

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Багмурова Назми Дюдинович
Должность: Врио ректора
Дата подписания: 28.03.2022 12:05:35
Уникальный программный ключ:
b261c06f25acbb0d1e6de5fc04abdfed0091d138


МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГБОУ ВО «ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

РЕКОМЕНДОВАНО

К УТВЕРЖЕНИЮ

Декан факультета

магистерской подготовки



Ашуралиева Р.К.
« 02 » 03 2020 г.

УТВЕРЖДАЮ

Врио ректора ДГТУ,

Председатель методического

совета ДГТУ


Суракатов Н.С.
« 03 » 03 2020 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Дисциплина М1.В.ДВ.2 Автоматизированные информационно-поисковые системы
код и наименование дисциплины по ООП

для направления 09.04.01 «Информатика и вычислительная техника»
код и направление направления подготовки

по профилю Сети ЭВМ и телекоммуникации
наименование магистерской программы

факультет Магистерской подготовки
наименование факультета, где ведется дисциплина (практика)

кафедра Управление и информатика в технических системах и вычислительной техники
наименование кафедры, за которой закреплена дисциплина (практика)

Квалификация выпускника (степень) Магистр
бакалавр, магистр (специалист)

Форма обучения очная курс 1 семестр (ы) 1
очная, заочная, др.

Всего трудоемкость в зачетных единицах (часах) 3 ЗЕТ (108)

лекции 9 экзамен -
час семестр

практические (семинарские) занятия 17 зачет 1
час семестр

лабораторные занятия 17 самостоятельная работа 65
час час

курсовой проект (работа, РГР) -
семестр

И.о. зав. кафедрой


подпись

Асланов Т.Г.

Начальник УО


подпись

Магомаева Э.В.

Программа составлена в соответствии с требованиями ФГОС ВО с учетом рекомендаций ООП ВО по направлению подготовки 09.04.01 «Информатика и вычислительная техника»

Программа одобрена на заседании выпускающей кафедры от «28» 02 2020 года, протокол № 6.

И.о. зав. кафедрой по данному направлению



Асланов Т.Г.

подпись

ОДОБРЕНО


**Методической комиссией
по УГС(Н)
09.00.00 – Информатика и
вычислительная техника**

АВТОР ПРОГРАММЫ
К.т.н., ст. преп. Т.Г. Асланов



подпись

Председатель М.К.



Абдулгалимов А.М.

подпись

«28» 02 2020г.

1. Цели освоения дисциплины

Целью дисциплины является ознакомление обучающихся с предметной областью теории информационного поиска и лингвистической информатики. Лекционный курс охватывает её основные разделы: теорию индексирования, теорию реферирования, теорию кластеризации цитирования, теорию логико-смыслового моделирования.

2. Место дисциплины в структуре ООП магистратуры

Дисциплина «Автоматизированные информационно-поисковые системы» представляет собой вариативную часть дисциплин по выбору учебного плана.

Дисциплина «Автоматизированные информационно-поисковые системы» является предшествующей для дисциплины «Методы администрирования вычислительных сетей».

3. Компетенции обучающегося, формируемые в результате освоения дисциплины «Автоматизированные информационно-поисковые системы»

Выпускник, освоивший программу магистратуры, должен обладать следующими компетенциями:

общекультурными (ОК):

– способностью совершенствовать и развивать свой интеллектуальный и общекультурный уровень (ОК-1);

– способностью понимать роль науки в развитии цивилизации, соотношение науки и техники, иметь представление о связанных с ними современных социальных и этических проблемах, понимать ценность научной рациональности и ее исторических типов (ОК-2);

общепрофессиональными (ОПК):

– владением, по крайней мере, одним из иностранных языков на уровне социального и профессионального общения, способностью применять специальную лексику и профессиональную терминологию языка (ОПК-4);

– владением методами и средствами получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях (ОПК-5);

– способностью анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями (ОПК-6)

профессиональными компетенциями (ПК):

– способностью выбирать методы и разрабатывать алгоритмы решения задач управления и проектирования объектов автоматизации (ПК-12);

– способностью к программной реализации распределенных информационных систем (ПК-13);

– способностью к программной реализации систем с параллельной обработкой данных и высокопроизводительных систем (ПК-14);

В результате освоения дисциплины обучающийся должен:

Знать: способы совершенствования и развития своего интеллектуального и

общекультурного уровня; роль науки в развитии цивилизации; виды научных исследований; виды исследовательских и проектных работ; виды рисков; методы проектирования объектов автоматизации; теоретические основы построения распределенных информационных систем; методы программной реализации систем с параллельной обработкой данных и высокопроизводительных систем.

