

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования
«**Российский государственный гуманитарный университет**»
(ФГБОУ ВО «РГГУ»)

ИНСТИТУТ ФИЛОЛОГИИ И ИСТОРИИ
ИСТОРИКО-ФИЛОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ
Кафедра германской филологии

ИНТЕЛЛЕКТУАЛЬНЫЙ ИНФОРМАЦИОННЫЙ ПОИСК
РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

45.04.01 - Филология

Код и наименование направления подготовки/специальности

Цифровое германское лингволитературоведение

Наименование направленности (профиля)/ специализации

Уровень высшего образования: *магистратура*

Форма обучения: *очная*

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2024

Интеллектуальный информационный поиск
Рабочая программа дисциплины (модуля)

Составитель(и):

к. техн. наук, руководитель Отделение интеллектуальных систем в гуманитарной сфере
Д.В. Стефановский

УТВЕРЖДЕНО

Протокол заседания кафедры

№ 5 от 29.03.2024

1. Пояснительная записка.....	4
1.1. Цель и задачи дисциплины.....	4
1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций.....	4
1.3. Место дисциплины в структуре образовательной программы.....	5
2. Структура дисциплины.....	5
3. Содержание дисциплины.....	6
4. Образовательные технологии.....	7
5. Оценка планируемых результатов обучения.....	9
5.1 Система оценивания.....	9
5.2 Критерии выставления оценки по дисциплине.....	10
5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине.....	11
6. Учебно-методическое и информационное обеспечение дисциплины.....	13
6.1 Список источников и литературы.....	13
6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».....	14
6.3 Профессиональные базы данных и информационно-справочные системы.....	14
7. Материально-техническое обеспечение дисциплины.....	14
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов.....	15
9. Методические материалы.....	16
9.1 Планы семинарских/ практических/ лабораторных занятий.....	16
9.2 Иные материалы.....	18
Приложение 1. Аннотация рабочей программы дисциплины.....	21
1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций.....	21

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Цель учебной дисциплины «Интеллектуальный информационный поиск» состоит в том, чтобы помочь обучающимся овладеть возможностями современного цифрового пространства для решения коммуникативных и прикладных задач лингвистики и литературоведения.

Курс включает в себя лекционные, семинарские и практические занятия. Общая трудоемкость освоения дисциплины (модуля) составляет 108 часов (3 з.е.)

Основными **задачами** изучения дисциплины являются:

- изучение методов компьютерной обработки текстов в электронном виде, обеспечивающие поддержку работы лингвистов и литературоведов при обработке информации, анализе данных и интерпретации результатов;
- формирование умения работать с корпусами, лингвистическими корпусами, электронными библиотеками, поисковыми системами; справочными информационными ресурсами цифровой гуманитаристики;
 - овладение практическими навыками по поиску, анализу, разработке, созданию информационных ресурсов;
- овладение технологиями поиска, анализа, и извлечения мнений из различных корпусов текстов.

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
УК-1	<i>УК-1.1 Обладает готовностью действовать в нестандартных ситуациях</i>	Знать: различие форм, методов, приемов и последовательности действий в стандартных и нестандартных коммуникативных и педагогических ситуациях; варианты альтернативных путей выхода из различных ситуаций; Уметь: проявлять инициативу и избегать автоматического применения стандартных форм и приемов при решении задач в нестандартных коммуникативных и педагогических ситуациях; Владеть: системой навыков действий в нестандартных коммуникативных и педагогических ситуациях.
	<i>УК-1.2 Способен нести социальную и этическую ответственность за принятые решения</i>	Знать: основы социальной и этической ответственности при принятии решений; Уметь: выделять и систематизировать основные представления о социальной и этической ответственности за принятые решения; прогнозировать возможные нестандартные ситуации; принимать и критически оценивать принятые решения;

		Владеть: умением прогнозировать результаты социальной и этической ответственности за принятые решения.
	УК-1.3 Демонстрирует умение критически оценивать и анализировать информацию, приобретенную самостоятельно, в том числе с помощью информационных технологий, и использовать ее в практической деятельности	Знать: основные требования информационной безопасности. Уметь: решать задачи по поиску источников и научной литературы. Владеть: навыками поиска научной литературы и составления списка источников и литературы для научной работы.
ПК-3	ПК-1.3 Способен определить стилистическую, жанровую, формальную и содержательную специфику текста	Знать: основные требования информационной безопасности. Уметь: решать задачи по поиску источников и научной литературы. Владеть: навыками поиска научной литературы и составления списка источников и литературы для научной работы.

1.3. Место дисциплины в структуре образовательной программы

Дисциплина (*модуль*) «Интеллектуальный информационный поиск» относится к части, формируемой участниками образовательных отношений блока 1 дисциплин учебного плана.

