

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

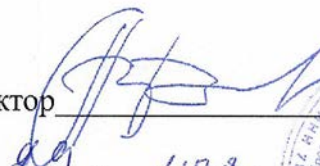
Дата подписания: 13.09.2024 11:40:38


Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a70e51513f08591

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

Факультет социологии

УТВЕРЖДАЮ:
Ректор  В.В. Волков
« 29 » мая 2024 г.
Протокол УС № 5 от 29.05.2024 г.



Рабочая программа дисциплины
Компьютерный анализ текста

образовательная программа
направление подготовки
39.04.01 Социология

направленность (профиль)
«Социальные исследования: смыслы и числа»
программа подготовки – магистратура

язык обучения – русский
форма обучения - очная

квалификация выпускника
Магистр

Санкт-Петербург

Автор:

Левшун Д.С., к.т.н., доцент факультета социологии АНООВО «ЕУСПб»

Рецензент:

Тушканова О.Н., к.т.н., доцент факультета социологии АНООВО «ЕУСПб»

Рабочая программа дисциплины **«Компьютерный анализ текста»**, входящей в образовательную программу уровня магистратуры «Социальные исследования: смыслы и числа», утверждена на заседании Совета факультета социологии.

Протокол заседания № 11 от 26.04.2024 года

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ «Компьютерный анализ текста»

Дисциплина **«Компьютерный анализ текста»** является дисциплиной обязательной части основной профессиональной образовательной программы высшего образования «Социальные исследования: смыслы и числа» по направлению подготовки 39.04.01 Социология.

Дисциплина **«Компьютерный анализ текста»** дает магистрантам представление о теоретических подходах к количественному анализу текстов в общественных науках. Дисциплина также знакомит магистрантов с ключевыми источниками текстовых данных в общественных науках, дает введение в корпусные исследования и проблемы вычислительной лингвистики, магистранты развивают навыки по созданию массивов структурированных текстов из неструктурированных данных.

Общая трудоемкость освоения дисциплины составляет 3 зачетных единицы, 108 часов.

Содержание

1.	4	
2.	5	
3.	МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	8
4.	ОБЪЕМ ДИСЦИПЛИНЫ	8
5.	СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ	8
5.1	Содержание дисциплины	8
5.2	Структура дисциплины	11
6.	УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ	12
6.1	Общие положения	12
6.2	Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины	12
6.3	Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине	14
6.4	Перечень литературы для самостоятельной работы обучающегося:	14
6.5	Перечень учебно-методического обеспечения для самостоятельной работы	14
7.	ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ	15
7.1	Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации	15
7.2	Контрольные задания для текущей аттестации	18
7.3	Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации	19
7.4	Типовые задания к промежуточной аттестации	20
7.5	Средства оценки индикаторов достижения компетенций	20
8.	ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА	23
8.1.	Основная литература	23
8.2	Дополнительная литература	24
9.	ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА	24
9.1	Программное обеспечение	24
9.2	Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:	24
9.3	Лицензионные электронные ресурсы библиотеки Университета	25
9.4	Электронная информационно-образовательная среда Университета	25
10.	МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА	26
	ПРИЛОЖЕНИЕ 1	27

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины «Компьютерный анализ текста» — изучить подходы к количественному анализу текстов в общественных науках.

Задачи:

1. Знакомство с ключевыми источниками текстовых данных в общественных науках, введение в корпусные исследования;
2. Получение навыков по созданию массивов структурированных текстов из неструктурированных данных;
3. Введение в проблемы вычислительной лингвистики.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В результате изучения учебной дисциплины обучающийся должен овладеть следующими компетенциями: 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)
ПК-4 Способен проводить социологическую экспертизу стратегий, мероприятий, качества исследований и консультировать по вопросам применения результатов фундаментальных или прикладных социологических исследований	ИД.ПК-4.1 Применение результатов анализа и интерпретация данных социологического исследования ИД.ПК-4.2 Дает социологическую оценку вариантов управленческих решений, разрабатывает предложения по их совершенствованию и перспективам развития	Знать: Обучающийся должен знать методологию проведения социологической экспертизы и критерии оценки качества социологических исследований. З (ПК-4)
		Уметь: Обучающийся должен уметь анализировать результаты социологического исследования с точки зрения их социальной значимости и соответствия целям и задачам. У (ПК-4)
		Владеть: Обучающийся должен владеть навыками формулирования обоснованных выводов и рекомендаций по результатам социологической экспертизы и применению результатов исследований. В (ПК-4)