Уметь: совершенствовать знания в выбранной области для решения профессиональных задач; анализировать социальные и этические проблемы науки и техники; заниматься научными исследованиями; использовать на практике умения и навыки в организации исследовательских и проектных работ, в управлении коллективом; проявлять инициативу, в том числе в ситуациях риска, брать на себя всю полноту ответственности; применять методы проектирования объектов автоматизации; применять аппарат распределенных информационных систем для решения задач в конкретной предметной области; программным путем реализовывать системы с параллельной обработкой данных.

Владеть: технологиями анализа, синтеза и оценки общекультурных и профессиональных знаний; навыками рационального мышления; навыками проведения научных исследований; навыками проведения исследовательских и проектных работ; навыками нивелирования рисков; средами проектирования; навыками программной реализации распределенных информационных систем; навыками реализации систем с параллельной обработкой данных.

4. Содержание дисциплины «Автоматизированные информационно-поисковые системы»

4.1 Содержание дисциплины

№	Раздел дисциплины Тема лекции и вопросы	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)				Формы текущего контроля успеваемости (по срокам текущих аттестаций в семестре) Форма промежуточной аттестации (по семестрам)
				ЛК	ПЗ	ЛБ	СР	
а	б	в	г	д	е	ж	з	и
1	Лекция 1 ТЕМА: Алгоритмы и программы автоматического анализа текста. Морфологический анализ и аннотирование. Аннотирование и взвешивание терминов. Декомпозиция и синтаксический парсинг. 1. Алгоритмы и программы автоматического анализа текста и уровни языковой системы. 2. Морфологический анализ. 3. Стемминг и стеммеры. 4. Алгоритмические и словарные	1	1	2	4	4	13	Входная контрольная работа

<p> стеммеры. 5. Недостаточное и избыточное стеммирование. Y- стеммер. 6. Отличие стемминга от лемматизации. 7. Понятие лексической декомпозиции и токена. 8. Особенности токенизации. 9. Необходимость распознавания единиц больше и меньше чем слово. 10. Аннотирование. 11. Понятие POS-тегов. 12. Теггеры, основание на правилах и стохастические теггеры. 13. Алгоритм двунаправленной инференции. 14. Семантические и когнитивные теги. 15. Использование семантических тегов в фактографических ИПС. 16. Использование когнитивных тегов в системах интеллектуального анализа текста. 17. Обобщённая архитектура POS-теггера. 18. Семантические и когнитивные теги. 19. Использование семантических тегов в фактографических ИПС. 20. Использование когнитивных тегов в системах интеллектуального анализа текста. 21. Алгоритмы взвешивания терминов и фильтры. 22. Интертекстуальные и интратекстуальные методы взвешивания. 23. Определение вероятностных величин. 24. Проблема сопоставления с эталонным корпусом. 25. Алгоритм TF*IDF и возможности его применения для фильтрации стоп слов и классификации и категоризации текстов. 26. Понятие n-gram: биграмм, триграмм, тетраграмм. 27. Алгоритм распознавания n-gram в тексте. 28. Возможности использования n-gram для автоматической классификации текстов. 29. Понятие синтаксического парсинга. 30. Распознавание иерархической структуры предложения на основе выделения словосочетаний. 31. Lexparser (Стэнфордский университет). Значение парсинга для моделирования структуры текста. 32. Программы-чанкеры. 33. Значение распознавания словосочетаний различных типов. 34. Понятие клаузы. 35. Алгоритмы распознавания клауз. 36. Значение разбивки текста на клаузы </p>									
---	--	--	--	--	--	--	--	--	--

