том числе на подготовку к экзамену 54 час.  
контрольные работы (количество) не предусмотрены  
курсовая работа / курсовой проект не предусмотрены  
зачет не предусмотрен  
экзамен 3 семестр

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта по направлению подготовки 45.04.01 Филология, утвержденного приказом Министерства науки и высшего образования РФ от 12 августа 2020 г. №980

Рабочая программа обсуждена на заседании кафедры лингвистики и межкультурной коммуникации, протокол № 3 от «25» ноября 2021 г.  
Заведующий кафедрой лингвистики и межкультурной коммуникации доктор филол. н. Ловцевич Г.Н.  
Составители: канд. филол. н., доцент Спицына Н.А.

Владивосток  
2022

**Оборотная сторона титульного листа РПД**

**I. Рабочая программа пересмотрена на заседании кафедры лингвистики и межкультурной коммуникации:**

Протокол от «\_\_\_\_\_» \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Заведующий кафедрой \_\_\_\_\_ Ловцевич Г.Н.\_\_\_\_  
(подпись) (И.О. Фамилия)

**II. Рабочая программа пересмотрена на заседании кафедры лингвистики и межкультурной коммуникации:**

Протокол от «\_\_\_\_\_» \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Заведующий кафедрой \_\_\_\_\_ Ловцевич Г.Н.\_\_\_\_  
(подпись) (И.О. Фамилия)

**III. Рабочая программа пересмотрена на заседании кафедры лингвистики и межкультурной коммуникации:**

Протокол от «\_\_\_\_\_» \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Заведующий кафедрой \_\_\_\_\_ Ловцевич Г.Н.\_\_\_\_  
(подпись) (И.О. Фамилия)

**IV. Рабочая программа пересмотрена на заседании кафедры лингвистики и межкультурной коммуникации:**

Протокол от «\_\_\_\_\_» \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Заведующий кафедрой \_\_\_\_\_ Ловцевич Г.Н.\_\_\_\_  
(подпись) (И.О. Фамилия)

Общая трудоемкость дисциплины составляет 4 зачётных единицы / 144 академических часа. Является дисциплиной части, формируемой участниками образовательных отношений ОП, изучается на 2 курсе и завершается экзаменом. Учебным планом предусмотрено проведение практических занятий, а также выделены часы на самостоятельную работу студента - 134 часа.

(1 зачетная единица соответствует 36 академическим часам)

## 2. Цели и задачи освоения дисциплины:

Цель: сформировать у обучающихся системное представление о методах автоматической обработки естественно-языкового текста на всех уровнях лингвистического анализа; ознакомление студентов с практическими методами автоматической обработки естественно-языковых текстов.

### Задачи:

1. Освоить общие принципы построения систем обработки речи (устной и письменной).
2. Дать представление о возможностях компьютерной лингвистики.
3. Ознакомить с лингвистическим обеспечением различных прикладных систем (автоматического индексирования, реферирования, и т.д.).
4. Раскрыть возможности лингвистических процессоров.

Результаты обучения по дисциплине должны быть соотнесены с установленными в ОПОП индикаторами достижения компетенций.

Совокупность запланированных результатов обучения по дисциплине должна обеспечивать формирование у выпускника всех компетенций, установленных ОПОП.

Наименование категории (группы) универсальных компетенций	Код и наименование универсальной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции
Системная аналитика	УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий	УК-1.1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий

Наименование категории (группы) универсальных компетенций	Код и наименование универсальной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции
Межкультурная коммуникация	УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	УК-4.1 Способен работать с аутентичными текстами академического и профессионального характера.

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения по дисциплине)
УК-1.1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий	Знает принципы разработки системного подхода
	Умеет выработать обоснованную стратегию в анализе лингвистических данных
	Владеет навыками критического рассмотрения эмпирического материала на естественном языке
УК-4.1 Способен работать с аутентичными текстами академического и профессионального характера.	Знает как устроена современная электронная языковая среда, основные уровни обработки текста на ЕЯ и существующие модели статистического, морфологического и синтаксического анализа текстов;
	Умеет проектировать процесс обработки языковых данных, работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;
	Владеет навыками и методами поиска информации по работе с естественным языком;

**Профессиональные компетенции выпускников и индикаторы их достижения:**

Тип задач	Код и наименование профессиональной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции
Анализ естественного языка	ПК-1 Способен самостоятельно проводить научные исследования в области системы языка и основных закономерностей функционирования фольклора и литературы в синхроническом и	ПК-1.2 Самостоятельно отбирает, систематизирует, анализирует материал исследования
		ПК-1.5 Использует современные информационно-коммуникационные технологии в процессе выполнения научного проекта

Тип задач	Код и наименование профессиональной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции
	диахроническом аспектах, в сфере устной и письменной коммуникации	
Текстовая аналитика	ПК-5 Способен проводить дискурсивный мониторинг межнациональных и межэтнических отношений в профессиональной сфере, консультативную и организаторскую деятельность по гармонизации	ПК-5.2 Под руководством специалиста более высокой квалификации разрабатывает методические рекомендации для руководящих структур образовательных и иных организаций социо-гуманитарной направленности по гармонизации устного и письменного межкультурного профессионального общения, профилактики межкультурных конфликтов

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения по дисциплине)
ПК-1.2 Самостоятельно отбирает, систематизирует, анализирует материал исследования	Знает особенности реализации естественно-языковых систем, назначение лингвистического процессора;
	Умеет эксплуатировать и сопровождать информационные системы и сервисы
	Владеет навыками методикой использования компьютерных технологий в предметной области;
ПК-1.5 Использует современные информационно-коммуникационные технологии в процессе выполнения научного проекта	Знает лингвистические технологии в сфере АОТ; виды программной продукции АОТ;
	Умеет осуществлять и обосновывать выбор проектных решений по видам обеспечения информационных систем; понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;
	Владеет навыками вести исследовательскую лингвистическую работу с применением самых современных методов получения и анализа данных (корпусные методы, статистическая обработка, работа с экспертными оценками и т.д.);
ПК-5.2 Под руководством специалиста более высокой квалификации разрабатывает методические рекомендации для руководящих структур образовательных и иных организаций социо-гуманитарной направленности по гармонизации	Знает ключевые направления в области компьютерной лингвистики; возможности использования программных технологий анализа и обработки текстовых данных;
	Умеет применять лингвистические технологии для проектирования систем анализа и синтеза естественного языка в сфере АОТ; использовать в работе и адаптировать под конкретную задачу имеющиеся программные средства обработки текста и существующие в мире лингвистические

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения по дисциплине)
устного и письменного межкультурного профессионального общения, профилактики межкультурных конфликтов	ресурсы;
	Владеет навыками работы со специальными программными средствами АОТ;

### 3. Трудоёмкость дисциплины и видов учебных занятий по дисциплине

Общая трудоёмкость дисциплины составляет 4 зачётных единиц (144 академических часа).

(1 зачетная единица соответствует 36 академическим часам)

Видами учебных занятий и работы обучающегося по дисциплине могут являться:

Обозначение	Виды учебных занятий и работы обучающегося
Пр	Практические занятия
СР	Самостоятельная работа обучающегося в период теоретического обучения
Контроль	Самостоятельная работа обучающегося и контактная работа обучающегося с преподавателем в период промежуточной аттестации

### Структура дисциплины:

Форма обучения – очная.

