

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

Дата подписания: 25.08.2025 12:11:10

Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a70e51517ff6d591

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

УТВЕРЖДАЮ:

Ректор

/В.В. Волков

«24 » сентябрь 2025 г.

Протокол УС № 3 от 27 марта 2025 г.



**Рабочая программа дисциплины
Текстовые данные**

образовательная программа
направление подготовки
09.04.03 Прикладная информатика

направленность (профиль)
«Прикладной анализ данных»
программа подготовки – магистратура

язык обучения – русский
форма обучения - очная

квалификация (степень) 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 «**Текстовые данные**», входящей в образовательную программу уровня магистратуры «Прикладной анализ данных», утверждена на заседании Совета факультета социологии.

Протокол заседания № 8 от 29 января 2024 года.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ **«Текстовые данные»**

Дисциплина «**Текстовые данные**» является дисциплиной обязательной части основной профессиональной образовательной программы высшего образования «Прикладной анализ данных» по направлению подготовки 09.04.03 Прикладная информатика.

Дисциплина «**Текстовые данные**» 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 развиваются навыки по созданию массивов структурированных текстов из неструктурированных данных.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточный контроль в форме зачета.

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

Содержание

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ	5
2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ	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	21
7.5 Средства оценки индикаторов достижения компетенций	32
8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА	35
8.1. Основная литература	35
8.2. Дополнительная литература	35
9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА	35
9.1 Программное обеспечение	35
9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:	36
9.3 Лицензионные электронные ресурсы библиотеки Университета	36
9.4 Электронная информационно-образовательная среда Университета	37
10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА	37
ПРИЛОЖЕНИЕ 1	39

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины «Текстовые данные» — 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: универсальными (УК) и общепрофессиональными (ОПК). 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. Анализирует проблемную ситуацию как систему, выявляя её составляющие и связи между ними ИД.УК-1.2. Определяет пробелы в информации, необходимой для решения проблемной ситуации, и проектирует процессы по их устранению ИД.УК-1.3. Критически оценивает надёжность источников информации, работает с противоречивой информацией из разных источников ИД.УК-1.4. Разрабатывает и содержательно аргументирует стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов ИД.УК-1.5. Строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устрани	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 социальную 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)
УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	ИД.УК-4.1. Устанавливает контакты и организует общение в соответствии с потребностями совместной деятельности, используя современные коммуникационные технологии ИД.УК-4.2. Составляет в соответствии с нормами русского языка деловую документацию разных жанров ИД.УК-4.3. Составляет типовую деловую документацию для академических и профессиональных целей на иностранном языке	Знать: принципы эффективной коммуникации в устной и письменной формах на русском и иностранном языке, принятые в академическом и профессиональном сообществе З (УК-4) Уметь: использовать разные коммуникативные роли и стратегии для решения задач профессиональной деятельности на русском и иностранном языке, принятые в академическом и профессиональном сообществе У (УК-4)

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
	<p>ИД.УК-4.4. Составляет академические и (или) профессиональные тексты на иностранном языке</p> <p>ИД.УК-4.5. Организует обсуждение результатов исследовательской и проектной деятельности на различных публичных мероприятиях на русском языке, выбирая подходящий формат</p> <p>ИД.УК-4.6. Представляет результаты исследовательской и проектной деятельности на различных публичных мероприятиях, участвует в академических и профессиональных дискуссиях на иностранном языке</p>	<p>Владеть:</p> <p>навыками использования разных коммуникативных ролей и стратегий для решения профессиональных задач на русском и иностранном языке, принятые в академическом и профессиональном сообществе</p> <p>В (УК-4)</p>
УК-5 Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия	<p>ИД.УК-5.1. Анализирует важнейшие идеологические и ценностные системы, сформировавшиеся в ходе исторического развития</p> <p>ИД.УК-5.2. Выстраивает социальное и профессиональное взаимодействие с учётом особенностей деловой и общей культуры представителей других этносов и конфессий, различных социальных групп</p> <p>ИД.УК-5.3. Обеспечивает создание недискриминационной среды для участников межкультурного взаимодействия при личном общении и при выполнении профессиональных задач</p>	<p>Знать:</p> <p>особенности социальных, этнических, конфессиональных, культурных различий, встречающихся среди членов коллектива</p> <p>З (УК-5)</p> <p>Уметь: выстраивать взаимодействие с членами межкультурного профессионального сообщества, на основе анализа социально-культурных особенностей, этнических и конфессиональных различий отдельных членов межкультурной группы</p> <p>У (УК-5)</p> <p>Владеть:</p> <p>навыками анализа социально-культурных особенностей, этнических и конфессиональных различий отдельных членов межкультурной группы с целью эффективного взаимодействия</p> <p>В (УК-5)</p>
ОПК-1 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	<p>ИД.ОПК-1.1. Решает нестандартные задачи профессиональной деятельности с применением математических, естественнонаучных, социально-экономических методов</p> <p>ИД.ОПК-1.2. В условиях неопределенности и риска способен выработать эффективную стратегию действий, используя математические, естественнонаучные, социально-экономические методы науки</p>	<p>Знать:</p> <p>математические, естественнонаучные и социально-экономические методы для использования в профессиональной деятельности</p> <p>З (ОПК-1)</p> <p>Уметь: решать нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных, социально-экономических и профессиональных знаний</p> <p>У (ОПК-1)</p> <p>Владеть:</p> <p>навыком применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач</p> <p>В (ОПК-1)</p>
		Знать:

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
ОПК-3 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами рекомендациями и	<p>ИД.ОПК-3.1. Анализирует и структурирует профессиональные данные с использованием современных методов прикладного анализа данных</p> <p>ИД.ОПК-3.2. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для решения профессиональных задач</p> <p>ИД.ОПК-3.3. Выявляет значимые проблемы и разрабатывает рекомендации по их решению</p> <p>ИД.ОПК-3.4. Оформляет и представляет результаты анализа в виде аналитических обзоров</p>	<p>принципы, методы и средства анализа и структурирования профессиональной информации З (ОПК-3)</p> <p>Уметь: интерпретировать данные и формулировать выводы и теоретические подходы для решения профессиональных задач, представляя результаты работы в виде аналитических обзоров У (ОПК-3)</p> <p>Владеть: навыками разработки рекомендаций по результатам анализа профессиональной информации В (ОПК-3)</p>
ОПК-4 Способен применять на практике новые научные принципы и методы исследований	<p>ИД.ОПК-4.1. На основе современных теорий и концепций обосновывает актуальность постановки целей и задач научных исследований в профессиональной области знаний</p> <p>ИД.ОПК-4.2. Анализирует новые научные принципы и методы исследований в профессиональной области знаний</p> <p>ИД.ОПК-4.3. Применяет новые научные принципы и методы исследований в профессиональной области знаний</p> <p>ИД.ОПК-4.4. Разрабатывает предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний</p>	<p>Знать: актуальные направления применения новых научных принципов и методов исследований в профессиональной деятельности З (ОПК-4)</p> <p>Уметь: самостоятельно формировать планы и программы научных исследований с применением новых принципов и методов, характерных для выбранной отрасли науки У (ОПК-4)</p> <p>Владеть: навыками системного использования различных новых научных принципов и методов исследований для различных направлений науки В (ОПК-4)</p>

В результате освоения дисциплины магистрант должен:

— **знать:** современные научные представления о мире и науке, методологию научного исследования и экспертно-аналитической работы; методы сбора и обработки данных; особенности применения различных теоретико-методологических концепций с использованием технологий прикладного анализа данных; общие правила ведения научных дискуссий;

— **уметь:** синтезировать новое профессиональное знание на базе применения знаний и аналитических навыков с использованием технологий прикладного анализа данных; использовать полученные знания и умения в профессиональной деятельности, деловой коммуникации и межличностном общении; использовать навыки научных исследований общественных процессов и отношений; разрабатывать программу научного исследования, правильно оформлять и представлять результаты исследований; анализировать и оценивать общественные процессы; выявлять необходимую информацию из текстов различной тематики и направленности, а также из иных источников; использовать имеющиеся знания для целей проведения научных дискуссий и участия в них;

— **владеть:** передовыми приёмами построения аналитического дискурса и аргументированного представления его результатов; навыками научных исследований общественных процессов и отношений, методами сбора и обработки данных, в том числе с использованием технологий прикладного анализа данных; углублёнными теоретическими знаниями и практическими навыками организации научных исследований; способностью к

самостоятельному обучению новым методам исследования, к изменению научного и научно-производственного профиля своей профессиональной деятельности, к изменению социокультурных и социальных условий деятельности; навыками использования полученных знаний для формулировки собственной позиции по актуальным проблемам общественных наук; приёмами и методами ведения дискуссии по проблемам современной науки.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «**Текстовые данные**» является обязательной дисциплиной Блока 1 «Дисциплины (модули)» обязательной части образовательной программы «Прикладной анализ данных». Код дисциплины по Учебному плану Б1.О.08. Курс читается в третьем модуле, форма промежуточной аттестации – зачет.

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 Б2.О.01(У) Технологической (проектно-технологической) практики и выполнения выпускной квалификационной работы.