	для моделирования его логико-семантической структуры.							
2	<p>Лекция 2</p> <p>ТЕМА: Индексирование как вид информационно-лингвистического моделирования. ИПС и ИПЯ</p> <ol style="list-style-type: none"> 1. Дискурсивный анализ текста. 2. Понятие дискурса в системах автоматической обработки текста. 3. Значение разрешение анафоры для систем реферирования и ИПС. 4. Алгоритмы и правила разрешения анафоры. 5. Семантическая структура текста и концепция У. Манна. 6. Программа RST-Tool. 7. Лексикографические ресурсы для систем NLP. 8. Онтологии, словари и тезаурусы. 9. Структура тезауруса WordNet. 10. Значение использования тезаурусов в ИПС и системах реферирования. 11. Закономерности предметной области. 12. Закон Ципфа, его предсказательная сила. 13. Закон Брэдфорда. 14. Трактовка информации в кибернетике и информатике; различия между объёмным и вероятностными подходами к определению количества информации. 15. Методика расчёта количества информации для текстов. 	5	2	4	4	13	Аттестационная контрольная работа 1	
3	<p>Лекция 3</p> <p>ТЕМА: Особенности информационно-поисковых систем Интернета</p> <ol style="list-style-type: none"> 1. Историческое развитие предметной области. 2. Т. Кун и закономерности исторического развития научных дисциплин. 3. Историческое развитие лингвистической информатики. 4. Александрийская библиотека и первые рефераты; появление реферативных журналов в XIX веке. М. Дюи и универсальная десятичная классификация (1873). 5. Г. Тейлор и механизация информационного поиска с помощью перфокарт. 6. Появление первых систем автоматического индексирования. 7. Г. Лун и системы автоматического реферирования. А.И. Михайлов, А.И. Черный, Р.С. Гиляревский и термин "информатика" в 60-е гг. XX века. 8. Кластеризация социтирования и логико-смысловое моделирование в 70-х гг. XX века. 9. Информационный взрыв конца XX 	9	2	4	4	13	Аттестационная контрольная работа 2	

	века, его особенности. 10. Интеграционные тенденции в современной лингвистической информатике.						
4	<p>Лекция 4</p> <p>ТЕМА: Лексикографические ресурсы и закономерности</p> <ol style="list-style-type: none"> 1. Предметная область лингвистической информатики. 2. Признаки информационно-лингвистических моделей (ИЛМ). 3. Отличие информационно-лингвистических моделей от лингвистических и информационных моделей. 4. Виды ИЛМ: лексико-семантические, логико-грамматические, дискретные, непрерывные. 5. Виды информационно-лингвистического моделирования. 6. Гносеологический и онтологический планы лингвистической информатики. 7. Единство онтологического плана лингвистической информатики. 8. Соотношение лингвистической информатики со смежными дисциплинами. 9. Значение терминов "компьютерная лингвистика", "прикладная лингвистика", "корпусная лингвистика", "информатика", "прикладная информатика". 	13	2	2	2	13	Аттестационная контрольная работа 3
5	<p>Лекция 5</p> <p>ТЕМА: Системы автоматического реферирования.</p> <ol style="list-style-type: none"> 1. Оценка эффективности функционирования АИПС. 2. Понятия пертинентности, релевантности, полноты и точности поиска, информационного шума, потери информации. 3. Количественные методы определения информационного шума и потерь информации; коэффициенты полноты и точности поиска. 4. Информационно-поисковые системы Интернета (ИАИПС). 5. Особенности электронных баз данных и поведения пользователей Интернета. 6. Виды ИАИПС: предметные каталоги и индексные ИАИПС, их особенности; распределённые ИАИПС типа <i>Comernic</i>; мета- ИАИПС. 7. Архитектура индексных ИАИПС и её отличие от архитектуры традиционных АИПС; особенности функционирования поисковой машины и робота-индексировщика. 8. Полнотекстовые базы данных; понятия инвертированного файла и пост-листов. 9. Способы повышения эффективности 	17	1	3	3	13	

<p>функционирования АИПС: двухступенчатость выдачи, эшелонирование выдачи, приписывание весовых коэффициентов дескрипторам.</p> <p>10. Кластеризация социтирования; понятия кластера и социтирования, примеры кластеров социтирования.</p> <p>11. Значение кластеризации социтирования для отслеживания развития научных идей и научной дисциплины Г.Смолл, Ю.Гарфилд, И.В.Маршакова.</p> <p>12. Гипертекстовые системы.</p> <p>13. Примеры гипертекстов.</p> <p>14. М.М.Субботин и логико-смысловое моделирование.</p>							
Итого:			9	17	17	65	Зачет

4.2 Содержание практических занятий

№ п/п	№ по содержанию дисциплины	Наименование практического занятия	Количество часов	Рекомендуемая литература и методические разработки (№ источника из списка литературы)
1	1	Алгоритмы и программы автоматического анализа текста и уровни языковой системы.	4	1-5
2	2	Методика расчёта количества информации для текстов.	4	1-5
3	3	Интеграционные тенденции в современной лингвистической информатике.	4	1-5
4	4	Виды информационно-лингвистического моделирования.	2	1-5
5	5	Информационно-поисковые системы Интернета (ИАИПС).	3	1-5
Итого:			17	