№	Наименование раздела дисциплины	Семестр	Количество часов по видам учебных занятий и работы обучающегося						Формы промежуточной аттестации
			Лек	Лаб	Пр	ОК	СР	Контроль	
1	Практическое занятие 1. Знакомство с технологией NLP (Natural Language Processing).	3	0	0	2	0	10	10	УО-1 экзамен
2	Практическое занятие 2. Морфологический анализ текстов.	3	0	0	2	0	10	10	УО-1 экзамен
3	Практическое занятие 3. Инструментальные системы разработки приложений по автоматической обработке текстов на естественном языке. Системы обработки ЕЯ-текстов.	3	0	0	2	0	20	14	УО-1 экзамен
4	Практическое занятие 4. Извлечение информации из	3	0	0	2	0	20	10	УО-1 экзамен

	текстов: портрет направления.								
5	Практическое занятие 5. Текстовая аналитика. Мониторинг и анализ тональности.	3	0	0	2	0	20	10	УО-1 экзамен
	Итого:	3	0	0	10	0	80	54	

### 3. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

Не предусмотрено

### 4. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА

#### ПРАКТИЧЕСКИЕ ЗАНЯТИЯ

#### ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 1. Знакомство с технологией NLP (Natural Language Processing). (2 ч.)

Естественный язык в сравнении с языком программирования. Определение термина «Обработка естественного языка». Трудности обработки естественного языка машинами. Практическое применение. Общие этапы и модули обработки текстов. Лингвистические ресурсы: построение и применение (словари, тезаурусы, онтологии). Текстовые корпуса.

Задания: 1. Ознакомьтесь с системами Wordnet <https://wordnet.princeton.edu/>, проектом Russnet [http://project.phil.spbu.ru/RussNet/index\\_ru.shtml](http://project.phil.spbu.ru/RussNet/index_ru.shtml), EuroWordNet <https://archive.illc.uva.nl/EuroWordNet/>. Подготовьте сообщение о возможностях каждой из систем.

2. Какие возможности обработки текста имеют текстовые корпуса: Национальный корпус русского языка <https://ruscorpora.ru/new/>.

3. Подготовьте проектное задание по использованию программы анализа и лингвистической обработки текстов (на выбор): OpenNLP <https://opennlp.apache.org/index.html>, СУБД StarLing <https://starlingdb.org/downl.php?lan=en>, WordSmith Tools <https://lexically.net/wordsmith/>, TextAnalyst 2.0 <http://www.analyst.ru/index.php?lang=eng&dir=content/downloads>.

#### ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 2. Морфологический анализ текстов. (2 ч.)

Основные понятия. Функции систем морфологического анализа.  
Представление текста в виде вектора. Представление слов в словаре.  
Морфологическая омонимия. Подходы к морфологическому анализу.  
Бессловарная морфология. Словарная морфология на основе словаря основ.  
Словарная морфология на основе словаря словоформ. Морфемный разбор.  
Обзор модулей морфологического анализа.

Задания: 1. Ознакомьтесь с системами Диалинг, Морфологический процессор АОТ <https://nlpub.ru/%D0%90%D0%9E%D0%A2>

2. Проведите обзор возможностей компьютерного анализа текста с использованием графематических и морфологических анализаторов <http://lab314.brsu.by/kmp-lite/CL/CL-Lect/КАТ.htm>;

3. Подготовьте проектное задание по использованию программы преобразования текстов: WinMerge project <https://winmerge.org/>, психолингвистические программы ВААЛ <http://www.vaal.ru/>, генераторы текстов SCIgen <https://pdos.csail.mit.edu/archive/scigen/>, программа для генерации русскоязычных стихоподобных текстов <https://fresheye.herokuapp.com/>, генератор письменных жалоб Скотта Пейкина <https://www.pakin.org/complaint>, зоопарк ботов <http://www.simonlaven.com/>, робот оракул <http://www.necromancers.ru/>, коллекция программ собеседников <http://netnotes.narod.ru/talkerus/index.html>.

### **ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 3. Извлечение информации из текстов: портрет направления. ( 2 ч.)**

Специфика задач, подходы к решению, извлекаемая информация. Методы оценки качества извлечения. Именованные сущности и особенности их извлечения. Особенности извлечения атрибутов, отношений и фактов. Лингвистические шаблоны и правила. Машинное обучение в задачах извлечения информации. Инструментальные системы для извлечения информации. Извлечение терминологической информации.

Задания: 1. Ознакомьтесь с системами AutoSlog, OntosMiner, GATE, Томита Парсер, пакет OpenNLP, Stanford CoreNLP и выделите какие типы информации можно извлечь из текста при помощи данных систем.

2. Проведите обзор возможностей компьютерного анализа текста с использованием семантических и синтаксических анализаторов <http://lab314.brsu.by/kmp-lite/CL/CL-Lect/КАТ.htm>;

3. Подготовьте проектное задание по использованию систем обработки естественного языка и машинного перевода: Webclopedia - проект, связанный с созданием информационных агентов, способных генерировать ответы пользователю на основе анализа разнородных коллекций ресурсов, доступных в сети.

ONTOSAURUS - онтологический тезаурус, включающий 90,000 концептуальных сущностей (словарная база взята из проекта WordNet).

SUMMARIST - мультязычная система реферирования текстовой информации.

ReWrite - система двуязычного перевода, основанная на статистическом анализе пар текстовых корпусов

[https://www.isi.edu/projects/nlg/software\\_1](https://www.isi.edu/projects/nlg/software_1)

#### **ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 4. Текстовая аналитика. (2 ч.)**

Извлечение данных из множества различных источников. Смещение акцента от поиска к обработке извлеченных данных. Автоматический разбор неструктурированных документов, выделение заголовка, основной части документа, анализ структуры документа и ссылок. Определение типов данных. Выбор обработчика в зависимости от типа данных. Обработка аудио, и видео информации. Извлечение значимой информации из текста. Выделение собственных имен, людей, организаций, географических мест, наименований, ссылок на сайты, e-mail, сообщения на форумах, сообщения в соц. сетях и пр. Поиск и наполнение данными о профилях клиентов, отношениях между клиентами, выявление родственных связей, выявление степени влияния между клиентами.

Задания: 1. Ознакомьтесь со списком продуктов в сфере корпоративного поиска [https://en.wikipedia.org/wiki/Comparison\\_of\\_enterprise\\_search\\_software](https://en.wikipedia.org/wiki/Comparison_of_enterprise_search_software), сделайте анализ возможностей данных продуктов;

2. Проведите обзор возможностей продуктов и систем команды Megaputer <https://www.megaputer.com/ru/polyanalyst/>;

3. Сделайте обзор систем текстовой аналитики в сферах: e-commerce, бренд-менеджмента, маркетинга, конкурентной разведки, управления опытом клиента, информационной безопасности, виртуальных ассистентов (технология Polaris <https://techcrunch.com/2012/08/30/in-battle-with-amazon-walmart-unveils-polaris-a-semantic-search-engine-for-products/> , Cloud Search <https://workspace.google.com/products/cloud-search/#commerce>).