4. ОБЪЕМ ДИСЦИПЛИНЫ

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

Таблица 2

Объем дисциплины

Типы учебных занятий и самостоятельная работа		Объем дисциплины									
		Всего		Модуль							
		1	2	3	4	5	6	7	8	9	10
Контактная работа обучающихся с преподавателем в соответствии с УП:											
Лекции (Л)	28	-	-	28	-	-	-	-	-	-	-
Практические занятия (ПЗ)	14	-	-	14	-	-	-	-	-	-	-
Самостоятельная работа (СР)	80	-	-	80	-	-	-	-	-	-	-
Промежуточная аттестация	форма	Zачет	-	-	Zачет	-	-	-	-	-	-
	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 компетенций	Индикаторы компетенций (в соотв. с Таблицей 1)	Коды ЗУВ (в соотв. с Таблицей 1)
1	Корпусы текстовых данных	Корpusные исследования. Описание корпусов текстовых данных: КРЯ, Wikipedia, CommonCrawl, Taiga.	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1.	3 (УК-1) У (УК-1) В (УК-1) 3 (УК-4) У (УК-4) В (УК-4)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Коды ЗУВ (в соот.с Таблицей 1)
		Разбор принципов их организации и причин создания. Примеры использования в исследованиях.		ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
2	Распознавание текстов. Создание процедуры превращения изображений или PDF в корпус	Организация корпуса. Проблемы источников данных и способы их преодоления.	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
3	Закон Ципфа. Издержки токенизации	Введение количественное представление текстов. Разбор процедур лемматизации, эвалюация их результатов	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Коды ЗУВ (в соот.с Таблицей 1)
				ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	В (ОПК-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.	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
5	Разреженное векторное представление текстовых данных	Строковые расстояния. Модель «мешок слов». Byte Pair Encoding. Анализ коллокаций.	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
6	Уплотненное векторное представление	Факторизация матриц. SVD, LSA, LDA. Дистрибутивная семантика. От word2vec	УК-1 УК-4 УК-5 ОПК-1	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 (УК-4)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соотв.с Таблицей 1)	Коды ЗУВ (в соотв.с Таблицей 1)
	текстовых данных	к контекстуальным эмбеддингам.	ОПК-3 ОПК-4	ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
7	Поиск по представлениям текстовых данных	Разреженные матрицы. Виды метрик и расстояний. Метод ближайших средних, его ограничения. Эвалюация результатов семантического поиска. Возможности и ограничения семантического поиска.	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)

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		
Очная форма обучения						
Тема 1	Корпусы текстовых данных	15	2	2	11 О	

№ 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		
Очная форма обучения					
Тема 2	Распознавание текстов. Создание процедуры превращения изображений или PDF в корпус	15	2	2	11
Тема 3	Закон Ципфа. Издержки токенизации	15	2	2	11
Тема 4	Извлечение сущностей из текстовых данных	15	2	2	11
Тема 5	Разреженное векторное представление текстовых данных	15	2	2	11
Тема 6	Уплотненное векторное представление текстовых данных	15	2	2	11
Тема 7	Поиск по представлениям текстовых данных	18	2	2	14
Промежуточная аттестация		-	-	-	-
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 оценки 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 «Текстовые данные» 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	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	3 (УК-1) У (УК-1) В (УК-1) 3 (УК-4) У (УК-4) В (УК-4) 3 (УК-5) У (УК-5) В (УК-5) 3 (ОПК-1) У (ОПК-1) В (ОПК-1) 3 (ОПК-3) У (ОПК-3) В (ОПК-3) 3 (ОПК-4) У (ОПК-4) В (ОПК-4)	Опрос 1	зачтено/ не зачтено
Распознавание текстов. Создание процедуры превращения изображений или PDF в корпус	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5.	3 (УК-1) У (УК-1) В (УК-1) 3 (УК-4) У (УК-4) В (УК-4) 3 (УК-5) У (УК-5) В (УК-5) 3 (ОПК-1)	Практическое задание 1	зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
		ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)		
Закон Ципфа. Издержки токенизации	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Практическое задание 2	зачтено/ не зачтено
Извлечение сущностей из текстовых данных	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Практическое задание 3	зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
		ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.			
Разреженное векторное представление текстовых данных	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Практическое задание 4	зачтено/ не зачтено
Уплотненное векторное представление текстовых данных	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Практическое задание 5	зачтено/ не зачтено
Поиск по представлениям текстовых данных	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5)	Практическое задание 6 Диспут 1	зачтено/ не зачтено зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
		ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	В (УК-5) 3 (ОПК-1) У (ОПК-1) В (ОПК-1) 3 (ОПК-3) У (ОПК-3) В (ОПК-3) 3 (ОПК-4) У (ОПК-4) В (ОПК-4)		

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 зачленено, 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) – зачленено
Опрос	ответ отсутствует или является односложным, или содержит существенные ошибки – не зачленено 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 — зачленено
Диспут	пассивность, участие без представления аргументов и обоснования точки зрения, несформированность навыков профессиональной коммуникации в группе — не зачленено представление аргументированной научной позиции, обоснование точки зрения в диспут, демонстрация навыков профессиональной коммуникации в группе — зачленено

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 Циффа. 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, выставляемый на основе оценки тестирования.

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

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Тест включает 20 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте -1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балла; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Задания открытого типа

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F = \frac{100}{K} * \left(\frac{x_1}{k_1} + \frac{x_2}{k_2} + \dots + \frac{x_n}{k_n} \right),$$

где F – итоговое количество баллов за тест,
 K – количество осваиваемых в рамках дисциплины компетенций,
 k_n – максимально возможное количество баллов за вопросы по компетенции,
 x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

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 / тест	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	3 (УК-1) У (УК-1) В (УК-1) 3 (УК-4) У (УК-4) В (УК-4) 3 (УК-5) У (УК-5) В (УК-5) 3 (ОПК-1) У (ОПК-1) В (ОПК-1) 3 (ОПК-3) У (ОПК-3) В (ОПК-3) 3 (ОПК-4) У (ОПК-4) В (ОПК-4)	41-100% правильных ответов 0-40% правильных ответов	Зачтено Не засчитано

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 засчитано

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «зачтено», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «не зачтено», показывают несформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

7.4 Типовые задания к промежуточной аттестации

Задания закрытого типа

Базовый уровень сложности

1. Прочитайте задание, выберите правильный ответ.

Что такое токенизация в обработке естественного языка?

- 1) Объединение слов в предложения
- 2) Перевод текста с одного языка на другой
- 3) Разделение текста на части, например, слова или предложения
- 4) Удаление всех гласных из текста

Поле для ответа:

2. Прочитайте задание, выберите правильный ответ.

Что такое стемминг?

- 1) Преобразование чисел в текст
- 2) Перевод текста в верхний регистр
- 3) Приведение слова к его корню
- 4) Разделение текста на абзацы

Поле для ответа:

3. Прочитайте задание, выберите правильный ответ.

Какую задачу решает морфологический разбор?

- 1) Определение темпа речи
- 2) Определение части речи и формы слова
- 3) Перевод текста на другой язык
- 4) Анализ эмоциональной окраски текста

Поле для ответа:

Повышенный уровень сложности

4. Прочитайте задание и установите соответствие.

Соотнесите задачи обработки естественного языка (слева) с их примерами (справа).

- A) Классификация текста
B) Машинный перевод
C) Извлечение информации
D) Распознавание именованных сущностей

- 1) Перевод с английского на русский

- 2) Определение категории текста
- 3) Поиск фактов в тексте
- 4) Выделение имен, локаций, организаций

Поле для ответа:

5. Прочитайте задание и установите соответствие.

Соотнесите этапы лингвистического анализа (слева) с их описаниями (справа).

- A) Морфологический анализ
B) Синтаксический анализ
C) Семантический анализ

- 1) Определение структуры предложения
- 2) Определение частей речи и форм слов
- 3) Анализ смысловых связей между словами

Поле для ответа:

6. Прочитайте задание и установите соответствие.

Соотнесите этапы предварительной обработки текста (слева) с их функциями (справа).

- A) Токенизация
B) Лемматизация
C) Удаление стоп-слов

- 1) Разделение текста на слова или фразы
- 2) Приведение слов к начальной форме
- 3) Исключение часто встречающихся, но малозначимых слов

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

7. Прочитайте текст и напишите обоснованный ответ.

В чём заключается задача морфологического анализа?

Поле для ответа:

8. Прочитайте текст и напишите обоснованный ответ.

Что такое токенизация и зачем она используется в NLP?

Поле для ответа:

УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия

Задания закрытого типа

Базовый уровень сложности

9. Прочтайте задание, выберите правильный ответ.

Что такое стоп-слова в обработке текста?

Самые важные слова в тексте

Слова, которые подчеркиваются

Часто употребляемые слова, не несущие смысловой нагрузки

Слова, содержащие ошибки

Поле для ответа:

10. Прочтайте задание, выберите правильный ответ.

Какую задачу решает лемматизация?

- 1) Добавление новых слов в текст
- 2) Перевод текста в аудио
- 3) Преобразование слова к его нормальной форме
- 4) Удаление всех прилагательных из текста

Поле для ответа:

11. Прочтайте задание, выберите правильный ответ.

Какая из перечисленных задач относится к обработке естественного языка?

- 1) Обработка изображений
- 2) Сегментация клиентов
- 3) Синтаксический разбор текста
- 4) Предсказание движения цены акций

Поле для ответа:

Повышенный уровень сложности

12. Прочтайте задание и установите соответствие.

Соотнесите типы векторных моделей (слева) с их описаниями (справа).

- | | |
|---------------------|--|
| A) One-hot encoding | |
| B) Frequency model | |
| B) TF-IDF | |

- | | |
|---|--|
| 1) Отображает количество вхождений слов в документ | |
| 2) Отражает важность слова с учетом его встречаемости в корпусе | |
| 3) Представляет слова в виде векторов с одним значением "1" | |

Поле для ответа:

13. Прочтайте задание и установите соответствие.

Соотнесите модели представления текста (слева) с их описаниями (справа).

- | | |
|-----------------|--|
| A) TF-IDF | |
| B) N-gram model | |

- Б) word2vec
Г) Bag of Words

- 1) Учитывает частотность слов и их уникальность в корпусе
- 2) Представляет текст как последовательности из N слов
- 3) Использует плотные векторы, обученные на контексте слов
- 4) Учитывает только количество вхождений слов без контекста

Поле для ответа:

Задания открытого типа
Высокий уровень сложности

14. Прочтите текст и напишите обоснованный ответ.

Что такое обработка естественного языка и для чего она применяется?

Поле для ответа:

15. Прочтите текст и напишите обоснованный ответ.

Что происходит на этапе синтаксического анализа текста?

Поле для ответа:

УК-5 Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия

Задания закрытого типа
Базовый уровень сложности

16. Прочтайте задание, выберите правильный ответ.

Что выполняется на этапе токенизации текста?

- 1) Объединение документов в один
- 2) Преобразование слов в числа
- 3) Разделение текста на слова или фразы
- 4) Классификация текста по категориям

Поле для ответа:

17. Прочтайте задание, выберите правильный ответ.

Какой из методов представляет слова в виде плотных векторов, обученных на контексте?

- 1) Bag of Words
- 2) TF-IDF
- 3) word2vec
- 4) One-hot encoding

Поле для ответа:

18. Прочтайте задание, выберите правильный ответ.

Что учитывает модель TF-IDF при расчёте весов слов?

- 1) Порядок слов в предложении
- 2) Обратную частоту слова в корпусе
- 3) Количество предложений в тексте
- 4) Синтаксическую структуру

Поле для ответа:

Повышенный уровень сложности

19. Прочтайте задание и установите соответствие.

