

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

Дата подписания: 18.08.2025 14:24:44

Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f022591d7e779870e51517f6d591

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

Международная школа искусств и культурного наследия

УТВЕРЖДАЮ:

Ректор

В.В. Волков

« 26 » *августа* 2025 г.

Протокол УС № *2* от *26.02* 2025 г.



Рабочая программа дисциплины
Инструменты обработки естественного языка

образовательная программа
направление подготовки
51.04.04 Музеология и охрана объектов культурного и природного наследия

направленность (профиль)
«Музейные исследования и кураторские стратегии»
программа подготовки – магистратура

язык обучения – русский
форма обучения – очная

квалификация выпускника
Магистр

Санкт-Петербург

Автор:

Котельников Е.В., доктор технических наук, профессор Школы вычислительных социальных наук АНООВО «ЕУСПб»

Рецензент

Басс В. Г., кандидат искусствоведения, доцент Международной школы искусств и культурного наследия АНООВО «ЕУСПб»

Рабочая программа дисциплины **«Инструменты обработки естественного языка»**, входящей в состав основной профессиональной образовательной программы высшего образования — программы магистратуры «Музейные исследования и кураторские стратегии», утверждена на заседании Совета Международной школы искусств и культурного наследия.

Протокол заседания № 7 от 13.02.2025 года.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ **«Инструменты обработки естественного языка»**

Дисциплина **«Инструменты обработки естественного языка»** является дисциплиной по выбору части, формируемой участниками образовательных отношений, образовательной программы, Блока 1 «Дисциплины (модули)» образовательной программы «Музейные исследования и кураторские стратегии» по направлению подготовки 51.04.04 Музеология и охрана объектов культурного и природного наследия.

Курс **«Инструменты обработки естественного языка»** знакомит магистрантов с основными методами и подходами к обработке естественного языка, в ходе изучения дисциплины проводится анализ принципов оценки качества методов обработки естественного языка, а также магистрантам предоставляется возможность овладеть практическими навыками обработки больших коллекций текстов.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточный контроль в форме зачета с оценкой (в конце 3 семестра).

Общая трудоемкость освоения дисциплины составляет 2 зачетных единицы, 72 часа.

Содержание

| | |
|---|----|
| 1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ | 5 |
| 2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ | 5 |
| 3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ | 6 |
| 4. ОБЪЕМ ДИСЦИПЛИНЫ | 6 |
| 5. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ | 7 |
| 5.1 Содержание дисциплины | 7 |
| 5.2 Структура дисциплины | 9 |
| 6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ | 9 |
| 6.1 Общие положения | 9 |
| 6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины | 10 |
| 6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине | 10 |
| 6.4 Перечень литературы для самостоятельной работы | 12 |
| 6.5 Перечень учебно-методического обеспечения для самостоятельной работы | 12 |
| 7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ | 12 |
| 7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации | 12 |
| 7.2 Контрольные задания для текущей аттестации | 15 |
| 7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации | 20 |
| 7.4 Типовые задания к промежуточной аттестации | 24 |
| 7.5 Средства оценки индикаторов достижения компетенций | 25 |
| 8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА | 26 |
| 8.1 Основная литература | 26 |
| 8.2 Дополнительная литература | 26 |
| 9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА | 26 |
| 9.1 Программное обеспечение | 27 |
| 9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины: | 27 |
| 9.3 Лицензионные электронные ресурсы библиотеки Университета | 27 |
| 9.4 Электронная информационно-образовательная среда Университета | 28 |
| 10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА | 28 |
| ПРИЛОЖЕНИЕ 1 | 30 |

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цели освоения дисциплины «Инструменты обработки естественного языка» - знакомство с основными методами и приложениями автоматической обработки естественного языка (ОЕЯ), получение практических навыков работы с инструментами ОЕЯ.

Задачи:

1. Знакомство с основными методами и подходами к ОЕЯ.
2. Понимание принципов оценки качества методов ОЕЯ.
3. Овладение практическими навыками обработки больших коллекций текстов.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В результате изучения учебной дисциплины обучающийся должен овладеть следующими компетенциями: универсальными (УК) и профессиональными (ПК). Планируемые результаты формирования компетенций и индикаторы их достижения в результате освоения дисциплины представлены в Таблице 1.

Таблица 1

Планируемые результаты освоения дисциплины, соотнесенные с индикаторами достижения компетенций обучающихся

| Код и наименование компетенции | Индикаторы достижения компетенции | Результаты обучения (знать, уметь, владеть) |
|---|---|--|
| УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий | ИД.УК-1.1. Знать методологию и методику системного и критического анализа проблемных ситуаций, стратегического управления ИД.УК-1.2. Уметь осуществлять системный и критический анализ проблемных ситуаций, вырабатывать стратегию действий ИД.УК-1.3. Уметь обосновывать, формулировать и решать задачи, возникающие в процессе профессиональной деятельности ИД.УК-1.4. Владеть методами системного и критического анализа, стратегического управления | Знать: методы научного познания, в основе которых лежит рассмотрение объекта как системы: целостного комплекса взаимосвязанных элементов, методы и модели стратегического планирования З (УК-1) |
| | | Уметь: с использованием методов системного подхода анализировать альтернативные варианты решения исследовательских задач, вырабатывать стратегию действий и оценивать эффективность реализации стратегических планов У (УК-1) |
| | | Владеть: целостной системой навыков методологического использования системного подхода при решении проблем, возникающих при выполнении исследовательских работ, навыками отстаивания своей точки зрения при выработке стратегических планов выполнения исследовательских работ В (УК-1) |
| ПК-3 Способен использовать современные методы обработки и интерпретации информации в профессиональной сфере | ИД.ПК-3.1. Знать современные методы накопления, обработки, передачи, поиска и использования информации о культурном и природном наследии ИД.ПК-3.2. Уметь вести результативный поиск информации с использованием современных информационно-коммуникационных технологий ИД.ПК-3.4. Уметь обрабатывать, анализировать и использовать | Знать: принципы и методы ведения самостоятельных научных исследований в профессиональной области и смежных областях З (ПК-3) |
| | | Уметь: выстраивать последовательную работу с информацией по актуальным проблемам сохранения культурного и природного наследия с использованием современных |

| Код и наименование компетенции | Индикаторы достижения компетенции | Результаты обучения (знать, уметь, владеть) |
|--------------------------------|--|--|
| | информацию в соответствии с научными и познавательными задачами ИД.ПК-3.5. Владеть приемами использования научной информации в научно-исследовательской и профессиональной деятельности ИД.ПК-3.6. Владеть основными способами поиска и представления информации | информационно-коммуникационных технологий У (ПК-3) Владеть: навыками самостоятельного проведения научных исследований в сфере профессиональных интересов В (ПК-3) |

В результате освоения дисциплины обучающийся должен:

знать: современные методологические принципы и методические приемы исторического исследования;

уметь: выявлять различия в методологических принципах и методических приемах исторического исследования; использовать на практике различные методики работы;

владеть: навыками применения современных методических приемов исторического исследования.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Инструменты обработки естественного языка» является дисциплиной по выбору части, формируемой участниками образовательных отношений, Блока 1 «Дисциплины (модули)» учебного плана основной профессиональной образовательной программы высшего образования — программы магистратуры «Музейные исследования и кураторские стратегии» по направлению подготовки 51.04.04 Музеология и охрана объектов культурного и природного наследия. Курс читается в третьем семестре, форма промежуточной аттестации — зачет с оценкой.

Для успешного освоения материала данной дисциплины требуются знания, умения и навыки, полученные в ходе изучения бакалавриата/специалитета.

Знания, умения и навыки, полученные при освоении данной дисциплины, применяются магистрантами в процессе выполнения научно-исследовательской работы и подготовки к защите и защиты выпускной квалификационной работы.

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 2 зачетных единиц, 72 часа.

