

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

Дата подписания: 14.06.2024 17:50:15

Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a70e51517ff6d591

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

УТВЕРЖДАЮ:

Ректор

/В.В. Волков

« 24 » мая 2024 г.

Протокол УС № 3 от 24 мая 2024 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 развивают 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 | 20 |
| 7.5 Средства оценки индикаторов достижения компетенций | 20 |
| 8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА | 23 |
| 8.1. Основная литература | 23 |
| 8.2 Дополнительная литература | 24 |
| 9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА | 24 |
| 9.1 Программное обеспечение | 24 |
| 9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины: | 24 |
| 9.3 Лицензионные электронные ресурсы библиотеки Университета | 25 |
| 9.4 Электронная информационно-образовательная среда Университета..... | 25 |
| 10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА | 26 |
| ПРИЛОЖЕНИЕ 1 | 27 |

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины «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 | Знать: методы научного познания, в основе которых лежит рассмотрение объекта как системы: целостного комплекса взаимосвязанных элементов, методы и модели стратегического планирования 3 (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) |
| | 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 устранения | |
| 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 | Знать: принципы эффективной коммуникации в устной и письменной формах на русском и иностранном языке, принятые в академическом и профессиональном сообществе 3 (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) |
| | 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 | |

| 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> |
| <p>УК-5 Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия</p> | <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> |
| <p>ОПК-1 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач</p> | <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 неопределённости 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> |
| | | <p>Знать:</p> |

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

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: | 28 | - | - | 28 | - | - | - | - | - | - | - |
| Лекции (Л) | 14 | - | - | 14 | - | - | - | - | - | - | - |
| Практические занятия (ПЗ) | 14 | - | - | 14 | - | - | - | - | - | - | - |
| Самостоятельная работа (СР) | 80 | - | - | 80 | - | - | - | - | - | - | - |
| Промежуточная аттестация | форма | Зачет | - | - | Зачет | - | - | - | - | - | - |
| | 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 соот.с Таблицей 1) | Коды ЗУВ (в соот.с Таблицей 1) |
|-------|-----------------------------|--|---|--|--|
| 1 | Корпусы текстовых данных | Корпусные исследования. Описание корпусов текстовых данных: КРЯ, Wikipedia, CommonCrawl, Taiga. | 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. | 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) |
| 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. | 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) |
| 3 | Закон Ципфа. Издержки токенизации | Введение в 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. | 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) |

| № 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. | 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) |
| 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. | 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) |
| 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. | 3 (УК-1) У (УК-1) В (УК-1) 3 (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) |
|-------|--|--|---|--|---|
| | текстовых данных | к контекстуальным эмбедингам. | 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) 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) |
| 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. | 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) |

5.2 Структура дисциплины

Таблица 4

Структура дисциплины

| № п/п | Наименование тем (разделов) | Объем дисциплины, час. | | | Форма текущего контроля успеваемости*, промежуточной аттестации | |
|-----------------------------|-----------------------------|------------------------|---|----|---|---|
| | | Всего | Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП | СР | | |
| Л | ПЗ | | | | | |
| <i>Очная форма обучения</i> | | | | | | |
| Тема 1 | Корпусы текстовых данных | 15 | 2 | 2 | 11 | О |

| № п/п | Наименование тем (разделов) | Объем дисциплины, час. | | | | Форма текущего контроля успеваемости*, промежуточной аттестации |
|---------------------------------|--|------------------------|---|-----------|-----------|---|
| | | Всего | Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП | | СР | |
| | | | Л | ПЗ | | |
| <i>Очная форма обучения</i> | | | | | | |
| Тема 2 | Распознавание текстов. Создание процедуры превращения изображений или PDF в корпус | 15 | 2 | 2 | 11 | ПЗ |
| Тема 3 | Закон Ципфа. Издержки токенизации | 15 | 2 | 2 | 11 | ПЗ |
| Тема 4 | Извлечение сущностей из текстовых данных | 15 | 2 | 2 | 11 | ПЗ |
| Тема 5 | Разреженное векторное представление текстовых данных | 15 | 2 | 2 | 11 | ПЗ |
| Тема 6 | Уплотненное векторное представление текстовых данных | 15 | 2 | 2 | 11 | ПЗ |
| Тема 7 | Поиск по представлениям текстовых данных | 18 | 2 | 2 | 14 | ПЗ, Д |
| Промежуточная аттестация | | - | - | - | - | Зачет |
| Всего: | | 108/3 | 14 | 14 | 80 | - |

