



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
федеральное государственное бюджетное образовательное учреждение
высшего образования
«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ИГУ»)
Институт математики и информационных технологий

«УТВЕРЖДАЮ»
Директор ИМИТ ИГУ
М. В. Фалалеев
М. В. Фалалеев
«19» мая 2021 г.

Рабочая программа дисциплины (модуля)

Б1.О.28 Интернет-программирование

Направление подготовки	01.03.02 Прикладная математика и информатика
Направленность (профиль) подготовки	Системная и бизнес-аналитика
Квалификация выпускника	бакалавр
Форма обучения	очная

Иркутск 2021 г.

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цели: ознакомление студентов с приемами решения практических задач в сети Интернет с помощью алгоритмов поиска и индексации данных, формирование практических умений и навыков, необходимых для формирования ключевых профильных компетенций.

Задачи: обучение студентов алгоритмам информационного поиска, методам классификации в информационном поиске и знакомство с проблемами поиска в web-e.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Учебная дисциплина Б1.О.28 Интернет-программирование относится к обязательной части Блока 1 образовательной программы.

3. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс освоения дисциплины направлен на формирование следующих компетенций в соответствии с ФГОС ВО и ОП ВО по направлению подготовки 01.03.02 Прикладная математика и информатика:

ОПК-2 Способен использовать и адаптировать существующие математические методы и системы программирования для разработки и реализации алгоритмов решения прикладных задач;

ОПК-4 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности;

ОПК-5 Способен разрабатывать алгоритмы и компьютерные программы, пригодные для практического применения.

4. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Объем дисциплины составляет 4 зачетных ед., 144 час.

Форма промежуточной аттестации: экзамен.

4.1. Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов

Раздел дисциплины / тема	Виды учебной работы			Самост. работа	Формы текущего контроля; Формы промежут. аттестации
	Контактная работа преподавателя с обучающимися				
	Лекции	Лаб. занятия	Практ. занятия		
РАЗДЕЛ 1. ОСНОВЫ ИНФОРМАЦИОННОГО ПОИСКА		24		6	
РАЗДЕЛ 2. КЛАССИФИКАЦИЯ В ИНФОРМАЦИОННОМ ПОИСКЕ		24		6	
РАЗДЕЛ 3. ПРОБЛЕМЫ ПОИСКА В ВЕБЕ		24		6	
Итого (2 семестр):		72		18	экс.

4.2. Содержание учебного материала

РАЗДЕЛ 1. ОСНОВЫ ИНФОРМАЦИОННОГО ПОИСКА

ТЕМА 1.1. Введение

Проблемы развития Интернета. Структурированные и неструктурированные данные. Термин информационный поиск. Запрос. Объект запроса. Парадигмы поиска. Этапы информационного поиска. Виды поиска.

ТЕМА 1.2. Булева модель информационного поиска

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

ТЕМА 1.3. Лексикон и списки словопозиций

Схематизация документа и декодирование последовательности символов.

Определение

лексикона терминов. Быстрое пересечение инвертированных списков с помощью указателей пропусков. Словопозиции с координатами и фразовые запросы.

ТЕМА 1.4. Словари и нечеткий поиск

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

ТЕМА 1.5. Сжатие индекса

Статистические характеристики термов в информационном поиске. Сжатие словаря.

Сжатие инвертированного файла.

ТЕМА 1.6. Ранжирование, взвешивание термов и модель векторного пространства

Параметрические и зонные индексы. Частота терма и взвешивание. Модель векторного

пространства для ранжирования. Варианты функции tf-idf.

ТЕМА 1.7. Обратная связь по релевантности и расширение запроса

Обратная связь по релевантности и псевдорелевантности. Глобальные методы для переформулировки запроса.

РАЗДЕЛ 2. КЛАССИФИКАЦИЯ В ИНФОРМАЦИОННОМ ПОИСКЕ

ТЕМА 2.1. Классификация текстов и наивный байесовский подход

Классификация текстов. Наивная байесовская классификация текстов. Модель Бернулли. Свойства наивной байесовской модели. Выбор признаков. Оценка классификации текстов.