Таблица 2

Объем дисциплины

| Типы учебных занятий и самостоятельная работа | | Объем дисциплины | | | | |
|---|-------|------------------|---------|---|-----------------|---|
| | | Всего | Семестр | | | |
| | | | 1 | 2 | 3 | 4 |
| Контактная работа обучающихся с преподавателем в соответствии с УП: | | 28 | - | - | 28 | - |
| Лекции (Л) | | 14 | - | - | 14 | - |
| Семинарские занятия (СЗ) | | 14 | - | - | 14 | - |
| Самостоятельная работа (СР) | | 44 | - | - | 44 | - |
| Промежуточная аттестация | форма | Зачет с оценкой | - | - | Зачет с оценкой | - |
| | час. | - | - | - | - | - |
| Общая трудоемкость дисциплины (час./з.е.) | | 72/2 | - | - | 72/2 | - |

5. 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) по средствам индикаторов достижения компетенций в соответствии с Таблицей 3.

5.1 Содержание дисциплины

Таблица 3

| Содержание дисциплины | | | | | |
|------------------------------|------------------------------------|---|-------------------------|--|--|
| № п/п | Наименование тем (разделов) | Содержание тем (разделов) | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) |
| 1 | Введение | Краткая история дисциплины, основные инструменты и приложения | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) |
| 2 | Морфологический анализ | Основные подходы, данные, инструменты. | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) |
| 3 | Языковые модели | Приложения, данные, оценки качества. Языковые модели на основе n-грамм, сглаживание. Нейронные языковые модели. | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) |
| 4 | Анализ тональности | Постановка задачи, приложения, данные, методы решения | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) |

| № п/п | Наименование тем (разделов) | Содержание тем (разделов) | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) |
|----------|--------------------------------|---|---------------------|--|--|
| 5 | Синтаксический анализ | Формализмы для представления синтаксической структуры, данные, подходы к решению, оценка качества. | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) |
| 6 | Извлечение информации | Именованные сущности, данные для обучения и тестирования. Методы на основе машинного обучения; методы на основе рекуррентных нейронных сетей. | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) |
| 7 | Автоматическое реферирование | Различные постановки задачи и методы, данные для обучения, автоматическое тестирование на основе метрики ROUGE, нейронные сети для автоматического реферирования. | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) |
| 8 | Вычислительная семантика | Семантические ресурсы (WordNet), дистрибутивные семантические модели, контекстуализированные векторные представления. | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) |
| 9 | Ответы на вопросы | Понимание прочитанного (reading comprehension) и ответы по базе знаний. Методы и данные. | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) |

| № п/п | Наименование тем (разделов) | Содержание тем (разделов) | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) |
|-------|-----------------------------|---|------------------|--|--|
| 10 | Машинный перевод (МП) | На основе правил, статистический МП, нейронный МП. Оценка качества (BLEU и другие метрики). | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) |

5.2 Структура дисциплины

Таблица 4

| Структура дисциплины | | | | | | |
|--------------------------|---------------------------------|------------------------|---|----|----|---|
| № п/п | Наименование тем (разделов) | Объем дисциплины, час. | | | | Форма текущего контроля успеваемости*, промежуточной аттестации |
| | | Всего | Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП | | СР | |
| | | | Л | СЗ | | |
| Очная форма обучения | | | | | | |
| Тема 1 | Введение | 7 | 1 | 2 | 4 | О, Д |
| Тема 2 | Морфологический анализ | 7 | 1 | 2 | 4 | КЗ, О |
| Тема 3 | Языковые модели | 7 | 2 | 1 | 4 | КЗ |
| Тема 4 | Анализ тональности | 7 | 2 | 1 | 4 | ПЗ |
| Тема 5 | Синтаксический анализ | 7 | 2 | 1 | 4 | КЗ, ПЗ |
| Тема 6 | Извлечение информации | 7 | 1 | 2 | 4 | ПЗ |
| Тема 7 | Автоматическое реферирование | 7 | 1 | 2 | 4 | КЗ |
| Тема 8 | Вычислительная семантика | 6 | 1 | 1 | 4 | ПЗ, Д |
| Тема 9 | Ответы на вопросы | 6 | 1 | 1 | 4 | ПЗ |
| Тема 10 | Машинный перевод (МП) | 11 | 2 | 1 | 8 | КЗ |
| Промежуточная аттестация | | - | - | - | - | Зачет с оценкой |
| ИТОГО: | | 72/2 | 14 | 14 | 44 | - |

*Примечание: формы текущего контроля успеваемости: диспут (Д), опрос (О), практическое задание (ПЗ), контрольное задание (КЗ).

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

6.1 Общие положения

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

Самостоятельная работа является важнейшей частью процесса высшего образования. Ее следует осознанно организовать, выделив для этого необходимое время и соответствующим образом организовав рабочее пространство. Важнейшим элементом самостоятельной работы является проработка материалов прошедших занятий (анализ конспектов, чтение рекомендованной литературы) и подготовка к следующим лекциям/семинарам. Литературу, рекомендованную в программе курса, следует, по

возможности, читать в течение всего семестра, концентрируясь на обусловленных программой курса темах.

Существенную часть самостоятельной работы магистранта представляет самостоятельное изучение учебно-методических изданий, лекционных конспектов, интернет-ресурсов и пр. Подготовка к семинарским занятиям, опросам также является важной формой работы магистранта. Самостоятельная работа может вестись как индивидуально, так и при содействии преподавателя.

6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины

Тема 1. Введение:

1.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.

1.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 2 часа. Итого: 4 часа.

Тема 2. Морфологический анализ:

2.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.

2.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 2 часа. Итого: 4 часа.

Тема 3. Языковые модели:

3.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.

3.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 2 часа. Итого: 4 часа.

Тема 4. Анализ тональности:

4.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.

4.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 2 часа. Итого: 4 часа.

Тема 5. Синтаксический анализ:

5.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.

5.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 2 часа. Итого: 4 часа.

Тема 6. Извлечение информации:

6.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.

6.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 2 часа. Итого: 4 часа.

Тема 7. Автоматическое реферирование:

7.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.

7.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 2 часа. Итого: 4 часа.

Тема 8. Вычислительная семантика:

8.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.

8.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 2 часа. Итого: 4 часа.

Тема 9. Ответы на вопросы:

9.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.

9.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 2 часа. Итого: 4 часа.

Тема 10. Машинный перевод (МП):

10.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 2 часа.

10.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 2 часа. Итого: 4 часа.

6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине

Вопросы для самостоятельной подготовки по темам дисциплины:

1. Исправление опечаток.
2. Распознавание звучащей речи.
3. Морфологический анализ: основные подходы.
4. Нейронные языковые модели.
5. Анализ тональности: постановка задачи, приложения, данные, методы решения.
6. Синтаксический анализ: формализмы для представления синтаксической структуры, данные, подходы к решению, оценка качества.
7. Методы извлечения информации на основе машинного обучения.

8. Нейронные сети для автоматического реферирования.
9. Дистрибутивные семантические модели.
10. Машинный перевод: методы оценки качества.

6.4 Перечень литературы для самостоятельной работы

1. 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>.
2. 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>
3. 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>
4. 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>

6.5 Перечень учебно-методического обеспечения для самостоятельной работы

Для обеспечения самостоятельной работы магистрантов по дисциплине «Инструменты обработки естественного языка» разработано учебно-методическое обеспечение в составе:

1. Контрольные задания для подготовки к процедурам текущего контроля (п. 7.2 Рабочей программы).
2. Типовые задания для подготовки к промежуточной аттестации (п. 7.4 Рабочей программы).
3. Рекомендуемые основная, дополнительная литература, Интернет-ресурсы и справочные системы (п. 8, 9 Рабочей программы).
4. Рабочая программа дисциплины размещена в электронной информационно-образовательной среде Университета на электронном учебно-методическом ресурсе АНООВО «ЕУСПб» — образовательном портале LMS Sakai — Sakai@EU.

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому лабораторному занятию, участие в опросах, диспутах, подготовку практических заданий, выполнение контрольных заданий, активное слушание на лекциях. Магистрант должен присутствовать на семинарских занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по обсуждаемым вопросам.

Текущий контроль проводится в форме устных опросов и оценивания участия магистрантов в проходящих диспутах, оценивания выполненных практических заданий, контрольных работ, демонстрирующих степень знакомства с дополнительной литературой.