Соотнесите этапы лингвистического анализа (слева) с их описаниями (справа).

- A) Морфологический анализ
- B) Синтаксический анализ
- V) Семантический анализ
- G) Прагматический анализ

- 1) Определение смысла текста
- 2) Определение частей речи и форм слов
- 3) Анализ структуры предложения
- 4) Учет контекста и целей общения

Поле для ответа:

20. Прочтайте задание и установите соответствие.

Соотнесите методы представления текста (слева) с их характеристиками (справа).

- A) Bag of Words
- B) TF-IDF
- V) word2vec
- G) One-hot encoding

- 1) Векторное представление с учетом контекста
- 2) Подсчет количества слов без учета порядка
- 3) Взвешивание слов с учетом частоты и редкости
- 4) Вектор с единицей в позиции слова и нулями в остальных

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

21. Прочтайте текст и напишите обоснованный ответ.

В чем заключается основная идея модели BERT и почему она эффективна для обработки естественного языка?

Поле для ответа:

22. Прочтайте текст и напишите обоснованный ответ.

Что такое лемматизация и чем она отличается от стемминга?

Поле для ответа:

ОПК-1 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач

Задания закрытого типа

Базовый уровень сложности

23. Прочтайте задание, выберите правильный ответ.

Какой подход используется в модели seq2seq?

- 1) Кодировщик-декодировщик
- 2) Кластеризация признаков
- 3) Случайное разбиение данных
- 4) Формирование деревьев решений

Поле для ответа:

24. Прочтайте задание, выберите правильный ответ.

Что такое "токен" в контексте языковой модели?

- 1) Сжатый архив текста
- 2) Минимальная единица текста, например, слово или часть слова
- 3) Часть синтаксического дерева
- 4) Категория документа

Поле для ответа:

25. Прочтайте задание, выберите правильный ответ.

Что делает механизм внимания (attention) в модели Transformer?

- 1) Упорядочивает слова по алфавиту
- 2) Сравнивает документы между собой
- 3) Выделяет важные слова в контексте
- 4) Проводит синтаксический анализ

Поле для ответа:

Повышенный уровень сложности

26. Прочтайте задание и установите соответствие.

Соотнесите архитектурные компоненты Transformer (слева) с их функциями (справа).

- A) Механизм внимания (Attention)
- Б) Позиционные кодировки
- В) Многоголовое внимание (Multi-head Attention)
- Г) Нормализация слоя (Layer Normalization)

- 1) Добавление информации о порядке слов
- 2) Выделение релевантных частей входа
- 3) Параллельное внимание по разным подпространствам
- 4) Стабилизация обучения сети

Поле для ответа:

27. Прочтайте задание и установите соответствие.

Соотнесите виды векторных моделей текста (слева) с их характеристиками (справа).

- A) One-hot encoding
- Б) Frequency model
- В) Bag of Words
- Г) TF-IDF

- 1) Представляет слова как уникальные векторы с единицей и нулями
- 2) Учитывает количество вхождений слов
- 3) Не учитывает порядок слов и контекст
- 4) Взвешивает слова по частоте и редкости в корпусе

Поле для ответа:

28. Прочтайте задание и установите соответствие.

Соотнесите понятия, связанные с моделями seq2seq (слева), с их функциями (справа).

- A) Кодировщик (Encoder)
- Б) Декодировщик (Decoder)
- В) Контекстный вектор
- Г) Механизм внимания

- 1) Преобразует входную последовательность в вектор
- 2) Генерирует выходную последовательность
- 3) Представляет сжатую информацию входа для декодера
- 4) Позволяет фокусироваться на разных частях входа при генерации

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

29. Прочтайте текст и напишите обоснованный ответ.

Что такое токенизация и зачем она нужна в обработке естественного языка?

Поле для ответа:

30. Прочтайте текст и напишите обоснованный ответ.

В чем разница между синтаксическим и морфологическим анализом?

Поле для ответа:

ОПК-3 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

Задания закрытого типа

Базовый уровень сложности

31. Прочтайте задание, выберите правильный ответ.

Что такое токен в обработке естественного языка?

- 1) Единица текста, например слово или символ
- 2) Название алгоритма
- 3) Тип языковой модели
- 4) Вид нейронной сети

Поле для ответа:

32. Прочтайте задание, выберите правильный ответ.

Какую задачу решает морфологический анализ?

- 1) Определение синтаксической структуры предложения
- 2) Определение частей речи и форм слов
- 3) Построение семантических связей
- 4) Токенизация текста

Поле для ответа:

33. Прочтайте задание, выберите правильный ответ.

Что такое N-грамма в NLP?

- 1) Последовательность из N символов или слов
- 2) Метод классификации текста
- 3) Алгоритм для обучения языковых моделей
- 4) Вид нейронной сети

Поле для ответа:

Повышенный уровень сложности

34. Прочтайте задание и установите соответствие.

Соотнесите методы векторного представления текста (слева) с их характеристиками (справа).

А) One-hot encoding
Б) Bag of Words

В) TF-IDF
Г) word2vec

- 1) Взвешивание слов с учетом частоты в документе и редкости в корпусе
- 2) Вектор с единицей на позиции слова и нулями в остальных
- 3) Представление слов в плотном векторном пространстве с учетом контекста
- 4) Модель, учитывающая количество слов без порядка

Поле для ответа:

35. Прочтайте задание и установите соответствие.

Соотнесите этапы лингвистического анализа (слева) с их описаниями (справа).

А) Токенизация
Б) Морфологический анализ
В) Синтаксический анализ
Г) Семантический анализ

- 1) Анализ смысловых отношений между словами
- 2) Разбиение текста на слова и знаки препинания
- 3) Определение частей речи и форм слов
- 4) Определение структуры предложения и связей между словами

Поле для ответа:

36. Прочтайте задание и установите соответствие.

Соотнесите задачи NLP (слева) с примерами (справа).

А) Классификация текста
Б) Машинный перевод
В) Распознавание речи
Г) Анализ тональности

- 1) Определение положительной или отрицательной оценки отзыва
- 2) Преобразование аудио в текст
- 3) Определение категории новости
- 4) Перевод текста с английского на русский

Поле для ответа:

Задания открытого типа
Высокий уровень сложности

37. Прочтайте текст и напишите обоснованный ответ.

Что такое токенизация и почему она важна для обработки естественного языка?

Поле для ответа:

38. Прочтайте текст и напишите обоснованный ответ.

В чем разница между стеммингом и лемматизацией?

Поле для ответа:

ОПК-4 Способен применять на практике новые научные принципы и методы исследований

Задания закрытого типа
Базовый уровень сложности

39. Прочтайте задание, выберите правильный ответ.

Что такое лемматизация?

- 1) Приведение слова к начальной форме
- 2) Разбиение текста на предложения
- 3) Обучение модели на большом корпусе
- 4) Удаление стоп-слов

Поле для ответа:

40. Прочтайте задание, выберите правильный ответ.

Что означает термин "галлюцинации" в больших языковых моделях?

- 1) Генерация неверной информации
- 2) Процесс токенизации
- 3) Обучение модели
- 4) Оптимизация гиперпараметров

Поле для ответа:

41. Прочтайте задание, выберите правильный ответ.

Что такое модель word2vec?

- 1) Модель для получения распределенных представлений слов
- 2) Метод токенизации
- 3) Алгоритм кластеризации

- 4) Синтаксический парсер

Поле для ответа:

Повышенный уровень сложности

42. Прочитайте задание и установите соответствие.

Соотнесите понятия, связанные с большими языковыми моделями (слева), с их определениями (справа).

- A) Токенизация
- B) Transfer learning
- В) Галлюцинации
- Г) Prompt engineering

- 1) Создание и оптимизация запросов к модели
- 2) Использование заранее обученной модели для новой задачи
- 3) Разбиение текста на минимальные единицы
- 4) Генерация модели неверной информации

Поле для ответа:

43. Прочитайте задание и установите соответствие.

Соотнесите компоненты архитектуры Transformer (слева) с их функциями (справа).

- A) Многоголовое внимание (Multi-head Attention)
- Б) Позиционное кодирование
- В) Нормализация слоя (Layer Normalization)
- Г) Кодировщик (Encoder)

- 1) Добавляет информацию о порядке слов
- 2) Обрабатывает входные данные и извлекает признаки
- 3) Улучшает обучение, стабилизируя распределение активаций
- 4) Позволяет модели фокусироваться на разных частях входа одновременно

Поле для ответа:

44. Прочитайте задание и установите соответствие.

Соотнесите виды моделей представления текста (слева) с их особенностями (справа).

- A) Модель частоты (Frequency model)
- Б) One-hot encoding
- В) N-граммная модель
- Г) Распределенные представления

- 1) Учитывает последовательность из N слов

- 2) Представляет слова плотными векторами
- 3) Использует векторы с единичным элементом
- 4) Опирается на частотные характеристики слов

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

45. Прочтите текст и напишите обоснованный ответ.

Что такое языковая модель и какую задачу она решает?

Поле для ответа:

46. Прочтите текст и напишите обоснованный ответ.

Объясните, что такое transfer learning и зачем он используется в NLP.