4.3 Содержание лабораторных занятий

№ п/п	№ по содержанию дисциплины	Наименование лабораторного занятия	Количество часов	Рекомендуемая литература и методические разработки (№ источника из списка литературы)
1	1	Лабораторная работа по интерпретации понятия информации в кибернетике и информатике. Подсчёт количества информации с точки зрения двух подходов.	4	1-5
2	2	Лабораторная работа по методам	4	1-5

		взвешивания терминов, статистические и позиционные параметры. Взвешивание по алгоритму tf/idf. Взвешивание терминов и распределения закона Ципфа.		
3	3	Лабораторная работа по системам автоматического индексирования.	4	1-5
4	4	Лабораторная работа по общей характеристике и алгоритмам функционирования современных систем автоматического реферирования текста. Метод симметричного реферирования.	2	1-5
5	5	Лабораторная работа по методам оценки качества современных систем автоматического реферирования текста. Метод сопоставления с эталонным словарём. Метод сопоставления с эталонным рефератом.	3	1-5
Итого:			17	

4.4 Тематика для самостоятельной работы студента

№ п/п	Тематика по содержанию дисциплины, выделенная для самостоятельного изучения	Количество часов из содержания дисциплины	Рекомендуемая литература и источники информации	Формы контроля СРС
1	Алгоритмы и программы автоматического анализа текста и уровни языковой системы. Морфологический анализ. Стемминг и стеммеры. Алгоритмические и словарные стеммеры. Недостаточное и избыточное стеммирование. У-стеммер. Отличие стемминга от лемматизации. Понятие лексической декомпозиции и токена. Особенности токенизации. Необходимость распознавания единиц больше и меньше чем слово. Аннотирование. Понятие POS-тегов. Теггеры, основание на правилах и стохастические теггеры. Алгоритм двунаправленной инференции. Семантические и когнитивные теги. Использование семантических тегов в фактографических ИПС.	13	1-5	Опрос

	<p>Использование когнитивных тегов в системах интеллектуального анализа текста. Обобщённая архитектура POS-теггера. Семантические и когнитивные теги. Использование семантических тегов в фактографических ИПС. Использование когнитивных тегов в системах интеллектуального анализа текста. Алгоритмы взвешивания терминов и фильтры. Интертекстуальные и интратекстуальные методы взвешивания. Определение вероятностных величин. Проблема сопоставления с эталонным корпусом. Алгоритм TF*IDF и возможности его применения для фильтрации стоп слов и классификации и категоризации текстов. Понятие n-gram: биграмм, триграмм, тетраграм. Алгоритм распознавания n-gram в тексте. Возможности использования n-gram для автоматической классификации текстов. Понятие синтаксического парсинга. Распознавание иерархической структуры предложения на основе выделения словосочетаний. Lxparser (Стэнфордский университет). Значение парсинга для моделирования структуры текста. Программы-чанкеры. Значение распознавания словосочетаний различных типов. Понятие клаузы. Алгоритмы распознавания клауз. Значение разбивки текста на клаузы для моделирования его логико-семантической структуры.</p>			
2	<p>Дискурсивный анализ текста. Понятие дискурса в системах автоматической обработки текста. Значение разрешение</p>	13	1-5	Опрос

	<p>анафоры для систем реферирования и ИПС. Алгоритмы и правила разрешения анафоры. Семантическая структура текста и концепция У. Манна. Программа RST-Tool. Лексикографические ресурсы для систем NLP. Онтологии, словари и тезаурусы. Структура тезауруса WordNet. Значение использования тезаурусов в ИПС и системах реферирования. Закономерности предметной области. Закон Ципфа, его предсказательная сила. Закон Брэдфорда. Трактовка информации в кибернетике и информатике; различия между объёмным и вероятностными подходами к определению количества информации. Методика расчёта количества информации для текстов.</p>			
3	<p>Историческое развитие предметной области. Т. Кун и закономерности исторического развития научных дисциплин. Историческое развитие лингвистической информатики. Александрийская библиотека и первые рефераты; появление реферативных журналов в XIX веке. М. Дюи и универсальная десятичная классификация (1873). Г. Тейлор и механизация информационного поиска с помощью перфокарт. Появление первых систем автоматического индексирования. Г. Лун и системы автоматического реферирования. А.И. Михайлов, А.И. Черный, Р.С. Гиляревский и термин "информатика" в 60-е гг. XX века. Кластеризация цитирования и логико-смысловое моделирование в 70-х гг. XX века.</p>	13	1-5	Опрос

