

МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ  
Н.Г. ЧЕРНЫШЕВСКОГО»

Механико-математический факультет

УТВЕРЖДАЮ  
Декан механико-математического  
факультета А.М. Захаров  
" 12 " \_\_\_\_\_ 20 21 г.



Рабочая программа дисциплины (модуля)

АВТОМАТИЗАЦИЯ ПОИСКА В ИНТЕРНЕТЕ

Направление подготовки  
09.03.03 - Прикладная информатика

Профиль подготовки  
Прикладная информатика в экономике

Квалификация (степень) выпускника  
Бакалавр

Форма обучения  
заочная

Саратов,  
2021

Статус	ФИО	Подпись	Дата
Преподаватель-разработчик	Коробченко Е.В.		12.11.2021 г.
Председатель НМК	Тышкевич С.В.		12.11.2021 г.
Заведующий кафедрой	Водолазов А.М.		12.11.2021 г.
Специалист Учебного управления			

## 1. Цели освоения дисциплины

Цели освоения дисциплины «Автоматизация поиска в интернете» заключаются в получении представления о современных проблемах информационного поиска и поиска в веб, включая смежные задачи классификации и кластеризации текстов; в знакомстве с ключевыми аспектами проектирования и реализации систем сбора, индексирования и поиска документов, методами оценки таких систем, а также с методами машинного обучения на базе коллекций текстов.

## 2. Место дисциплины в структуре ООП бакалавриата

Дисциплина «Автоматизация поиска в интернете» включена в вариативную часть блока «Факультативы» ООП бакалавриата. На ее изучение отводится 72 часа (38 часов аудиторной работы, 34 часа самостоятельной работы). Согласно учебному плану бакалавриата данный курс в пятом семестре заканчивается зачетом.

Изучение дисциплины «Автоматизация поиска в интернете» основывается на базе знаний, полученных студентами на первом курсе в ходе освоения дисциплины «Информатика и программирование», а также дисциплин «Математика», «Дискретная математика» и «Теория вероятностей и математическая статистика».

## 3. Результаты обучения по дисциплине

Код и наименование компетенции	Код и наименование индикатора (индикаторов) достижения компетенции	Результаты обучения
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	1.1_Б.УК-1. Анализирует задачу, выделяя ее базовые составляющие. Осуществляет декомпозицию задачи.	<b>Знать:</b> – сущность поисковых систем, принципы индексации и ранжирования <b>Уметь:</b> – составлять и обрабатывать простейшие булевы запросы
	2.1_Б.УК-1. Находит и критически анализирует информацию, необходимую для решения поставленной задачи.	<b>Знать:</b> – понятие инвертированного (обратного) индекса и принцип его построения, основы нечеткого поиска, взвешенное зонное ранжирования, понятие обратной частоты документа, модель векторного пространства для ранжирования <b>Уметь:</b> – создавать инвертированный индекс
	3.1_Б.УК-1. Рассматривает различные варианты решения задачи, оценивая их достоинства и недостатки.	<b>Знать:</b> – различные варианты индексирования (пересечение инвертированных списков с

		<p>помощью указателей пропусков, двухсловные индексы, координатные индексы, параметрические и зонные индексы)</p> <p><b>Уметь:</b></p> <ul style="list-style-type: none"> <li>– оценивать достоинства и недостатки различных вариантов индексирования</li> </ul> <p><b>Владеть:</b></p> <ul style="list-style-type: none"> <li>– навыками решения задачи индексирования различными способами</li> </ul>
	<p><b>4.1_Б.УК-1.</b> Грамотно, логично, аргументированно формирует собственные суждения и оценки. Отличает факты от мнений, интерпретаций, оценок и т.д. в рассуждениях других участников деятельности.</p>	<p><b>Уметь:</b></p> <ul style="list-style-type: none"> <li>– грамотно, логично, аргументированно формулировать собственные суждения и оценки, связанные с информационным поиском</li> </ul> <p><b>Владеть:</b></p> <ul style="list-style-type: none"> <li>– способностью отличать факты от мнений, интерпретаций, оценок и т.д. в рассуждениях других участников деятельности</li> </ul>
	<p><b>5.1_Б.УК-1.</b> Определяет и оценивает практические последствия возможных решений задачи.</p>	<p><b>Знать:</b></p> <ul style="list-style-type: none"> <li>– практические последствия возможных решений задачи индексирования документов</li> </ul> <p><b>Уметь:</b></p> <ul style="list-style-type: none"> <li>– определять и оценивать практические последствия возможных решений задачи индексирования</li> </ul> <p><b>Владеть:</b></p> <ul style="list-style-type: none"> <li>– навыками оценивания практических последствий возможных решений задачи индексирования</li> </ul>
<p><b>УК-2</b> Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений</p>	<p><b>1.1_Б.УК-2.</b> Формулирует в рамках поставленной цели проекта совокупность взаимосвязанных задач, обеспечивающих ее достижение. Определяет ожидаемые результаты решения выделенных задач.</p>	<p><b>Знать:</b></p> <ul style="list-style-type: none"> <li>– совокупность взаимосвязанных задач с индексированием документов</li> </ul> <p><b>Уметь:</b></p> <ul style="list-style-type: none"> <li>– формулировать в рамках индексации документа совокупность взаимосвязанных задач: схематизации документа, разделение текста на лексемы, нормализации термов, стемминга и лемматизации</li> </ul> <p><b>Владеть:</b></p> <ul style="list-style-type: none"> <li>– способностью определять ожидаемые результаты решения</li> </ul>