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

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

6.1 Общие положения

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

Самостоятельная работа является важнейшей частью процесса высшего образования. Ее следует осознанно организовать, выделив для этого необходимое время и соответствующим образом организовав рабочее пространство. Важнейшим элементом самостоятельной работы является проработка материалов прошедших занятий (анализ конспектов, чтение рекомендованной литературы) и подготовка к следующим лекциям/семинарским занятиям. Литературу, рекомендованную в программе курса, следует, по возможности, читать в течение всего семестра, концентрируясь на обусловленных программой курса темах.

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

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

Тема 1. Корпусы текстовых данных:

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

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

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

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

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

Тема 3. Закон Ципфа. Издержки токенизации:

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

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

Тема 4. Извлечение сущностей из текстовых данных:

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

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

Тема 5. Разреженное векторное представление текстовых данных:

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

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

Тема 6. Уплотненное векторное представление текстовых данных:

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

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

Тема 7. Поиск по представлениям текстовых данных:

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

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

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

Вопросы для самостоятельной подготовки по темам дисциплины:

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

6.4 Перечень литературы для самостоятельной работы обучающегося:

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

6.5 Перечень учебно-методического обеспечения для самостоятельной работы

Для обеспечения самостоятельной работы магистрантов по дисциплине «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. | 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. | 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) | 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. | 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) | 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. | 3 (УК-1) У (УК-1) В (УК-1) 3 (УК-4) У (УК-4) В (УК-4) 3 (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) 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) | | |

Таблица 6

Критерии оценивания

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

7.2 Контрольные задания для текущей аттестации

Примерный материал опросов, диспутов, практических заданий:

Тема 1. Корпусы текстовых данных.

Опрос 1:

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

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

Практическое задание 1:

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

Тема 3. Закон Ципфа. Издержки токенизации.

Практическое задание 2:

Оценить качество сегментации разными методами, сравнить результаты.

Тема 4. Извлечение сущностей из текстовых данных.

Практическое задание 3:

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

Тема 5. Разреженное векторное представление текстовых данных.

Практическое задание 4:

Оценить на терм-документарной матрице вектора документов предложенными алгоритмами.

Тема 6. Уплотненное векторное представление текстовых данных.

Практическое задание 5:

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

Тема 7. Поиск по представлениям текстовых данных.

Практическое задание 6, диспут 1:

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

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

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

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

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Таблица 7

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

| Форма промежуточной аттестации/вид промежуточной аттестации | Коды компетенций | Индикаторы компетенций (в соот. с Таблицей 1) | Коды ЗУВ (в соот. с Таблицей 1) | Критерии оценивания | Оценка |
|--|---|--|---|---|---------------|
| Зачет / проект | 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. | 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-1) У (ОПК-3) | Магистрант демонстрирует полную самостоятельность в подборе фактического материала и аналитическое отношение к нему, умение рассматривать примеры и факты во взаимосвязи и взаимообусловленности, отбирать наиболее существенные из них; а также показывает грамотное использование методов описания и презентации исследования | Зачтено |

| 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.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 (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4) | Магистрант не демонстрирует аналитическое отношение к материалу, не видит взаимосвязь примеров и фактов; а также использует методы описания и презентации исследования с большим количеством существенных ошибок | Не зачтено |

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

Таблица 7а

Система оценки знаний обучающихся

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

Результаты промежуточного контроля по дисциплине, выраженные в 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