	Информационный взрыв конца XX века, его особенности. Интеграционные тенденции в современной лингвистической информатике.			
4	Предметная область лингвистической информатики. Признаки информационно-лингвистических моделей (ИЛМ). Отличие информационно-лингвистических моделей от лингвистических и информационных моделей. Виды ИЛМ: лексико-семантические, логико-грамматические, дискретные, непрерывные. Виды информационно-лингвистического моделирования. Гносеологический и онтологический планы лингвистической информатики. Единство онтологического плана лингвистической информатики. Соотношение лингвистической информатики со смежными дисциплинами. Значение терминов "компьютерная лингвистика", "прикладная лингвистика", "корпусная лингвистика", "информатика", "прикладная информатика".	13	1-5	Опрос
5	Оценка эффективности функционирования АИПС. Понятия пертинентности, релевантности, полноты и точности поиска, информационного шума, потери информации. Количественные методы определения информационного шума и потерь информации; коэффициенты полноты и точности поиска. Информационно-поисковые системы Интернета (ИАИПС). Особенности электронных баз	13	1-5	Опрос

<p>данных и поведения пользователей Интернета. Виды ИАИПС: предметные каталоги и индексные ИАИПС, их особенности; распределённые ИАИПС типа Copernic; мета- ИАИПС. Архитектура индексных ИАИПС и её отличие от архитектуры традиционных АИПС; особенности функционирования поисковой машины и робота-индексировщика. Полнотекстовые базы данных; понятия инвертированного файла и пост-листов. Способы повышения эффективности функционирования АИПС: двухступенчатость выдачи, эшелонирование выдачи, приписывание весовых коэффициентов дескрипторам. Кластеризация социтирования; понятия кластера и социтирования, примеры кластеров социтирования. Значение кластеризации социтирования для отслеживания развития научных идей и научной дисциплины Г. Смолл, Ю. Гарфилд, И.В. Маршакова. Гипертекстовые системы. Примеры гипертекстов. М.М. Субботин и логико-смысловое моделирование.</p>			
Итого:	65		

5. Образовательные технологии

В ходе проведения занятий используются такие методы обучения как презентация, применение компьютерной техники.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

6.1 Перечень вопросов по проверке входных знаний студентов

1. Технология (определение)

2. Информационная технология (определения)
3. Этапы развития информационных технологий
4. Классификация информационных технологий
5. Информационная система (определения)
6. Классификация информационных систем
7. Этапы цикла разработки информационных систем и их целевая продукция
8. Этапы цикла разработки информационных систем. Системные исследования
9. Этапы цикла разработки информационных систем. Системный анализ
10. Этапы цикла разработки информационных систем. Системное проектирование
11. Этапы цикла разработки информационных систем. Внедрение
12. Этапы цикла разработки информационных систем. Сопровождение
13. Управление проектом разработки информационной системы
14. CASE-технология
15. Назначение и область применения CASE-технологии
16. Программные продукты, реализующие возможности CASE-технологии
17. IDEF-диаграмма и ее элементы в программном продукте BPWin
18. Данные, база данных, банк данных
19. Система управления базами данных
20. Виды моделей данных базы данных
21. Структура базы данных
22. Уровни представления данных базы данных, архитектура баз данных
23. Технологии «Файл-сервер» и «Клиент-сервер»
24. Хранилище данных

6.2. Задания для текущих аттестаций

6.2.1. Контрольные вопросы для первой аттестации

1. Алгоритмы и программы автоматического анализа текста и уровни языковой системы.
2. Морфологический анализ.
3. Стемминг и стеммеры.
4. Алгоритмические и словарные стеммеры.
5. Недостаточное и избыточное стеммирование.
6. Y-стеммер.
7. Отличие стемминга от лемматизации.
8. Понятие лексической декомпозиции и токена.
9. Особенности токенизации.
10. Необходимость распознавания единиц больше и меньше чем слово.
11. Аннотирование.
12. Понятие POS-тегов.
13. Теггеры, основание на правилах и стохастические теггеры.
14. Алгоритм двунаправленной инференции.
15. Семантические и когнитивные теги.
16. Использование семантических тегов в фактографических ИПС.