4. Подготовьте проектное задание по использованию систем синтеза и распознавания речи Exiton <https://www.softportal.com/software-4924-text-to-speech-converter-for-ms-word.html>, Говорилка <https://www.vector-ski.ru/vecs/govorilka/>.

## **ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 5. Текстовая аналитика. Мониторинг и анализ тональности. (2 ч.)**

Мониторинг упоминаний объектов в медиасреде. Анализ социальных медиа: Meltwater <https://www.meltwater.com/en>, измерение эффективности компаний по продвижению бренда Cision <https://www.cision.com/>. Измерение эффективности компаний по продвижению бренда: Sysomos, Luminoso Dashboard, Radian6, NewsWhip. Российские системы мониторинга социальных медиа — YouScan, Крибрум, BrandAnalytics, SemanticForce. Сложности анализа тональности текстов. Словари оценочной лексики : Ру СентиЛекс <https://www.labinform.ru/pub/rusentilex/index.htm>, словарь ProductSentiRus, словарь Linis-crowd. Анализ тональности документов в целом.

Задания: 1. Ознакомьтесь с программами анализа и лингвистической обработки текстов: Mystem <https://yandex.ru/dev/mystem>, Лингвоанализатор <http://www.rusf.ru/books/analysis/>, Морфологический анализатор <https://starlingdb.org/morph.htm>, MonoConc/ParaConc <http://www.athel.com/>. Подготовьте аналитику своего материала с использованием одного из сервисов.

2. Представьте анализ текста с использованием одной из программ: <https://software.sil.org/products/>, <http://www.textanalysis.info/>

3. Проведите анализа текста новостей с использованием программ : [https://www.isi.edu/projects/nlg/software\\_1](https://www.isi.edu/projects/nlg/software_1).

## **5. СТРУКТУРА, СОДЕРЖАНИЕ, УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ**

### **Рекомендации по самостоятельной работе студентов**

Самостоятельная работа студентов состоит в работе с литературой, подготовке к лабораторным работам и выполнении индивидуальных заданий по темам.

### **Работа с литературой**

В процессе подготовки к практическим занятиям студентам необходимо обратить особое внимание на самостоятельное изучение

рекомендованной учебно-методической литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной литературой, материалами Интернета является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у студентов свое отношение к конкретной проблеме.

### **К практическим занятиям**

Подготовку к каждому практическому занятию каждый студент должен начать с изучения теоретического материала и ознакомления с имеющимися реализациями. Внимательно знакомиться с предлагаемым программным обеспечением.

### **Индивидуальные задания**

По теме практического занятия студентам выдаются задания, предназначенные для закрепления умений и навыков, полученных при выполнении практических заданий. Задания выполняются самостоятельно. По каждому заданию предполагается подготовка индивидуального документа (проекта). Выполненный проект демонстрируется преподавателю в начале следующего практического занятия. Структура индивидуального задания определяется темой практического занятия.

### **Критерии оценки индивидуальных заданий (проектов)**

– 100-86 баллов выставляется, если содержание и составляющие части соответствуют выданному заданию. Продемонстрировано владение навыками подготовки документа по теме. Фактических ошибок, связанных с пониманием структуры и содержания задания нет.

– 85-76 - баллов выставляется, если при выполнении задания допущено не более одной ошибки. Продемонстрировано владение навыками подготовки документа по теме. Фактических ошибок, связанных с пониманием структуры и содержания задания нет.

– 75-61 балл выставляется, если при выполнении задания допущено не более двух ошибок. Продемонстрировано знание и владение навыками подготовки документа по теме. Допущено не более 2 ошибок, связанных с пониманием структуры и содержания задания.

– 60-50 баллов - если структура и содержание задания не соответствуют требуемым.

## Шкала оценивания

Менее 60 баллов	незачтено	неудовлетворительно
От 61 до 75 баллов	зачтено	удовлетворительно
От 76 до 85 баллов	зачтено	хорошо
От 86 до 100 баллов	зачтено	отлично

## План-график выполнения самостоятельной работы по дисциплине

№ п/п	Вид самостоятельной работы	Дата/сроки выполнения	Примерные нормы времени на выполнение	Форма контроля
1	Задания для самостоятельной работы к теме 1.	1 - 2 недели	5	Устный опрос
2	Задания для самостоятельной работы к теме 2.	3 - 4 недели	5	Устный опрос
3	Задания для самостоятельной работы к теме 3.	5-6 недели	10	Устный опрос
4	Задания для самостоятельной работы к теме 4.	7-8 недели	10	Устный опрос
5	Задания для самостоятельной работы к теме 1.	9-10 недели	10	Устный опрос
6	Задания для самостоятельной работы к практическому занятию 1.	1 - 2 недели	5	Устный опрос
7	Задания для самостоятельной работы к практическому занятию 2.	3 - 4 недели	5	Устный опрос
8	Задания для самостоятельной работы к практическому занятию 3.	5-6 недели	10	Устный опрос
9	Задания для самостоятельной работы к практическому занятию 4.	7-8 недели	10	Устный опрос
10	Задания для самостоятельной работы к практическому занятию 4.	9-10 недели	10	Устный опрос

### Задания для самостоятельной работы к теме 1-5.

1. Какие типы электронных лингвистических ресурсов вам известны?  
На каком критерии основана данная типология?

2. Приведите по 4 примера ресурсов, относящихся к ресурсам-продуктам и ресурсам-средствам обработки.
3. К какому типу ресурсов, по-вашему, относится семантическая сеть WordNet? Электронный текстовый корпус НКРЯ? БД Русского ассоциативного словаря? Конкордансер SCP? Морфологический анализатор MyStem? Фонетический анализатор Praat?
4. Выполнение этапа нарезки аудиофрагментов для корпусной выдачи:

Инструкция по нарезке диалектных аудиофайлов для корпуса

1. Файлы должны быть в формате WAV. Конвертировать файлы можно в программе Switch (ярлык на рабочем столе). Для этого необходимо нажать кнопку + Add file(s), внизу слева выбрать нужный формат (wav) и нажать кнопку Convert.
2. Каждому аудиофайлу в задании соответствует текстовый файл с расшифровкой. Внимательно прочитывая текст и прослушивая запись, необходимо нарезать файл на фрагменты длиной до 30 секунд (но не менее 10 секунд!). Необходимо при этом следить, чтобы кусочки были смысловыми (т.е. если высказывание о чем-л. короткое, то лучше сделать его отдельным кусочком в 10 секунд, чтобы не начинать новый смысловой кусок, который может в 30 секунд и не уместиться). Ни в коем случае кусочек не должен обрываться на полуслове!!!
3. Параллельно с нарезкой аудиофрагментов необходимо также размечать текстовые файлы с расшифровкой: в текст вставляются маркеры границ отрывков, соответствующих звучащим кусочкам. Символ начала фрагмента должен иметь вид «& (без пробела) + N (где N – номер фрагмента, полностью имя файла указывать не нужно) + пробел». Конец фрагмента не маркируется. При разметке текстовой расшифровки опираться нужно на файл исходной расшифровки (он имеет имя в виде фамилии информанта + (необяз.) числовой индекс), затем те же обозначения вставляются в файл с размеченной (т.е. выполнена грамматическая разметка) расшифровкой (имя файла – имя информанта + индекс + Razm).
4. Нарезку файлов можно осуществлять в любом редакторе: Audacity, Sony SoundForge – в них работают стандартные горячие клавиши: - открыть нужный аудиофайл Ctrl + O; - для удобства работы масштабировать график: Ctrl + 1 / Ctrl + 3 (ближе, дальше) в Audacity, колесиком мышки в Sony SoundForge; - выделить фрагмент записи: Shift + ←или→ (должен получиться цельный смысловой кусочек, паузы видны по графикам, это плоские линии – там и должна