		выделенных задач
	<p><b>2.1_Б.УК-2.</b> Проектирует решение конкретной задачи проекта, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений.</p>	<p><b>Знать:</b> – список характеристик аппаратного обеспечения, влияющих на архитектуру систем информационного поиска</p> <p><b>Уметь:</b> – проектировать построение индекса, выбирая оптимальный способ, исходя из имеющихся ресурсов и ограничений</p>
	<p><b>3.1_Б.УК-2.</b> Решает конкретные задачи проекта заявленного качества и за установленное время</p>	<p><b>Уметь:</b> – решать задачу построения индекса за установленное время</p>
	<p><b>4.1_Б.УК-2.</b> Публично представляет результаты решения конкретной задачи проекта.</p>	<p><b>Владеть:</b> – способностью публично представлять результаты решения задачи построения индекса</p>
<p><b>УК-4</b> Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном (-ых) языке (ах)</p>	<p><b>1.1_Б.УК-4.</b> Выбирает на государственном и иностранном (-ых) языках коммуникативно приемлемые стиль делового общения, вербальные и невербальные средства взаимодействия с партнерами.</p>	<p><b>Знать:</b> – стиль делового общения, вербальные и невербальные средства взаимодействия с партнерами</p> <p><b>Уметь:</b> – выбирать на государственном и иностранном (-ых) языках коммуникативно приемлемые стиль делового общения, вербальные и невербальные средства взаимодействия с партнерами</p> <p><b>Владеть:</b> – коммуникативно приемлемыми стилем делового общения, вербальными и невербальными средствами взаимодействия с партнерами на государственном и иностранном (-ых) языках</p>
	<p><b>2.1_Б.УК-4.</b> Использует информационно-коммуникационные технологии при поиске необходимой информации в процессе решения стандартных коммуникативных задач на государственном и иностранном (-ых) языках.</p>	<p><b>Знать:</b> – информационно-коммуникационные технологии при поиске необходимой информации в процессе решения стандартных коммуникативных задач на государственном и иностранном (-ых) языках</p> <p><b>Уметь:</b> – использовать информационно-коммуникационные технологии при поиске необходимой информации в процессе решения стандартных</p>