Поле для ответа:

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. ИД.УК-1.5.	Опрос, диспут, практическое задание, тест
УК-4	ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6.	Опрос, диспут, практическое задание, тест
УК-5	ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3.	Опрос, диспут, практическое задание, тест
ОПК-1	ИД.ОПК-1.1. ИД.ОПК-1.2.	Опрос, диспут, практическое задание, тест
ОПК-3	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4.	Опрос, диспут, практическое задание, тест
ОПК-4	ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	Опрос, диспут, практическое задание, тест

Таблица 9

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
Практическое задание	<p>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:</p> <ol style="list-style-type: none"> 1. Анализировать проблемную ситуацию, определять пробелы в информации, оценивать надёжность источников информации, разрабатывать стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строить сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения 2. Устанавливать контакты и организовывать общение в соответствии с потребностями совместной деятельности, используя современные коммуникационные технологии, составлять в соответствии с нормами русского языка деловую документацию разных жанров, в том числе для академических и профессиональных целей на иностранном языке. 3. Обсуждать и представлять результаты исследовательской и проектной деятельности, участвовать в дискуссиях на различных публичных мероприятиях, выбирая подходящий формат 4. Анализировать важнейшие идеологические и ценностные системы, сформировавшиеся в ходе исторического развития. Выстраивать социальное и профессиональное взаимодействие с учётом особенностей деловой и общей культуры представителей других этносов и конфессий, различных социальных групп на основе идеологических и ценностных систем, обеспечивать создание недискриминационной среды для участников межкультурного взаимодействия 5. Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий 6. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров 7. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний
Опрос	<p>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:</p> <ol style="list-style-type: none"> 1. Анализировать проблемную ситуацию, определять пробелы в информации, оценивать надёжность источников информации, разрабатывать стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строить сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения 2. Устанавливать контакты и организовывать общение в соответствии с потребностями совместной деятельности, используя современные коммуникационные технологии, составлять в соответствии с нормами русского языка деловую документацию разных жанров, в том числе для академических и профессиональных целей на иностранном языке. 3. Обсуждать и представлять результаты исследовательской и проектной деятельности, участвовать в дискуссиях на различных публичных мероприятиях, выбирая подходящий формат 4. Анализировать важнейшие идеологические и ценностные системы, сформировавшиеся в ходе исторического развития. Выстраивать социальное и профессиональное взаимодействие с учётом особенностей деловой и общей культуры представителей других этносов и конфессий, различных социальных групп на основе идеологических и ценностных систем, обеспечивать создание недискриминационной среды для участников межкультурного взаимодействия

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
	<p>5. Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий</p> <p>6. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров</p> <p>7. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний</p>
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> <p>1. Анализировать проблемную ситуацию, определять пробелы в информации, оценивать надёжность источников информации, разрабатывать стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строить сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения</p> <p>2. Устанавливать контакты и организовывать общение в соответствии с потребностями совместной деятельности, используя современные коммуникационные технологии, составлять в соответствии с нормами русского языка деловую документацию разных жанров, в том числе для академических и профессиональных целей на иностранном языке.</p> <p>3. Обсуждать и представлять результаты исследовательской и проектной деятельности, участвовать в дискуссиях на различных публичных мероприятиях, выбирая подходящий формат</p> <p>4. Анализировать важнейшие идеологические и ценностные системы, сформировавшиеся в ходе исторического развития. Выстраивать социальное и профессиональное взаимодействие с учётом особенностей деловой и общей культуры представителей других этносов и конфессий, различных социальных групп на основе идеологических и ценностных систем, обеспечивать создание недискриминационной среды для участников межкультурного взаимодействия</p> <p>5. Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий</p> <p>6. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров</p> <p>7. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний</p>
Тест	<p>Магистрант в ходе подготовки и выполнения теста показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности:</p> <p>1. Анализировать проблемную ситуацию, определять пробелы в информации, оценивать надёжность источников информации, разрабатывать стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строить сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения</p> <p>2. Устанавливать контакты и организовывать общение в соответствии с потребностями совместной деятельности, используя современные коммуникационные технологии, составлять в соответствии с нормами русского языка деловую</p>

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
	<p>документацию разных жанров, в том числе для академических и профессиональных целей на иностранном языке.</p> <p>3. Обсуждать и представлять результаты исследовательской и проектной деятельности, участвовать в дискуссиях на различных публичных мероприятиях, выбирая подходящий формат</p> <p>4. Анализировать важнейшие идеологические и ценностные системы, сформировавшиеся в ходе исторического развития. Выстраивать социальное и профессиональное взаимодействие с учётом особенностей деловой и общей культуры представителей других этносов и конфессий, различных социальных групп на основе идеологических и ценностных систем, обеспечивать создание недискриминационной среды для участников межкультурного взаимодействия</p> <p>5. Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий</p> <p>6. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров</p> <p>7. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний</p>

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: <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. 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://proed.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. Google. Книги: <https://books.google.com>
2. Internet Archive: <https://archive.org>
3. Koob.ru. Электронная библиотека «Куб»: <http://www.koob.ru/philosophy/>
4. SOC.LIB.RU. Социология, психология, управление: <http://soc.lib.ru/>
5. Socioline.ru. Учебники, монографии по социологии: <http://socioline.ru>
6. Библиотека Гумер – гуманитарные науки: <http://www.gumer.info>
7. ЕНИП — Электронная библиотека «Научное наследие России»: <http://e-heritage.ru/>
8. Интеллектуальная Россия: <http://www.intelros.ru/>
9. Национальная электронная библиотека НЭБ: <http://www.rusneb.ru>
10. Неприкосновенный запас: <http://magazines.russ.ru/nz/>
11. Президентская библиотека: <http://www.prlib.ru>
12. Российская государственная библиотека: <http://www.rsl.ru/>
13. 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, научометрическая база данных: <http://elibrary.ru>;

2. **Университетская информационная система РОССИЯ** — база электронных ресурсов для учебных программ и исследовательских проектов в области социально-гуманитарных наук: <http://www.uisrussia.msu.ru/>;

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

Электронные библиотечные системы:

1. **Znaniум.com** — Электронная библиотечная система (ЭБС) — <http://znanium.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.

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

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 электронным ресурсам библиотеки Университета 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
«Текстовые данные»

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, 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 1
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	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Опрос 1	зачтено/ не зачтено
Распознавание текстов. Создание процедуры превращения изображений или PDF в корпус	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1)	Практическое задание 1	зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
		ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)		
Закон Ципфа. Издержки токенизации	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Практическое задание 2	зачтено/ не зачтено
Извлечение сущностей из текстовых данных	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Практическое задание 3	зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
		ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.			
Разреженное векторное представление текстовых данных	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Практическое задание 4	зачтено/ не зачтено
Уплотненное векторное представление текстовых данных	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5) В (УК-5) З (ОПК-1) У (ОПК-1) В (ОПК-1) З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Практическое задание 5	зачтено/ не зачтено
Поиск по представлениям текстовых данных	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3.	З (УК-1) У (УК-1) В (УК-1) З (УК-4) У (УК-4) В (УК-4) З (УК-5) У (УК-5)	Практическое задание 6 Диспут 1	зачтено/ не зачтено зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
		ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	В (УК-5) 3 (ОПК-1) У (ОПК-1) В (ОПК-1) 3 (ОПК-3) У (ОПК-3) В (ОПК-3) 3 (ОПК-4) У (ОПК-4) В (ОПК-4)		

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 зачленено, 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) – зачленено
Опрос	ответ отсутствует или является однозначным, или содержит существенные ошибки – не зачленено 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 — зачленено
Диспут	пассивность, участие без представления аргументов и обоснования точки зрения, несформированность навыков профессиональной коммуникации в группе — не зачленено представление аргументированной научной позиции, обоснование точки зрения в диспут, демонстрация навыков профессиональной коммуникации в группе — зачленено

2 Контрольные задания для текущей аттестации

Материал опросов, диспутов, практических заданий:

Тема 1. Корпусы текстовых данных.

Опрос 1:

Выбрать из корпуса Taiga статьи определенного интернет-сми, упоминающие определенное слово. Например, СМИ - интернет-издание Лента.ру, слово – интеллект.

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.

**DISTRICT WISE ESTIMATES OF MARKETABLE SURPLUS / DEFICIT OF RICE
DURING 2012-13**

Sl. No.	District	Projected Population for 2012-13 (In lakhs)	Adult Equivalent to 88% (In lakhs)	Total Consumption requirement (@ 400gms/adult/day) (In Lakh tonnes)	Total Requirement (Including seeds, feeds & wastage) (In Lakh tonnes)	Production (Rice) (In Lakh tonnes)			Surplus/Deficit (In Lakh tonnes)	
						Kharif	Rabi	Total	Rice	Paddy
1	Balasore	23.65	20.81	3.04	3.47	2.78	0.86	3.64	0.17	0.25
2	Bhadrak	15.34	13.50	1.97	2.25	3.50	0.05	3.55	1.30	1.94
3	Bolangir	17.01	14.97	2.19	2.50	6.23	0.10	6.33	3.83	5.72
4	Subarnapur	6.70	5.90	0.86	0.98	4.48	1.13	5.61	4.63	6.91
5	Cuttack	26.63	23.43	3.42	3.91	3.75	0.06	3.81	-0.10	-0.15

Тема 3. Закон Циффа. 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 сущности, удовлетворяющие определенному требованию. Корпус: параллельные переводы Всеобщей декларации о правах человека, тип сущности: VERB.

Тема 5. Разреженное векторное представление текстовых данных.

Практическое задание 4:

Оценить на терм-документарной матрице вектора документов предложенными алгоритмами: LSA, word2vec, fasttext.

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 дневниковых записей проекта «Прожито», преобразованном в формат терм-документной матрицы, ближайших соседей.

4.

т.к. служила лишь прикрытием батареи и непосредственного участия в бою не принимала. Оба офицера Алевский и Вандышев плачут. Дал им бром урану. Вандышев сел на моя лошадь, когда слез Бурмейстер. Офицеры так же не знают куда идти. Идем ночью, тихо. Подходят разбежавшиеся солдаты других рот. В час ночи приходим в Гостынин. Бурмейстер, как старший, отправился в штаб корпуса, уже складывающийся, узнать, куда нам идти. Нам командир бригады приказал вернуться назад и ждать отступающий полк, но какой же дорогой он будет отступать сказать в штабе не могли. Дивизионный врач приказал нам отходить по на-

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

Форма промежуточной аттестации – зачет, выставляемый на основе оценки тестирования.

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

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Тест включает 20 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте -1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балла; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Задания открытого типа

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F = \frac{100}{K} * \left(\frac{x_1}{k_1} + \frac{x_2}{k_2} + \dots + \frac{x_n}{k_n} \right),$$

где F – итоговое количество баллов за тест,

K – количество осваиваемых в рамках дисциплины компетенций,

k_n – максимально возможное количество баллов за вопросы по компетенции,

x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

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 соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)	Критерии оценивания	Оценка
Зачет / тест	УК-1 УК-4 УК-5 ОПК-1 ОПК-3 ОПК-4	ИД.УК-1.1. ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3.	3 (УК-1) У (УК-1) В (УК-1) 3 (УК-4) У (УК-4) В (УК-4) 3 (УК-5) У (УК-5)	41-100% правильных ответов 0-40% правильных ответов	Зачтено Не засчитано

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
		ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6. ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. ИД.ОПК-1.1. ИД.ОПК-1.2. ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	В (УК-5) 3 (ОПК-1) У (ОПК-1) В (ОПК-1) 3 (ОПК-3) У (ОПК-3) В (ОПК-3) 3 (ОПК-4) У (ОПК-4) В (ОПК-4)		

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 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	не зачтено
2 (неудовлетворительно)	40 и менее	

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «зачтено», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в бинарной системе «не зачтено», показывают несформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

4 Задания к промежуточной аттестации

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

Задания закрытого типа

Базовый уровень сложности

1. Прочитайте задание, выберите правильный ответ.

