

**Московский государственный университет
имени М.В.Ломоносова**

**Введение в теорию распределенных
информационных систем**

Introduction to the Theory of Distributed Information Systems

Учебно-методический комплекс

Для специальности:

Направление

010400 «Прикладная математика и информатика»

Москва 2013

Утверждено на заседании кафедры **вычислительной математики механико-математического факультета МГУ**, __.__.2013. **Протокол № _.**

Составитель:

_____ кафедры **вычислительной математики механико-математического факультета МГУ**,
к.ф.-м.н. С. А. Афонин, к.ф.-м.н. А.С. Козицын

Учебно-методический комплекс «Введение в теорию распределенных информационных систем» составлен в соответствии с требованиями Государственного образовательного стандарта высшего профессионального образования по специальности 010400 «Прикладная математика и информатика».

Дисциплина входит в региональный компонент цикла общепрофессиональных дисциплин (ОПД.Р.03) и является обязательной для изучения.

СОДЕРЖАНИЕ

1. Цель, задачи и предмет дисциплины

Основная цель дисциплины:

- получение теоретических знаний о принципах построения и архитектуре современных информационных систем и сред (в том числе распределенных систем обработки текстовых данных), обеспечивающих организацию информационного взаимодействия в программных комплексах экономического, управленческого, производственного, научного и другого назначения, а также практических навыков по созданию таких систем для реализации бизнес-процессов в корпоративных сетях предприятий.

Задачи дисциплины:

- формирование и развитие знаний, практических навыков и умений, обеспечивающих разработку автоматизированных информационных систем;

- изучение базовых принципов, моделей и алгоритмов обработки данных.

2. Требования к уровню освоения дисциплины

Изучение дисциплины предполагает знание студентами основ алгебры, математического анализа, информатики и программирования.

В результате изучения дисциплины студенты должны:

Знать:

- основные классические алгоритмы обработки данных, используемые в современных информационных системах;

- основные понятия, используемые при построении современных информационных систем;

- основные принципы организации обработки данных;

- основные модели, закладываемые при создании систем обработки данных;

- основные алгоритмы построения распределенных приложений, включая алгоритмы взаимного исключения, достижения согласия, организации многоадресной передачи данных.

Уметь:

ставить и решать задачи по разработке распределенных информационных систем и по обеспечению отказоустойчивости систем данных.

иметь представление о принципах функционирования и взаимодействия систем в распределенной среде (в том числе, в среде Интернет).

3. Объем дисциплины

3.1. Объем дисциплины и виды учебной работы

Вид учебной работы	Количество часов по очной форме обучения
№№ семестров	1,2
Аудиторные занятия:	72
Лекции	48
практические и семинарские занятия	24
Самостоятельная работа	56
ВСЕГО ЧАСОВ НА ДИСЦИПЛИНУ	128
Текущий контроль (количество, №№ семестров)	№1 Контрольная работа №2 Контрольная работа
Виды итогового контроля (экзамен, зачет)	Зачет по окончании каждого семестра

3.2 Распределение часов по темам и видам учебной работы

Форма обучения – очная

Тема	Виды учебных занятий	Аудиторные занятия, из них			самостоятельная работа
		Лекции	практические занятия		
Всего часов по учебному плану					
1 Системы управления базами данных (Database Management Systems)	20	6	4		10

2	Методы поиска текстовой информации	20	8	4	8
3	Методы поиска слабоструктурированной информации	20	8	4	8
4	Основы сетевых протоколов	18	6	4	8
5	Распределенные системы	26	10	4	10
6	Экспертные и интеллектуальные системы	24	10	4	12
Итого:		128	48	24	56

4. Содержание дисциплины

Тема 1. Системы управления базами данных(Database Management Systems and Text Search Methods).

Лекции: Реляционные, сетевые и иерархические СУБД. Области применимости и основные понятия. Данные, целостность данных, методы обеспечения целостности Реляционная алгебра и реляционное исчисление. Нормальные формы. Методы ускорения поиска: индексы, кластеры, распараллеливание, хэширование, построение и оценка планов запросов. Феномены, уровни изолированности и блокировок, транзакции, механизмы реализации транзакций. Вспомогательные объекты баз данных. Основные модули СУБД на примере ORACLE.