		коммуникативных задач на государственном и иностранном (-ых) языках
	<b>3.1_Б.УК-4.</b> Ведет деловую переписку, учитывая особенности стилистики официальных и неофициальных писем, социокультурные различия в формате корреспонденции на государственном и иностранном (-ых) языках.	<b>Владеть:</b> – способностью вести деловую переписку, учитывая особенности стилистики официальных и неофициальных писем, социокультурные различия в формате корреспонденции на государственном и иностранном (-ых) языках
	<b>4.1_Б.УК-4.</b> Умеет коммуникативно и культурно приемлемо вести устные деловые разговоры на государственном и иностранном (-ых) языках.	<b>Уметь:</b> – коммуникативно и культурно приемлемо вести устные деловые разговоры на государственном и иностранном (-ых) языках
	<b>5.1_Б.УК-4.</b> Демонстрирует умение выполнять перевод академических текстов с иностранного (-ых) языка (-ов) на государственный язык.	<b>Уметь:</b> – выполнять перевод академических текстов с иностранного (-ых) языка (-ов) на государственный язык <b>Владеть:</b> – навыками перевода академических текстов с иностранного (-ых) языка (-ов) на государственный язык
<b>УК-5</b> Способен воспринимать межкультурное разнообразие общества в социально-историческом, этическом и философском контекстах	<b>1.1_Б.УК-5.</b> Находит и использует необходимую для саморазвития и взаимодействия с другими информацию о культурных особенностях и традициях различных социальных групп.	<b>Знать:</b> – информацию о культурных особенностях и традициях различных социальных групп <b>Уметь:</b> – находить и использовать необходимую для саморазвития и взаимодействия с другими информацию о культурных особенностях и традициях различных социальных групп
	<b>2.1_Б.УК-5.</b> Демонстрирует уважительное отношение к историческому наследию и социокультурным традициям различных социальных групп, опирающееся на знание этапов исторического развития России (включая основные события, основных исторических деятелей) в контексте мировой истории и ряда культурных традиций мира (в зависимости от среды	<b>Уметь:</b> – демонстрировать уважительное отношение к историческому наследию и социокультурным традициям различных социальных групп, опирающееся на знание этапов исторического развития России (включая основные события, основных исторических деятелей) в контексте мировой истории и ряда культурных традиций мира (в зависимости от среды и задач образования), включая мировые

	и задач образования), включая мировые религии, философские и этические учения.	религии, философские и этические учения
	<b>3.1_Б.УК-5.</b> Умеет недискриминационно и конструктивно взаимодействовать с людьми с учетом их социокультурных особенностей в целях успешного выполнения профессиональных задач и усиления социальной интеграции.	<b>Уметь:</b> – недискриминационно и конструктивно взаимодействовать с людьми с учетом их социокультурных особенностей в целях успешного выполнения профессиональных задач и усиления социальной интеграции
<b>ПК-7</b> Способность настраивать, эксплуатировать и сопровождать информационные системы и сервисы	<b>1.1_Б.ПК-7.</b> Грамотно использует информацию о: возможностях ИС предметной области автоматизации; инструментах и методах проведения приемо-сдаточных испытаний (валидации) ИС; инструментах и методах выдачи и контроля поручений; устройстве и функционировании современных ИС; программных средствах и платформах инфраструктуры информационных технологий организаций; современных стандартах информационного взаимодействия систем; современных подходах и стандартах автоматизации организации (например, CRM, MRP, ERP, ITIL, ITSM); системах классификации и кодирования информации, в том числе присвоении кодов документам и элементам справочников; отраслевой нормативной технической документации; источниках информации, необходимой для профессиональной деятельности; современном отечественном и зарубежном опыте в профессиональной деятельности; основах управления торговлей, поставками и запасами; основах организации	<b>Знать:</b> – информацию о: возможностях поисковой системы; устройстве и функционировании поиска в интернете; классификации и кластеризации текстов; источниках информации, необходимой для поиска в интернете; современном отечественном и зарубежном опыте в создании и сжатии индекса <b>Уметь:</b> – грамотно использовать информацию о: возможностях поисковой системы; устройстве и функционировании поиска в интернете; классификации и кластеризации текстов; источниках информации, необходимой для поиска в интернете; современном отечественном и зарубежном опыте в создании и сжатии индекса <b>Владеть:</b> – навыками использования информации о: возможностях поисковой системы; устройстве и функционировании поиска в интернете; классификации и кластеризации текстов; источниках информации, необходимой для поиска в интернете; современном отечественном и зарубежном опыте в создании и сжатии индекса