**Показатели, критерии и оценивание компетенций и индикаторов их
достижения в процессе текущей аттестации**

| Наименование темы (раздела) | Код компетенц ии | Индикаторы компетенций | Коды ЗУВ (в соотв. с табл. 1) | Формы текущего контроля | Результаты текущего контроля |
|--------------------------------|------------------------|--|--|---|--|
| Введение | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Опрос 1 Диспут 1 | зачтено/ не зачтено зачтено/ не зачтено |
| Морфологический анализ | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Контрольное задание 1 Контрольное задание 2 Опрос 2 Контрольное задание 3 | зачтено/ не зачтено зачтено/ не зачтено зачтено/ не зачтено зачтено/ не зачтено |
| Языковые модели | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Контрольное задание 4 Контрольное задание 5 Контрольное задание 6 | зачтено/ не зачтено зачтено/ не зачтено зачтено/ не зачтено |
| Анализ тональности | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Практическо е задание 1 | зачтено/ не зачтено |
| Синтаксический анализ | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Контрольное задание 7 Контрольное задание 8 Практическо е задание 2 | зачтено/ не зачтено зачтено/ не зачтено зачтено/ не зачтено |
| Извлечение информации | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) | Практическо е задание 3 | зачтено/ не зачтено |

| Наименование темы (раздела) | Код компетенции | Индикаторы компетенций | Коды ЗУВ (в соотв. с табл. 1) | Формы текущего контроля | Результаты текущего контроля |
|--------------------------------|-----------------|--|--|--|--|
| | | ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | У (ПК-3) В (ПК-3) | | |
| Автоматическое реферирование | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Контрольное задание 9 | зачтено/ не зачтено |
| Вычислительная семантика | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Диспут 2 Практическое задание 4 | зачтено/ не зачтено зачтено/ не зачтено |
| Ответы на вопросы | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Практическое задание 5 | зачтено/ не зачтено зачтено/ не зачтено |
| Машинный перевод (МП) | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Контрольное задание 10 | зачтено/ не зачтено |

Таблица 6

Критерии оценивания

| Формы текущего контроля успеваемости | Критерии оценивания |
|--------------------------------------|---|
| Опрос | <p>Ответ отсутствует или является односложным, или содержит существенные ошибки – не зачтено</p> <p>Магистрант 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</p> |

| Формы текущего контроля успеваемости | Критерии оценивания |
|--------------------------------------|--|
| 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 | магистрант 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 |
| Контрольное задание | магистрант 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 |

7.2. Контрольные задания для текущей аттестации

Примерный материал опросов, диспутов, практических заданий:

Тема 1. Введение.

Опрос 1.

1. Какие основные инструменты и приложения используются при обработке естественного языка?
2. История развития методов обработки естественного языка?

Диспут 1.

Как технологии обработки естественного языка меняют повседневную жизнь людей.

Тема 2. Морфологический анализ.

Контрольное задание 1.

Три слова в списке имеют одинаковый тип словоизменения, одно — отличается. Отметьте это слово.

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

Контрольное задание 2.

Какие термины относятся к морфологической обработке? Выберите номера правильных ответов:

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

Опрос 2.

1. Каковы основные подходы к морфологическому анализу?
2. Какова основная методика отбора данных для морфологического анализа?
3. Каковы основные инструменты морфологического анализа?

Контрольное задание 3.

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 3. Языковые модели.

Контрольное 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>

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>.

Контрольное задание 5.

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).

Контрольное задание 6.

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 4. Анализ тональности.

Практическое задание 1

Реализуйте анализатор тональности одним из способов. Данные – английские предложений из отзывов о фильмах из Stanford Sentiment Treebank (из датасета взяты только целые предложения; 5-уровневая разметка приведена к трехуровневой – негативный, нейтральный, позитивный). Вы можете использовать подход на основе словаря тонально окрашенных слов (например, SentiWords) или обучить классификатор на тренировочных данных. Вы можете использовать тренировочные данные и в первом случае – чтобы подобрать пороги для классификации предложений на основе весов словаря. Постройте матрицу ошибок (confusion matrix) на тестовом наборе. Оцените правильность (ассурасу, доля правильно классифицированных предложений) классификатора на тестовом наборе. Проанализируйте неверно классифицированные предложения, сделайте предположения о причинах неверной классификации, предложите улучшения.

Тема 5. Синтаксический анализ.

Контрольное задание 7.

Выберите правильный разбор на составляющие предложения:

Советник губернатора Чукотского автономного округа Романа Абрамовича 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]]]]]]]]]]]

Контрольное задание 8.

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)

Практическое задание 2

Постройте деревья зависимостей для 2000 предложений «Войны и Мира» и «Братьев Карамазовых» с помощью библиотеки Stanza.

Оцените производительность библиотеки (предложения/с).

Проведите ручную оценку качества разбора на случайных 20 предложениях (по 10 из каждой книги).

Посчитайте среднюю глубину дерева разбора для каждого из романов. Посчитайте корреляцию между длиной предложения в словах и глубиной дерева разбора. Оцените на нескольких примерах, насколько глубина дерева соответствует субъективной сложности понимания предложения.

Тема 6. Извлечение информации.

Практическое задание 3

Примените и оцените модуль извлечения именованных сущностей Natasha (<https://github.com/natasha/natasha>). Для тестирования используйте текст и соответствующую разметку. Тестовые данные содержат только разметку для людей (PER) и организаций (ORG). Рассчитайте F1 для каждого типа сущностей и общее значение F1.

Тема 7. Автоматическое реферирование.

Контрольное задание 9.

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 8. Вычислительная семантика.

Диспут 2.

Возможные негативные последствия использования больших предобученных моделей для генерации текста.

Практическое задание 4

Реализуйте генератор юмора по мотивам работы Alessandro Valitutti et al. “Let Everything Turn Well in Your Wife”: Generation of Adult Humor Using Lexical Constraints, 2013. На входе генератора – новостной заголовок, в котором надо заменить одно слово. Предлагаемый алгоритм:

1. Проведите разбор предложения с помощью библиотеки Stanza. На основе результатов разбора выберете слово-кандидат на замену.
2. Найдите антоним для слова в WordNet (используйте интерфейс библиотеки NLTK).
3. Если антоним не нашелся, то найдите несколько слов, близких по звучанию или рифму с помощью datamuse api.
4. Получите вектора fasttext для начального слова и вариантов замены. Среди этих слов найдите самое далекое по косинусному расстоянию.

Оцените 20 модификаций по шкале от 0 (совсем не смешно) до 3 (очень смешно), приведите среднюю оценку.

Тема 9. Ответы на вопросы.

Практическое задание 5

Обучите модель для выделения ответа на вопрос из параграфа на тренировочной части русскоязычных данных TyDi QA. Оцените качество модели на данных для настройки (development set, EM/F1). Проанализируйте случаи, с которыми модель справилась хуже всего. Сделайте предположение, в чем сложность этих случаев.

Тема 10. Машинный перевод (МП).

Контрольное задание 10

Для фразы

Call me what instrument you will, though you can fret me, you cannot play upon me.

Есть два образцовых перевода (для простоты знаки препинания в образцах и вариантах перевода удалены):

- *назовите меня каким угодно инструментом вы хоть и можете меня терзать но играть на мне не можете*
- *объявите меня каким угодно инструментом вы можете расстроить меня но играть на мне нельзя*

Упорядочите варианты перевода ниже по убыванию BLEU-2 (метрика на основе униграмм и биграмм). Помните, что стандартная метрика BLEU не предполагает лемматизацию текстов.

- *позвони мне на каком инструменте вы будете хотя вы можете беспокоиться меня но вы не можете играть на мне*
- *назовите мне какой инструмент вы хотите хотя можете меня беспокоить но вы не можете играть на меня*
- *позвони мне какой инструмент ты будешь хотя ты можешь меня волновать но ты не можешь играть на меня*
- *назовите меня какой инструмент вы будете хотя вы можете раздражать меня все же вы не можете играть на меня*
- *считай меня чем тебе угодно ты можешь мучить меня но не играть мною*

• *назови меня каким угодно инструментом ты можешь меня расстроить но не играть на мне.*

7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации — зачет с оценкой, выставляемый на основе 2-х письменных работ (эссе).