В результате освоения дисциплины магистрант должен:

знать:

- основные подходы к количественному анализу текстов в общественных науках;
- основные понятия и принципы корпусной лингвистики и её роль в общественных науках;
- основные методы и подходы к анализу текстовых данных в социальных науках, таких как контент-анализ, дискурс-анализ и другие.
- проблемы и ограничения, связанные с использованием текстовых данных для научных исследований, и способы их преодоления.

уметь:

- Создавать массивы структурированных текстов из неструктурированных данных;
- создавать и настраивать алгоритмы для обработки текстовых данных, включая токенизацию, лемматизацию и стемминг;
- анализировать результаты работы алгоритмов и интерпретировать полученные данные;

- применять методы статистического анализа для оценки значимости результатов исследования.
- **владеть:**
 - навыками анализа результатов компьютерного анализа текста и интерпретации их в контексте общественных наук;
 - основными методами и инструментами для сбора, обработки и анализа текстовых данных;
 - навыками работы с корпусами текстов и использования их для проведения исследований;
 - способностью применять методы машинного обучения для анализа текстовых данных и извлечения информации.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Компьютерный анализ текста» является дисциплиной части, формируемой участниками образовательных отношений Блока 1 «Дисциплины (модули)» обязательной части образовательной программы «Социальные исследования: смыслы и числа». Курс читается в третьем семестре, форма промежуточной аттестации – зачет с оценкой.

Для успешного освоения данной дисциплины требуются знания, полученные в рамках прохождения обучения на уровне бакалавриата/ специалитета.

Знания, умения и навыки, полученные при освоении данной дисциплины, применяются магистрантами в процессе прохождения производственной практики и выполнения выпускной квалификационной работы.

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 3 зачетных единицы, 108 часов.

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 под руководством преподавателя	-	-	-	-	-
Самостоятельная работа (СР)	80	-	-	80	-
Промежуточная аттестация	форма	Зачет с оценкой	-	-	Зачет с оценкой
	час.	-	-	-	-
Общая трудоемкость дисциплины (час./з.е.)	108/3	-	-	108/3	-

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 компетенции	Индикаторы компетенций (в соот.с Таблицей 1)	Коды ЗУВ (в соот.с Таблицей 1)
1	Корпусы текстовых данных	Корпусные исследования. Описание корпусов текстовых данных: КРЯ, Wikipedia, CommonCrawl, Taiga. Разбор принципов их организации и причин создания. Примеры использования в исследованиях.	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)
2	Распознавание текстов. Создание процедуры превращения изображений или PDF в корпус	Организация корпуса. Проблемы источников данных и способы их преодоления.	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)
3	Закон Ципфа. Издержки токенизации	Введение в количественное представление текстов. Разбор процедур лемматизации, эвалюация их результатов	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)
4	Извлечение сущностей из текстовых данных	Введение в классические задачи обработки естественных языков. Детальный разбор задачи извлечения именованных сущностей. Обзор существующих решений. Проблема эвалюирования.	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)
5	Разреженное векторное представление текстовых данных	Строковые расстояния. Модель «мешок слов». Byte Pair Encoding. Анализ коллокаций.	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)
6	Уплотненное векторное представление текстовых данных	Факторизация матриц. SVD, LSA, LDA. Дистрибутивная семантика. От word2vec к контекстуальным эмбедингам.	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)
7	Поиск по представлениям текстовых данных	Разреженные матрицы. Виды метрик и расстояний. Метод ближайших средних, его ограничения.	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенции	Индикаторы компетенций (в соот.с Таблицей 1)	Коды ЗУВ (в соот.с Таблицей 1)
		Эвалюация результатов семантического поиска. Возможности и ограничения семантического поиска.			