Для освоения дисциплины (*модуля*) необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин и прохождения практик: «Информационная среда цифровой гуманитаристики», «Современная корпусная лингвистика», «Информационные технологии в филологии».

В результате освоения дисциплины (*модуля*) формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: исследовательская работа, информационно-коммуникационная практика», «технологии искусственного интеллекта».

2. Структура дисциплины¹

Общая трудоёмкость дисциплины составляет 3 з.е., 108 академических часа (ов).

Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
	Лекции	14
	Семинары	16
	Всего:	30

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 78 академических часа(ов).

3. Содержание дисциплины

№	Наименование раздела дисциплины	Содержание
1.	Тема 1 Словари. Логический и нечеткий поиск.	Булев поиск. Инвертированный индекс. Обработка булевых запросов. Сравнение расширенной булевой модели и ранжированного поиска. Лексикон и списки словопозиций. Схематизация документа и декодирование последовательности символов. Определение лексикона терминов. Словопозиции с координатами и фразовые запросы. Словари и нечеткий поиск. Поисковые структуры для словарей. Исправления
2.	Тема 2. Построение и сжатие индекса поисковой системе анализ данных. Ранжирование, взвешивание терминов и модель векторного пространства	Построение индекса. Основы аппаратного обеспечения. Блочное индексирование, основанное на сортировке. Однопроходное индексирование в оперативной памяти. Распределенное индексирование. Динамическое индексирование. Сжатие индекса. Статистические характеристики терминов в информационном поиске. Сжатие словаря. Сжатие инвертированного файла. Ранжирование, взвешивание терминов и модель векторного пространства. Параметрические и зонные индексы. Частота термина и взвешивание. Модель векторного пространства для ранжирования. Варианты функций tf-idf
3.	Тема 3. Ранжирование в полнофункциональной поисковой системе. Оценка информационного поиска. Обратная связь по релевантности и расширение запроса	Эффективное ранжирование. Компоненты информационно-поисковой системы. Влияние операторов языка запросов на ранжирование в векторном пространстве. Оценка информационно-поисковой системы. Стандартные тестовые коллекции. Оценка неранжированных результатов поиска. Оценка ранжированных результатов поиска. Оценка релевантности. Более широкая точка зрения: качество системы и ее полезность для пользователя. Обратная связь по релевантности и псевдорелевантности. Глобальные методы для переформулирования запроса
4.	Тема 4. Основные концепции языка XML. Вероятностная модель информационного поиска. Языковые модели для информационного поиска	Основные концепции языка XML. XML-поиск. Проблемы, связанные с XML-поиском. Модель векторного пространства для XML-поиска. Оценка XML-поиска. Методы XML-поиска, ориентированные на текст и на данные. Вероятностная модель информационного поиска. Принцип вероятностного ранжирования. Бинарная модель независимости. Вероятностные модели и некоторые модификации. Языковые модели. Модель правдоподобия запроса. Сравнение языкового моделирования с другими подходами к

		информационному поиску. Расширения языковых моделей.
5.	<i>Тема 5. Классификация текстов и наивный байесовский подход . Классификация в векторном пространстве. Машинное обучение на текстах</i>	Классификация текстов. Наивная байесовская классификация текстов. Модель Бернулли. Свойства наивной байесовской модели. Выбор признаков. Оценка классификации текстов 287. Представление документов и меры близости в векторном пространстве. Метод Роккио. Метод k ближайших соседей. Линейные и нелинейные классификаторы. Классификация с несколькими классами. Компромисс между смещением и дисперсией. Метод опорных векторов: случай линейно разделимых классов. Расширения модели опорных векторов. Проблемы, связанные с классификацией текстовых документов. Методы машинного обучения для поиска по запросу.
6.	<i>Тема 6. Кластеризация. Обзор сведений из линейной алгебры. Разложение матриц и латентно-семантическое индексирование</i>	Кластеризация в информационном поиске. Оценивание кластеризации. Метод K-средних. Кластеризация, основанная на моделях. Иерархическая кластеризация. Агломеративная иерархическая кластеризация. Кластеризация методами одиночной и полной связи. Агломеративная кластеризация на основе усреднения по группе. Кластеризация методом центроидов. Оптимальность агломеративной иерархической кластеризации. Нисходящая кластеризация. Именованые кластеры. Матрицы “термин–документ” и сингулярные разложения. Малоранговые аппроксимации. Латентно-семантическое индексирование.
7.	<i>Тема 7. Основы поиска в вебе. Обход и индексирование веба. Анализ ссылок</i>	Опыт пользователей поисковых систем 428 19.5. Размер индекса и оценка его размера. Нечеткие дубликаты. Обход и индексирование веба. Распределение индексов. Серверы проверки ссылочной связности. Анализ ссылок. Веб как граф. Метод PageRank. Порталы и авторитетные источники

4. Образовательные технологии

№ п/п	Наименование раздела	Виды учебных занятий	Образовательные технологии
1	2	3	4
1	<i>Тема 1 Словари. Логический и нечеткий поиск.</i>	Лекция 1 Семинар 1	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнение задач с преподавателем, в том числе с использованием Python.