17. Использование когнитивных тегов в системах интеллектуального анализа текста.
18. Обобщённая архитектура POS-теггера.
19. Семантические и когнитивные теги.
20. Использование семантических тегов в фактографических ИПС.
21. Использование когнитивных тегов в системах интеллектуального анализа текста.
22. Алгоритмы взвешивания терминов и фильтры.
23. Интертекстуальные и интратекстуальные методы взвешивания.
24. Определение вероятностных величин.
25. Проблема сопоставления с эталонным корпусом.
26. Алгоритм $TF*IDF$ и возможности его применения для фильтрации стоп слов и классификации и категоризации текстов.
27. Понятие n-gram: биграмм, триграмм, тетраграм.
28. Алгоритм распознавания n-gram в тексте.
29. Возможности использования n-gram для автоматической классификации текстов.
30. Понятие синтаксического парсинга.

6.2.2. Контрольные вопросы для второй аттестации

1. Распознавание иерархической структуры предложения на основе выделения словосочетаний.
2. Lexparser (Стэнфордский университет).
3. Значение парсинга для моделирования структуры текста.
4. Программы-чанкеры.
5. Значение распознавания словосочетаний различных типов.
6. Понятие клаузы.
7. Алгоритмы распознавания клауз.
8. Значение разбивки текста на клаузы для моделирования его логико-семантической структуры.
9. Дискурсивный анализ текста.
10. Понятие дискурса в системах автоматической обработки текста.
11. Значение разрешение анафоры для систем реферирования и ИПС.
12. Алгоритмы и правила разрешения анафоры.
13. Семантическая структура текста и концепция У. Манна.
14. Программа RST-Tool.
15. Лексикографические ресурсы для систем NLP.
16. Онтологии, словари и тезаурусы.
17. Структура тезауруса WordNet.
18. Значение использования тезаурусов в ИПС и системах реферирования.
19. Закономерности предметной области.
20. Закон Ципфа, его предсказательная сила.
21. Закон Брэдфорда.
22. Трактовка информации в кибернетике и информатике; различия между объёмным и вероятностными подходами к определению количества информации.

23. Методика расчёта количества информации для текстов.
24. Историческое развитие предметной области.
25. Т. Кун и закономерности исторического развития научных дисциплин.
26. Историческое развитие лингвистической информатики.
27. Александрийская библиотека и первые рефераты; появление реферативных журналов в XIX веке.
28. М. Дюи и универсальная десятичная классификация (1873).
29. Г. Тейлор и механизация информационного поиска с помощью перфокарт.
30. Появление первых систем автоматического индексирования.

6.2.3. Контрольные вопросы для третьей аттестации

1. Г. Лун и системы автоматического реферирования.
2. А.И. Михайлов, А.И. Черный, Р.С. Гиляревский и термин "информатика" в 60-е гг. XX века.
3. Кластеризация цитирования и логико-смысловое моделирование в 70-х гг. XX века.
4. Информационный взрыв конца XX века, его особенности.
5. Интеграционные тенденции в современной лингвистической информатике.
6. Предметная область лингвистической информатики.
7. Признаки информационно-лингвистических моделей (ИЛМ).
8. Отличие информационно-лингвистических моделей от лингвистических и информационных моделей.
9. Виды ИЛМ: лексико-семантические, логико-грамматические, дискретные, непрерывные.
10. Виды информационно-лингвистического моделирования.
11. Гносеологический и онтологический планы лингвистической информатики.
12. Единство онтологического плана лингвистической информатики.
13. Соотношение лингвистической информатики со смежными дисциплинами.
14. Значение терминов "компьютерная лингвистика", "прикладная лингвистика", "корпусная лингвистика", "информатика", "прикладная информатика".
15. Оценка эффективности функционирования АИПС.
16. Понятия пертинентности, релевантности, полноты и точности поиска, информационного шума, потери информации.
17. Количественные методы определения информационного шума и потерь информации; коэффициенты полноты и точности поиска.
18. Информационно-поисковые системы Интернета (ИАИПС).
19. Особенности электронных баз данных и поведения пользователей Интернета.
20. Виды ИАИПС: предметные каталоги и индексные ИАИПС, их особенности; распределённые ИАИПС типа *Соретіс*; мета- ИАИПС.
21. Архитектура индексных ИАИПС и её отличие от архитектуры традиционных АИПС; особенности функционирования поисковой машины и робота-индексировщика.

22. Полнотекстовые базы данных; понятия инвертированного файла и пост-листов.