Примерные требования к проекту для промежуточной аттестации по дисциплине:

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

7.5 Средства оценки индикаторов достижения компетенций

Таблица 8

Средства оценки индикаторов достижения компетенций

| Коды компетенций | Индикаторы компетенций (в соот.с Таблицей 1) | Средства оценки (в соот. с Таблицами 5, 7) |
|------------------|---|---|
| УК-1 | ИД.УК-1.1. | Опрос, диспут, практическое задание, проект |

| Коды компетенций | Индикаторы компетенций (в соот.с Таблицей 1) | Средства оценки (в соот. с Таблицами 5, 7) |
|-------------------------|--|---|
| | ИД.УК-1.2. ИД.УК-1.3. ИД.УК-1.4. ИД.УК-1.5. | |
| 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, проект |
| УК-5 | ИД.УК-5.1. ИД.УК-5.2. ИД.УК-5.3. | Опрос, диспут, практическое задание, проект |
| ОПК-1 | ИД.ОПК-1.1. ИД.ОПК-1.2. | Опрос, диспут, практическое задание, проект |
| ОПК-3 | ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. | Опрос, диспут, практическое задание, проект |
| ОПК-4 | ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4. | Опрос, диспут, практическое задание, проект |

Таблица 9

Описание средств оценки индикаторов достижения компетенций

| Средства оценки (в соот. С Таблицами 5, 7) | Рекомендованный план выполнения работы |
|---|--|
| Практическое задание | <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 |

| Средства оценки (в соот. С Таблицами 5, 7) | Рекомендованный план выполнения работы |
|--|---|
| | <p>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> |
| 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 |

| Средства оценки (в соот. С Таблицами 5, 7) | Рекомендованный план выполнения работы |
|--|--|
| | <p>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> |
| Проект | <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> |

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://npoed.ru>
4. Официальная Россия. Сервер органов государственной власти Российской Федерации: <http://www.gov.ru>
5. Официальный интернет-портал правовой информации. Государственная система правовой информации: <http://pravo.gov.ru>
6. Правовой сайт КонсультантПлюс: <http://www.consultant.ru/sys>
7. Российское образование. Федеральный портал: <http://www.edu.ru>

Профессиональные базы данных информационно-телекоммуникационной сети «Интернет»:

1. 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. Интелрос. 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. **East View** – 100 ведущих российских журналов по гуманитарным наукам (архив и текущая подписка): <https://dlib.eastview.com/browse;>
2. **eLIBRARY.RU** — Российский информационно-аналитический портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты научных статей и публикаций, наукометрическая база данных: <http://elibrary.ru;>
3. **Университетская информационная система РОССИЯ** — база электронных ресурсов для учебных программ и исследовательских проектов в области социально-гуманитарных наук: [http://www.uisrussia.msu.ru/;](http://www.uisrussia.msu.ru/)
4. Электронные журналы по подписке (текущие номера научных зарубежных журналов)

Электронные библиотечные системы:

1. **Znaniium.com** – Электронная библиотечная система (ЭБС) – [http://znaniium.com/;](http://znaniium.com/)
2. Университетская библиотека онлайн – Электронная библиотечная система (ЭБС) – <http://biblioclub.ru/>

9.4 Электронная информационно-образовательная среда Университета

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

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

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

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

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

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

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

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

Проведение занятий лекционного типа обеспечивается демонстрационным оборудованием.

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

Для лиц с ограниченными возможностями здоровья и инвалидов предоставляется возможность присутствия в аудитории вместе с ними ассистента (помощника). Для слабовидящих предоставляется возможность увеличения текста на экране ПК. Для самостоятельной работы лиц с ограниченными возможностями здоровья в помещении для самостоятельной работы организовано одно место (ПК) с возможностями бесконтактного ввода информации и управления компьютером (специализированное лицензионное программное обеспечение – Camera Mouse, веб камера). Библиотека университета предоставляет удаленный доступ к 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).