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	ПЗ		
<i>Очная форма обучения</i>						
Тема 1	Корпусы текстовых данных	15	2	2	11	О
Тема 2	Распознавание текстов. Создание процедуры превращения изображений или PDF в корпус	15	2	2	11	ПЗ
Тема 3	Закон Ципфа. Издержки токенизации	15	2	2	11	ПЗ
Тема 4	Извлечение сущностей из текстовых данных	15	2	2	11	ПЗ
Тема 5	Разреженное векторное представление текстовых данных	15	2	2	11	ПЗ
Тема 6	Уплотненное векторное представление текстовых данных	15	2	2	11	ПЗ
Тема 7	Поиск по представлениям текстовых данных	18	2	2	14	ПЗ, Д
Промежуточная аттестация		-	-	-	-	Зачет
Всего:		108/3	14	14	80	-

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

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 соответственным образом организовав рабочее пространство. Важнейшим элементом самостоятельной работы является проработка материалов прошедших занятий (анализ конспектов, чтение рекомендованной литературы) и подготовка к следующим лекциям/семинарским занятиям. 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 вспомогательных учебно-методических изданий, лекционных конспектов, интернет-ресурсов и пр. Подготовка к семинарским занятиям, контрольному тесту 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. Корпусы текстовых данных:

1.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 5 часов.

1.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов. Итого: 11 часов.

Тема 2. Распознавание текстов. Создание процедуры превращения изображений или PDF в корпус:

2.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 5 часов.

2.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов. Итого: 11 часов.

Тема 3. Закон Ципфа. Издержки токенизации:

3.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 5 часов.

3.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов. Итого: 11 часов.

Тема 4. Извлечение сущностей из текстовых данных:

4.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 5 часов.

4.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов. Итого: 11 часов.

Тема 5. Разреженное векторное представление текстовых данных:

5.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 5 часов.

5.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов. Итого: 11 часов.

Тема 6. Уплотненное векторное представление текстовых данных:

6.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 5 часов.

6.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 6 часов. Итого: 11 часов.

Тема 7. Поиск по представлениям текстовых данных:

7.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 7 часов.

7.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 7 часов. Итого: 14 часов.

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

Вопросы для самостоятельной подготовки по темам дисциплины:

1. Почему закон Ципфа имеет разные эмпирические параметры распределения при разных единицах сегментации корпуса?
2. В каких случаях модель «мешок слов» не применима для представления текстовых данных?
3. Перечислите основные проблемы при оптическом распознавании изображений отсканированных документов.
4. Какие «золотые стандарты» разметки текстов для эвалюации качества распознавания именованных сущностей на русском языке вам известны?
5. Почему расчет строкового расстояния — вычислительно емкая операция?
6. Какие алгоритмы извлечения коллокаций вы знаете?
7. В чем разница между SVD и LSA?
8. Как использовать разреженность терм-документарных матриц для ускорения вычислений?
9. Опишите задачи, где метод ближайших соседей неприменим.

6.4 Перечень литературы для самостоятельной работы обучающегося:

1. Хименко, В.И. Случайные данные: структура и анализ / В.И. Хименко. – Москва : Техносфера, 2017. – 424 с. : ил.,табл., схем. – (Мир фотоники). – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=496479>
2. Основы научных исследований / Б.И. Герасимов, В.В. Дробышева, Н.В. Злобина и др. - М.: Форум: НИЦ Инфра-М, 2013. - 272 с. - [Электронный ресурс]. -URL: <http://znanium.com/bookread2.php?book=390595>
3. Общая теория статистики: Учебное пособие / С.Н. Лысенко, И.А. Дмитриева. - Изд., испр. и доп. - М.: Вузовский учебник: НИЦ ИНФРА-М, 2014. - 219 с. - [Электронный ресурс]. URL: <http://znanium.com/bookread2.php?book=397795>
4. Маркин, А.В. Построение запросов и программирование на SQL : учебное пособие / А.В. Маркин. – 3-е изд., перераб. и доп. – Москва : Диалог-МИФИ, 2014. – 384 с. : ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=89077>