			Консультирование и проверка домашних заданий через социальные сети и электронную почту.
2	Тема 2. Построение и сжатие индекса поисковой системе анализ данных. Ранжирование, взвешивание терминов и модель векторного пространств	Лекция 2 Семинар 2	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнение задач с преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
3	Тема 3. Ранжирование в полнофункциональной поисковой системе. Оценка информационного поиска. Обратная связь по релевантности и расширение запроса	Лекция 3 Семинар 3	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнение задач с преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
4	Тема 4. Основные концепции языка XML. Вероятностная модель информационного поиска. Языковые модели для информационного поиска	Лекция 4 Семинар 4	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнение задач с преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
5	Тема 5. Классификация текстов и наивный байесовский подход . Классификация в векторном пространстве. Машинное обучение на текстах	Лекция 5 Семинар 5	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятия обучение проводятся с использованием программного продукта Полианалист 6.5 или выше. Выполнение задач с преподавателем, в том числе с использованием Python.

			Консультирование и проверка домашних заданий через социальные сети и электронную почту.
6	Тема 6. Кластеризация. Обзор сведений из линейной алгебры. Разложение матриц и латентно-семантическое индексирование	Лекция 6 Семинар 6	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятиях обучение проводится с использованием программного продукта Полианалист 6.5 или выше. Выполнение задач с преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
7	Тема 7. Основы поиска в вебе. Обход и индексирование веба. Анализ ссылок	Лекция 7 Семинар 7	Лекция/семинар с использованием видеоматериалов, проекционной техники. На семинарских занятиях обучение проводится с использованием программного продукта Полианалист 6.5 или выше. Выполнение задач с преподавателем, в том числе с использованием Python. Консультирование и проверка домашних заданий через социальные сети и электронную почту.
8	Консультация	Семинар 8	Консультирование и проверка домашних заданий. Ответы на вопросы

В период временного приостановления посещения обучающимися помещений и территории РГГУ для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

5. Оценка планируемых результатов обучения

5.1 Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего

Текущий контроль:		
- опрос	5 баллов	30 баллов
- участие в дискуссии на семинаре	5 баллов	10 баллов
- контрольная работа (темы 1-3)	10 баллов	10 баллов
- контрольная работа (темы 4-5)	10 баллов	10 баллов
Промежуточная аттестация – зачет/зачет с оценкой/экзамен		40 баллов
Итого за семестр		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49			неудовлетворительно
0 – 19	не зачтено	F	

5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	отлично/ зачтено	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ C	хорошо/ зачтено	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	удовлетворительно/ зачтено	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		<p>направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами. Демонстрирует достаточный уровень знания учебной литературы по дисциплине. Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	неудовлет- ворительно/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации. Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

Вопросы для подготовки к экзамену:

1. Расскажите об алгоритме и особенности булева поиска? Приведите примеры и поясните их.
2. Как создать инвертированный индекс? Приведите примеры и поясните их.
3. Каким образом происходит обработка булевых запросов? ? Приведите примеры и поясните их.
4. Сравните расширенной булевой модели и ранжированного поиска, расскажите об отличиях? Приведите примеры и поясните их.
5. Расскажите о списках словопозиций? Приведите примеры и поясните их.
6. Как происходит схематизация документа и декодирование последовательности символов? Приведите примеры и поясните их.
7. Как определяется лексикон терминов? Приведите примеры и поясните их.
8. Как осуществляется быстрое пересечение инвертированных списков с помощью указателей пропусков? Приведите примеры и поясните их.
9. Расскажите о словопозициях с координатами и фразовых запросах? Приведите примеры и поясните их.
10. Расскажите о словарях и нечётком поиске? Приведите примеры и поясните их.
11. Как устроены поисковые структуры для словарей? Приведите примеры и поясните их.
12. Что такое запросы с джокером? Приведите примеры и поясните их.
13. Опишите основы аппаратного обеспечения важные для информационного поиска? Приведите примеры и поясните их.
14. Как устроено блочное индексирование, основанное на сортировке? Приведите примеры и поясните их.
15. Как устроено однопроходное индексирование в оперативной памяти? Приведите примеры и поясните их.
16. Что такое распределенное индексирование? Опишите основные принципы и подходы? Приведите примеры и поясните их.