	<p>производства; основах управления взаимоотношениями с клиентами и заказчиками (CRM); основах теории управления; методологии ведения документооборота в организациях; инструментах и методах определения финансовых и производственных показателей деятельности организациями; управлении качеством: контрольных списках, верификации, валидации (приемо-сдаточных испытаниях); культуре речи; правилах деловой переписки.</p>	
	<p><b>2.1_Б.ПК-7.</b> Планирует работы; распределяет работы и выделяет ресурсы; контролирует исполнение поручений.</p>	<p><b>Уметь:</b> – планировать работы; распределять работы и выделять ресурсы; контролировать исполнение поручений</p>
	<p><b>3.1_Б.ПК-7.</b> Организует проведение приемо-сдаточных испытаний ИС; организует подписание документов по результатам приемо-сдаточных испытаний.</p>	<p><b>Знать:</b> – приемо-сдаточные испытания систем информационного поиска <b>Уметь:</b> – организовать проведение приемо-сдаточных испытаний систем информационного поиска <b>Владеть:</b> – навыками организации подписания документов по результатам приемо-сдаточных испытаний</p>
<p><b>ПК-10</b> Способность принимать участие в организации ИТ-инфраструктуры и управлении информационной безопасностью.</p>	<p><b>1.1_Б.ПК-10.</b> Грамотно использует информацию о: методах и средства сбора модулей и компонентов программного обеспечения; методах и программных интерфейсах взаимодействия с внешними программными компонентами; методах проектирования и разработки программных интерфейсов взаимодействия внутренних модулей системы; методах и средствах разработки процедур для развертывания программного обеспечения;</p>	<p><b>Знать:</b> – информацию о: распределенном индексировании; о методах и средства сбора модулей и компонентов программного обеспечения; методах и программных интерфейсах взаимодействия с внешними программными компонентами; методах проектирования и разработки программных интерфейсов взаимодействия внутренних модулей системы; методах и средствах разработки процедур для развертывания программного обеспечения; методах</p>

	<p>методах и средствах миграции и преобразования данных; методах проверки работоспособности программного продукта; интерфейсах взаимодействия с внешней средой; интерфейсах взаимодействия внутренних модулей системы; языках, утилитах и средах программирования, средствах пакетного выполнения процедур; методах принятия управленческих решений; основных принципах и методах управления персоналом.</p>	<p>и средствах миграции и преобразования данных; интерфейсах взаимодействия с внешней средой; интерфейсах взаимодействия внутренних модулей системы; языках, утилитах и средах программирования, средствах пакетного выполнения процедур информационного поиска.</p> <p><b>Уметь:</b> – грамотно использовать описанную информацию</p> <p><b>Владеть:</b> – описанными методами</p>
	<p><b>2.1_Б.ПК-10.</b> Может: писать программный код процедур интеграции программных модулей; использовать выбранную среду программирования для разработки процедур интеграции программных модулей; применять методы и средства сборки модулей и компонентов программного обеспечения, разработки процедур для развертывания программного обеспечения, миграции и преобразования данных, создания программных интерфейсов.</p>	<p><b>Уметь:</b> – писать программный код процедур интеграции программных модулей; использовать выбранную среду программирования для разработки процедур интеграции программных модулей</p> <p><b>Владеть:</b> – методами и средствами сборки модулей и компонентов программного обеспечения, разработки процедур для развертывания программного обеспечения, миграции и преобразования данных, создания программных</p>
	<p><b>3.1_Б.ПК-10.</b> Применяет коллективную среду разработки программного обеспечения и систему контроля версий; выявляет соответствие требований заказчиков существующим продуктам; оценивает работоспособность программного продукта; применяет методы принятия управленческих решений.</p>	<p><b>Уметь:</b> – применять коллективную среду разработки программного обеспечения и систему контроля версий</p> <p><b>Владеть:</b> – способностью оценивает работоспособность программного продукта</p>
	<p><b>4.1_Б.ПК-10.</b> Формулирует задания на разработку процедур интеграции, сборку, подключение к внешней среде, проверку работоспособности выпусков программного</p>	<p><b>Уметь:</b> – формулировать задания на разработку процедур интеграции, сборку, подключение к внешней среде, проверку работоспособности выпусков программного продукта</p>

	<p>продукта; оценивает результаты выполнения назначенных заданий на разработку процедур интеграции, сборку, подключение к внешней среде, проверку работоспособности выпусков программного продукта; принимает управленческие решения по результатам проверки работоспособности выпусков программного продукта (решения о выпуске/невыходе версии, отправке задач на доработку, добавление новых задач, передачу на тестирование).</p>	<p><b>Владеть:</b> – способностью оценивать результаты выполнения назначенных заданий на разработку процедур интеграции, сборку, подключение к внешней среде, проверку работоспособности выпусков программного продукта</p>
--	---	---

#### 4. Структура и содержание дисциплины (модуля)

Общая трудоемкость дисциплины составляет 2 зачетные единицы (72 часа).