6.5 Перечень учебно-методического обеспечения для самостоятельной работы

Для обеспечения самостоятельной работы магистрантов по дисциплине «Компьютерный анализ текста» 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, 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 оценивания участия в опросах и диспутах, выполненных практических заданий, демонстрирующих степень знакомства с дополнительной литературой.

Таблица 5

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

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
Корпусы текстовых данных	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)	Опрос 1	зачтено/ не зачтено
Распознавание текстов. Создание процедуры превращения изображений или PDF в корпус	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)	Практическое задание 1	зачтено/ не зачтено
Закон Ципфа. Издержки токенизации	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)	Практическое задание 2	зачтено/ не зачтено
Извлечение сущностей из текстовых данных	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)	Практическое задание 3	зачтено/ не зачтено
Разреженное векторное представление текстовых данных	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)	Практическое задание 4	зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
Уплотненное векторное представление текстовых данных	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)	Практическое задание 5	зачтено/ не зачтено
Поиск по представлениям текстовых данных	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)	Практическое задание 6 Диспут 1	зачтено/ не зачтено зачтено/ не зачтено

Таблица 6

Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
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
Опрос	ответ отсутствует или является односложным, или содержит существенные ошибки – не зачтено 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
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 диспут, демонстрация навыков профессиональной коммуникации в группе — зачтено

7.2 Контрольные задания для текущей аттестации

Примерный материал опросов, диспутов, практических заданий:

Тема 1. Корпусы текстовых данных.

Опрос 1:

Выбрать из корпуса Taiga статьи определенного интернет-сми, упоминающие определенное слово.

Тема 2. Распознавание текстов. Создание процедуры превращения изображений или PDF в корпус.

Практическое задание 1:

Разработать конвейер по превращению заданных PDF-изображений в текстовые файлы с заданной структурой.

Тема 3. Закон Ципфа. Издержки токенизации.

Практическое задание 2:

Оценить качество сегментации разными методами, сравнить результаты.

Тема 4. Извлечение сущностей из текстовых данных.

Практическое задание 3:

Извлечь из заданного корпуса все именованные сущности, удовлетворяющие определенному требованию.

Тема 5. Разреженное векторное представление текстовых данных.

Практическое задание 4:

Оценить на терм-документарной матрице вектора документов предложенными алгоритмами.

Тема 6. Уплотненное векторное представление текстовых данных.

Практическое задание 5:

Рассчитать уплотненные вектора слов на заданном корпусе, сравнить результаты с результатами проекта RusVectōrēs.

Тема 7. Поиск по представлениям текстовых данных.

Практическое задание 6, диспут 1:

Найти в заданном корпусе, преобразованном в формат терм-документной матрицы, ближайших соседей.

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

Форма промежуточной аттестации – зачет с оценкой, выставляемый на основе оценки проекта, подготовленного магистрантом.

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 соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)	Критерии оценивания	Оценка
Зачет с оценкой / проект	ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	З (ПК-4) У (ПК-4) В (ПК-4)	Магистрант демонстрирует полную самостоятельность в подборе фактического материала и аналитическое отношение к нему, умение рассматривать примеры и факты во взаимосвязи и взаимообусловленности, отбирать наиболее существенные из них; а также показывает грамотное использование методов описания и презентации исследования	Зачтено, отлично
				Магистрант демонстрирует самостоятельность в подборе фактического материала и аналитическое отношение к нему, в большинстве случаев видит взаимосвязь примеров и фактов, в целом отбирает существенные из них; а также использует правильные методы описания и презентации исследования с небольшими ошибками	Зачтено, хорошо