17. Что такое Динамическое индексирование? Опишите основные принципы и подходы? Приведите примеры и поясните их.
18. Как устроено сжатие индекса? Приведите примеры и поясните их.
19. Что такое Статистические характеристики терминов в информационном поиске? Приведите примеры и поясните их.
20. Как устроено сжатие словаря? Приведите примеры и поясните их.
21. Как устроено сжатие инвертированного файла? Приведите примеры и поясните их.
22. Что такое ранжирование, взвешивание терминов и модель векторного пространства? Приведите примеры и поясните их.
23. Как устроены параметрические и зонные индексы? Приведите примеры и поясните их.
24. Что такое модель векторного пространства для ранжирования? Приведите примеры и поясните их.
25. Как устроено эффективное ранжирование? Приведите примеры и поясните их.
26. Опишите компоненты информационно-поисковой системы? Приведите примеры и поясните их.
27. Как влияют операторы языка запросов на ранжирование в векторном пространстве? Приведите примеры и поясните их.
28. Как устроены стандартные тестовые коллекции? Приведите примеры и поясните их.
29. Как осуществляется оценка неранжированных результатов поиска? Приведите примеры и поясните их.
30. Как осуществляется оценка ранжированных результатов поиска? Приведите примеры и поясните их.
31. Как осуществляется оценка релевантности? Приведите примеры и поясните их.
32. Что такое сниппеты? Приведите примеры и поясните их.
33. Зачем нужна обратная связь по релевантности и псевдорелевантности? Приведите примеры и поясните их.
34. Опишите глобальные методы для переформулирования запроса? Приведите примеры и поясните их.
35. Опишите основные концепции языка XML? Приведите примеры и поясните их.
36. Опишите проблемы, связанные с XML -поиском? Приведите примеры и поясните их.
37. Как устроена модель векторного пространства для XML поиска? Приведите примеры и поясните их.
38. Опишите основные методы XML поиска, ориентированные на текст и на данные? Приведите примеры и поясните их.
39. Опишите как устроена вероятностная модель информационного поиска? Приведите примеры и поясните их.
40. Опишите принцип вероятностного ранжирования? Приведите примеры и поясните их.
41. Опишите бинарная модель независимости? Приведите примеры и поясните их.
42. Опишите языковые модели для информационного поиска
43. Зачем нужна модель правдоподобия запроса
44. Сравните языковые модели с другими подходами к информационному поиску
45. Опишите как делается классификация текстов в целом? Приведите примеры и поясните их.
46. Как устроена наивная байесовская классификация текстов? Приведите примеры и поясните их.
47. Что такое модель Бернулли? Приведите примеры и поясните их.
48. Расскажите о свойствах наивной байесовской модели? Приведите примеры и поясните их.
49. Как осуществляется выбор признаков? Приведите примеры и поясните их.
50. Как осуществляется оценка классификации текстов? Приведите примеры и поясните их.
51. Опишите способы представления документов и меры близости в векторном пространстве? Приведите примеры и поясните их.

52. Опишите как устроен метод Роккио? Приведите примеры и поясните их.
53. Опишите как устроен метод ближайших соседей? Приведите примеры и поясните их.
54. Опишите как устроены линейные и нелинейные классификаторы? Приведите примеры и поясните их.
55. Опишите как устроена классификация с несколькими классами? Приведите примеры и поясните их.
56. Опишите как устроен метод опорных векторов? Приведите примеры и поясните их.
57. Опишите проблемы, связанные с классификацией текстовых документов? Приведите примеры и поясните их.
58. Расскажите о методах машинного обучения для поиска по запросу? Приведите примеры и поясните их.
59. Опишите как устроена плоская кластеризация в информационном поиске? Приведите примеры и поясните их.
60. Опишите как устроена оценка кластеризации? Приведите примеры и поясните их.
61. Опишите как устроен метод К-средних? Приведите примеры и поясните их.
62. Опишите как устроена агломеративная иерархическая кластеризация? Приведите примеры и поясните их.
63. Опишите как устроена кластеризация методом центроидов? Приведите примеры и поясните их.
64. Что такое оптимальность агломеративной иерархической кластеризации? Приведите примеры и поясните их.
65. Что такое нисходящая кластеризация? Приведите примеры и поясните их.
66. Опишите как устроена матрица “термин-документ” и сингулярные разложения? Приведите примеры и поясните их.
67. Что такое латентно-семантическое индексирование ? Приведите примеры и поясните их.
68. Как представить Веб виде графа? Приведите примеры и поясните их.
69. Опишите как устроен Метод PageRank? Приведите примеры и поясните их.