проходить граница фрагмента); - вырезать фрагмент: Ctrl + X; - открыть новый файл Ctrl + N; При открытии (по сути, создании) нового файла необходимо следить за тем, чтобы качество записи совпадало с исходным (если в исходной записи, например, 64 Кбит/с, то если редактор предлагает качество выше, то это не улучшает качество фрагмента на выходе, а лишь увеличивает занимаемое место на диске); - вставить вырезанный из исходного файла фрагмент Ctrl + V; - далее новый файл закрываем, сохраняя его с именем исходного файла + порядковый номер: Например, был файл Краснова.wav – на выходе должна получиться папка «Краснова», в которой находятся файлы Krasnova\_01.wav, Krasnova\_02.wav, Krasnova\_03.wav и т.д. Называть нарезанные файлы нужно латинскими символами. Если исходный файл имеет числовой индекс, то его нужно сохранять и в именах фрагментов. К примеру, имеем аудиофайл Trininal – режем его на кусочки с именами Trininal\_1, Trininal\_2 и т.д. Итак, материалы с выполненным заданием должны иметь вид папки с именем «Имя информанта+индекс», в этой папке должны содержаться:

1) Папка с аудиофрагментами продолжительностью от 10 до 30 секунд и битрейтом, не превышающим исходный;

2) Текст расшифровки с маркерами начальных границ фрагментов: ему присваивается имя «Имя информанта+индекс+&&&»;

3) Текст расшифровки с разметкой с теми же маркерами: ему присваивается имя «Имя информанта+индекс+Razm+&&&».

5: Пример задания для работы с созданной БД (пример искусственный, поскольку реальная работа будет проводиться над индивидуальными научно исследовательскими проектами): Имеется файл базы данных MS Access «Зоопарк», в котором находится две не связанных между собой таблицы.

Задача – с опорой на эти таблицы построить БД, выполнив следующие задания:

1. Заполнить таблицу «Питомцы», создав в ней поля «Кличка», «Вид» (связать с таблицей «План зоопарка»), «Пол» (значение из списка), «Дата рождения» (формат даты), «Дата поступления», «Зоологнаблюдатель» (связать с таблицей «Зоологи»), «Пища», «Прививки» (да/нет).

2. Создать простую форму для таблицы «Питомцы»; воспользовавшись этой формой, внести в таблицу данные о поступлении 3 новых зверушек. 3. Составить запрос, по которому выводятся только клички и вид животных, содержащихся в отделе хищников (Запрос\_1).

4. Составить запрос, по которому выводятся клички животных, находящихся под присмотром биологов или докторов наук, с указанием фамилии наблюдающего зоолога (Запрос\_2).
5. Составить запрос, по которому в БД создается новая таблица «Звери\_тропики», в которую помещены записи о питомцах, чья среда обитания расположена в тропическом климате (Запрос\_создание).
6. Составить запрос на удаление из таблицы «Питомцы» записей о животных, которые поступили в зоопарк ранее 2022 года (Запрос\_удаление). Примечание: для тестирования запроса можно создать копию таблицы «Питомцы».

#### 6: Порядок работы с синсетам YARN

1. Ознакомиться с инструкцией, представленной на сайте <http://nlpub.ru/YARN/Инструкция>.
2. Выполнить отбор 5 лексических единиц, нуждающихся в представлении в тезаурусе.
3. Привести их толкования (МАС, словарь под ред. С.И. Ожегова-Н.Ю. Шведовой, С.А. Кузнецова);
4. Найти возможные синонимы для каждого из слов.
5. Просмотрев толкования этих слов в ряде толковых словарей, выбрать наиболее нейтральное слово для ряда синонимов – т.е. заглавие синсета.
6. Привести частотные характеристики для каждого из слов-синонимов в обоих рядах (Новый частотный словарь русской лексики О.Н. Ляшевской и С.А. Шарова) – цифры в столбце ipm.
- Пересмотреть порядок синонимов в синсете с учетом их употребительности.
7. Если необходимо, снабдить слова функционально-стилистическими пометами (в скобочках).
8. Снабдить слова и их синонимы контекстами-иллюстрациями из НКРЯ

## 6. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

№ п/п	Контролируемые разделы / темы дисциплины	Код и наименование индикатора достижения компетенции	Результаты обучения	Оценочные средства	
				текущий контроль	промежуточная аттестация
	Тема №1, Знакомство с технологией NLP (Natural Language Processing).	УК-1.1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	Знает принципы разработки системного подхода	УО-1	УО-1 экзамен
			Умеет вырабатывать обоснованную стратегию в анализе лингвистических данных		
Владеет навыками критического рассмотрения эмпирического материала на естественном языке					
	Тема №2, Морфологический анализ текстов.	УК-4.1 Способен работать с аутентичными текстами академического и профессионального характера.	Знает как устроена современная электронная языковая среда, основные уровни обработки текста на ЕЯ и существующие модели статистического, морфологического и синтаксического анализа текстов;	УО-1	УО-1 экзамен
			Умеет проектировать процесс обработки языковых данных, работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;		
			Владеет навыками и методами поиска информации по работе с естественным языком;		
	Тема №3, Инструментальные системы разработки	ПК-1.2 Самостоятельно отбирает, систематизирует, анализирует материал	Знает особенности реализации естественно-языковых систем, назначение лингвистического процессора;	УО-1	УО-1 экзамен

	приложений по автоматической обработке текстов на естественном языке.	исследования	<p>Умеет эксплуатировать и сопровождать информационные системы и сервисы</p> <p>Владеет навыками методикой использования компьютерных технологий в предметной области;</p>		
	Тема №4, Извлечение информации из текстов: портрет направления.	ПК-1.5 Использует современные информационно-коммуникационные технологии в процессе выполнения научного проекта	<p>Знает лингвистические технологии в сфере АОТ; виды программной продукции АОТ;</p> <p>Умеет осуществлять и обосновывать выбор проектных решений по видам обеспечения информационных систем; понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;</p> <p>Владеет навыками вести исследовательскую лингвистическую работу с применением самых современных методов получения и анализа данных (корпусные методы, статистическая обработка, работа с экспертными оценками и т.д.);</p>	УО-1	УО-1 экзамен
	Тема №5, Текстовая аналитика. Мониторинг и анализ тональности.	ПК-5.2 Под руководством специалиста более высокой квалификации разрабатывает методические рекомендации для руководящих структур образовательных и иных организаций социо-гуманитарной направленности	<p>Знает ключевые направления в области компьютерной лингвистики; возможности использования программных технологий анализа и обработки текстовых данных;</p> <p>Умеет применять лингвистические технологии для проектирования систем анализа и синтеза естественного языка в сфере АОТ; использовать в работе и адаптировать под</p>	УО-1	УО-1 экзамен