Перед зачетом с оценкой проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Таблица 7

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

| Форма промежуточной аттестации/вид промежуточной аттестации | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) | Критерии оценивания | Оценка |
|---|------------------|--|--|---|------------------|
| Зачет с оценкой/ Письменная работа (эссе) | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Эссе соответствует следующим требованиям: 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., магистрант демонстрирует: глубокое усвоение программного материала, изложение его исчерпывающе, последовательно, четко, умение делать обоснованные выводы, соблюдение норм устной и письменной литературной речи/ Эссе успешно представлено на защите. Магистрант дает правильный ответ на теоретический вопрос, при условии, что отдельные неточности, допускаемые в ходе ответа, никак не снижают общего качества ответа. Для ответа характерно: • глубокое усвоение программного материала, • изложение его исчерпывающе, последовательно, четко, | Зачтено, отлично |

| Форма промежуточной аттестации/вид промежуточной аттестации | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) | Критерии оценивания | Оценка |
|---|------------------|---|--|--|----------------------------|
| | | | | <ul style="list-style-type: none"> • умение делать обоснованные выводы, • соблюдение норм устной и письменной литературной речи. | |
| | | | | <p>В эссе не соблюдены некоторые требования к работе: возможно несоблюдении одного-двух требований и допущении некоторых неточностей, магистрант демонстрирует: твердое знание материала курса, последовательное изложение материала, знание теоретических положений без обоснованной их аргументации, соблюдение норм устной и письменной литературной речи;</p> <p>Эссе успешно представлено на защите.</p> <p>Магистрант верно отвечает на вопрос, указанный в билете, при условии, что ответ на вопрос характеризуется отсутствием серьезных, значимых неточностей, при следующих характеристиках ответа:</p> <ul style="list-style-type: none"> • твердое знание материала курса, • последовательное изложение материала, • знание теоретических положений без обоснованной их аргументации, • соблюдение норм устной и письменной литературной речи. | Зачтено, хорошо |
| | | | | <p>Эссе содержит существенные оплошности: нарушено сразу несколько требований, например, выводы плохо обоснованы, есть фактические ошибки, магистрант при защите демонстрирует: знание основного материала, но владение им не в полном объеме,</p> | Зачтено, удовлетворительно |

| Форма промежуточной аттестации/вид промежуточной аттестации | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) | Критерии оценивания | Оценка |
|---|------------------|---|--|---|----------------------------------|
| | | | | <p>допущение существенных неточностей, недостаточно правильных формулировок, допущение нарушения логической последовательности в изложении материала, наличие нарушений норм литературной устной и письменной речи.</p> <p>Эссе представлено на защите.</p> <p>Магистрант представляет правильный ответ на теоретический вопрос, указанный в билете, при условии, что ответ на вопрос характеризуется значительными неточностями, при следующих параметрах ответа:</p> <ul style="list-style-type: none"> • знание основного материала, но владение им не в полном объеме, • допущение существенных неточностей, недостаточно правильных формулировок, • допущение нарушения логической последовательности в изложении материала, • наличие нарушений норм литературной устной и письменной речи. | |
| | | | | <p>Представленное эссе не отвечает предъявляемым требованиям (либо не предоставление эссе); имеет место:</p> <p>незнание значительной части программного материала, наличие существенных ошибок в определениях, формулировках, понимании теоретических положений;</p> <p>бессистемность при ответе на поставленный вопрос, отсутствие в ответе логически корректного анализа, аргументации, классификации,</p> <p>наличие нарушений норм устной и письменной литературной речи.</p> | Не зачтено, не удовлетворительно |

| Форма промежуточной аттестации/вид промежуточной аттестации | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) | Критерии оценивания | Оценка |
|---|------------------|---|--|--|--------|
| | | | | <p>Магистрант представляет ответ на вопрос, характеризующийся наличием существенных ошибок в определениях, формулировках, понимании теоретических положений, свидетельствующий о некомпетентности магистранта, при следующих параметрах ответа:</p> <ul style="list-style-type: none"> • незнание значительной части программного материала, • наличие существенных ошибок в определениях, формулировках, понимании теоретических положений; • бессистемность при ответе на поставленный вопрос, • отсутствие в ответе логически корректного анализа, аргументации, классификации, наличие нарушений норм устной и письменной литературной речи. | |

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются по стобалльной системе оценки в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 7а.

Таблица 7а

Система оценки знаний обучающихся

| Пятибалльная (стандартная) система | Стобалльная система оценки | Бинарная система оценки |
|------------------------------------|----------------------------|-------------------------|
| 5 (отлично) | 100-81 | зачтено |
| 4 (хорошо) | 80-61 | |
| 3 (удовлетворительно) | 60-41 | |
| 2 (неудовлетворительно) | 40 и менее | не зачтено |

Результаты промежуточного контроля по дисциплине, выраженные в оценках «зачтено, удовлетворительно», «зачтено, хорошо», «зачтено, отлично», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций основной профессиональной образовательной программы высшего образования — программы магистратуры «Музейные исследования и кураторские стратегии» по направлению подготовки 51.04.04 Музеология и охрана объектов культурного и природного наследия.

Результаты промежуточного контроля по дисциплине, выраженные в оценках «не зачтено, неудовлетворительно», показывают несформированность у обучающегося

компетенций по дисциплине в соответствии с картами компетенций основной профессиональной образовательной программы высшего образования — программы магистратуры «Музейные исследования и кураторские стратегии» по направлению подготовки 51.04.04 Музеология и охрана объектов культурного и природного наследия.

7.4 Типовые задания к промежуточной аттестации

УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий

В рамках выбранной темы магистранту необходимо показать наличие знаний и умения применять различные методологии и методики системного и критического анализа проблемных ситуаций, решение которых возможно по средствам применения инструментов обработки естественного языка, для чего необходимо смоделировать ситуацию, возможную в профессиональной деятельности, сформулировать, обосновать и решить задачи, которые характерны для данной ситуации, по средствам соответствующих инструментов обработки естественного языка.

Примерный перечень тем 1-й письменной работы (эссе):

1. Сравнение качества и производительности морфологических библиотек для русского языка.
2. Сравнение качества и производительности библиотек анализа тональности для русского языка.
3. Сравнение качества и производительности библиотек для выделения именованных сущностей для русского языка.
4. Сравнение качества онлайн сервисов машинного перевода для пары русский-английский (одно направление).
5. Сравнительный анализ произведений Льва Толстого и Федора Достоевского с помощью стилометрических методов.
6. Гендерное смещение (gender bias) в дистрибутивных моделях русского языка (проанализировать 2-3 статические модели отсюда: <https://rusvectors.org/ru/models/>).
7. Сравнительный анализ двух моделей вопросно-ответного поиска для русского языка с помощью инструмента CheckList (<https://github.com/marcotcr/checklist>).
8. Анализ качества кросс-языкового переноса моделей вопросно-ответного поиска на данных без дообучения. Исходный английский набор данных – SquAD, тестирование – на данных TyDi QA.
9. Исследование переносимости моделей распознавания юмора: исследовать 2-3 метода классификации на 2-3 англоязычных наборах данных.
10. Анализ существующих систем вопросно-ответного поиска по базам знаний (deeppavlov, Qanswer) с помощью тестового набора данных RuBQ.
11. Систематический анализ качества генерации текстов с помощью модели ruGPT-3 для различных сценариев (<https://github.com/sberbank-ai/ru-gpts>).

ПК-3 Способен использовать современные методы обработки и интерпретации информации в профессиональной сфере

В рамках выбранной темы магистранту необходимо показать наличие знаний современных методов обработки изображений в области исследований культурного и природного наследия, последовательно описать процедуру применения различных методов и инструментов обработки изображений на примере конкретного научно-исследовательского проекта или возможной ситуации в работе музея по выбору магистранта.

Примерный перечень тем 2-й письменной работы:

1. Методы обработки естественного языка.
2. Машинный перевод.
3. Голосовые помощники.