Практические занятия: Проектирование логических и физических схем баз данных. Составление запросов. Приведение схем к требуемым нормальным формам.

Самостоятельная работа: Изучение рекомендованной литературы по теме Системы управления базами данных.

Тема 2. Методы поиска текстовой информации(Text Search Methods).

Лекции: Поиск по ключевым словам. Методы ранжирования результатов поиска. Элементы интеллектуальной обработки текстов, выделение значимой части текста, анатирование, рефереирование. Поиск структурированных документов. Классификация и кластеризация (вероятностные, геометрические, физические и нейросетевые методы). Методы представления логической структуры документов в виде графа и

упорядоченного дерева. Языки запросов к слабоструктурированным данным: регулярные путевые запросы, XPath, XQuery.

Практические занятия: Реализация алгоритма кластеризации.

Самостоятельная работа: Изучение рекомендованной литературы по теме Методы поиска текстовой информации. Поиск информации в сети Интернет о результатах соревнований различных модификаций базовых алгоритмов обработки текстовых данных.

Тема 3. Методы поиска слабоструктурированной информации(XML Search Methods).

Лекции: Поиск структурированных документов. Методы представления логической структуры документов в виде графа и упорядоченного дерева. Языки запросов к слабоструктурированным данным: регулярные путевые запросы, XPath, XQuery.

Практические занятия: Реализация алгоритма кластеризации.

Самостоятельная работа: Изучение рекомендованной литературы по теме Методы поиска слабоструктурированной информации.

Тема 4. Основы сетевых протоколов(Introduction to Network Protocols).

Лекции: Основные уровни сетевых протоколов. Понятие мак-адреса и IP-адреса, ARP-запросы. Способы маршрутизации, DNS, DHCP, Gateway, маски сети, классы сети. Проблемы оптимизации передачи в TCP. Обзор протоколов верхнего уровня. Преимущества объединения. Типы сетей. Терминология компьютерных сетей.

Практические занятия: Мониторинг сетевого взаимодействия. Решение задач на управление размером окна TCP протокола.

Самостоятельная работа: Изучение рекомендованной литературы по теме Основы сетевых протоколов. Поиск информации в сети Интернет об уязвимостях основных используемых протоколов.

Тема 5. Распределенные системы(Distributed Systems).

Лекции: Одно-, двух- и трехзвенные архитектуры. Распределенные объекты. Технологии промежуточного слоя (CORBA, COM, RMI). Распределенные вычисления, статическое и динамическое распараллеливание алгоритмов. Базовые распределенные алгоритмы в синхронных и асинхронных распределенных системах: выбор ведущего процесса, взаимное исключение, принятие согласованного решения, построение остоного дерева, построение контрольной точки. Математические методы анализа распределенных алгоритмов. Отладка распределенных алгоритмов. Методы построения отказоустойчивых приложений: перераспределения нагрузки, распределенные и кластерные базы данных, распределенные транзакции.

Практические занятия: Построения архитектуры распределенной системы для решения конкретной практической задачи.

Самостоятельная работа: Изучение рекомендованной литературы по теме Распределенные системы.

Тема 6. Экспертные и интеллектуальные системы (Expert and Intelligent Systems).

Лекции: Правила вывода, нечеткая логика. Понятие мутации и эволюции. Понятие нейрона, способы построения и обучения нейронной сети. Проблема автоматического обучения. Генетические алгоритмы. Нейро-сетевая модель. ДСМ-метод автоматического порождения гипотез.

Практические занятия: Решение задач с нечеткой логикой.

Самостоятельная работа: Изучение рекомендованной литературы по теме сетевые Экспертные и интеллектуальные системы.