		по гармонизации устного и письменного межкультурного профессионального общения, профилактики межкультурных конфликтов	конкретную задачу имеющиеся программные средства обработки текста и существующие в мире лингвистические ресурсы; Владеет навыками работы со специальными программными средствами АОТ;		
--	--	---	--	--	--

Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков и (или) опыта деятельности, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие результаты обучения, представлены в Приложении

## 7. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### Основная литература

1. Введение в прикладную лингвистику: учебное пособие / А. Н. Баранов; Московский государственный университет, Филологический факультет. А. Н. Баранов. - М.: URSS [Либроком], 2019. - 367 с. <https://lib.dvfu.ru:8443/lib/item?id=chamo:779256&theme=FEFU>
2. Корпусная лингвистика / Е.В. Грудева. —М.: ФЛИНТА, 2018. - 165 с.
3. Автоматическое понимание текстов: Системы, модели, ресурсы: Учебное пособие / Н.Н. Леонтьева. – М.: Академия, 2019. - 304с. <https://lib.dvfu.ru:8443/lib/item?id=chamo:235710&theme=FEFU>
4. Тезаурусы в задачах информационного поиска. / Н.В. Лукашевич. - М.: Изд-во Московского университета, 2017. - 512с. <https://lib.dvfu.ru:8443/lib/item?id=chamo:403174&theme=FEFU>
5. Щипицина, Л. Ю. Информационные технологии в лингвистике [Электронный ресурс]: учеб. пособие / Л. Ю. Щипицина. — М.: ФЛИНТА, 2018. — 128 с. - ISBN 978-5-9765-1431-7 <http://znanium.com/catalog.php?bookinfo=462989>
6. Информатика и лингвистика: Учебное пособие/ ВолосатоваТ.М., Чичварин Н.В. - М.: НИЦ ИНФРА-М, 2019. - 196 с.: 60x90 1/16. - (Высшее образование: Бакалавриат) (Переплёт 7БЦ) ISBN 978-5-16-010977- <http://znanium.com/catalog.php?bookinfo=508097>
7. Онтологическая семантика. Знания. Онтологии. Онтологически ориентированные методы информационного анализа текстов [Электронный ресурс] / Рубашкин В.Ш. - М.: ФИЗМАТЛИТ, 2017. - <http://www.studentlibrary.ru/book/ISBN9785922114394.html>
8. Компьютерная лингвистика. Учебное пособие / Ю.Н. Марчук.- М.:Изд.АСТ, 2017. - 319 с.

### Дополнительная литература

1. Интернет-коммуникация как новая речевая формация / [С. И. Агаюлова, Е. Н. Галичкина, Е. И. Горошко и др.; науч. ред. Т.Н.Колокольцева, О.В. Лутовинова] – М.: Флинта: Наука, 2018.- 323 с. <https://lib.dvfu.ru:8443/lib/item?id=chamo:675726&theme=FEFU>
2. Информационные технологии в лингвистике / А.В. Зубов, И.И. Зубова. – М.: Академия, 2014. - 207с. <https://lib.dvfu.ru:8443/lib/item?id=chamo:245160&theme=FEFU>

3. Современные информационные технологии для гуманитария: практическое руководство / А.Т. Хроленко, А.В. Денисов. - М.: Флинта: Наука, 2018. - 128 с.  
<https://lib.dvfu.ru:8443/lib/item?id=chamo:308500&theme=FEFU>
4. Интернет-коммуникация как новая речевая формация / [Агагюлова С. И., Галичкина Е. Н., Горошко Е. И. и др.; научн. ред.: Т.Н.Колокольцева, О.В. Луговина] М.: Флинта Наука, 2019. - 323 с.  
<https://lib.dvfu.ru:8443/lib/item?id=chamo:675726&theme=FEFU>
5. Компьютерная лингвистика. Учебное пособие / Марчук Ю.Н. - М.:Изд.АСТ, 2017. - 317 с.
6. Корпусная лингвистика / Е.В. Грудева. - М.: ФЛИНТА, 2016. - 165 с.

### **Перечень ресурсов информационно-телекоммуникационной сети «Интернет»**

1. <http://window.edu.ru/resource/387/80387> Введение в компьютерную лингвистику. Учебное пособие. / К.К Боярский. - СПб: НИУ ИТМО, 2013. - 72 с.
2. <http://window.edu.ru/library/pdf2txt/465/78465/59324> Автоматическая обработка текстов на естественном языке и компьютерная лингвистика / Е.И. Большакова, Э.С.Клышинский, Д.В. Ландэ, А.А.Носков, О.В. Пескова, Е.В.- Ягунова М.: МИЭМ, 2011. -272с.
3. <http://prog-school.ru/> Школа программирования
4. <http://www.hardforum.ru/f22/> Форум программистов
5. <http://bib.com.ua/> ВІВлиотека программиста
6. Научный форум dxdy>> Моделирование русского языка  
<http://dxdy.ru/topic4210-60.html>
7. Лингвофорум >> Компьютерная лингвистика  
<http://lingvoforum.net/index.php?topic=29965.0>
8. Форум Диалог <http://forum.dialog-21.ru/actualthread.aspx?tid=363>
9. Научно-образовательный портал "Лингвистика в России: ресурсы для исследователей"  
[http://uisrussia.msu.ru/linguist/\\_B\\_comput\\_ling.jsp](http://uisrussia.msu.ru/linguist/_B_comput_ling.jsp)
10. Форум my TTS <http://mytts.forum2x2.ru/f45-forum>
11. Книги по лингвистике <http://www.antic-r.ru/bib15.htm>

### **Перечень информационных технологий и программного обеспечения**

1. Операционная система Windows.
2. Microsoft Office.

3. Интернет.

4. Лингвистические программы: Clear TextReader's Edition, Словогрыз, xReplacer; Russian Morphological Dictionary, Рабочее Место Лингвиста, Galaktika-ZOOM; Sakrament Text-to-Speech Engine v 2.0, Аки Росс, Письмовник; Neur OK Semantic Suite, ARM Engine 4.0, Следопыт 3.0; Гиперсловарь Ариадна, Словарь сокращений русского языка, Обратные и частотные словари русского языка; Гиперсловарь Ариадна, Словарь сокращений русского языка, Обратные и частотные словари русского языка.

При осуществлении образовательного процесса студентами и профессорско-преподавательским составом используются следующие информационно-справочные системы:

1. Научная электронная библиотека eLIBRARY.
2. Электронно-библиотечная система издательства «Лань».
3. Электронная библиотека "Консультант студента".
4. Электронно-библиотечная система IPRbooks.
5. Информационная система "ЕДИНОЕ ОКНО доступа к образовательным ресурсам".

6. Доступ к электронному заказу книг в библиотеке ДВФУ, доступ к нормативным документам ДВФУ, расписанию, рассылке писем.

Лекции проводятся с использованием проектора и мультимедийного комплекса для проведения лекций внутренней системы портала ДВФУ. Лабораторные занятия проводятся в специализированном компьютерном классе. Для составления документации используется текстовый процессор (Libre Office или Microsoft Word).