4. Анализ текстов.
5. Распознавание и синтез речи.
6. Обработка естественного языка на Java.
7. Обработка естественного языка в контексте Data Science фреймворков.
8. Обработка естественного языка с TensorFlow.
9. Глубокое обучение при обработке естественного языка.
10. NLP на Python.
11. Интеграция Spark и библиотек машинного обучения.
12. Нейросетевые методы в обработке естественного языка.
13. Анализ моделей, основанных на механизме внимания и архитектуре Transformer.
14. Анализ задач музейной работы, которые может решить NLP.
15. Предобработка текста.
16. Стемминг.
17. Лемматизация.
18. Векторизация.
19. Дедубликация.
20. Семантический анализ.
21. Распознавание именованных сущностей и извлечение отношений.
22. Использование N-грамм.
23. Частеречная разметка.
24. Библиотеки для NLP.
25. Анализ примера кода на языке Scala.
26. Обзор подходов и методов к задаче автоматического извлечения именованных сущностей.

7.5 Средства оценки индикаторов достижения компетенций

Таблица 8

Средства оценки индикаторов достижения компетенций

| Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Средства оценки (в соотв. с Таблицами 5, 7) |
|------------------|--|--|
| УК-1 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. | Опрос, диспут, практическое задание, контрольное задание, письменная работа (эссе) |
| ПК-3 | ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | Опрос, диспут, практическое задание, контрольное задание, письменная работа (эссе) |

Таблица 9

Описание средств оценки индикаторов достижения компетенций

| Средства оценки (в соотв. с Таблицами 5, 7) | Рекомендованный план выполнения работы |
|---|---|
| Опрос | Магистрант в ходе подготовки и участия в опросе показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — анализирует проблемную ситуацию, определяет пробелы в информации, оценивать надёжность источников информации, разрабатывает стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения; — применяет современные методы, поиска, обработки, анализа и использования информации в рамках проведения научно-исследовательских и организационных работ в области музейных исследований и кураторских стратегий. |

| Средства оценки (в соотв. с Таблицами 5, 7) | Рекомендованный план выполнения работы |
|--|--|
| Диспут | <p>Магистрант в ходе подготовки и участия в диспуте показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности:</p> <ul style="list-style-type: none"> — анализирует проблемную ситуацию, определяет пробелы в информации, оценивать надёжность источников информации, разрабатывает стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения; — применяет современные методы, поиска, обработки, анализа и использования информации в рамках проведения научно-исследовательских и организационных работ в области музейных исследований и кураторских стратегий. |
| Практическое задание | <p>Магистрант в ходе подготовки и выполнения практического задания показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности:</p> <ul style="list-style-type: none"> — анализирует проблемную ситуацию, определяет пробелы в информации, оценивать надёжность источников информации, разрабатывает стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения; — применяет современные методы, поиска, обработки, анализа и использования информации в рамках проведения научно-исследовательских и организационных работ в области музейных исследований и кураторских стратегий. |
| Контрольное задание | <p>Магистрант в ходе подготовки и выполнения контрольного задания показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности:</p> <ul style="list-style-type: none"> — анализирует проблемную ситуацию, определяет пробелы в информации, оценивать надёжность источников информации, разрабатывает стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения; — применяет современные методы, поиска, обработки, анализа и использования информации в рамках проведения научно-исследовательских и организационных работ в области музейных исследований и кураторских стратегий. |
| Письменная работа (эссе) | <p>Магистрант в ходе подготовки и написания письменной работы (эссе), показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности:</p> <ul style="list-style-type: none"> — анализирует проблемную ситуацию, определяет пробелы в информации, оценивать надёжность источников информации, разрабатывает стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения; — применяет современные методы, поиска, обработки, анализа и использования информации в рамках проведения научно-исследовательских и организационных работ в области музейных исследований и кураторских стратегий. |

8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

8.1 Основная литература

1. Бутусова, А. С. Машинный и автоматизированный перевод : учебное пособие / А. С. Бутусова, Ю. В. Бец ; Южный федеральный университет. - Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2021. - 106 с. - ISBN 978-5-9275-3982-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2057597> . – Режим доступа: по подписке.

8.2 Дополнительная литература

1 Богданов, Е. П. Интеллектуальный анализ данных : практикум для магистрантов направления 09.04.03 «Прикладная информатика» профиль подготовки «Информационные системы и технологии корпоративного управления» / Е. П. Богданов. - Волгоград : ФГБОУ ВО Волгоградский ГАУ, 2019. - 112 с. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1087885> . – Режим доступа: по подписке

9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

9.1 Программное обеспечение

При осуществлении образовательного процесса магистрантами и профессорско-преподавательским составом используется следующее лицензионное программное обеспечение:

1. ABBYY FineReader 11 Corporate Edition
2. ABBYY Lingvo x5
3. Adobe Acrobat Professional 11.0 MLP AOO License RU
4. Adobe CS5.5 Design Standart Win IE EDU CLP
5. Adobe Acrobat Reader – бесплатно
6. Git (версия 2.40 и выше)
7. Google Chrome
8. Mozilla – бесплатно
9. MS Office (OVS Office Platform)
10. Opera – бесплатно
11. OS Microsoft Windows (OVS OS Platform)
12. VLC – бесплатно
13. Яндекс.Браузер (Yandex Browser) – бесплатно

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

Информационно-справочные системы

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

Профессиональные базы данных информационно-телекоммуникационной сети «Интернет»:

1. Национальная электронная библиотека НЭБ: <http://www.rusneb.ru>
2. Неприкосновенный запас: <http://magazines.russ.ru/nz/>
3. Президентская библиотека: <http://www.prilib.ru>
4. Российская государственная библиотека: <http://www.rsl.ru/>
5. Российская национальная библиотека: <http://www.nlr.ru/poisk/>

9.3 Лицензионные электронные ресурсы библиотеки Университета

Профессиональные базы данных:

Полный перечень доступных обучающимся профессиональных баз данных представлен на официальном сайте Университета <https://eusp.org/library/electronic-resources>, включая следующие базы данных:

1. **eLIBRARY.RU** — Российский информационно-аналитический портал в

области науки, технологии, медицины и образования, содержащий рефераты и полные тексты научных статей и публикаций, наукометрическая база данных: <http://elibrary.ru>;

2. Электронные журналы по подписке (текущие номера научных зарубежных журналов).

Электронные библиотечные системы:

1. **Znaniy.com** – Электронная библиотечная система (ЭБС) – <http://znaniy.com/>;

2. Университетская библиотека онлайн – Электронная библиотечная система (ЭБС) – <http://biblioclub.ru/>

9.4 Электронная информационно-образовательная среда Университета

Образовательный процесс по дисциплине поддерживается средствами электронной информационно-образовательной среды Университета, которая включает в себя электронный учебно-методический ресурс АНООВО «ЕУСПб» — образовательный портал LMS Sakai — Sakai@EU, лицензионные электронные ресурсы библиотеки Университета, официальный сайт Университета (Европейский университет в Санкт-Петербурге [<https://euspb.org/>]), локальную сеть Университета и корпоративную электронную почту и обеспечивает:

— доступ к учебным планам, рабочим программам дисциплин (модулей), практик и к изданиям электронных библиотечных систем и электронным образовательным ресурсам, указанным в рабочих программах;

— фиксацию хода образовательного процесса, результатов промежуточной аттестации и результатов освоения основной образовательной программы;

— формирование электронного портфолио обучающегося, в том числе сохранение работ обучающегося, рецензий и оценок за эти работы со стороны любых участников образовательного процесса;

— взаимодействие между участниками образовательного процесса, в том числе синхронное и (или) асинхронное взаимодействие посредством сети «Интернет» (электронной почты и т.д.).

Каждый обучающийся в течение всего периода обучения обеспечен индивидуальным неограниченным доступом к электронным ресурсам библиотеки Университета, содержащей издания учебной, учебно-методической и иной литературы по изучаемой дисциплине.

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

В ходе реализации образовательного процесса используются специализированные многофункциональные аудитории для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Проведение занятий лекционного типа обеспечивается демонстрационным оборудованием.

Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Для лиц с ограниченными возможностями здоровья и инвалидов предоставляется возможность присутствия в аудитории вместе с ними ассистента (помощника). Для слабовидящих предоставляется возможность увеличения текста на экране (ПК). Для самостоятельной работы лиц с ограниченными возможностями здоровья в помещении для самостоятельной работы организовано одно место (ПК) с возможностями

бесконтактного ввода информации и управления компьютером (специализированное лицензионное программное обеспечение – Camera Mouse, веб камера). Библиотека университета предоставляет удаленный доступ к ЭБ с возможностями для слабовидящих увеличения текста на экране ПК. Лица с ограниченными возможностями здоровья могут при необходимости воспользоваться имеющимся в университете креслом-коляской. В учебном корпусе имеется адаптированный лифт. На первом этаже оборудован специализированный туалет. У входа в здание университета для инвалидов оборудована специальная кнопка, входная среда обеспечена информационной доской о режиме работы университета, выполненной рельефно-точечным тактильным шрифтом (азбука Брайля).

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ
«Инструменты обработки естественного языка»

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому лабораторному занятию, участие в опросах, диспутах, подготовку практических заданий, выполнение контрольных заданий, активное слушание на лекциях. Магистрант должен присутствовать на семинарских занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по обсуждаемым вопросам.

Текущий контроль проводится в форме устных опросов и оценивания участия магистрантов в проходящих диспутах, оценивания выполненных практических заданий, контрольных работ, демонстрирующих степень знакомства с дополнительной литературой.

Таблица 1

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

| Наименование темы (раздела) | Код компетенции | Индикаторы компетенций | Коды ЗУВ (в соотв. с табл. 1) | Формы текущего контроля | Результаты текущего контроля |
|-----------------------------|-----------------|--|--|--|--|
| Введение | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Опрос 1 Диспут 1 | зачтено/ не зачтено зачтено/ не зачтено |
| Морфологический анализ | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Контрольное задание 1 Контрольное задание 2 Опрос 2 Контрольное задание 3 | зачтено/ не зачтено зачтено/ не зачтено зачтено/ не зачтено зачтено/ не зачтено |
| Языковые модели | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Контрольное задание 4 Контрольное задание 5 Контрольное задание 6 | зачтено/ не зачтено зачтено/ не зачтено зачтено/ не зачтено |
| Анализ тональности | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. | З (УК-1) У (УК-1) | Практическое задание 1 | зачтено/ не зачтено |

| Наименование темы (раздела) | Код компетенц ии | Индикаторы компетенций | Коды ЗУВ (в соотв. с табл. 1) | Формы текущего контроля | Результаты текущего контроля |
|---------------------------------|------------------------|--|--|--|--|
| | | ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | | |
| Синтаксический анализ | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Контрольное задание 7 Контрольное задание 8 Практическо е задание 2 | зачтено/ не зачтено зачтено/ не зачтено зачтено/ не зачтено |
| Извлечение информации | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Практическо е задание 3 | зачтено/ не зачтено |
| Автоматическое реферирование | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Контрольное задание 9 | зачтено/ не зачтено |
| Вычислительная семантика | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Диспут 2 Практическо е задание 4 | зачтено/ не зачтено зачтено/ не зачтено |
| Ответы на вопросы | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Практическо е задание 5 | зачтено/ не зачтено зачтено/ не зачтено |
| Машинный перевод (МП) | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. | З (УК-1) У (УК-1) | Контрольное задание 10 | зачтено/ не зачтено |

| Наименование темы (раздела) | Код компетенции | Индикаторы компетенций | Коды ЗУВ (в соотв. с табл. 1) | Формы текущего контроля | Результаты текущего контроля |
|-----------------------------|-----------------|--|--|-------------------------|------------------------------|
| | | ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | | |

Таблица 2

Критерии оценивания

| Формы текущего контроля успеваемости | Критерии оценивания |
|--------------------------------------|---|
| Опрос | <p>Ответ отсутствует или является односложным, или содержит существенные ошибки – не зачтено</p> <p>Магистрант 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</p> |
| Диспут | <p>Пассивность, участие без представления аргументов и обоснования точки зрения, несформированность навыков профессиональной коммуникации в группе — не зачтено</p> <p>Представление аргументированной научной позиции, обоснование точки зрения в диспуте, демонстрация навыков профессиональной коммуникации в группе — зачтено</p> |
| Практическое задание | <p>магистрант 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,</p> <p>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</p> |
| Контрольное задание | <p>магистрант 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,</p> <p>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</p> |

2. Контрольные задания для текущей аттестации

Материалы опросов, диспутов, практических заданий:

Тема 1. Введение.

Опрос 1.

1. Какие основные инструменты и приложения используются при обработке естественного языка?
2. История развития методов обработки естественного языка?

Диспут 1.

Как технологии обработки естественного языка меняют повседневную жизнь людей.

Тема 2. Морфологический анализ.

Контрольное задание 1.

Три слова в списке имеют одинаковый тип словоизменения, одно – отличается.
Отметьте это слово.

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

Контрольное задание 2.

Какие термины относятся к морфологической обработке? Выберите номера правильных ответов:

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

Опрос 2.

1. Каковы основные подходы к морфологическому анализу?
2. Какова основная методика отбора данных для морфологического анализа?
3. Каковы основные инструменты морфологического анализа?

Контрольное задание 3.

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 3. Языковые модели.

Контрольное 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>

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>.

Контрольное задание 5.

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).

Контрольное задание 6.

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 4. Анализ тональности.

Практическое задание 1

Реализуйте анализатор тональности одним из способов. Данные – английские предложения из отзывов о фильмах из Stanford Sentiment Treebank (из датасета взяты только целые предложения; 5-уровневая разметка приведена к трехуровневой – негативный, нейтральный, позитивный). Вы можете использовать подход на основе словаря тонально окрашенных слов (например, SentiWords) или обучить классификатор на тренировочных данных. Вы можете использовать тренировочные данные и в первом случае – чтобы подобрать пороги для классификации предложений на основе весов словаря. Постройте матрицу ошибок (confusion matrix) на тестовом наборе. Оцените правильность (accuracy, доля правильно классифицированных предложений) классификатора на тестовом наборе. Проанализируйте неверно классифицированные предложения, сделайте предположения о причинах неверной классификации, предложите улучшения.

Тема 5. Синтаксический анализ.

Контрольное задание 7.

Выберите правильный разбор на составляющие предложения:

Советник губернатора Чукотского автономного округа Романа Абрамовича 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]]]]]]]]]

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

Контрольное задание 8.

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)

Практическое задание 2

Постройте деревья зависимостей для 2000 предложений «Войны и Мира» и «Братьев Карамазовых» с помощью библиотеки Stanza.

Оцените производительность библиотеки (предложения/с).

Проведите ручную оценку качества разбора на случайных 20 предложениях (по 10 из каждой книги).

Посчитайте среднюю глубину дерева разбора для каждого из романов. Посчитайте корреляцию между длиной предложения в словах и глубиной дерева разбора. Оцените на нескольких примерах, насколько глубина дерева соответствует субъективной сложности понимания предложения.

Тема 6. Извлечение информации.

Практическое задание 3

Примените и оцените модуль извлечения именованных сущностей Natasha (<https://github.com/natasha/natasha>). Для тестирования используйте текст и соответствующую разметку. Тестовые данные содержат только разметку для людей (PER) и организаций (ORG). Рассчитайте F1 для каждого типа сущностей и общее значение F1.

Тема 7. Автоматическое реферирование.

Контрольное задание 9.

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 8. Вычислительная семантика.

Диспут 2.

Возможные негативные последствия использования больших предобученных моделей для генерации текста.

Практическое задание 4

Реализуйте генератор юмора по мотивам работы Alessandro Valitutti et al. “Let Everything Turn Well in Your Wife”: Generation of Adult Humor Using Lexical Constraints, 2013. На входе генератора – новостной заголовок, в котором надо заменить одно слово. Предлагаемый алгоритм:

1. Проведите разбор предложения с помощью библиотеки Stanza. На основе результатов разбора выберете слово-кандидат на замену.
2. Найдите антоним для слова в WordNet (используйте интерфейс библиотеки NLTK).
3. Если антоним не нашелся, то найдите несколько слов, близких по звучанию или рифму с помощью datamuse api.
4. Получите вектора fasttext для начального слова и вариантов замены. Среди этих слов найдите самое далекое по косинусному расстоянию.