23. Способы повышения эффективности функционирования АИПС: двухступенчатость выдачи, эшелонирование выдачи, приписывание весовых коэффициентов дескрипторам.

24. Кластеризация цитирования; понятия кластера и цитирования, примеры кластеров цитирования.

25. Значение кластеризации цитирования для отслеживания развития научных идей и научной дисциплины Г. Смолл, Ю. Гарфилд, И.В. Маршакова.

26. Гипертекстовые системы.

27. Примеры гипертекстов.

28. М.М. Субботин и логико-смысловое моделирование.

6.3. Перечень вопросов по проверке остаточных знаний

1. Соотношение между законами Бредфорда и Ципфа

2. Структура и функционирование АИПС

3. Системный подход в информатике

4. Индекс цитирования и ранжирование страниц Интернета

5. Различные подходы к интерпретации информации

6. Проблемы категоризации текстов в процессе информационного поиска

7. Оценка эффективности функционирования АИПС Интернета

8. Проблемы разработки информационно-поисковых тезаурусов

9. АИПС "Коперник" и поиск через интранет.

10. Структурно-тематическое реферирование в концепции Н.В. Лукашевич

11. Некоторые особенности современных систем автоматического реферирования.

12. Индикативные и информативные рефераты

6.4. Задания для промежуточной аттестации

6.4.1 Контрольные вопросы для проведения зачета

1. Алгоритмы и программы автоматического анализа текста и уровни языковой системы.

2. Морфологический анализ.

3. Стемминг и стеммеры.

4. Алгоритмические и словарные стеммеры.

5. Недостаточное и избыточное стеммирование.

6. Y-стеммер.

7. Отличие стемминга от лемматизации.

8. Понятие лексической декомпозиции и токена.

9. Особенности токенизации.

10. Необходимость распознавания единиц больше и меньше чем слово.

11. Аннотирование.

12. Понятие POS-тегов.

13. Теггеры, основание на правилах и стохастические теггеры.

14. Алгоритм двунаправленной инференции.
15. Семантические и когнитивные теги.
16. Использование семантических тегов в фактографических ИПС.
17. Использование когнитивных тегов в системах интеллектуального анализа текста.
18. Обобщённая архитектура POS-теггера.
19. Семантические и когнитивные теги.
20. Использование семантических тегов в фактографических ИПС.
21. Использование когнитивных тегов в системах интеллектуального анализа текста.
22. Алгоритмы взвешивания терминов и фильтры.
23. Интертекстуальные и интратекстуальные методы взвешивания.
24. Определение вероятностных величин.
25. Проблема сопоставления с эталонным корпусом.
26. Алгоритм $TF*IDF$ и возможности его применения для фильтрации стоп слов и классификации и категоризации текстов.
27. Понятие n-gram: биграмм, триграмм, тетраграм.
28. Алгоритм распознавания n-gram в тексте.
29. Возможности использования n-gram для автоматической классификации текстов.
30. Понятие синтаксического парсинга.
31. Распознавание иерархической структуры предложения на основе выделения словосочетаний.
32. Lxparser (Стэнфордский университет).
33. Значение парсинга для моделирования структуры текста.
34. Программы-чанкеры.
35. Значение распознавания словосочетаний различных типов.
36. Понятие клаузы.
37. Алгоритмы распознавания клауз.
38. Значение разбивки текста на клаузы для моделирования его логико-семантической структуры.
39. Дискурсивный анализ текста.
40. Понятие дискурса в системах автоматической обработки текста.
41. Значение разрешение анафоры для систем реферирования и ИПС.
42. Алгоритмы и правила разрешения анафоры.
43. Семантическая структура текста и концепция У. Манна.
44. Программа RST-Tool.
45. Лексикографические ресурсы для систем NLP.
46. Онтологии, словари и тезаурусы.
47. Структура тезауруса WordNet.
48. Значение использования тезаурусов в ИПС и системах реферирования.
49. Закономерности предметной области.
50. Закон Ципфа, его предсказательная сила.
51. Закон Брэдфорда.