ТЕМА 2.2. Классификация в векторном пространстве

Представление документов и меры близости в векторном пространстве. Метод Роккио. Метод k-ближайших соседей. Линейные и нелинейные классификаторы. Классификация с несколькими классами.

РАЗДЕЛ 3. ПРОБЛЕМЫ ПОИСКА В ВЕБЕ

ТЕМА 3.1. Задачи информационного поиска

Введение в Data Mining. Вопросно-ответные системы. Рекомендательные системы и совместная фильтрация. Поиск в полуструктурированных данных. Индексация изображений. Принципы и технические методы работы со спамом. Способы определения географической привязки IP адресов. Поисковая оптимизация и спамдексирование.

Индексация звука (Shazam). Система полнотекстового поиска Sphinx.

ТЕМА 3.2. Основы поиска в вебе

Основы и история. Характеристики веба. Реклама как экономическая модель. Опыт пользователей поисковых систем. Размер индекса и оценка его размера. Нечеткие дубликаты и алгоритм шинглов.

ТЕМА 3.3. Анализ ссылок

Веб как граф. Метод PageRank. Порталы и авторитетные источники.

4.3. Методические указания по организации самостоятельной работы студентов

Самостоятельная работа студентов всех форм и видов обучения является одним из обязательных видов образовательной деятельности, обеспечивающей реализацию требований Федеральных государственных стандартов высшего образования. Согласно требованиям нормативных документов самостоятельная работа студентов является обязательным компонентом образовательного процесса, так как она обеспечивает закрепление получаемых на лекционных занятиях знаний путем приобретения навыков осмысления и расширения их содержания, навыков решения актуальных проблем формирования общекультурных и профессиональных компетенций, научно-исследовательской деятельности, подготовки к семинарам, лабораторным работам, сдаче зачетов и экзаменов. Самостоятельная работа студентов представляет собой совокупность аудиторных и внеаудиторных занятий и работ. Самостоятельная работа в рамках образовательного процесса в вузе решает следующие задачи:

- закрепление и расширение знаний, умений, полученных студентами во время аудиторных и внеаудиторных занятий, превращение их в стереотипы умственной и физической деятельности;
- приобретение дополнительных знаний и навыков по дисциплинам учебного плана;
- формирование и развитие знаний и навыков, связанных с научно-исследовательской деятельностью;
- развитие ориентации и установки на качественное освоение образовательной программы;
- развитие навыков самоорганизации;
- формирование самостоятельности мышления, способности к саморазвитию, самосовершенствованию и самореализации;

– выработка навыков эффективной самостоятельной профессиональной теоретической, практической и учебно-исследовательской деятельности.

Подготовка к лекции. Качество освоения содержания конкретной дисциплины прямо зависит от того, насколько студент сам, без внешнего принуждения формирует у себя установку на получение на лекциях новых знаний, дополняющих уже имеющиеся по данной дисциплине. Время на подготовку студентов к двухчасовой лекции по нормативам составляет не менее 0,2 часа.

Подготовка к практическому занятию. Подготовка к практическому занятию включает следующие элементы самостоятельной деятельности: четкое представление цели и задач его проведения; выделение навыков умственной, аналитической, научной деятельности, которые станут результатом предстоящей работы. Выработка навыков осуществляется с помощью получения новой информации об изучаемых процессах и с помощью знания о том, в какой степени в данное время студент владеет методами исследовательской деятельности, которыми он станет пользоваться на практическом занятии. Подготовка к практическому занятию нередко требует подбора материала, данных и специальных источников, с которыми предстоит учебная работа. Студенты должны дома подготовить к занятию 3–4 примера формулировки темы исследования, представленного в монографиях, научных статьях, отчетах. Затем они самостоятельно осуществляют поиск соответствующих источников, определяют актуальность конкретного исследования процессов и явлений, выделяют основные способы доказательства авторами научных работ ценности того, чем они занимаются. В ходе самого практического занятия студенты сначала представляют найденные ими варианты формулировки актуальности исследования, обсуждают их и обосновывают свое мнение о наилучшем варианте. Время на подготовку к практическому занятию по нормативам составляет не менее 0,2 часа.