№ п/п	Раздел дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)					Формы текущего контроля успеваемости (по неделям семестра) Формы промежуточной аттестации (по семестрам)
				Лекции	Практические занятия	КСР	СР	Контроль	
1	Введение. Основные понятия. Булев поиск. Индекс.	5	1,2	2	2		4		Выполнение практических заданий на ПК, консультации
2	Схематизация документа	5	3,4	2	2		2		Выполнение практических заданий на ПК, консультации
3	Разделение текста на лексемы. «Стоп-слова»	5	5,6	1	1		2		Выполнение практических заданий на ПК, консультации

4	Нормализация, стемминг и лемматизация	5	5,6	1	1		2	Выполнение практических заданий на ПК, консультации
5	Указатели пропусков	5	7,8	1	1		2	Выполнение практических заданий на ПК, консультации
6	Двухсловные и координатные индексы.	5	7,8	1			2	Выполнение практических заданий на ПК, консультации
7	Нечеткий поиск	5	7,8		1		2	Выполнение практических заданий на ПК, консультации
8	Построение и сжатие индекса	5	9,10	2	2		4	Выполнение практических заданий на ПК, консультации
9	Ранжирование	5	11,12	2	2		2	Выполнение практических заданий на ПК, консультации
10	Частота термина и взвешивание	5	13,14	2	2		4	Выполнение практических заданий на ПК, консультации
11	Модель векторного пространства для ранжирования	5	15,16	2	2		4	Выполнение практических заданий на ПК, консультации
12	XML-поиск	5	17,18	2			2	Опрос
13	Оценка информационного поиска	5	17,18		2	2	2	Консультации, опрос
	<b>Промежуточная аттестация</b>	<b>5</b>					<b>2</b>	<b>Зачет</b>
	<b>Итого за 5 семестр</b>			<b>18</b>	<b>18</b>	<b>2</b>	<b>34</b>	<b>72 ч</b>
	<b>Общая трудоемкость дисциплины</b>			<b>18</b>	<b>18</b>	<b>2</b>	<b>34</b>	<b>72 ч</b>

## Содержание дисциплины

### 1. Булев Поиск

Информационный поиск. Создание инвертированного индекса. Сравнение расширенной булевой модели и ранжированного поиска. Обработка булевых запросов.

### 2. Лексикон и списки словопозиций

Схематизация документа и декодирование последовательности символов. Выделение последовательности символов в документе. Выбор структурной

единицы документа. Определение лексикона терминов. Разделение текста на лексемы. Игнорирование распространенных терминов: стоп-слова. Нормализация (классификация терминов по классам эквивалентности). Стемминг и лемматизация. Быстрое пересечение инвертированных списков с помощью указателей пропусков. Словопозиции с координатами и фразовые запросы. Двухсловные индексы. Координатные индексы. Комбинированные схемы.

### **3. Нечеткий поиск**

Поисковые структуры для словарей. Запросы с джокером. Исправление опечаток. Фонетические исправления.

### **4. Построение индекса**

Основы аппаратного обеспечения. Блочное индексирование, основанное на сортировке. Однопроходное индексирование в оперативной памяти. Распределенное индексирование.

### **5. Сжатие индекса**

Статистические характеристики терминов в информационном поиске. Сжатие словаря. Сжатие инвертированного файла.

### **6. Ранжирование, взвешивание терминов и модель векторного пространства**

Параметрические и зонные индексы. Взвешенное зонное ранжирование. Определение весов на основе машинного обучения. Частота термина и взвешивание. Обратная документная частота. Взвешивание на основе комбинации частоты и обратной документной частоты термина. Модель векторного пространства для ранжирования. Скалярное произведение. Запросы как векторы. Ранжирование в векторной модели.

### **7. XML-поиск**

Основные концепции языка XML. Проблемы, связанные с XML-поиском. Оценка XML-поиска.

### **8. Оценка информационного поиска.**

Понятие сниппета, его структура. Метрики сниппетов. Оценка ассессорами. Метрики качества поисковой системы. Качество поиска. Точность/полнота информационного поиска. Критика чистой релевантности. Маркерные тесты. Поиск периферийных сайтов. Региональная навигация. Тематический поиск. Общее качество поиска. Ассессорская служба. Оценка релевантности документа. Кросс-валидация. Автопоиск ошибок. Онлайн-метрики. Оценка гипотез. Кликовые метрики. Корреляция с ассессорами.