Оцените 20 модификаций по шкале от 0 (совсем не смешно) до 3 (очень смешно), приведите среднюю оценку.

Тема 9. Ответы на вопросы.

Практическое задание 5

Обучите модель для выделения ответа на вопрос из параграфа на тренировочной части русскоязычных данных TyDi QA. Оцените качество модели на данных для настройки (development set, EM/F1). Проанализируйте случаи, с которыми модель справилась хуже всего. Сделайте предположение, в чем сложность этих случаев.

Тема 10. Машинный перевод (МП).

Контрольное задание 10

Для фразы

Call me what instrument you will, though you can fret me, you cannot play upon me.

Есть два образцовых перевода (для простоты знаки препинания в образцах и вариантах перевода удалены):

- *назовите меня каким угодно инструментом вы хоть и можете меня терзать но играть на мне не можете*

- *объявите меня каким угодно инструментом вы можете расстроить меня но играть на мне нельзя*

Упорядочите варианты перевода ниже по убыванию BLEU-2 (метрика на основе униграмм и биграмм). Помните, что стандартная метрика BLEU не предполагает лемматизацию текстов.

- *позвони мне на каком инструменте вы будете хотя вы можете беспокоиться меня но вы не можете играть на мне*
- *назовите мне какой инструмент вы хотите хотя можете меня беспокоить но вы не можете играть на меня*
- *позвони мне какой инструмент ты будешь хотя ты можешь меня волновать но ты не можешь играть на меня*
- *назовите меня какой инструмент вы будете хотя вы можете раздражать меня все же вы не можете играть на меня*
- *считай меня чем тебе угодно ты можешь мучить меня но не играть мною*
- *назови меня каким угодно инструментом ты можешь меня расстроить но не играть на мне.*

3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации — **зачет с оценкой**, выставляемый на основе 2-х письменных работ (эссе).

Перед зачетом с оценкой проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Таблица 3

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

| Форма промежуточной аттестации/вид промежуточной аттестации | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) | Критерии оценивания | Оценка |
|---|------------------|--|--|---|------------------|
| Зачет с оценкой/ Письменная работа (эссе) | УК-1 ПК-3 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. ИД.ПК-3.5. ИД.ПК-3.6. | З (УК-1) У (УК-1) В (УК-1) З (ПК-3) У (ПК-3) В (ПК-3) | Эссе соответствует следующим требованиям: 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., магистрант демонстрирует: глубокое усвоение программного материала, изложение его исчерпывающе, последовательно, четко, умение делать обоснованные выводы, | Зачтено, отлично |

| Форма промежуточной аттестации/вид промежуточной аттестации | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) | Критерии оценивания | Оценка |
|---|------------------|---|--|--|-----------------|
| | | | | <p>соблюдение норм устной и письменной литературной речи/ Эссе успешно представлено на защите. Магистрант дает правильный ответ на теоретический вопрос, при условии, что отдельные неточности, допускаемые в ходе ответа, никак не снижают общего качества ответа. Для ответа характерно:</p> <ul style="list-style-type: none"> • глубокое усвоение программного материала, • изложение его исчерпывающе, последовательно, четко, • умение делать обоснованные выводы, • соблюдение норм устной и письменной литературной речи. | |
| | | | | <p>В эссе не соблюдены некоторые требования к работе: возможно несоблюдении одного-двух требований и допущении некоторых неточностей, магистрант демонстрирует: твердое знание материала курса, последовательное изложение материала, знание теоретических положений без обоснованной их аргументации, соблюдение норм устной и письменной литературной речи; Эссе успешно представлено на защите. Магистрант верно отвечает на вопрос, указанный в билете, при условии, что ответ на вопрос характеризуется отсутствием серьезных, значимых неточностей, при следующих характеристиках ответа:</p> <ul style="list-style-type: none"> • твердое знание материала курса, • последовательное изложение материала, | Зачтено, хорошо |

| Форма промежуточной аттестации/вид промежуточной аттестации | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) | Критерии оценивания | Оценка |
|---|------------------|---|--|--|----------------------------|
| | | | | <ul style="list-style-type: none"> • знание теоретических положений без обоснованной их аргументации, • соблюдение норм устной и письменной литературной речи. | |
| | | | | <p>Эссе содержит существенные оплошности: нарушено сразу несколько требований, например, выводы плохо обоснованы, есть фактические ошибки, магистрант при защите демонстрирует:</p> <ul style="list-style-type: none"> • знание основного материала, но владение им не в полном объеме, • допущение существенных неточностей, недостаточно правильных формулировок, • допущение нарушения логической последовательности в изложении материала, • наличие нарушений норм литературной устной и письменной речи. <p>Эссе представлено на защите.</p> <p>Магистрант представляет правильный ответ на теоретический вопрос, указанный в билете, при условии, что ответ на вопрос характеризуется значительными неточностями, при следующих параметрах ответа:</p> <ul style="list-style-type: none"> • знание основного материала, но владение им не в полном объеме, • допущение существенных неточностей, недостаточно правильных формулировок, • допущение нарушения логической последовательности в изложении материала, • наличие нарушений норм литературной устной и письменной речи. | Зачтено, удовлетворительно |

| Форма промежуточной аттестации/вид промежуточной аттестации | Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Коды ЗУВ (в соответствии с Таблицей 1) | Критерии оценивания | Оценка |
|---|------------------|---|--|--|----------------------------------|
| | | | | <p>Представленное эссе не отвечает предъявляемым требованиям (либо не предоставление эссе); имеет место:</p> <ul style="list-style-type: none"> незнание значительной части программного материала, наличие существенных ошибок в определениях, формулировках, понимании теоретических положений; бессистемность при ответе на поставленный вопрос, отсутствие в ответе логически корректного анализа, аргументации, классификации, наличие нарушений норм устной и письменной литературной речи. <p>Магистрант представляет ответ на вопрос, характеризующийся наличием существенных ошибок в определениях, формулировках, понимании теоретических положений, свидетельствующий о некомпетентности магистранта, при следующих параметрах ответа:</p> <ul style="list-style-type: none"> • незнание значительной части программного материала, • наличие существенных ошибок в определениях, формулировках, понимании теоретических положений; • бессистемность при ответе на поставленный вопрос, • отсутствие в ответе логически корректного анализа, аргументации, классификации, наличие нарушений норм устной и письменной литературной речи. | Не зачтено, не удовлетворительно |

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются по столбальной системе оценки в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 3а.

Система оценки знаний обучающихся

| Пятибалльная (стандартная) система | Стобалльная система оценки | Бинарная система оценки |
|------------------------------------|----------------------------|-------------------------|
| 5 (отлично) | 100-81 | зачтено |
| 4 (хорошо) | 80-61 | |
| 3 (удовлетворительно) | 60-41 | |
| 2 (неудовлетворительно) | 40 и менее | не зачтено |

Результаты промежуточного контроля по дисциплине, выраженные в оценках «зачтено, удовлетворительно», «зачтено, хорошо», «зачтено, отлично», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций основной профессиональной образовательной программы высшего образования — программы магистратуры «Музейные исследования и кураторские стратегии» по направлению подготовки 51.04.04 Музеология и охрана объектов культурного и природного наследия.

Результаты промежуточного контроля по дисциплине, выраженные в оценках «не зачтено, неудовлетворительно», показывают несформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций основной профессиональной образовательной программы высшего образования — программы магистратуры «Музейные исследования и кураторские стратегии» по направлению подготовки 51.04.04 Музеология и охрана объектов культурного и природного наследия.

4 Задания к промежуточной аттестации

УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий

В рамках выбранной темы магистранту необходимо показать наличие знаний и умения применять различные методологии и методики системного и критического анализа проблемных ситуаций, решение которых возможно по средствам применения инструментов обработки естественного языка, для чего необходимо смоделировать ситуацию, возможную в профессиональной деятельности, сформулировать, обосновать и решить задачи, которые характерны для данной ситуации, по средствам соответствующих инструментов обработки естественного языка.