Что такое токенизация в обработке естественного языка?

- 1) Объединение слов в предложения
- 2) Перевод текста с одного языка на другой
- 3) Разделение текста на части, например, слова или предложения

- 4) Удаление всех гласных из текста

Поле для ответа:

2. Прочтайте задание, выберите правильный ответ.

Что такое стемминг?

- 1) Преобразование чисел в текст
- 2) Перевод текста в верхний регистр
- 3) Приведение слова к его корню
- 4) Разделение текста на абзацы

Поле для ответа:

3. Прочтайте задание, выберите правильный ответ.

Какую задачу решает морфологический разбор?

- 1) Определение темпа речи
- 2) Определение части речи и формы слова
- 3) Перевод текста на другой язык
- 4) Анализ эмоциональной окраски текста

Поле для ответа:

4. Прочтайте задание, выберите правильный ответ.

Что такое частотный словарь в NLP?

- 1) Список редких слов
- 2) Таблица с переводами слов
- 3) Перечень слов с указанием частоты их употребления
- 4) Список синонимов слов

Поле для ответа:

5. Прочтайте задание, выберите правильные ответы.

Какие задачи входят в область обработки естественного языка (NLP)?

- 1) Распознавание речи
- 2) Кластеризация изображений
- 3) Машинный перевод
- 4) Генерация числовых таблиц

Поле для ответа:

6. Прочтайте задание, выберите правильные ответы.

Какие этапы входят в лингвистический анализ текста?

- 1) Морфологический анализ
- 2) Сжатие изображений
- 3) Синтаксический анализ
- 4) Предсказание временного ряда

Поле для ответа:

7. Прочтайте задание, выберите правильные ответы.

Что включает в себя этап предварительной обработки текста?

- 1) Удаление пунктуации
- 2) Объединение видеофайлов
- 3) Приведение слов к начальной форме
- 4) Генерация случайных чисел

Поле для ответа:

8. Прочтайте задание, выберите правильные ответы.

Какие операции выполняются при морфологическом анализе текста?

- 1) Определение частей речи
- 2) Определение синтаксических зависимостей
- 3) Нормализация слов
- 4) Построение деревьев решений

Поле для ответа:

Повышенный уровень сложности

9. Прочтайте задание и установите соответствие.

Соотнесите задачи обработки естественного языка (слева) с их примерами (справа).

- A) Классификация текста
B) Машинный перевод
В) Извлечение информации
Г) Распознавание именованных сущностей

- 1) Перевод с английского на русский
- 2) Определение категории текста
- 3) Поиск фактов в тексте
- 4) Выделение имен, локаций, организаций

Поле для ответа:

10. Прочтайте задание и установите соответствие.

Соотнесите этапы лингвистического анализа (слева) с их описаниями (справа).

- A) Морфологический анализ
Б) Синтаксический анализ
В) Семантический анализ

- 1) Определение структуры предложения
- 2) Определение частей речи и форм слов
- 3) Анализ смысловых связей между словами

Поле для ответа:

11. Прочтайте задание и установите соответствие.

Соотнесите этапы предварительной обработки текста (слева) с их функциями (справа).

- A) Токенизация
- Б) Лемматизация
- В) Удаление стоп-слов

- 1) Разделение текста на слова или фразы
- 2) Приведение слов к начальной форме
- 3) Исключение часто встречающихся, но малозначимых слов

Поле для ответа:

12. Прочтайте задание и установите соответствие.

Соотнесите модели представления текста (слева) с их характеристиками (справа).

- A) One-hot encoding
- Б) Частотная модель Bag of Words
- В) TF-IDF
- Г) word2vec

- 1) Учитывает частотность и редкость слов в корпусе
- 2) Представляет слово единичным вектором без учета контекста
- 3) Учитывает только количество вхождений слов
- 4) Использует распределённые представления слов

Поле для ответа:

13. Прочтайте задание и установите последовательность.

Расположите этапы лингвистического анализа текста в правильном порядке:

- 1) Морфологический анализ
- 2) Синтаксический анализ
- 3) Семантический анализ
- 4) Прагматический анализ

Поле для ответа:

14. Прочтайте задание и установите последовательность.

Расположите этапы морфологического анализа в правильном порядке:

- 1) Определение границ слов
- 2) Определение начальной формы слова
- 3) Определение части речи
- 4) Определение морфологических признаков (число, падеж и т. д.)

Поле для ответа:

15. Прочтите задание и установите последовательность.

Расположите шаги работы частотной модели Bag of Words в правильном порядке:

- 1) Составление словаря корпуса
- 2) Подсчёт количества вхождений слов
- 3) Преобразование текста в вектор
- 4) Сбор текстов корпуса

Поле для ответа:

16. Прочтите задание и установите последовательность.

Расположите шаги построения модели TF-IDF в правильном порядке:

- 1) Расчёт частоты слова в документе
- 2) Расчёт обратной частоты документа
- 3) Умножение TF на IDF
- 4) Формирование вектора документа

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

17. Прочтайте текст и напишите обоснованный ответ.

В чём заключается задача морфологического анализа?

Поле для ответа:

18. Прочтайте текст и напишите обоснованный ответ.

Что такое токенизация и зачем она используется в NLP?

Поле для ответа:

19. Прочтайте текст и напишите обоснованный ответ.

Чем модель Bag of Words отличается от модели word2vec?

Поле для ответа:

20. Прочтайте текст и напишите обоснованный ответ.

Для чего используется модель TF-IDF при анализе текста?

Поле для ответа:

21. Прочтайте текст и напишите обоснованный ответ.

Что делает архитектура Transformer важной для языкового моделирования?

Поле для ответа:

22. Прочтите текст и напишите обоснованный ответ.

Что такое prompt engineering и зачем он нужен при работе с LLM?

Поле для ответа:

УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия

Задания закрытого типа

Базовый уровень сложности

23. Прочтайте задание, выберите правильный ответ.

Что такое стоп-слова в обработке текста?

Самые важные слова в тексте

Слова, которые подчеркиваются

Часто употребляемые слова, не несущие смысловой нагрузки

Слова, содержащие ошибки

Поле для ответа:

24. Прочтайте задание, выберите правильный ответ.

Какую задачу решает лемматизация?

- 5) Добавление новых слов в текст
- 6) Перевод текста в аудио
- 7) Преобразование слова к его нормальной форме
- 8) Удаление всех прилагательных из текста

Поле для ответа:

25. Прочтайте задание, выберите правильный ответ.

Какая из перечисленных задач относится к обработке естественного языка?

- 5) Обработка изображений
- 6) Сегментация клиентов
- 7) Синтаксический разбор текста
- 8) Предсказание движения цены акций

Поле для ответа:

26. Прочтайте задание, выберите правильный ответ.

Какой подход к представлению текста использует подсчёт количества вхождений слов?

- 1) One-hot encoding
- 2) TF-IDF
- 3) Transformer
- 4) word2vec

Поле для ответа:

27. Прочтайте задание, выберите правильные ответы.

Какие утверждения верны для синтаксического анализа?

- 1) Выявляет структуру предложения
- 2) Используется для кодирования изображений
- 3) Находит связи между словами
- 4) Работает только с аудиоданными

Поле для ответа:

28. Прочтайте задание, выберите правильные ответы.

Какие характеристики справедливы для модели представления текста Bag of Words?

- 1) Не учитывает порядок слов
- 2) Использует плотные вектора
- 3) Учитывает частотность слов
- 4) Использует контекст слов

Поле для ответа:

29. Прочтайте задание, выберите правильные ответы.

Что верно для модели TF-IDF?

- 1) Учитывает важность слова в документе и корпусе
- 2) Игнорирует частотность слов
- 3) Использует веса, а не бинарные значения
- 4) Основана на последовательности пикселей

Поле для ответа:

30. Прочтайте задание, выберите правильные ответы.

Какие понятия связаны с большими языковыми моделями?

- 1) Токенизация
- 2) Сжатие звуковых сигналов
- 3) Prompt engineering
- 4) Построение гистограмм

Поле для ответа:

Повышенный уровень сложности

31. Прочтайте задание и установите соответствие.

Соотнесите типы векторных моделей (слева) с их описаниями (справа).

- A) One-hot encoding
B) Frequency model
C) TF-IDF

- 1) Отображает количество вхождений слов в документ
- 2) Отражает важность слова с учетом его встречаемости в корпусе
- 3) Представляет слова в виде векторов с одним значением "1"

Поле для ответа:

32. Прочтайте задание и установите соответствие.

Соотнесите модели представления текста (слева) с их описаниями (справа).

- A) TF-IDF
B) N-gram model
B) word2vec
Г) Bag of Words

- 1) Учитывает частотность слов и их уникальность в корпусе
- 2) Представляет текст как последовательности из N слов
- 3) Использует плотные векторы, обученные на контексте слов
- 4) Учитывает только количество вхождений слов без контекста

Поле для ответа:

33. Прочтайте задание и установите соответствие.

Соотнесите компоненты архитектуры Transformer (слева) с их функциями (справа).

- A) Механизм внимания
Б) Многоголовое внимание
В) Позиционные кодировки

- 1) Позволяет учитывать порядок слов
- 2) Усиливает способность модели обращать внимание на разные части входа
- 3) Вычисляет важность слов друг относительно друга

Поле для ответа:

34. Прочтайте задание и установите соответствие.

Соотнесите понятия, связанные с большими языковыми моделями (слева), с их определениями (справа).

- A) Transfer learning
Б) Prompt engineering
В) Токенизация
Г) Галлюцинации

- 1) Разделение текста на минимальные единицы
- 2) Передача знаний от одной задачи к другой
- 3) Настройка запросов к языковой модели
- 4) Генерация правдоподобного, но вымышленного ответа

Поле для ответа:

35. Прочтите задание и установите последовательность.

Расположите этапы работы модели N-gram в правильном порядке:

- 1) Разделение текста на токены
- 2) Формирование последовательностей по N токенов
- 3) Подсчёт частот появления N-грамм
- 4) Построение вероятностной модели

Поле для ответа:

36. Прочтите задание и установите последовательность.