## **5. Образовательные технологии**

Для реализации компетентностного подхода в учебном процессе применяются следующие образовательные технологии:

1) при проведении лекционных занятий: информационные лекции, проблемные лекции, лекции беседы, лекции дискуссии, лекции с заранее запланированными ошибками;

2) при проведении практических занятий: традиционные занятия, занятия исследования, проблемные ситуации, ситуации с ошибкой;

3) при организации самостоятельной работы студентов: поиск и обработка информации, в том числе с использованием информационно-телекоммуникационных технологий; исследование проблемной ситуации; постановка и решение задач из предметной области; отработка навыков применения стандартных методов к решению задач предметной области.

Успешное освоение материала курса предполагает большую самостоятельную работу студентов и руководство этой работой со стороны преподавателей. Применяются следующие формы контроля: устный опрос, проверка решения практических задач, контрольная работа.

При обучении лиц с ограниченными возможностями здоровья и инвалидов используются подходы, способствующие созданию безбарьерной образовательной среды: технологии дифференциации и индивидуального обучения, применение соответствующих методик по работе с инвалидами, использование средств дистанционного общения, проведение дополнительных индивидуальных консультаций по изучаемым теоретическим вопросам и практическим занятиям, оказание помощи при подготовке к промежуточной и итоговой аттестации. Подготовка, при необходимости, учебных и контрольно-измерительных материалов в формах, доступных для изучения студентами с особыми образовательными потребностями (для студентов с нарушениями зрения учебные материалы подготавливаются с применением укрупненного шрифта, используются аудиозаписи занятий; для студентов с нарушением слуха предоставляются электронные лекции, печатные раздаточные материалы с заданиями для самостоятельной работы).

При необходимости, для подготовки к ответу на практическом занятии, студентам с инвалидностью и студентам с ограниченными возможностями здоровья среднее время увеличивается в 1,5–2 раза по сравнению со средним временем подготовки обычного студента.

Для студентов с инвалидностью или с ограниченными возможностями здоровья форма промежуточной аттестации устанавливается с учетом индивидуальных психофизических особенностей (устно, письменно на бумаге, письменно на компьютере, в форме тестирования и т.п.). Промежуточная аттестация по дисциплине может проводиться в несколько этапов в форме рубежного контроля по завершению изучения отдельных тем дисциплины.

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

**Самостоятельная внеаудиторная работа** студентов проводится в форме изучения и анализа лекционного материала, изучения отдельных теоретических вопросов по предлагаемой литературе, подбора дополнительных источников для извлечения научно-технической информации, связанной с проблемами, изучаемыми в рамках данной дисциплины и решения задач с дальнейшим их разбором или обсуждением на аудиторных занятиях, подготовки к промежуточной аттестации.

**Самостоятельная аудиторная работа** студентов проводится в форме самостоятельного выполнения заданий на практических занятиях с дальнейшим их разбором и обсуждением; проведения контрольной работы; поиска решений проблемных ситуаций, предложенных на лекциях и практических занятиях; поиска и устранения ошибок, заложенных в представлении материала преподавателем и допущенных другими студентами.

**Текущий контроль** усвоения дисциплины «Автоматизация поиска в интернете» проводится в форме устных опросов на лекционных и практических занятиях, разбора и обсуждения выполняемых заданий на практических занятиях, контрольных работ. Примерные варианты контрольных работ содержатся в фонде оценочных средств текущего контроля и промежуточной аттестации по дисциплине.

**Промежуточная аттестация** по дисциплине «Автоматизация поиска в интернете» проводится в форме зачета. Контрольные вопросы готовятся к каждому разделу.

*Примерные вопросы к зачёту:*

1. Построение инвертированного индекса
2. Обработка булевых запросов
3. Сравнение расширенной булевой модели и ранжированного поиска
4. Выделение последовательности символов в документе
5. Выбор структурной единицы документа
6. Разделение текста на лексемы
7. Игнорирование распространенных терминов: стоп-слова
8. Нормализация (классификация терминов по классам эквивалентности)
9. Ударения и диакритические символы. Использование заглавных букв и обработка без учета регистра
10. Стемминг и лемматизация
11. Быстрое пересечение инвертированных списков с помощью указателей пропусков
12. Двухсловные индексы
13. Координатные индексы
14. Комбинированные схемы
15. Основы аппаратного обеспечения
16. Блочное индексирование, основанное на сортировке
17. Однопроходное индексирование в оперативной памяти
18. Распределенное индексирование
19. Параметрические и зонные индексы
20. Взвешенное зонное ранжирование
21. Определение весов на основе машинного обучения
22. Обратная документная частота
23. Взвешивание на основе комбинации частоты и обратной документной частоты термина
24. Скалярное произведение
25. Запросы как векторы
26. Ранжирование в векторной модели

27. Основные концепции языка XML
28. Проблемы, связанные с XML поиском
29. Оценка XML-поиска
30. Понятие сниппета, его структура. Метрики сниппетов
31. Качество информационного поиска, его точность и полнота

### 7. Данные для учета успеваемости студентов в БАРС

Таблица максимальных баллов по видам учебной деятельности.