Образец билета

1. Что такое латентно-семантическое индексирование? Приведите примеры и поясните их
2. Как устроена наивная байесовская классификация текстов? Приведите примеры и поясните их.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы

Литература основная

1. Маннинг К. Введение в информационный поиск. / К. Маннинг, П. Рагхаван, Х. Шютце. – М.: Вильямс, 2014. – 528 с. Режим доступа:
<https://libramedia.ru/books/%D0%92%D0%B2%D0%B5%D0%B4%D0%B5%D0%BD%D0%B8%D0%B5%20%D0%B2%20%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D0%B9%20%D0%BF%D0%BE%D0%B8%D1%81%D0%BA.pdf>
2. Buettcher S. Information Retrieval: Implementing and Evaluating Search Engines / S. Buettcher, C. Clarke, G. Cormack. – Massachusetts Institute of Technology, 2010. – 632 p. Режим доступа:
<https://mitmecsept.files.wordpress.com/2018/05/stefan-bc3bcttcher-charles-l-a-clarke-gordon-v-cormack-information-retrieval-implementing-and-evaluating-search-engines-2010-mit.pdf>

Литература дополнительная

1. Химстра Д. Модели информационного поиска. Режим доступа:

<https://intuit.ru/studies/courses/580/436/info>

2. Воронцов К.В. Машинное обучение (курс лекций). Режим доступа:

<https://www.machinelearning.ru/wiki/index.php?title=%D0%9C>

%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BE

%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%28%D0%BA

%D1%83%D1%80%D1%81_%D0%BB%D0%B5%D0%BA%D1%86%D0%B8%D0%B9%2C_

%D0%9A.%D0%92.%D0%92%D0%BE%D1%80%D0%BE%D0%BD%D1%86%D0%BE

%D0%B2%29

6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

Необходимо добавить то, что необходимо для изучения дисциплины

Национальная электронная библиотека (НЭБ) www.rusneb.ru

ELibrary.ru Научная электронная библиотека www.elibrary.ru

Электронная библиотека Grebennikon.ru www.grebennikon.ru

Cambridge University Press

ProQuest Dissertation & Theses Global

SAGE Journals

Taylor and Francis

JSTOR

6.3 Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

Информационные справочные системы:

1. Консультант Плюс

2. Гарант

7. Материально-техническое обеспечение дисциплины

Для обеспечения дисциплины используется материально-техническая база образовательного учреждения: учебные аудитории, оснащённые компьютером и проектором для демонстрации учебных материалов.

Состав программного обеспечения:

1. Windows

2. Microsoft Office

3. Kaspersky Endpoint Security

4. Adobe Master Collection

5. Полиналист 6.5. и выше

6. JupyterLab.

7. ОС «Альт Образование»

8. VSCode

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением или могут быть заменены устным ответом; обеспечивается индивидуальное равномерное освещение не менее 300 люкс; для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств; письменные задания оформляются увеличенным шрифтом; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

- для глухих и слабослышащих: лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования; письменные задания выполняются на компьютере в письменной форме; экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата: лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением; письменные задания выполняются на компьютере со специализированным программным обеспечением; экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих: в печатной форме увеличенным шрифтом, в форме электронного документа, в форме аудиофайла.

- для глухих и слабослышащих: в печатной форме, в форме электронного документа.

- для обучающихся с нарушениями опорно-двигательного аппарата: в печатной форме, в форме электронного документа, в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих: устройством для сканирования и чтения с камерой SARA SE; дисплеем Брайля PAC Mate 20; принтером Брайля EmBraille ViewPlus;

- для глухих и слабослышащих: автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих; акустический усилитель и колонки;

- для обучающихся с нарушениями опорно-двигательного аппарата: передвижными, регулируемые эргономическими партами СИ-1; компьютерной техникой со специальным программным обеспечением.

9. Методические материалы

9.1 Планы семинарских/ практических/ лабораторных занятий

Пример описания семинарского занятия:

Тема 1. Вопросы для обсуждения:

- Расскажите об алгоритме и особенности булева поиска? Приведите примеры и поясните их.
 Как создать инвертированный индекс? Приведите примеры и поясните их.
 Каким образом происходит обработка булевых запросов? ? Приведите примеры и поясните их.
 Сравните расширенной булевой модели и ранжированного поиска, расскажите об отличиях?
 Приведите примеры и поясните их.
 Расскажите о списках словопозиций? Приведите примеры и поясните их.
 Как происходит схематизация документа и декодирование последовательности символов?
 Приведите примеры и поясните их.
 Как определяется лексикон терминов? Приведите примеры и поясните их.
 Как осуществляется быстрое пересечение инвертированных списков с помощью указателей пропусков? Приведите примеры и поясните их.
 Расскажите о словопозициях с координатами и фразовых запросах? Приведите примеры и поясните их.
 Расскажите о словарях и нечётком поиске? Приведите примеры и поясните их.
 Как устроены поисковые структуры для словарей? Приведите примеры и поясните их.
 Что такое запросы с джокером? Приведите примеры и поясните их.