5. Учебно-методическое обеспечение самостоятельной работы студентов, оценочные средства контроля успеваемости и промежуточной аттестации:

5.1. Типовые практические домашние задания:

- 1) Проектирование структуры базы данных для решения конкретной практической задачи.
- 2) Составление SQL-запросов по заданной схеме базы данных.
- 3) Исследование изменения результатов поиска в поисковых системах Яндекс и Google в зависимости от контекста запроса.
- 4) Разработки и тестирование результатов кластеризации методом К-средних на линейно неразделимых выборках.
- 5) Составление XPath и XQuery запросов по заданной схеме XML-документа.
- 6) Настройка сетевого обмена в среде Windows или Linux.
- 7) Исследование путей передачи пакетов на различные адреса утилитой tracerf.
- 8) Реализация гетерогенной распределенной системы на основе открытых стандартов CORBA, HTTP/JSON.

5.2. Типы задач для самостоятельного выполнения.

Отчет о выполненной домашней работе рекомендуется представлять в форме презентации. Отчёт должен содержать материалы о сути решаемой задачи, в том числе краткую постановку задачи; подход к ее решению (известные подходы /алгоритмы решения задачи с пояснениями, выбранный автором работы алгоритм/подход/ метод решения задачи); результаты эксперимента.

Задачи:

- 1) Вычисление требуемой пропускной способности сети для заданной информационной системы.
- 2) Определение размера дискового пространства для хранения резервной информации при заданной политике резервирования.
- 3) Определение коэффициентов многокритериального ранжирования по заданной обучающей выборке.
- 4) Разработка архитектуры распределенной системы, обеспечивающей устойчивость по отношению к заданному количеству сбоев.
- 5) Построение и анализ формальной модели алгоритма сбора «мусора» в распределенной системе на основе подсчета ссылок.

6. Учебно-методическое и информационное обеспечение дисциплины

Литература для выполнения домашнего задания

1. Подбельский В.В., Фомин С.С. Программирование на языке Си., М.: Финансы и статистика, 2003, с. 369 – 381.
2. Хусаинов Б.С. Структуры и алгоритмы обработки данных. Примеры на языке Си (+CD). М.: Финансы и статистика, 2004, с. 400 – 440.

Основная литература

1. Тель Ж. Введение в распределенные алгоритмы. - М.:МЦНМО, 2009
2. Э. Таненбаум, М. Ван Стеен, Распределенные системы. Принципы и парадигмы, 2003, «Питер»
3. Грабер М. Введение в SQL. - М.: Лори, 1996. - 379 с.
4. Дейт К., Введение в системы баз данных, М. Наука, 1980
5. Джексон Г. Проектирование реляционных баз данных для использования с микроЭВМ. - М.: Мир, 1991. - 252 с.
6. Калининченко Л.А., Методы и средства интеграции неоднородных баз данных, М. Наука, 1983
7. Кнут Д. Искусство программирования для ЭВМ, Т.3., М. Мир, 1978
8. Кодд Е.Ф. Реляционная модель данных для больших совместно используемых [банков](#) данных //СУБД. - 1995. - №1. - С.145-160
9. Мейер М. Теория реляционных баз данных. - М.: Мир, 1987. - 608 с.

10. Падучева Е.В. О семантике синтаксиса, М. Наука, 1974
11. Апресян Ю.Д. Лексическая семантика, М. Наука, 1974
12. Сэлтон Г. Автоматическая обработка, хранение и поиск информации. М. Сов. Радио, 1973
13. Гасанов Э.Э., Кудрявцев В.Б., «Теория хранения и поиска информации», Физматлит, 2002.
14. Ульман Дж, Основа систем баз данных, М. Финансы и статистика, 1983

7. Материально-техническое обеспечение дисциплины

(помещения, оборудование, иные материалы.)

Для проведения занятий по дисциплине «Введение в теорию распределенных информационных систем» предоставляется компьютерный класс

- с возможностью подключения проектора.
- с локальным сетевым оборудованием и выходом в сеть Интернет.
- Системное программное обеспечение: операционная система Microsoft Windows XP, операционная система Ubuntu Linux, служебные утилиты.