52. Трактовка информации в кибернетике и информатике; различия между объёмным и вероятностными подходами к определению количества информации.
53. Методика расчёта количества информации для текстов.
54. Историческое развитие предметной области.
55. Т. Кун и закономерности исторического развития научных дисциплин.
56. Историческое развитие лингвистической информатики.
57. Александрийская библиотека и первые рефераты; появление реферативных журналов в XIX веке.
58. М. Дюи и универсальная десятичная классификация (1873).
59. Г. Тейлор и механизация информационного поиска с помощью перфокарт.
60. Появление первых систем автоматического индексирования.
61. Г. Лун и системы автоматического реферирования.
62. А.И. Михайлов, А.И. Черный, Р.С. Гиляревский и термин "информатика" в 60-е гг. XX века.
63. Кластеризация цитирования и логико-смысловое моделирование в 70-х гг. XX века.
64. Информационный взрыв конца XX века, его особенности.
65. Интеграционные тенденции в современной лингвистической информатике.
66. Предметная область лингвистической информатики.
67. Признаки информационно-лингвистических моделей (ИЛМ).
68. Отличие информационно-лингвистических моделей от лингвистических и информационных моделей.
69. Виды ИЛМ: лексико-семантические, логико-грамматические, дискретные, непрерывные.
70. Виды информационно-лингвистического моделирования.
71. Гносеологический и онтологический планы лингвистической информатики.
72. Единство онтологического плана лингвистической информатики.
73. Соотношение лингвистической информатики со смежными дисциплинами.
74. Значение терминов "компьютерная лингвистика", "прикладная лингвистика", "корпусная лингвистика", "информатика", "прикладная информатика".
75. Оценка эффективности функционирования АИПС.
76. Понятия pertinence, релевантности, полноты и точности поиска, информационного шума, потери информации.
77. Количественные методы определения информационного шума и потерь информации; коэффициенты полноты и точности поиска.
78. Информационно-поисковые системы Интернета (ИАИПС).
79. Особенности электронных баз данных и поведения пользователей Интернета.
80. Виды ИАИПС: предметные каталоги и индексные ИАИПС, их особенности; распределённые ИАИПС типа Copernic; мета- ИАИПС.
81. Архитектура индексных ИАИПС и её отличие от архитектуры традиционных АИПС; особенности функционирования поисковой машины и работа-индексировщика.

82. Полнотекстовые базы данных; понятия инвертированного файла и пост-листов.

83. Способы повышения эффективности функционирования АИПС: двухступенчатость выдачи, эшелонирование выдачи, приписывание весовых коэффициентов дескрипторам.

84. Кластеризация социтирования; понятия кластера и социтирования, примеры кластеров социтирования.

85. Значение кластеризации социтирования для отслеживания развития научных идей и научной дисциплины Г. Смолл, Ю. Гарфилд, И.В. Маршакова.

86. Гипертекстовые системы.

87. Примеры гипертекстов.

88. М.М. Субботин и логико-смысловое моделирование.

7. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Рекомендуемая литература и источники информации (основная и дополнительная)

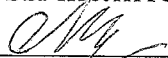
Зав. библиотекой



№ п/п	Необходимая учебная, учебно-методическая (основная и дополнительная) литература, программное обеспечение и Интернет ресурсы	Автор(ы)	Издательство и год издания	Количество изданий	
				В библиотеке	На кафедре
ОСНОВНАЯ					
1	Информационно-поисковые системы : учебно-методическое пособие	Коваленко Ю.В.	Омск : Омская юридическая академия, 2017	IPR BOOKS iprbookshop.ru/ 66817.html	
2	Обработка изображений в пассивных обзорно-поисковых оптико-электронных системах	Артемьев В.М.	Минск : Белорусская наука, 2014	IPR BOOKS iprbookshop.ru/ 29486.html	
ДОПОЛНИТЕЛЬНАЯ					
3	Поисковые системы сети Internet : курс лекций	Королева О.Н.	Москва : Московский гуманитарный университет, 2012	IPR BOOKS iprbookshop.ru/ 14523.html	
4	Информационные системы и технологии. Часть 1 : монография	Корзаченко О.В.	Москва : Перо, Центр научной мысли, 2011	IPR BOOKS iprbookshop.ru/ 8982.html	
5	Информационные системы и технологии. Часть 2 : монография	Корзаченко О.В.	Москва : Перо, Центр научной мысли, 2012	IPR BOOKS iprbookshop.ru/ 8983.html	

8. Материально-техническое обеспечение дисциплины (модуля)

Семинарские занятия по дисциплине проводятся в аудитории с презентационной техникой и учебной мебелью.

Программа составлена в соответствии с требованиями ФГОС ВО с учетом рекомендаций ООП ВО по направлению и профилю подготовки 09.04.01 – Информатика и вычислительная техника. Рецензент от выпускающей кафедры по направлению  Меркухин Е.Н.