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)	Критерии оценивания	Оценка
				Магистрант демонстрирует аналитическое отношение к материалу, видит взаимосвязь некоторых примеров и фактов; а также использует методы описания и презентации исследования с большим количеством несущественных ошибок	Зачтено, удовлетворительно
				Магистрант не демонстрирует аналитическое отношение к материалу, не видит взаимосвязь примеров и фактов; а также использует методы описания и презентации исследования с большим количеством существенных ошибок	Зачтено, неудовлетворительно

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, отлично», «зачтено, хорошо», «зачтено, удовлетворительно», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Социальные исследования: смыслы и числа» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценках «не зачтено, неудовлетворительно», показывают несформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Социальные исследования: смыслы и числа» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

7.4 Типовые задания к промежуточной аттестации

Примерные требования к проекту для промежуточной аттестации по дисциплине:

1. Тему проекта магистрант выбирает, основываясь на своих научных интересах, и согласовывает с преподавателем заранее.
2. Проект представляется в виде созданного текстового корпуса и сопроводительного документа-описания его источников, структуры и ключевых характеристик.

7.5 Средства оценки индикаторов достижения компетенций

Таблица 8

Средства оценки индикаторов достижения компетенций

Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Средства оценки (в соот. с Таблицами 5, 7)
ПК-4	ИД.ПК-4.1. ИД.ПК-4.2.	Опрос, диспут, практическое задание, проект

Таблица 9

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
Практическое задание	Магистрант в ходе подготовки и выполнения практических заданий показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: Применять результаты анализа и интерпретация данных социологического исследования. Давать социологическую оценку вариантов управленческих решений, разрабатывает предложения по их совершенствованию и перспективам развития.
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: Применять результаты анализа и интерпретация данных социологического исследования. Давать социологическую оценку вариантов управленческих решений, разрабатывает предложения по их совершенствованию и перспективам развития.
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: Применять результаты анализа и интерпретация данных социологического исследования. Давать социологическую оценку вариантов управленческих решений, разрабатывает предложения по их совершенствованию и перспективам развития.
Проект	Магистрант в ходе подготовки и защиты проекта показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: Применять результаты анализа и интерпретация данных социологического исследования. Давать социологическую оценку вариантов управленческих решений, разрабатывает предложения по их совершенствованию и перспективам развития.

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.

2. Тапскотт, Д. Викиномика: как массовое сотрудничество изменяет все / Д. Тапскотт, Э. Д. Уильямс. - Москва : Интеллектуальная Литература, 2020. - 456 с. - ISBN 978-5-6042878-7-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1220225> . – 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:

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. OS Microsoft Windows (OVS OS Platform)
2. MS Office (OVS Office Platform)
3. Adobe Acrobat Professional 11.0 MLP AOO License RU
4. Adobe CS5.5 Design Standart Win IE EDU CLP
5. ABBYY FineReader 11 Corporate Edition
6. ABBYY Lingvo x5
7. Adobe Photoshop Extended CS6 13.0 MLP AOO License RU
8. Adobe Acrobat Reader DC /Pro – бесплатно
9. Google Chrome – бесплатно
10. Opera – бесплатно
11. Mozilla – бесплатно
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://npoed.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. Президентская библиотека: <http://www.prlib.ru>
3. Российская государственная библиотека: <http://www.rsl.ru/>
4. 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. **East View** – 100 ведущих российских журналов по гуманитарным наукам (архив и текущая подписка): <https://dlib.eastview.com/browse;>
2. **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;>
3. **Университетская информационная система РОССИЯ** — база электронных ресурсов для учебных программ и исследовательских проектов в области социально-гуманитарных наук: [http://www.uisrussia.msu.ru/;](http://www.uisrussia.msu.ru/)
4. Электронные журналы по подписке (текущие номера научных зарубежных журналов)

Электронные библиотечные системы:

1. **Znaniium.com** – Электронная библиотечная система (ЭБС) – [http://znaniium.com/;](http://znaniium.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://eusp.org/>]), локальную сеть и корпоративную электронную почту Университета, 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 (практических занятий, лабораторных работ), 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.