## **8.МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

Описание последовательности действий обучающихся, алгоритм изучения дисциплины:

- Систематизировать литературные источники,
- проанализировать и обобщить представленные в них концепции
- Из всего материала выделить существующие точки зрения на проблему,
- Проанализировать их, сравнить, дать им оценку.

Итогом этой работы должна стать логически выстроенная система сведений по существу исследуемого вопроса.

## 9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебные занятия по дисциплине могут проводиться в следующих помещениях, оснащенных соответствующим оборудованием и программным обеспечением, расположенных по адресу Приморский край, г. Владивосток, Фрунзенский р-н г., Русский Остров, ул. Аякс, п, д. 10:

Перечень материально-технического и программного обеспечения дисциплины приведен в таблице.

№ помещения по плану БТИ	Наименование специальных помещений и помещений для самостоятельной работы <sup>1</sup>	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
Учебные аудитории для проведения учебных занятий:			
	D208/347, D303, D313а, D401, D453, D461, D518, D708, D709, D758, D761, D762, D765, D766, D771, D917, D918, D920, D925, D576, D807	Лекционная аудитория оборудована маркерной доской, аудиопроигрывателем	ЗДЕСЬ ДОПОЛНИТСЯ ЛИЦЕНЗИОННЫМ ПО
	D229, D304, D306, D349, D350, D351, D352, D353, D403, D404, D405, D414, D434, D435, D453, D503, D504, D517, D522, D577, D578, D579, D580, D602, D603, D657, D658, D702, D704, D705, D707, D721, D722, D723, D735, D736, D764, D769, D770, D773, D810, D811, D906, D914, D921, D922, D923, D924, D926	Мультимедийная аудитория: Проектор Mitsubishi EW330U, Экран проекционный ScreenLineTrimWhiteIce, профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG, подсистема видеоисточников документ-камера CP355AF Avervision; подсистема видеокоммутации; подсистема аудиокоммутации и звукоусиления; подсистема интерактивного управления	
	D207/346	Мультимедийная аудитория: Проектор 3-chip DLP, 10 600 ANSI-лм, WUXGA 1 920x1 200 (16:10) PT-DZ110XE Panasonic; экран 316x500 см, 16:10 с эл. приводом; крепление настенно-потолочное ElproLargeElectrolProjecta; профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG; подсистема видеоисточников документ-камера CP355AF Avervision; подсистема видеокоммутации; подсистема аудиокоммутации и звукоусиления; подсистема	

<sup>1</sup> В соответствии с п.4.3.1 ФГОС

		интерактивного управления),	
	D226	Мультимедийная аудитория: Проектор Mitsubishi EW330U, Экран проекционный ScreenLineTrimWhiteIce, профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG, подсистема видеокоммутации; подсистема аудиокоммутации и звукоусиления; подсистема интерактивного управления), D362 (профессиональная ЖК- панель 47", 500 Кд/м2, Full HD M4716CCBA LG, подсистема аудиокоммутации и звукоусиления; Компьютерный класс на 15 посадочных мест	
	D447, D448, D449, D450, D451, D452, D502, D575	Мультимедийная аудитория: Проектор Mitsubishi EW330U, Экран проекционный ScreenLineTrimWhiteIce, подсистема видеоисточников документ-камера CP355AF Avervision; подсистема видеокоммутации; подсистема аудиокоммутации и звукоусиления; подсистема интерактивного управления	
	D446, D604, D656, D659, D737, D808, D809, D812	Мультимедийная аудитория: Проектор Mitsubishi EW330U, Экран проекционный ScreenLineTrimWhiteIce, профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG, подсистема видеоисточников документ-камера CP355AF Avervision; подсистема видеокоммутации; подсистема аудиокоммутации и звукоусиления; подсистема интерактивного управления; Компьютерный класс; Рабочее место: Компьютеры (Твердотельный диск - объемом 128 ГБ; Жесткий диск - объем 1000 ГБ; Форм-фактор – Tower); комплектуется клавиатурой, мышью. Монитором АОС i2757Fm; комплектом шнуров эл. питания) Модель - M93p 1; Лингафонный класс, компьютеры оснащены программным комплексом Sanakostudy 1200	
	D501, D601	Мультимедийная аудитория: Проектор Mitsubishi EW330U, Экран проекционный ScreenLineTrimWhiteIce, профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG, подсистема видеоисточников документ-камера CP355AF	

		<p>Avervision; подсистема видеокмутации; подсистема аудиокмутации и звукоусиления; подсистема интерактивного управления; Компьютерный класс на 26 рабочих мест. Рабочее место: Моноблок Lenovo C360G-i34164G500UDK</p>	
	Помещения для самостоятельной работы:		
	<p>A1042 аудитория для самостоятельной работы студентов</p>	<p>Моноблок Lenovo C360G-i34164G500UDK – 115 шт.; Интегрированный сенсорный дисплей PolymediaFlipBox; Копир-принтер-цветной сканер в e-mail с 4 лотками XeroxWorkCentre 5330 (WC5330C; Полноцветный копир-принтер-сканер XeroxWorkCentre 7530 (WC7530CPS Оборудование для инвалидов и лиц с ограниченными возможностями здоровья: Дисплей Брайля Focus-40 Blue – 3 шт.; Дисплей Брайля Focus-80 Blue; Рабочая станция LenovoThinkCentre E73z – 3 шт.; Видео увеличитель ONYX Swing-Arm PC edition; Маркер-диктофон TouchМемо цифровой; Устройство портативное для чтения плоскочечатных текстов PEarl; Сканирующая и читающая машина для незрячих и слабовидящих пользователей SARA; Принтер Брайля EmprintSpotDot - 2 шт.; Принтер Брайля Everest - D V4; Видео увеличитель ONYX Swing-Arm PC edition; Видео увеличитель Topaz 24" XL стационарный электронный; Обучающая система для детей тактильно-речевая, либо для людей с ограниченными возможностями здоровья; Увеличитель ручной видео RUBY портативный – 2 шт.; Экран Samsung S23C200B; Маркер-диктофон TouchМемо цифровой.</p>	<p>MicrosoftWindows 7 ProMAGic 12.0 Pro, JawsforWindows 15.0 Pro, Openbook 9.0, DuxburyBrailleTranslator, DolphinGuide (контракт № А238-14/2); Неисключительные права на использование ПО Microsoft рабочих станций пользователей (контракт ЭА-261-18 от 02.08.2018); - лицензия на клиентскую операционную систему; - лицензия на пакет офисных продуктов для работы с документами включая формат.docx , .xlsx , .vsd , .ppt.; - лицензия на право подключения пользователя к серверным операционным системам , используемым в ДВФУ : MicrosoftWindowsServer 2008/2012; - лицензия на право подключения к серверу MicrosoftExchangeServerEnterprise; - лицензия на право подключения к внутренней информационной системе документооборота и portalу с возможностью поиска информации во множестве удаленных и локальных хранилищах, ресурсах, библиотеках информации, включая порталные хранилища, используемой в ДВФУ: MicrosoftSharePoint; - лицензия на право подключения к системе централизованного управления рабочими станциями, используемой в ДВФУ: MicrosoftSystemCenter.</p>

## X. ФОНДЫ ОЦЕНОЧНЫХ СРЕДСТВ

Фонды оценочных средств представлены в Приложении.