Расположите основные этапы архитектуры Transformer при обработке входного текста:

1. Преобразование слов в эмбеддинги
2. Добавление позиционных кодировок
3. Пропуск через блоки внимания
4. Получение выходного вектора

Поле для ответа:

37. Прочтите задание и установите последовательность.

Расположите этапы работы модели seq2seq в правильном порядке:

1. Кодирование входной последовательности
2. Передача скрытого вектора
3. Генерация выходной последовательности
4. Обработка выходных данных

Поле для ответа:

38. Прочтите задание и установите последовательность.

Расположите этапы применения большой языковой модели при генерации ответа в правильном порядке:

1. Ввод запроса (prompt)
2. Токенизация запроса
3. Генерация ответа моделью
4. Декодирование токенов в текст

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

39. Прочтите текст и напишите обоснованный ответ.

Что такое обработка естественного языка и для чего она применяется?

Поле для ответа:

40. Прочтите текст и напишите обоснованный ответ.
Что происходит на этапе синтаксического анализа текста?

Поле для ответа:

41. Прочтите текст и напишите обоснованный ответ.
Что такое языковая модель и какую задачу она решает?

Поле для ответа:

42. Прочтите текст и напишите обоснованный ответ.
Что означает термин "галлюцинации" в контексте больших языковых моделей?

Поле для ответа:

43. Прочтите текст и напишите обоснованный ответ.

Как работает модель seq2seq и для каких задач она обычно применяется?

Поле для ответа:

44. Прочтите текст и напишите обоснованный ответ.

Что такое размер контекста в больших языковых моделях и как он влияет на качество работы?

Поле для ответа:

УК-5 Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия

Задания закрытого типа

Базовый уровень сложности

45. Прочтайте задание, выберите правильный ответ.
Что выполняется на этапе токенизации текста?

- 1) Объединение документов в один
- 2) Преобразование слов в числа
- 3) Разделение текста на слова или фразы
- 4) Классификация текста по категориям

Поле для ответа:

46. Прочтайте задание, выберите правильный ответ.

Какой из методов представляет слова в виде плотных векторов, обученных на контексте?

- 1) Bag of Words
- 2) TF-IDF
- 3) word2vec
- 4) One-hot encoding

Поле для ответа:

47. Прочтайте задание, выберите правильный ответ.

Что учитывает модель TF-IDF при расчёте весов слов?

- 1) Порядок слов в предложении
- 2) Обратную частоту слова в корпусе
- 3) Количество предложений в тексте
- 4) Синтаксическую структуру

Поле для ответа:

48. Прочтайте задание, выберите правильный ответ.

Какая архитектура лежит в основе модели BERT?

- 1) LSTM
- 2) CNN
- 3) Transformer
- 4) RNN

Поле для ответа:

49. Прочтайте задание, выберите правильные ответы.

Какие задачи решаются с помощью синтаксического анализа?

- 1) Определение структуры предложения
- 2) Определение начальной формы слова
- 3) Выделение грамматических связей между словами
- 4) Определение эмоциональной окраски текста

Поле для ответа:

50. Прочтайте задание, выберите правильные ответы.

Какие этапы входят в предварительную обработку текста?

- 1) Токенизация
- 2) Сегментация документа
- 3) Лемматизация
- 4) Кластеризация

Поле для ответа:

51. Прочтайте задание, выберите правильные ответы.

Какие модели относятся к классическим способам представления текста?

- 1) Bag of Words
- 2) word2vec
- 3) One-hot encoding
- 4) Transformer

Поле для ответа:

52. Прочтайте задание, выберите правильные ответы.

Какие характеристики учитывает модель TF-IDF?

- 1) Частоту слова в документе
- 2) Порядок слов в предложении
- 3) Обратную частоту слова в корпусе
- 4) Семантическую связь слов

Поле для ответа:

Повышенный уровень сложности

53. Прочтайте задание и установите соответствие.

Соотнесите этапы лингвистического анализа (слева) с их описаниями (справа).

- A) Морфологический анализ
- B) Синтаксический анализ
- V) Семантический анализ
- G) Прагматический анализ

- 1) Определение смысла текста
- 2) Определение частей речи и форм слов
- 3) Анализ структуры предложения
- 4) Учет контекста и целей общения

Поле для ответа:

54. Прочтайте задание и установите соответствие.

Соотнесите методы представления текста (слева) с их характеристиками (справа).

- A) Bag of Words
- B) TF-IDF
- V) word2vec
- G) One-hot encoding

- 1) Векторное представление с учетом контекста
- 2) Подсчет количества слов без учета порядка
- 3) Взвешивание слов с учетом частоты и редкости

- 4) Вектор с единицей в позиции слова и нулями в остальных

Поле для ответа:

55. Прочтайте задание и установите соответствие.

Соотнесите этапы предварительной обработки текста (слева) с их действиями (справа).

- A) Токенизация
- Б) Лемматизация
- В) Удаление стоп-слов
- Г) Приведение к нижнему регистру

- 1) Разбиение текста на отдельные слова или фразы
- 2) Приведение слов к начальной форме
- 3) Исключение часто встречающихся служебных слов
- 4) Приведение всех букв к маленькому регистру

Поле для ответа:

56. Прочтайте задание и установите соответствие.

Соотнесите понятия, связанные с большими языковыми моделями (слева), с их определениями (справа).

- A) Токенизация
- Б) Transfer learning
- В) Prompt engineering
- Г) Галлюцинации

- 1) Разбиение текста на минимальные единицы
- 2) Использование заранее обученной модели для новой задачи
- 3) Создание эффективных запросов к модели
- 4) Генерация неправильной или вымышленной информации

Поле для ответа:

57. Прочтайте задание и установите последовательность.

Расположите этапы морфологического анализа в правильном порядке:

- 1) Определение части речи
- 2) Определение начальной формы слова
- 3) Определение морфологических признаков (число, падеж и т. д.)
- 4) Определение границ слов

Поле для ответа:

58. Прочтайте задание и установите последовательность.

Расположите этапы синтаксического анализа в правильном порядке:

- 1) Построение синтаксического дерева
- 2) Определение грамматических связей между словами
- 3) Токенизация текста
- 4) Определение структуры предложения

Поле для ответа:

59. Прочтайте задание и установите последовательность.

Расположите этапы работы модели TF-IDF в правильном порядке:

- 1) Подсчет количества вхождений слова в документ
- 2) Подсчет обратной частоты слова в корпусе
- 3) Умножение TF на IDF
- 4) Преобразование текста в набор слов

Поле для ответа:

60. Прочтайте задание и установите последовательность.

Расположите этапы работы модели word2vec (CBOW) в правильном порядке:

- 1) Выбор контекстных слов
- 2) Прогнозирование центрального слова
- 3) Обновление весов модели
- 4) Повторение процесса для всего корпуса

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

61. Прочтайте текст и напишите обоснованный ответ.

В чем заключается основная идея модели BERT и почему она эффективна для обработки естественного языка?

Поле для ответа:

62. Прочтайте текст и напишите обоснованный ответ.

Что такое лемматизация и чем она отличается от стемминга?

Поле для ответа:

63. Прочтайте текст и напишите обоснованный ответ.

Как работает механизм внимания (attention) в архитектуре Transformer и зачем он нужен?

Поле для ответа:

64. Прочтайте текст и напишите обоснованный ответ.

Что такое N-грамма и как она используется в моделях обработки текста?

Поле для ответа:

65. Прочтайте текст и напишите обоснованный ответ.

Почему предварительная обработка текста важна для задач NLP?

Поле для ответа:

66. Прочтайте текст и напишите обоснованный ответ.

Что такое transfer learning и как он применяется в больших языковых моделях?

Поле для ответа:

ОПК-1 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач

Задания закрытого типа

Базовый уровень сложности

67. Прочтайте задание, выберите правильный ответ.

Какой подход используется в модели seq2seq?

- 1) Кодировщик-декодировщик
- 2) Кластеризация признаков
- 3) Случайное разбиение данных
- 4) Формирование деревьев решений

Поле для ответа:

68. Прочтайте задание, выберите правильный ответ.

Что такое "токен" в контексте языковой модели?

- 1) Сжатый архив текста
- 2) Минимальная единица текста, например, слово или часть слова
- 3) Часть синтаксического дерева
- 4) Категория документа

Поле для ответа:

69. Прочтайте задание, выберите правильный ответ.

Что делает механизм внимания (attention) в модели Transformer?

- 1) Упорядочивает слова по алфавиту
- 2) Сравнивает документы между собой
- 3) Выделяет важные слова в контексте
- 4) Проводит синтаксический анализ

Поле для ответа:

70. Прочтайте задание, выберите правильный ответ.

Что означает термин "transfer learning" в NLP?

- 1) Передача модели по сети
- 2) Повторное обучение модели с нуля
- 3) Использование предварительно обученной модели для новой задачи
- 4) Обучение модели на случайных данных

Поле для ответа:

71. Прочтайте задание, выберите правильные ответы.

Какие понятия связаны с большими языковыми моделями?

- 1) Transfer learning
- 2) Decision tree
- 3) Prompt engineering
- 4) K-means clustering

Поле для ответа:

72. Прочтайте задание, выберите правильные ответы.

Какие из перечисленных этапов относятся к обучению модели word2vec?

- 1) Определение контекстных слов
- 2) Преобразование текста в синтаксическое дерево
- 3) Обновление весов модели
- 4) Лемматизация текста

Поле для ответа:

73. Прочтайте задание, выберите правильные ответы.

Какие функции выполняет архитектура Transformer?

- 1) Механизм внимания
- 2) Случайное разбиение данных
- 3) Многоголовое внимание
- 4) Построение линейной регрессии

Поле для ответа:

74. Прочтайте задание, выберите правильные ответы.

Какие понятия относятся к этапам лингвистического анализа?

- 1) Морфологический анализ
- 2) Синтаксический анализ

- 3) Кластеризация
- 4) Регрессия

Поле для ответа:

Повышенный уровень сложности

75. Прочтайте задание и установите соответствие.

Соотнесите архитектурные компоненты Transformer (слева) с их функциями (справа).

- A) Механизм внимания (Attention)
- Б) Позиционные кодировки
- В) Многоголовое внимание (Multi-head Attention)
- Г) Нормализация слоя (Layer Normalization)

- 1) Добавление информации о порядке слов
- 2) Выделение релевантных частей входа
- 3) Параллельное внимание по разным подпространствам
- 4) Стабилизация обучения сети

Поле для ответа:

76. Прочтайте задание и установите соответствие.