Тема 2. Вопросы для обсуждения:

- Опишите основы аппаратного обеспечения важные для информационного поиска? Приведите примеры и поясните их.
 Как устроено блочное индексирование, основанное на сортировке? Приведите примеры и поясните их.
 Как устроено однопроходное индексирование в оперативной памяти? Приведите примеры и поясните их.
 Что такое распределенное индексирование? Опишите основные принципы и подходы?
 Приведите примеры и поясните их.
 Что такое Динамическое индексирование? Опишите основные принципы и подходы?
 Приведите примеры и поясните их.
 Как устроено сжатие индекса? Приведите примеры и поясните их.
 Что такое статистические характеристики терминов в информационном поиске? Приведите примеры и поясните их.
 Как устроено сжатие словаря? Приведите примеры и поясните их.
 Как устроено сжатие инвертированного файла? Приведите примеры и поясните их.
 Что такое ранжирование, взвешивание терминов и модель векторного пространства? Приведите примеры и поясните их.
 Как устроено параметрические и зонные индексы? Приведите примеры и поясните их.
 Что такое модель векторного пространства для ранжирования? Приведите примеры и поясните их.

Тема 3. Вопросы для обсуждения:

- Как устроено эффективное ранжирование? Приведите примеры и поясните их.
 Опишите компоненты информационно-поисковой системы? Приведите примеры и поясните их.
 Как влияют операторы языка запросов на ранжирование в векторном пространстве? Приведите примеры и поясните их.
 Как устроены стандартные текстовые коллекции? Приведите примеры и поясните их.

Как осуществляется оценка неранжированных результатов поиска? Приведите примеры и поясните их.

Как осуществляется оценка ранжированных результатов поиска? Приведите примеры и поясните их.

Как осуществляется оценка релевантности? Приведите примеры и поясните их.

Что такое сниппеты? Приведите примеры и поясните их.

Зачем нужна обратная связь по релевантности и псевдорелевантности? Приведите примеры и поясните их.

Опишите глобальные методы для переформулирования запроса? Приведите примеры и поясните их.

Тема 4. Вопросы для обсуждения:

Опишите основные концепции языка XML? Приведите примеры и поясните их.

Опишите проблемы, связанные с XML -поиском? Приведите примеры и поясните их.

Как устроена модель векторного пространства для XML поиска? Приведите примеры и поясните их.

Опишите основные методы XML поиска, ориентированные на текст и на данные? Приведите примеры и поясните их.

Опишите как устроена вероятностная модель информационного поиска? Приведите примеры и поясните их.

Опишите принцип вероятностного ранжирования? Приведите примеры и поясните их.

Опишите бинарная модель независимости? Приведите примеры и поясните их.

Опишите языковые модели для информационного поиска

Зачем нужна модель правдоподобия запроса

Сравните языковые модели с другими подходами к информационному поиску

Тема 5. Вопросы для обсуждения:

Опишите как делается классификация текстов в целом? Приведите примеры и поясните их.

Как устроена наивная байесовская классификация текстов? Приведите примеры и поясните их.

Что такое модель Бернулли? Приведите примеры и поясните их.

Расскажите о свойствах наивной байесовской модели? Приведите примеры и поясните их.

Как осуществляется выбор признаков? Приведите примеры и поясните их.

Как осуществляется оценка классификации текстов? Приведите примеры и поясните их.

Опишите способы представления документов и меры близости в векторном пространстве? Приведите примеры и поясните их.

Опишите как устроен метод Роккио? Приведите примеры и поясните их.

Опишите как устроен метод ближайших соседей? Приведите примеры и поясните их.

Опишите как устроены линейные и нелинейные классификаторы? Приведите примеры и поясните их.

Опишите как устроена классификация с несколькими классами? Приведите примеры и поясните их.

Опишите как устроен метод опорных векторов? Приведите примеры и поясните их.

Опишите проблемы, связанные с классификацией текстовых документов? Приведите примеры и поясните их.

Расскажите о методах машинного обучения для поиска по запросу? Приведите примеры и поясните их.

Тема 6. Вопросы для обсуждения:

Опишите как устроена плоская кластеризация в информационном поиске? Приведите примеры и поясните их.