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
**«Дальневосточный федеральный университет»**  
(ДВФУ)

---

**ВОСТОЧНЫЙ ИНСТИТУТ – ШКОЛА РЕГИОНАЛЬНЫХ И МЕЖДУНАРОДНЫХ  
ИССЛЕДОВАНИЙ**

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ**  
**по дисциплине «Обработка естественного языка и текстовая аналитика»**  
**Направление подготовки 45.04.01 Филология**  
магистерская программа «Европейские языки: коммуникация и  
лингвоаналитика»  
**Форма подготовки очная/ заочная**

**Владивосток**  
**2022**

**Перечень форм оценивания, применяемых на различных этапах формирования компетенций в ходе освоения дисциплины**

№ п/п	Контролируемые разделы / темы дисциплины	Код и наименование индикатора достижения компетенции	Результаты обучения	Оценочные средства	
				текущий контроль	промежуточная аттестация
	Тема №1, Знакомство с технологией NLP (Natural Language Processing).	УК-1.1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	Знает принципы разработки системного подхода	УО-1	УО-1 экзамен
			Умеет вырабатывать обоснованную стратегию в анализе лингвистических данных		
Владеет навыками критического рассмотрения эмпирического материала на естественном языке					
	Тема №2, Морфологический анализ текстов.	УК-4.1 Способен работать с аутентичными текстами академического и профессионального характера.	Знает как устроена современная электронная языковая среда, основные уровни обработки текста на ЕЯ и существующие модели статистического, морфологического и синтаксического анализа текстов;	УО-1	УО-1 экзамен
			Умеет проектировать процесс обработки языковых данных, работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;		
			Владеет навыками и методами поиска информации по работе с естественным языком;		
	Тема №3, Инструментальные системы разработки	ПК-1.2 Самостоятельно отбирает, систематизирует, анализирует материал	Знает особенности реализации естественно-языковых систем, назначение лингвистического процессора;	УО-1	УО-1 экзамен

	приложений по автоматической обработке текстов на естественном языке.	исследования	<p>Умеет эксплуатировать и сопровождать информационные системы и сервисы</p> <p>Владеет навыками методикой использования компьютерных технологий в предметной области;</p>		
	Тема №4, Извлечение информации из текстов: портрет направления.	ПК-1.5 Использует современные информационно-коммуникационные технологии в процессе выполнения научного проекта	<p>Знает лингвистические технологии в сфере АОТ; виды программной продукции АОТ;</p> <p>Умеет осуществлять и обосновывать выбор проектных решений по видам обеспечения информационных систем; понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;</p> <p>Владеет навыками вести исследовательскую лингвистическую работу с применением самых современных методов получения и анализа данных (корпусные методы, статистическая обработка, работа с экспертными оценками и т.д.);</p>	УО-1	УО-1 экзамен
	Тема №5, Текстовая аналитика. Мониторинг и анализ тональности.	ПК-5.2 Под руководством специалиста более высокой квалификации разрабатывает методические рекомендации для руководящих структур образовательных и иных организаций социо-гуманитарной направленности	<p>Знает ключевые направления в области компьютерной лингвистики; возможности использования программных технологий анализа и обработки текстовых данных;</p> <p>Умеет применять лингвистические технологии для проектирования систем анализа и синтеза естественного языка в сфере АОТ; использовать в работе и адаптировать под</p>	УО-1	УО-1 экзамен

		по гармонизации устного и письменного межкультурного профессионального общения, профилактики межкультурных конфликтов	конкретную задачу имеющиеся программные средства обработки текста и существующие в мире лингвистические ресурсы; Владеет навыками работы со специальными программными средствами АОТ;		
--	--	---	--	--	--

### Оценочные средства для текущего контроля

#### Вопросы для устного опроса

1. Автоматическая обработка звучащей речи.
2. Системы распознавания речи.
3. Механизмы распознавания речи.
4. Автоматический синтез речи.
5. Автоматическое индексирование.
6. Создание словарей, конкордансов, частотных словарей.
7. Законы Зипфа.
8. Автоматическое реферирование.
9. Автоматическое порождение текста.
10. Структура информационно-поискового тезауруса.
11. Построение информационно-поисковых тезаурусов.
12. «Мягкое» понимание текста. Семантической структуры.
13. Получение компрессированного текста.
14. Основные характеристики текста.

### Оценочные средства для промежуточной аттестации

Код и наименование индикатора компетенции	Результаты обучения	Шкала оценивания промежуточной аттестации			
		Неудовлетворительн о	Удовлетворительн о	Хорошо	Отлично

УК-1.1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	Знает	Мыслит не системно	Не знает методы критического мышления	Не в полной мере применяет методiku критического мышления и системного анализа	Знает принципы разработки системного подхода
	Умеет	Не умеет вырабатывать обоснованную стратегию в анализе лингвистических данных	Неуверенно умеет вырабатывать обоснованную стратегию в анализе лингвистических данных	Не в полной мере умеет вырабатывать обоснованную стратегию в анализе лингвистических данных	Умеет вырабатывать обоснованную стратегию в анализе лингвистических данных
	Владеет навыками	Не владеет навыками критического рассмотрения эмпирического материала на естественном языке	Не имеет уверенного владения навыками критического рассмотрения эмпирического материала на естественном языке	Не в полной мере навыками критического рассмотрения эмпирического материала на естественном языке	Владеет навыками критического рассмотрения эмпирического материала на естественном языке
УК-4.1 Способен работать с аутентичными текстами академического и профессионального характера.	Знает	Не знает как устроена современная электронная языковая среда, основные уровни обработки текста на ЕЯ и существующие модели статистического, морфологического и синтаксического анализа текстов;	Имеет общее представление как устроена современная электронная языковая среда, основные уровни обработки текста на ЕЯ и существующие модели статистического,	Имеет базовое понимание о том как устроена современная электронная языковая среда, основные уровни обработки текста на ЕЯ и существующие модели статистического, морфологического и	Знает как устроена современная электронная языковая среда, основные уровни обработки текста на ЕЯ и существующие модели статистического, морфологического и синтаксического

			морфологического и синтаксического анализа текстов;	синтаксического анализа текстов;	анализа текстов;
	Умеет	Не умеет проектировать процесс обработки языковых данных, работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;	Имеет неглубокое представление о том, как проектировать процесс обработки языковых данных, работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;	Имеет хорошую базу о том, как проектировать процесс обработки языковых данных, работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;	Умеет проектировать процесс обработки языковых данных, работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;
	Владеет навыками	Не владеет навыками и методами поиска информации по работе с естественным языком;	Слабо владеет навыками и методами поиска информации по работе с естественным языком;	В достаточном объеме владеет навыками и методами поиска информации по работе с естественным языком;	Владеет навыками и методами поиска информации по работе с естественным языком;
ПК-1.2 Самостоятельно отбирает,	Знает	Не знает особенности реализации	Слабо знает особенности	Достаточно уверенно знает особенности	Знает особенности реализации