Соотнесите виды векторных моделей текста (слева) с их характеристиками (справа).

- A) One-hot encoding
- Б) Frequency model
- В) Bag of Words
- Г) TF-IDF

- 1) Представляет слова как уникальные векторы с единицей и нулями
- 2) Учитывает количество вхождений слов
- 3) Не учитывает порядок слов и контекст
- 4) Взвешивает слова по частоте и редкости в корпусе

Поле для ответа:

77. Прочтайте задание и установите соответствие.

Соотнесите понятия, связанные с моделями seq2seq (слева), с их функциями (справа).

- A) Кодировщик (Encoder)
- Б) Декодировщик (Decoder)
- В) Контекстный вектор
- Г) Механизм внимания

- 1) Преобразует входную последовательность в вектор
- 2) Генерирует выходную последовательность
- 3) Представляет сжатую информацию входа для декодера
- 4) Позволяет фокусироваться на разных частях входа при генерации

Поле для ответа:

78. Прочтайте задание и установите соответствие.

Соотнесите этапы обучения модели word2vec (слева) с их действиями (справа).

- A) Выбор контекстных слов
- B) Предсказание целевого слова
- C) Обновление весов нейронной сети
- D) Повторение для всех слов корпуса

- 1) Определение слов, окружающих центральное
- 2) Моделирование вероятности целевого слова по контексту
- 3) Корректировка параметров модели
- 4) Процесс итерации по всему тексту

Поле для ответа:

79. Прочтайте задание и установите последовательность.

Расположите этапы предварительной обработки текста в правильном порядке:

- 1) Удаление стоп-слов
- 2) Приведение к нижнему регистру
- 3) Токенизация
- 4) Лемматизация

Поле для ответа:

80. Прочтайте задание и установите последовательность.

Расположите этапы обучения модели seq2seq в правильном порядке:

- 1) Обработка входной последовательности кодировщиком
- 2) Генерация выходной последовательности декодировщиком
- 3) Вычисление ошибки предсказания
- 4) Обновление весов модели

Поле для ответа:

81. Прочтайте задание и установите последовательность.

Расположите этапы работы Transformer в правильном порядке:

- 1) Введение позиционных кодировок
- 2) Многоголовое внимание
- 3) Нормализация слова
- 4) Обработка входных данных

Поле для ответа:

82. Прочтите задание и установите последовательность.

Расположите этапы токенизации текста в правильном порядке:

- 1) Определение границ слов
- 2) Разделение текста на токены
- 3) Обработка знаков препинания
- 4) Преобразование токенов в числа

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

83. Прочтайте текст и напишите обоснованный ответ.

Что такое токенизация и зачем она нужна в обработке естественного языка?

Поле для ответа:

84. Прочтайте текст и напишите обоснованный ответ.

В чем разница между синтаксическим и морфологическим анализом?

Поле для ответа:

85. Прочтайте текст и напишите обоснованный ответ.

Как работает модель Bag of Words и какие у неё ограничения?

Поле для ответа:

86. Прочтайте текст и напишите обоснованный ответ.

Что такое галлюцинации в больших языковых моделях и почему они возникают?

Поле для ответа:

87. Прочтайте текст и напишите обоснованный ответ.

Как работает модель seq2seq и в каких задачах она применяется?

Поле для ответа:

88. Прочтайте текст и напишите обоснованный ответ.

Что такое prompt engineering и как оно влияет на работу больших языковых моделей?

Поле для ответа:

ОПК-3 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

Задания закрытого типа

Базовый уровень сложности

89. Прочтайте задание, выберите правильный ответ.

Что такое токен в обработке естественного языка?

- 1) Единица текста, например слово или символ
- 2) Название алгоритма
- 3) Тип языковой модели
- 4) Вид нейронной сети

Поле для ответа:

90. Прочтайте задание, выберите правильный ответ.

Какую задачу решает морфологический анализ?

- 1) Определение синтаксической структуры предложения
- 2) Определение частей речи и форм слов
- 3) Построение семантических связей
- 4) Токенизация текста

Поле для ответа:

91. Прочтайте задание, выберите правильный ответ.

Что такое N-грамма в NLP?

- 1) Последовательность из N символов или слов
- 2) Метод классификации текста
- 3) Алгоритм для обучения языковых моделей
- 4) Вид нейронной сети

Поле для ответа:

92. Прочтайте задание, выберите правильный ответ.

Для чего используется модель TF-IDF?

- 1) Для взвешивания слов с учетом частоты и редкости
- 2) Для определения частей речи
- 3) Для построения синтаксических деревьев
- 4) Для токенизации текста

Поле для ответа:

93. Прочтайте задание, выберите правильные ответы.

Какие из перечисленных этапов входят в предварительную обработку текста?

- 1) Токенизация
- 2) Построение синтаксического дерева
- 3) Удаление стоп-слов

4) Морфологический анализ

Поле для ответа:

94. Прочтайте задание, выберите правильные ответы.

Какие модели относятся к распределённым представлениям слов?

- 1) One-hot encoding
- 2) word2vec
- 3) Bag of Words
- 4) GloVe

Поле для ответа:

95. Прочтайте задание, выберите правильные ответы.

Какие задачи решает синтаксический анализ?

- 1) Определение грамматических связей между словами
- 2) Определение частей речи
- 3) Построение дерева зависимостей
- 4) Лемматизация слов

Поле для ответа:

96. Прочтайте задание, выберите правильные ответы.

Какие методы используются для снижения размерности текстовых данных?

- 1) TF-IDF
- 2) Word embedding
- 3) Стемминг
- 4) PCA

Поле для ответа:

Повышенный уровень сложности

97. Прочтайте задание и установите соответствие.

Соотнесите методы векторного представления текста (слева) с их характеристиками (справа).

- A) One-hot encoding
- Б) Bag of Words
- В) TF-IDF
- Г) word2vec

- 1) Взвешивание слов с учетом частоты в документе и редкости в корпусе
- 2) Вектор с единицей на позиции слова и нулями в остальных

- 3) Представление слов в плотном векторном пространстве с учетом контекста
- 4) Модель, учитываящая количество слов без порядка

Поле для ответа:

98. Прочтайте задание и установите соответствие.

Соотнесите этапы лингвистического анализа (слева) с их описаниями (справа).

- A) Токенизация
Б) Морфологический анализ
В) Синтаксический анализ
Г) Семантический анализ

- 1) Анализ смысловых отношений между словами
- 2) Разбиение текста на слова и знаки препинания
- 3) Определение частей речи и форм слов
- 4) Определение структуры предложения и связей между словами

Поле для ответа:

99. Прочтайте задание и установите соответствие.

Соотнесите задачи NLP (слева) с примерами (справа).

- A) Классификация текста
Б) Машинный перевод
В) Распознавание речи
Г) Анализ тональности

- 1) Определение положительной или отрицательной оценки отзыва
- 2) Преобразование аудио в текст
- 3) Определение категории новости
- 4) Перевод текста с английского на русский

Поле для ответа:

100. Прочтайте задание и установите соответствие.

Соотнесите виды лингвистического анализа (слева) с примерами (справа).

- A) Морфологический
Б) Синтаксический
В) Семантический
Г) Прагматический

- 1) Определение роли слова в предложении
- 2) Анализ смысла и значения текста
- 3) Определение части речи и формы слова

- 4) Анализ контекста и намерений говорящего

Поле для ответа:

101. Прочтайте задание и установите последовательность.

Расположите этапы морфологического анализа текста в правильном порядке:

- 1) Определение части речи
- 2) Токенизация
- 3) Определение грамматических форм слова
- 4) Лемматизация

Поле для ответа:

102. Прочтайте задание и установите последовательность.

Расположите этапы работы модели word2vec (CBOW) в правильном порядке:

- 1) Использование контекстных слов для предсказания центрального слова
- 2) Формирование обучающей выборки
- 3) Обучение нейронной сети
- 4) Представление слов в виде векторов

Поле для ответа:

103. Прочтайте задание и установите последовательность.

Расположите этапы работы Transformer в правильном порядке:

- 1) Позиционное кодирование входных данных
- 2) Многоголовое внимание
- 3) Нормализация слова
- 4) Генерация выходного текста

Поле для ответа:

104. Прочтайте задание и установите последовательность.

Расположите этапы предварительной обработки текста в правильном порядке:

- 1) Удаление стоп-слов
- 2) Токенизация
- 3) Стемминг или лемматизация
- 4) Очистка от лишних символов

Поле для ответа:

Задания открытого типа
Высокий уровень сложности

105. Прочтите текст и напишите обоснованный ответ.

Что такое токенизация и почему она важна для обработки естественного языка?

Поле для ответа:

106. Прочтите текст и напишите обоснованный ответ.

В чем разница между стеммингом и лемматизацией?

Поле для ответа:

107. Прочтайте текст и напишите обоснованный ответ.

Что такое модель BERT и в чем её основное отличие от классических моделей NLP?

Поле для ответа:

108. Прочтайте текст и напишите обоснованный ответ.

Опишите принцип работы attention-механизма в архитектуре Transformer.

Поле для ответа:

109. Прочтайте текст и напишите обоснованный ответ.

Что такое TF-IDF и как он помогает при анализе текста?

Поле для ответа:

110. Прочтайте текст и напишите обоснованный ответ.

Почему распределённые представления слов (например, word2vec) лучше, чем one-hot encoding?

Поле для ответа:

ОПК-4 Способен применять на практике новые научные принципы и методы исследований

Задания закрытого типа
Базовый уровень сложности

111. Прочтайте задание, выберите правильный ответ.

Что такое лемматизация?

- 1) Приведение слова к начальной форме
- 2) Разбиение текста на предложения
- 3) Обучение модели на большом корпусе
- 4) Удаление стоп-слов

Поле для ответа:

112. Прочтайте задание, выберите правильный ответ.

Что означает термин "галлюцинации" в больших языковых моделях?

- 1) Генерация неверной информации
- 2) Процесс токенизации
- 3) Обучение модели
- 4) Оптимизация гиперпараметров

Поле для ответа:

113. Прочтайте задание, выберите правильный ответ.

Что такое модель word2vec?

- 1) Модель для получения распределенных представлений слов
- 2) Метод токенизации
- 3) Алгоритм кластеризации
- 4) Синтаксический парсер

Поле для ответа:

114. Прочтайте задание, выберите правильный ответ.