Подготовка к семинарскому занятию. Самостоятельная подготовка к семинару направлена: на развитие способности к чтению научной и иной литературы; на поиск дополнительной информации, позволяющей глубже разобраться в некоторых вопросах; на выделение при работе с разными источниками необходимой информации, которая требуется для полного ответа на вопросы плана семинарского занятия; на выработку умения правильно выписывать высказывания авторов из имеющихся источников информации, оформлять их по библиографическим нормам; на развитие умения осуществлять анализ выбранных источников информации; на подготовку собственного выступления по обсуждаемым вопросам; на формирование навыка оперативного реагирования на разные мнения, которые могут возникать при обсуждении тех или иных научных проблем. Время на подготовку к семинару по нормативам составляет не менее 0,2 часа.

Подготовка к коллоквиуму. Коллоквиум представляет собой коллективное обсуждение раздела дисциплины на основе самостоятельного изучения этого раздела студентами. Подготовка к данному виду учебных занятий осуществляется в следующем порядке. Преподаватель дает список вопросов, ответы на которые следует получить при изучении определенного перечня научных источников. Студентам во внеаудиторное время необходимо прочитать специальную литературу, выписать из нее ответы на вопросы, которые будут обсуждаться на коллоквиуме, мысленно сформулировать свое мнение по каждому из вопросов, которое они выскажут на занятии. Время на подготовку к коллоквиуму по нормативам составляет не менее 0,2 часа.

Подготовка к контрольной работе. Контрольная работа назначается после изучения определенного раздела (разделов) дисциплины и представляет собой совокупность развернутых письменных ответов студентов на вопросы, которые они заранее получают от преподавателя. Самостоятельная подготовка к контрольной работе включает в себя: — изучение конспектов лекций, раскрывающих материал, знание которого проверяется контрольной работой; повторение учебного материала, полученного

при подготовке к семинарским, практическим занятиям и во время их проведения; изучение дополнительной литературы, в которой конкретизируется содержание проверяемых знаний; составление в мысленной форме ответов на поставленные в контрольной работе вопросы; формирование психологической установки на успешное выполнение всех заданий. Время на подготовку к контрольной работе по нормативам составляет 2 часа.

Подготовка к зачету. Самостоятельная подготовка к зачету должна осуществляться в течение всего семестра. Подготовка включает следующие действия: перечитать все лекции, а также материалы, которые готовились к семинарским и практическим занятиям в течение семестра, соотнести эту информацию с вопросами, которые даны к зачету, если информации недостаточно, ответы находят в предложенной преподавателем литературе. Рекомендуется делать краткие записи. Время на подготовку к зачету по нормативам составляет не менее 4 часов.

Подготовка к экзамену. Самостоятельная подготовка к экзамену схожа с подготовкой к зачету, особенно если он дифференцированный. Но объем учебного материала, который нужно восстановить в памяти к экзамену, вновь осмыслить и понять, значительно больше, поэтому требуется больше времени и умственных усилий. Важно сформировать целостное представление о содержании ответа на каждый вопрос, что предполагает знание разных научных трактовок сущности того или иного явления, процесса, умение раскрывать факторы, определяющие их противоречивость, знание имен ученых, изучавших обсуждаемую проблему. Необходимо также привести информацию о материалах эмпирических исследований, что указывает на всестороннюю подготовку студента к экзамену. Время на подготовку к экзамену по нормативам составляет 36 часов для бакалавров.