Опишите как устроена оценка кластеризации? Приведите примеры и поясните их.

Опишите как устроен метод К-средних? Приведите примеры и поясните их.

Опишите как устроена агломеративная иерархическая кластеризация? Приведите примеры и поясните их.

Опишите как устроена кластеризация методом центроидов? Приведите примеры и поясните их. Что такое оптимальность агломеративной иерархической кластеризации? Приведите примеры и поясните их.

Что такое нисходящая кластеризация? Приведите примеры и поясните их.

Тема 7. Вопросы для обсуждения:

Опишите как устроена матрица “термин-документ” и сингулярные разложения? Приведите примеры и поясните их.

Что такое латентно-семантическое индексирование? Приведите примеры и поясните их.

Как представить Веб виде графа? Приведите примеры и поясните их.

Опишите как устроен Метод PageRank? Приведите примеры и поясните их.

В процессе обсуждения вопросов на семинаре студенты должны собрать корпус текстов и с использованием программного продукта Полианалист, выполнить тематическое моделирование и кластеризацию корпуса текстов. Результат представляется потоком работ, реализованным в Полианалисте.

9.2 Иные материалы

Любой вид занятий, создающий условия для зарождения самостоятельной мысли, познавательной и творческой активности студента связан с самостоятельной работой. В широком смысле под самостоятельной работой понимают совокупность всей самостоятельной деятельности студентов как в учебной аудитории, так и вне ее, в контакте с преподавателем и в его отсутствие. Самостоятельная работа может реализовываться: непосредственно в процессе аудиторных занятий – на лекциях, практических и семинарских занятиях, при выполнении контрольных и лабораторных работ и др.; в контакте с преподавателем вне рамок аудиторных занятий – на консультациях по учебным вопросам, в ходе творческих контактов, при ликвидации задолженностей, при выполнении индивидуальных заданий и т.д.; в библиотеке, дома, в общежитии, на кафедре и других местах при выполнении студентом учебных и творческих заданий.

Лекции

Главное в период подготовки к лекционным занятиям – научиться методам самостоятельного умственного труда, сознательно развивать свои творческие способности и овладевать навыками творческой работы. Для этого необходимо строго соблюдать дисциплину учебы и поведения. Четкое планирование своего рабочего времени и отдыха является необходимым условием для успешной самостоятельной работы. В основу его нужно положить рабочие программы изучаемых в семестре дисциплин. Каждому студенту следует составлять еженедельный и семестровый планы работы, а также план на каждый рабочий день. С вечера всегда надо распределять работу на завтрашний день. В конце каждого дня целесообразно подводить итог работы: тщательно проверить, все ли выполнено по намеченному плану, не было ли каких-либо отступлений, а если были, по какой причине это произошло. Нужно осуществлять самоконтроль, который является необходимым условием успешной учебы. Если что-то осталось невыполненным, необходимо изыскать время для завершения этой части работы, не уменьшая объема недельного плана.

Самостоятельная работа на лекции Слушание и запись лекций – сложный вид вузовской аудиторной работы. Внимательное слушание и конспектирование лекций предполагает интенсивную умственную деятельность студента. Краткие записи лекций, их конспектирование помогает усвоить учебный материал. Конспект является полезным тогда, когда записано самое

существенное, основное и сделано это самим студентом. Не надо стремиться записать дословно всю лекцию. Такое

«конспектирование» приносит больше вреда, чем пользы. Запись лекций рекомендуется вести по возможности собственными формулировками. Желательно запись осуществлять на одной странице, а следующую оставлять для проработки учебного материала самостоятельно в домашних условиях. Конспект лекции лучше подразделять на пункты, параграфы, соблюдая красную строку. Этому в большой степени будут способствовать пункты плана лекции, предложенные преподавателям. Принципиальные места, определения, формулы и другое следует сопровождать замечаниями «важно», «особо важно», «хорошо запомнить» и т.п. Можно делать это и с помощью разноцветных маркеров или ручек. Лучше если они будут собственными, чтобы не приходилось просить их у однокурсников и тем самым не отвлекать их во время лекции. Целесообразно разработать собственную «маркографию» (значки, символы), сокращения слов. Не лишним будет и изучение основ стенографии. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть знаниями.