систематизирует, анализирует материал исследования		естественно-языковых систем, назначение лингвистического процессора;	реализации естественно-языковых систем, назначение лингвистического процессора;	реализации естественно-языковых систем, назначение лингвистического процессора;	естественно-языковых систем, назначение лингвистического процессора;
	Умеет	Не умеет эксплуатировать и сопровождать информационные системы и сервисы	Не в полной мере умеет эксплуатировать и сопровождать информационные системы и сервисы	Уверенно умеет эксплуатировать и сопровождать информационные системы и сервисы	Умеет эксплуатировать и сопровождать информационные системы и сервисы
	Владеет навыками	Не владеет навыками методикой использования компьютерных технологий в предметной области;	Не в полной мере владеет навыками методикой использования компьютерных технологий в предметной области;	Уверенно владеет навыками методикой использования компьютерных технологий в предметной области;	Владеет навыками методикой использования компьютерных технологий в предметной области;
ПК-1.5 Использует современные информационно-коммуникационные технологии в процессе выполнения научного проекта	Знает	Не знает лингвистические технологии в сфере АОТ; виды программной продукции АОТ;	Неуверенно разбирается в теории лингвистических технологий в сфере АОТ; виды программной продукции АОТ;	Уверенно использует лингвистические технологии в сфере АОТ; виды программной продукции АОТ;	Знает лингвистические технологии в сфере АОТ; виды программной продукции АОТ;
	Умеет	Не умеет осуществлять и обосновывать выбор	Неуверенно умеет осуществлять и обосновывать	Хорошо разбирается какие именно данные нужны для решения	Умеет осуществлять и обосновывать

		проектных решений по видам обеспечения информационных систем; понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;	выбор проектных решений по видам обеспечения информационных систем; понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;	поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;	выбор проектных решений по видам обеспечения информационных систем; понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;
	Владеет навыками	Не владеет навыками вести исследовательскую лингвистическую работу с применением самых современных методов получения и анализа данных (корпусные методы, статистическая обработка, работа с экспертными оценками и т.д.);	Слабо владеет навыками вести исследовательскую лингвистическую работу с применением самых современных методов получения и анализа данных (корпусные методы, статистическая обработка, работа с экспертными оценками и т.д.);	Уверенно владеет навыками вести исследовательскую лингвистическую работу с применением самых современных методов получения и анализа данных (корпусные методы, статистическая обработка, работа с экспертными оценками и т.д.);	Владеет навыками вести исследовательскую лингвистическую работу с применением самых современных методов получения и анализа данных (корпусные методы, статистическая обработка, работа с экспертными оценками и т.д.);
ПК-5.2 Под руководством	Знает	Не знает ключевые направления в области	Слабо знает ключевые	Уверенно знает ключевые	Знает ключевые направления в

<p>специалиста более высокой квалификации разрабатывает методические рекомендации для руководящих структур образовательных и иных организаций социо-гуманитарной направленности по гармонизации устного и письменного межкультурного профессионального общения, профилактики межкультурных конфликтов</p>		<p>компьютерной лингвистики; возможности использования программных технологий анализа и обработки текстовых данных;</p>	<p>направления в области компьютерной лингвистики; возможности использования программных технологий анализа и обработки текстовых данных;</p>	<p>направления в области компьютерной лингвистики; возможности использования программных технологий анализа и обработки текстовых данных;</p>	<p>области компьютерной лингвистики; возможности использования программных технологий анализа и обработки текстовых данных;</p>
	<p>Умеет</p>	<p>Не умеет применять лингвистические технологии для проектирования систем анализа и синтеза естественного языка в сфере АОТ; использовать в работе и адаптировать под конкретную задачу имеющиеся программные средства обработки текста и существующие в мире лингвистические ресурсы;</p>	<p>Слабо умеет применять лингвистические технологии для проектирования систем анализа и синтеза естественного языка в сфере АОТ; использовать в работе и адаптировать под конкретную задачу имеющиеся программные средства обработки текста и существующие в мире лингвистические ресурсы;</p>	<p>Уверенно умеет применять лингвистические технологии для проектирования систем анализа и синтеза естественного языка в сфере АОТ; использовать в работе и адаптировать под конкретную задачу имеющиеся программные средства обработки текста и существующие в мире лингвистические ресурсы;</p>	<p>Умеет применять лингвистические технологии для проектирования систем анализа и синтеза естественного языка в сфере АОТ; использовать в работе и адаптировать под конкретную задачу имеющиеся программные средства обработки текста и существующие в мире лингвистические ресурсы;</p>

	Владеет навыками	Не владеет навыками навыками работы со специальными программными средствами АОТ;	Слабо владеет навыками навыками работы со специальными программными средствами АОТ;	Уверенно владеет навыками навыками работы со специальными программными средствами АОТ;	Владеет навыками навыками работы со специальными программными средствами АОТ;
--	------------------	--	---	--	---

Примерные вопросы к экзамену:

1. Понятие набора символов (character set, UCS) и кодировки (encoding).
2. Принципы кодирования на примерах ASCII, CP-1251, UTF-8, UTF-16.
3. Основные принципы автоматического определения кодировки.
4. Основные принципы автоматического определения языка текста.
5. Элементы регулярных выражений.
6. Детерминированные и недетерминированные конечные автоматы.
7. Распознавание регулярных выражений на основе конечного автомата.
8. Регулярные языки.
9. Организация словаря на основе конечных автоматов (acceptor) и преобразователей (transducer).
10. Принципы работы модуля морфологического анализа в проекте aot.ru. Стеммер Портера
11. Понятие корпуса, разделение на обучающий, тестовый, "разработческий"
12. Языковая модель - вероятностный подход.
13. Предположение Маркова. N-граммы.

14. Оценка качества языковой модели (perplexity).

Примерные задания:

1. Разработать программный комплекс для сбора базы данных текстов русскоязычных песен из сети Интернет и их анализа. Примеры анализа: - получить распределение самых частотных слов в зависимости от жанра песни. Исключить служебные слова. - получить распределение самых частотных слов из определенной категории в зависимости от жанра песни. Примеры категорий: упоминаний времени суток, сезонов года, частей тела, грамматических категорий и т.п. - получить распределение количества уникальных (т.е. богатства лексикона) в зависимости от жанра песни.
2. Разработать программный комплекс с поддержкой следующих функций: - первоначальное обучение языковой модели на большом корпусе; - генерация случайных предложений; - восстановление пропущенного слова в предложении; - восстановление правильного порядка в последовательности слов, полученной произвольной перестановкой в предложении.
3. Разработка комплекса правил снятия морфологической дизамбигуации в для русскоязычных текстов на базе технологии Apache Ruta (компонента платформы Apache UIMA).

#### **Критерии выставления оценки студенту на экзамене**

<b>Баллы</b> (рейтинговой оценки)	<b>Оценка зачета</b> (стандартная)	<b>Требования к сформированным компетенциям</b>
--------------------------------------	---------------------------------------	---

86-100	«зачтено»/ «отлично»	Оценка «зачтено» / «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, правильно обосновывает принятое решение, владеет разносторонними навыками и приемами выполнения практических задач.
76-85	«зачтено»/ «хорошо»	Оценка «зачтено» / «хорошо» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения.
61-75	«зачтено»/ «удовлетворительно»	Оценка «зачтено» / «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении практических работ.
0-60	«не зачтено»/ «неудовлетворительно»	Оценка «не зачтено» / «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет практические работы. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.