В ФБГОУ ВО «ИГУ» организация самостоятельной работы студентов регламентируется Положением о самостоятельной работе студентов, принятым Ученым советом ИГУ 22 июня 2012 г.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Литература, базы данных, информационно-справочные и поисковые системы

1. Современные средства хранения и обработки информации [Текст] : учеб. пособие / В.И. Красов ; рец.: В. И. Сажин, А. Г. Ченский ; Иркутский гос. ун-т. - Иркутск : Изд-во ИГУ, 2013. - 115 с. ISBN 978-5-9624-0782-1 (31 экз.).
2. Советов Б. Я. Представление знаний в информационных системах : учебник для студ. учреждений высш. проф. образования / Советов Б. Я., Цехановский В. В., Чертовской В. Д. - Издат. центр "Академия, 2012. - 144 с. ISBN: 978-5-7685-92281-2 . - (ЭБС "БиблиоТех").

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебная аудитория для проведения:

- занятий лекционного типа,
- занятий семинарского (практического) типа,
- групповых и индивидуальных консультаций,
- текущего контроля и промежуточной аттестации.

Оснащение:

Учебная аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории, для проведения занятий лекционного типа, практических занятий (семинарского типа), курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Для проведения занятий лекционного типа обучающимся предлагаются наборы демонстрационного оборудования и учебно-наглядные пособия, обеспечивающие тематические иллюстрации.

6.2. Помещения для самостоятельной работы обучающихся.

Оснащение:

Помещения для самостоятельной работы обучающихся, оснащенные учебной мебелью. Рабочие места обучающихся оборудованы компьютерной техникой и подключены в локальную вычислительную сеть, в т.ч. с использованием беспроводного Wi-Fi подключения, с возможностью выхода в глобальную сеть Интернет и с доступом в электронную информационно-образовательную среду.

6.3. Программное обеспечение

Приложение для чтения PDF-файлов, браузер для просмотра интернет контента, приложение для создания PDF-файлов.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

7.1. Оценочные средства для промежуточной аттестации

Список вопросов для промежуточной аттестации:

1. Проблемы развития Интернета. Структурированные и неструктурированные данные. Термин информационный поиск. Запрос. Объект запроса.
2. Парадигмы поиска. Этапы информационного поиска. Виды поиска.
3. Булева модель информационного поиска. Создание инвертированного индекса. Обработка булевых запросов.
4. Булева модель информационного поиска. Сравнение расширенной булевой модели и ранжированного поиска.
5. Лексикон и списки словопозиций. Схематизация документа и декодирование последовательности символов. Определение лексикона терминов.
6. Быстрое пересечение инвертированных списков с помощью указателей пропусков.
7. Словопозиции с координатами и фразовые запросы.
8. Поискковые структуры для словарей. Фонетические исправления.
9. Запросы с джокером.
10. Исправление опечаток.
11. Статистические характеристики термов в информационном поиске. Сжатие словаря.
12. Сжатие инвертированного файла.
13. Параметрические и зонные индексы. Частота термина и взвешивание.
14. Модель векторного пространства для ранжирования.
15. Варианты функции tf-idf.
16. Обратная связь по релевантности и псевдорелевантности.
17. Глобальные методы для переформулировки запроса.
18. Классификация текстов. Наивная байесовская классификация текстов.
19. Модель Бернулли. Свойства наивной байесовской модели.
20. Выбор признаков. Оценка классификации текстов.
21. Представление документов и меры близости в векторном пространстве. Метод Роккио.

22. Метод k-ближайших соседей. Линейные и нелинейные классификаторы.
23. Классификация с несколькими классами.
24. Языковые модели информационного поиска.
25. Модель правдоподобного запроса.
26. Сравнение языкового моделирования с другими подходами к информационному поиску. Расширения языковых моделей.
27. Характеристики веба. Реклама как экономическая модель.
28. Опыт пользователей поисковых систем. Размер индекса и оценка его размера.
29. Нечеткие дубликаты и алгоритм шинглов.
30. Обход веба.
31. Распределение индексов. Серверы проверки ссылочной связности.
32. Веб как граф. Порталы и авторитетные источники.
33. Метод PageRank.