Перечень тем 1-й письменной работы (эссе) с эталонными ответами:

1. Сравнение качества и производительности морфологических библиотек для русского языка.
2. Сравнение качества и производительности библиотек анализа тональности для русского языка.
3. Сравнение качества и производительности библиотек для выделения именованных сущностей для русского языка.
4. Сравнение качества онлайн сервисов машинного перевода для пары русский-английский (одно направление).
5. Сравнительный анализ произведений Льва Толстого и Федора Достоевского с помощью стилометрических методов.
6. Гендерное смещение (gender bias) в дистрибутивных моделях русского языка (проанализировать 2-3 статические модели отсюда: <https://rusvectors.org/ru/models/>).
7. Сравнительный анализ двух моделей вопросно-ответного поиска для русского языка с помощью инструмента CheckList (<https://github.com/marcotcr/checklist>).
8. Анализ качества кросс-языкового переноса моделей вопросно-ответного поиска на данных без дообучения. Исходный английский набор данных – SquAD, тестирование – на данных TyDi QA.
9. Исследование переносимости моделей распознавания юмора: исследовать 2-3 метода классификации на 2-3 англоязычных наборах данных.

10. Анализ существующих систем вопросно-ответного поиска по базам знаний (deeppavlov, Qanswer) с помощью тестового набора данных RuBQ.
11. Систематический анализ качества генерации текстов с помощью модели ruGPT-3 для различных сценариев (<https://github.com/sberbank-ai/ru-gpts>).

ПК-3 Способен использовать современные методы обработки и интерпретации информации в профессиональной сфере

В рамках выбранной темы магистранту необходимо показать наличие знаний современных методов обработки изображений в области исследований культурного и природного наследия, последовательно описать процедуру применения различных методов и инструментов обработки изображений на примере конкретного научно-исследовательского проекта или возможной ситуации в работе музея по выбору магистранта.

Перечень тем 2-й письменной работы с эталонными ответами:

1. Методы обработки естественного языка.
2. Машинный перевод.
3. Голосовые помощники.
4. Анализ текстов.
5. Распознавание и синтез речи.
6. Обработка естественного языка на Java.
7. Обработка естественного языка в контексте Data Science фреймворков.
8. Обработка естественного языка с TensorFlow.
9. Глубокое обучение при обработке естественного языка.
10. NLP на Python.
11. Интеграция Spark и библиотек машинного обучения.
12. Нейросетевые методы в обработке естественного языка.
13. Анализ моделей, основанных на механизме внимания и архитектуре Transformer.
14. Анализ задач музейной работы, которые может решить NLP.
15. Предобработка текста.
16. Стемминг.
17. Лемматизация.
18. Векторизация.
19. Дедубликация.
20. Семантический анализ.
21. Распознавание именованных сущностей и извлечение отношений.
22. Использование N-грамм.
23. Частеречная разметка.
24. Библиотеки для NLP.
25. Анализ примера кода на языке Scala.
26. Обзор подходов и методов к задаче автоматического извлечения именованных сущностей.
27. Обзор подходов и методов к задаче морфологического анализа.
28. Способы векторного представления слов.
29. Автоматическое построение онтологий.
30. Постройте дерево зависимостей для следующего предложения: If you don't know the answer, guess.
31. Приведите примеры инструментов/приложений/программ обработки текстов.
32. Что такое корпусная лингвистика?
33. Изложить основные идеи теории речевых действий.
34. Пояснить принцип работы наивного Байесовского классификатора.

35. В чем отличие между классификацией и кластеризацией текстов?
36. Объяснить принципы использования регулярных выражений в поисковых системах.
37. Привести классификацию речевых действий.
38. В чем отличие между тезаурусами, семантическими сетями и онтологиями?
39. Описать методы анализа социальных сетей.
40. Привести примеры использования методов кластеризации и классификации для определения авторства текстов.
41. Дать определения лингвостатистических параметров, авторского инварианта и лингвистических спектров.
42. Перечислить методы задания синтаксической структуры предложений.
43. Привести примеры мер семантической близости.
44. Дать определения отмеченных последовательностей, контекста, дистрибутивных классов.
45. Перечислить методы классификации и кластеризации текстовой информации. Сформулировать основные принципы.
46. Применение частотных методов в компьютерной лингвистике.
47. Сравнительный анализ семантических сетей и фреймовых моделей.
48. Направления компьютерной лингвистики.
49. Принципы работы автоматических систем извлечения информации.
50. Определения центральностей разного типа.
51. Перечислить основные направления исследований социальных сетей.
52. Разъяснить принципы работы графематического и морфологического анализаторов.
53. Дать формальные определения частей речи, грамматического рода и категории падежа в терминах модели языка, предложенной С. Маркусом.
54. Перечислить основные этапы построения автоматизированных систем обработки текстов.
55. Разъяснить принципы работы фрагментационного и синтаксического анализаторов.
56. Изложить основную идею теоретико-модельного подхода к исследованию семантики текстов.
57. Проблемы автоматизации синтаксического анализа.
58. Сформулировать принципы обработки неструктурированной и плохо структурированной информации. Индексация текстов.
59. Изложить основные идеи подхода И. Мельчука к семантическому анализу.
60. Формальные методы атрибуции текстов.
61. Сформулировать принципы построения систем представления знаний для компьютерной обработки.

5 Средства оценки индикаторов достижения компетенций

Таблица 4

Средства оценки индикаторов достижения компетенций

| Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Средства оценки (в соотв. с Таблицами 5, 7) |
|------------------|--|--|
| УК-1 | ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. | Опрос, диспут, практическое задание, контрольное задание, письменная работа (эссе) |
| ПК-3 | ИД.ПК-3.1. ИД.ПК-3.2. ИД.ПК-3.3. ИД.ПК-3.4. | Опрос, диспут, практическое задание, контрольное задание, письменная работа (эссе) |

| Коды компетенций | Индикаторы компетенций (в соотв. с Таблицей 1) | Средства оценки (в соотв. с Таблицами 5, 7) |
|------------------|---|---|
| | ИД.ПК-3.5. ИД.ПК-3.6. | |

Таблица 5

Описание средств оценки индикаторов достижения компетенций

| Средства оценки (в соотв. с Таблицами 5, 7) | Рекомендованный план выполнения работы |
|---|--|
| Опрос | Магистрант в ходе подготовки и участия в опросе показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — анализирует проблемную ситуацию, определяет пробелы в информации, оценивать надёжность источников информации, разрабатывает стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения; — применяет современные методы, поиска, обработки, анализа и использования информации в рамках проведения научно-исследовательских и организационных работ в области музейных исследований и кураторских стратегий. |
| Диспут | Магистрант в ходе подготовки и участия в диспуте показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — анализирует проблемную ситуацию, определяет пробелы в информации, оценивать надёжность источников информации, разрабатывает стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения; — применяет современные методы, поиска, обработки, анализа и использования информации в рамках проведения научно-исследовательских и организационных работ в области музейных исследований и кураторских стратегий. |
| Практическое задание | Магистрант в ходе подготовки и выполнения практического задания показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — анализирует проблемную ситуацию, определяет пробелы в информации, оценивать надёжность источников информации, разрабатывает стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения; — применяет современные методы, поиска, обработки, анализа и использования информации в рамках проведения научно-исследовательских и организационных работ в области музейных исследований и кураторских стратегий. |
| Контрольное задание | Магистрант в ходе подготовки и выполнения контрольного задания показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — анализирует проблемную ситуацию, определяет пробелы в информации, оценивать надёжность источников информации, разрабатывает стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения; — применяет современные методы, поиска, обработки, анализа и использования информации в рамках проведения научно-исследовательских и организационных работ в области музейных исследований и кураторских стратегий. |
| Письменная работа (эссе) | Магистрант в ходе подготовки и написания письменной работы (эссе), показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — анализирует проблемную ситуацию, определяет пробелы в информации, оценивать надёжность источников информации, разрабатывает стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения; — применяет современные методы, поиска, обработки, анализа и использования информации в рамках проведения научно-исследовательских и организационных работ в области музейных исследований и кураторских стратегий. |