Семестр	Лекции	Лабораторные занятия	Практические занятия	Самостоятельная работа	Автоматизированное тестирование	Другие виды учебной деятельности	Промежуточная аттестация	Итого
5	18	0	18	44	0	0	20	100

#### 5 семестр

##### Лекции

Посещаемость, опрос, активность и др. от 0 до 18 баллов.

##### Лабораторные занятия

Не предусмотрены.

##### Практические занятия

Оценивается самостоятельность при выполнении работы, активность работы в аудитории, правильность выполнения индивидуальных заданий. От 0 до 18 баллов.

##### Самостоятельная работа

Контроль качества и количества выполненных домашних заданий – от 0 до 20 баллов, правильность выполнения – от 0 до 20 баллов, своевременность сдачи выполненных домашних заданий – от 0 до 4 баллов.

##### Автоматизированное тестирование

Не предусмотрено.

##### Другие виды учебной деятельности

Не предусмотрено.

##### Промежуточная аттестация

61 балл и более	«зачтено»
меньше 61 баллов	«не зачтено»

Таким образом, максимально возможная сумма баллов за все виды учебной деятельности студента за 5 семестр по дисциплине «Автоматизация поиска в интернете» составляет 100 баллов.

## **8. Учебно-методическое и информационное обеспечение дисциплины**

### ***а) литература:***

1. Кутовенко А. Профессиональный поиск в Интернете [Текст] / А. Кутовенко. - Москва ; Санкт-Петербург [и др.] : Питер, 2011. – 252 с.
2. Лукашевич Н.В. Тезаурусы в задачах информационного поиска [Текст] / Н.В. Лукашевич. - Москва : Издательство Московского университета, 2011. - 508, [4] с. : ил. - Библиогр.: с. 483-508.



### ***б) программное обеспечение и Интернет-ресурсы:***

#### *Лицензионное программное обеспечение:*

Операционная система Windows 7, или более поздняя версия  
Microsoft Office PowerPoint

#### *Интернет-ресурсы:*

1. Саратовской государственной университет им. Н.Г. Чернышевского. – Режим доступа: [www.sgu.ru/](http://www.sgu.ru/)
2. Зональная научная библиотека им. В.А. Артисевич Саратовского государственного университета им. Н.Г. Чернышевского. – Режим доступа: <http://library.sgu.ru/>
3. Каталог образовательных Интернет-ресурсов. – Режим доступа: <http://window.edu.ru/>

## 9. Материально-техническое обеспечение дисциплины

Материально-техническое обеспечение дисциплины «Автоматизация поиска в интернете» составляют:

- Учебная аудитория с обязательным наличием специализированной доски, мела (маркера), проектора, с возможностью размещения всех обучающихся по данной дисциплине.
- Сайт поддержки учебного процесса NTO.IMMPU.SGU.RU, позволяющий гибко формировать индивидуальную образовательную траекторию обучающихся.
- дисплейные классы (аудитории 111, 307, 308, 309, 310, 312 учебного корпуса 9), оборудованных компьютерами: по 10 компьютеров в каждом дисплейном класс, с источниками бесперебойного питания; компьютеры дисплейных классов объединены в единую локальную сеть с доступом к локальным информационным образовательным и рабочим ресурсам СГУ и к сети Интернет. Компьютеры дисплейных классов оборудованы видеокартами с поддержкой технологии CUDA для реализации задач по параллельному многопоточному программированию.
- Программное обеспечение: Gentoo Linux, Kate, Eclipse, Python.

Программа составлена в соответствии с требованиями ФГОС ВО с учетом рекомендаций по направлению 09.03.03 – Прикладная информатика и профилю подготовки «Прикладная информатика в экономике».

Автор: к.ф.-м.н., доцент кафедры компьютерной алгебры и теории чисел  
Коробченко Е.В.

Программа одобрена на заседании кафедры компьютерной алгебры и теории чисел от 12 ноября 2021 года, протокол № 4.