Что такое transfer learning?

- 1) Использование заранее обученной модели для новой задачи
- 2) Процесс токенизации
- 3) Алгоритм обучения с подкреплением
- 4) Метод предварительной обработки текста

Поле для ответа:

115. Прочтайте задание, выберите правильные ответы.

Какие из перечисленных относятся к задачам Natural Language Processing?

- 1) Машинный перевод
- 2) Рекомендательные системы
- 3) Распознавание речи
- 4) Анализ тональности текста

Поле для ответа:

116. Прочтайте задание, выберите правильные ответы.

Какие архитектуры моделей основаны на механизме внимания (attention)?

- 1) Transformer
- 2) RNN (рекуррентные нейронные сети)
- 3) BERT
- 4) Naive Bayes

Поле для ответа:

117. Прочтайте задание, выберите правильные ответы.

Какие задачи решают модели seq2seq?

- 1) Генерация текста
- 2) Классификация изображений
- 3) Машинный перевод
- 4) Кластеризация текстов

Поле для ответа:

118. Прочтайте задание, выберите правильные ответы.

Какие из перечисленных методов относятся к векторным моделям представления текста?

- 1) One-hot encoding
- 2) Лемматизация
- 3) Bag of Words
- 4) N-граммы

Поле для ответа:

Повышенный уровень сложности

119. Прочтайте задание и установите соответствие.

Соотнесите понятия, связанные с большими языковыми моделями (слева), с их определениями (справа).

- A) Токенизация
- B) Transfer learning
- В) Галлюцинации
- Г) Prompt engineering

- 1) Создание и оптимизация запросов к модели
- 2) Использование заранее обученной модели для новой задачи
- 3) Разбиение текста на минимальные единицы
- 4) Генерация модели неверной информации

Поле для ответа:

120. Прочтайте задание и установите соответствие.

Соотнесите компоненты архитектуры Transformer (слева) с их функциями (справа).

А) Многоголовое внимание (Multi-head Attention)

Б) Позиционное кодирование

В) Нормализация слоя (Layer Normalization)

Г) Кодировщик (Encoder)

1) Добавляет информацию о порядке слов

2) Обрабатывает входные данные и извлекает признаки

3) Улучшает обучение, стабилизируя распределение активаций

4) Позволяет модели фокусироваться на разных частях входа одновременно

Поле для ответа:

121. Прочтайте задание и установите соответствие.

Соотнесите виды моделей представления текста (слева) с их особенностями (справа).

А) Модель частоты (Frequency model)

Б) One-hot encoding

В) N-граммная модель

Г) Распределенные представления

1) Учитывает последовательность из N слов

2) Представляет слова плотными векторами

3) Использует векторы с единичным элементом

4) Опирается на частотные характеристики слов

Поле для ответа:

122. Прочтайте задание и установите соответствие.

Соотнесите этапы обучения модели seq2seq (слева) с их описаниями (справа).

А) Кодировщик

Б) Декодировщик

В) Вычисление ошибки

Г) Обновление весов

1) Преобразует входную последовательность в вектор

2) Генерирует выходную последовательность

3) Сравнивает предсказание с правильным ответом

4) Корректирует параметры модели на основе ошибки

Поле для ответа:

123. Прочтите задание и установите последовательность.

Расположите этапы обучения модели seq2seq в правильном порядке:

- 1) Кодирование входной последовательности
- 2) Декодирование с использованием предыдущих выходов
- 3) Вычисление ошибки предсказания
- 4) Обновление весов модели

Поле для ответа:

124. Прочтите задание и установите последовательность.

Расположите этапы построения модели TF-IDF в правильном порядке:

- 1) Подсчет частоты слова в документе (TF)
- 2) Подсчет обратной частоты документа (IDF)
- 3) Умножение TF на IDF для каждого слова
- 4) Нормализация вектора признаков

Поле для ответа:

125. Прочтите задание и установите последовательность.

Расположите этапы работы модели BERT в правильном порядке:

- 1) Токенизация входного текста
- 2) Применение маскирования слов
- 3) Обработка через слои трансформера
- 4) Выдача предсказания для задачи

Поле для ответа:

126. Прочтите задание и установите последовательность.

Расположите этапы работы N-граммной модели в правильном порядке:

- 1) Разбиение текста на N-граммы
- 2) Подсчет частоты каждой N-граммы
- 3) Использование частот для оценки вероятности следующего слова
- 4) Применение модели для генерации или анализа текста

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

127. Прочтите текст и напишите обоснованный ответ.

Что такое языковая модель и какую задачу она решает?

Поле для ответа:

128. Прочтайте текст и напишите обоснованный ответ.

Объясните, что такое transfer learning и зачем он используется в NLP.

Поле для ответа:

129. Прочтайте текст и напишите обоснованный ответ.

Что такое N-граммная модель и где её применяют?

Поле для ответа:

130. Прочтайте текст и напишите обоснованный ответ.

Какие задачи решает синтаксический анализ текста?

Поле для ответа:

131. Прочтайте текст и напишите обоснованный ответ.

Что такое галлюцинации в больших языковых моделях и почему это проблема?

Поле для ответа:

132. Прочтайте текст и напишите обоснованный ответ.

Опишите, что такое prompt engineering и как он помогает улучшить работу языковых моделей.

Поле для ответа:

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. ИД.УК-1.5.	Опрос, диспут, практическое задание, тест

Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Средства оценки (в соот. с Таблицами 5, 7)
УК-4	ИД.УК-4.1. ИД.УК-4.2. ИД.УК-4.3. ИД.УК-4.4. ИД.УК-4.5. ИД.УК-4.6.	Опрос, диспут, практическое задание, тест
УК-5	ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3.	Опрос, диспут, практическое задание, тест
ОПК-1	ИД.ОПК-1.1. ИД.ОПК-1.2.	Опрос, диспут, практическое задание, тест
ОПК-3	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4.	Опрос, диспут, практическое задание, тест
ОПК-4	ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	Опрос, диспут, практическое задание, тест

Таблица 5

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
Практическое задание	<p>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:</p> <ol style="list-style-type: none"> Анализировать проблемную ситуацию, определять пробелы в информации, оценивать надёжность источников информации, разрабатывать стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строить сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения Устанавливать контакты и организовывать общение в соответствии с потребностями совместной деятельности, используя современные коммуникационные технологии, составлять в соответствии с нормами русского языка деловую документацию разных жанров, в том числе для академических и профессиональных целей на иностранном языке. Обсуждать и представлять результаты исследовательской и проектной деятельности, участвовать в дискуссиях на различных публичных мероприятиях, выбирая подходящий формат Анализировать важнейшие идеологические и ценностные системы, сформировавшиеся в ходе исторического развития. Выстраивать социальное и профессиональное взаимодействие с учётом особенностей деловой и общей культуры представителей других этносов и конфессий, различных социальных групп на основе идеологических и ценностных систем, обеспечивать создание недискриминационной среды для участников межкультурного взаимодействия Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
	использованию новых научных принципов и методов исследований в профессиональной области знаний
Опрос	<p>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:</p> <ol style="list-style-type: none"> 1. Анализировать проблемную ситуацию, определять пробелы в информации, оценивать надёжность источников информации, разрабатывать стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строить сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения 2. Устанавливать контакты и организовывать общение в соответствии с потребностями совместной деятельности, используя современные коммуникационные технологии, составлять в соответствии с нормами русского языка деловую документацию разных жанров, в том числе для академических и профессиональных целей на иностранном языке. 3. Обсуждать и представлять результаты исследовательской и проектной деятельности, участвовать в дискуссиях на различных публичных мероприятиях, выбирая подходящий формат 4. Анализировать важнейшие идеологические и ценностные системы, сформировавшиеся в ходе исторического развития. Выстраивать социальное и профессиональное взаимодействие с учётом особенностей деловой и общей культуры представителей других этносов и конфессий, различных социальных групп на основе идеологических и ценностных систем, обеспечивать создание недискриминационной среды для участников межкультурного взаимодействия 5. Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий 6. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров 7. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний
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> <ol style="list-style-type: none"> 1. Анализировать проблемную ситуацию, определять пробелы в информации, оценивать надёжность источников информации, разрабатывать стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строить сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения 2. Устанавливать контакты и организовывать общение в соответствии с потребностями совместной деятельности, используя современные коммуникационные технологии, составлять в соответствии с нормами русского языка деловую документацию разных жанров, в том числе для академических и профессиональных целей на иностранном языке. 3. Обсуждать и представлять результаты исследовательской и проектной деятельности, участвовать в дискуссиях на различных публичных мероприятиях, выбирая подходящий формат 4. Анализировать важнейшие идеологические и ценностные системы, сформировавшиеся в ходе исторического развития. Выстраивать социальное и профессиональное взаимодействие с учётом особенностей деловой и общей культуры представителей других этносов и конфессий, различных социальных групп на основе идеологических и ценностных систем, обеспечивать создание недискриминационной среды для участников межкультурного взаимодействия

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
	<p>5. Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий</p> <p>6. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров</p> <p>7. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний</p>
Тест	<p>Магистрант в ходе подготовки и защиты проекта показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности:</p> <p>1. Анализировать проблемную ситуацию, определять пробелы в информации, оценивать надёжность источников информации, разрабатывать стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов, строить сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения</p> <p>2. Устанавливать контакты и организовывать общение в соответствии с потребностями совместной деятельности, используя современные коммуникационные технологии, составлять в соответствии с нормами русского языка деловую документацию разных жанров, в том числе для академических и профессиональных целей на иностранном языке.</p> <p>3. Обсуждать и представлять результаты исследовательской и проектной деятельности, участвовать в дискуссиях на различных публичных мероприятиях, выбирая подходящий формат</p> <p>4. Анализировать важнейшие идеологические и ценностные системы, сформировавшиеся в ходе исторического развития. Выстраивать социальное и профессиональное взаимодействие с учётом особенностей деловой и общей культуры представителей других этносов и конфессий, различных социальных групп на основе идеологических и ценностных систем, обеспечивать создание недискриминационной среды для участников межкультурного взаимодействия</p> <p>5. Делать обоснованный выбор методов разработки алгоритмов и программных средств для решения профессиональных задач, в том числе с использованием современных информационных технологий</p> <p>6. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров</p> <p>7. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний</p>