Семинар и проведение опроса

Каждый студент должен начать с ознакомления с планом занятия, который отражает содержание предложенной темы. Тщательное продумывание и изучение вопросов плана основывается на проработке текущего материала лекции, а затем изучения обязательной и дополнительной литературы, рекомендованную к данной теме. На основе индивидуальных предпочтений студенту необходимо самостоятельно выбрать тему доклада по проблеме семинара и по возможности подготовить по нему презентацию. Если программой дисциплины предусмотрено выполнение практического задания, то его необходимо выполнить с учетом предложенной инструкции (устно или письменно). Все новые понятия по изучаемой теме необходимо выучить наизусть и внести в глоссарий, который целесообразно вести с самого начала изучения курса. Результат такой работы должен проявиться в способности студента свободно ответить на теоретические вопросы семинара, его выступлении и участии в коллективном обсуждении вопросов изучаемой темы, правильном выполнении практических заданий и контрольных работ.

Работа с литературными источниками.

В процессе подготовки к семинарским занятиям, студентам необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также научной и популярной) литературы. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у студентов свое отношение к конкретной проблеме. Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем по каждой теме семинарского или практического занятия, что позволяет студентам проявить свою индивидуальность в рамках выступления на данных занятиях, выявить широкий спектр мнений по изучаемой проблеме.

Сессия

Каждый учебный семестр заканчивается зачетно-экзаменационной сессией. Подготовка к зачетно-экзаменационной сессии, сдача зачетов и экзаменов является также самостоятельной

работой студента. Основное в подготовке к сессии – повторение всего учебного материала дисциплины, по которому необходимо сдавать зачет или экзамен. Только тот студент успевает, кто хорошо усвоил учебный материал. Если студент плохо работал в семестре, пропускал лекции, не конспектировал, не изучал рекомендованную литературу, то в процессе подготовки к сессии ему придется не повторять уже знакомое, а заново в короткий срок изучать весь учебный материал. Все это зачастую невозможно сделать из-за нехватки времени. Для такого студента подготовка к зачету или экзамену будет трудным, а иногда и непосильным делом, а конечный результат – возможное отчисление из учебного заведения.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина (*модуль*) реализуется Отделением интеллектуальных систем в гуманитарной сфере.

Цель учебной дисциплины «Интеллектуальный информационный поиск» состоит в том, чтобы помочь обучающимся овладеть возможностями современного цифрового пространства для решения коммуникативных и прикладных задач лингвистики и литературоведения.

Курс включает в себя лекционные, семинарские и практические занятия. Общая трудоемкость освоения дисциплины (модуля) составляет 108 часов (3 з.е.)

Основными **задачами** изучения дисциплины являются:

- изучение методов компьютерной обработки текстов в электронном виде, обеспечивающие поддержку работы лингвистов и литературоведов при обработке информации, анализе данных и интерпретации результатов;
- формирование умения работать с корпусами, лингвистическими корпусами, электронными библиотеками, поисковыми системами; справочными информационными ресурсами цифровой гуманитаристики;
 - овладение практическими навыками по поиску, анализу, разработке, созданию информационных ресурсов;
 - овладение технологиями поиска, анализа, и извлечения мнений из различных корпусов текстов.

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
УК-1	<i>УК-1.1 Обладает готовностью действовать в нестандартных ситуациях</i>	Знать: различие форм, методов, приемов и последовательности действий в стандартных и нестандартных коммуникативных и педагогических ситуациях; варианты альтернативных путей выхода из различных ситуаций; Уметь: проявлять инициативу и избегать автоматического применения стандартных форм и приемов при решении задач в нестандартных коммуникативных и педагогических ситуациях; Владеть: системой навыков действий в нестандартных коммуникативных и педагогических ситуациях.
	<i>УК-1.2 Способен нести социальную и этическую ответственность за принятые решения</i>	Знать: основы социальной и этической ответственности при принятии решений; Уметь: выделять и систематизировать основные представления о социальной и

		этической ответственности за принятые решения; прогнозировать возможные нестандартные ситуации; принимать и критически оценивать принятые решения; Владеть: умением прогнозировать результаты социальной и этической ответственности за принятые решения.
	<i>УК-1.3 Демонстрирует умение критически оценивать и анализировать информацию, приобретенную самостоятельно, в том числе с помощью информационных технологий, и использовать ее в практической деятельности</i>	Знать: основные требования информационной безопасности. Уметь: решать задачи по поиску источников и научной литературы. Владеть: навыками поиска научной литературы и составления списка источников и литературы для научной работы.
<i>ПК-3</i>	<i>ПК-1.3 Способен определить стилистическую, жанровую, формальную и содержательную специфику текста</i>	Знать: основные требования информационной безопасности. Уметь: решать задачи по поиску источников и научной литературы. Владеть: навыками поиска научной литературы и составления списка источников и литературы для научной работы.

По дисциплине (модулю) предусмотрена промежуточная аттестация в форме экзамена .

Общая трудоемкость освоения дисциплины (модуля) составляет 3 ___ зачетных единиц.