

Программа учебной дисциплины «Теоретические основы информатики»

Утверждена

Академическим советом ООП

Протокол № 10 от «26» июня 2018 г.

Автор	А.П. Кирсанов, д.т.н., ki@hse.ru
Число кредитов	6
Контактная работа (час.)	72
Самостоятельная работа (час.)	156
Курс	1 курс, направление 38.03.05. Бизнес-информатика
Формат изучения дисциплины	Без он-лайн курса

I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целями освоения дисциплины являются: освоение теоретических основ информатики, необходимых для изучения, понимания и разработки прикладных информационных технологий и систем; формирование у студентов целеустремленности, организованности, трудолюбия, ответственности, готовности к ответственному и целеустремленному решению поставленных задач.

В результате освоения дисциплины студент должен:

Знать:

- понятие информации, основные свойства и виды (формы) информации, методы представления (кодирования) информации;
- основные модели процессов передачи, хранения, поиска и обработки информации;
- основные принципы криптографической защиты информации

Уметь:

- обосновывать применимость теоретических моделей для описания конкретных информационных систем и процессов;
- использовать теоретические методы и модели для исследования информационных процессов;
- самостоятельно разрабатывать математические модели информационных процессов

Владеть:

- навыками выбора теоретических моделей для описания и исследования конкретных информационных процессов;
- навыками моделирования процессов получения (сбора), хранения, поиска, передачи, и обработки данных.

Изучение дисциплины «Теоретические основы информатики» базируется на следующих дисциплинах:

Математический анализ;

Алгебра и геометрия;

Дискретная математика;

Для освоения учебной дисциплины, студенты должны знать основы линейной алгебры (матрицы и матричные операции, векторные пространства, системы линейных уравнений и методы их решения, пространство решений), дискретной математики (множества, отношения, операции над отношениями, графы, деревья), комбинаторики (основные комбинаторные объекты и комбинаторные тождества), теории вероятности (понятие вероятности, дискретные случайные величины и их вероятностные распределения, математическое ожидание случайной величины);

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин: Программирование, Информационные процессы, системы и сети, Управление данными, Моделирование процессов и систем, Базы, данных, Методы разработки и анализа алгоритмов, Высокопроизводительные вычисления, Распределенные информационные системы, Семантические информационные системы, Интеллектуальные системы.

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Тема 1. Информация.

Понятие информации, ее основные свойства и особенности. Понятие сообщения и его формы, знаки, алфавиты, понятие формального языка. Формула Хартли. Информация и данные. Конечный вероятностный источник сообщений. Энтропия источника.

Тема 2. Представление информации

Кодирование сообщений источника и текстов. Равномерное и неравномерное кодирование. Дерево кода. Однозначное декодирование, префиксные коды. Метод Маркова. Условия существования префиксного кода с заданными длинами слов, теорема Крафта. Методы построения префиксных кодов. Код Фано. Средняя длина кодового слова. Нижняя граница средней длины кодового слова. Оптимальное кодирование, свойства оптимальных кодов, построение оптимального кода методом Хаффмана. Сжатие данных.

Тема 3. Передача информации

Передача информации. Основные способы передачи сообщений (последовательный, параллельный, синхронный и асинхронный). Модель процесса передачи (двоичный симметричный канал). Надежность передачи сообщений, способы повышения надежности. Принципы использования кодов, обнаруживающих и исправляющих ошибки. Расстояние Хемминга. Связь минимального расстояния кода с его характеристиками. Корректирующие возможности кодов, границы Хэмминга и Варшавова-Гилберта. Понятие линейного группового кода. Построение линейного группового кода по заданной проверочной матрице. Свойства линейного группового кода. Декодирование с использованием синдрома.

Защита информации при передаче, основные угрозы и методы защиты от них. Основы цифровой стеганографии. Симметричная, асимметричная и комбинированная криптосистемы. Электронная цифровая подпись и принципы ее использования. Цифровые сертификаты.

Тема 4. Хранение и поиск информации.

Основные виды задач поиска. Описание запросов и объектов поиска. Модели информационного поиска. Структуры хранения данных и методы доступа. Взаимосвязь

способов хранения и эффективности поиска. Основы технологии баз данных. Модели данных, реляционная модель данных. Реляционная алгебра. Запросы в виде реляционных выражений. Эквивалентность, сложность и оптимизация запросов. Основы написания запросов на языке SQL.

Тема 5. Обработка информации

Понятие алгоритма и его свойства. Способы формальной записи алгоритмов. Моделирование процессов обработки данных конечными автоматами. Автоматы Мили и Мура. Преобразование автоматов. Минимизация состояний автомата. Распределенная обработка информации и проблемы взаимодействия параллельно выполняемых процессов обработки. Методы описания и анализа процессов распределенной обработки. Сети Петри. Основные задачи, решаемые с использованием сетей Петри (ограниченность, активность, достижимость, покрываемость). Дерево достижимости и матричный метод анализа сетей Петри. Язык сети Петри.

III. ОЦЕНИВАНИЕ

Тип контроля	Форма контроля	1 год			Параметры
		2	3		
Текущий (неделя)	Контрольная работа	+			Текущий (неделя)
Текущий (неделя)	Домашнее задание		+		Самостоятельная письменная работа на анализ частотных характеристик текстов и использование этих характеристик для дешифрирования текстов: 4 часа, оценка результатов – 1 неделя
Итоговый (неделя)	Экзамен		+		Письменная работа (неделя)

Итоговая оценка по учебной дисциплине складывается из оценок за:

- работу на аудиторных занятиях;
- контрольную работу;
- домашнее задание;
- экзамен.

В процессе экзамена студенты должны дать письменные ответы на теоретические вопросы и решить задачи.

Оценки по всем формам контроля выставляются по 10-ти балльной шкале.

Критерии оценивания

Накопленная оценка « $O_{\text{накоп}}$ » формируется следующим образом:

$$O_{\text{накоп}} = 0,2 \times O_{\text{ауд}} + 0,5 \times O_{\text{кр}} + 0,3 \times O_{\text{дз}},$$

где:

$O_{\text{ауд}}$ – оценка за аудиторную работу (посещение занятий);

$O_{\text{кр}}$ – оценка за контрольную работу;

$O_{\text{дз}}$ – оценка за домашнее задание.

Итоговая оценка « O » формируется следующим образом:

$$O = 0,6 \times O_{\text{накоп}} + 0,4 \times O_{\text{экз}},$$

где $O_{\text{экз}}$ – оценка за экзамен.

Критерии выставления оценки за аудиторную работу (посещение занятий):

100% посещаемости – 10 баллов

90% посещаемости – 9 баллов

80% посещаемости – 8 баллов

70% посещаемости – 7 баллов

60% посещаемости – 6 баллов

50% посещаемости – 5 баллов

40% посещаемости – 4 балла

30% посещаемости – 3 балла

20% посещаемости – 2 балла

10% посещаемости – 1 балл

При формировании результирующей оценки на основе весовых коэффициентов применяется округление до целого числа за счет отбрасывания дробной части.

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Тематика заданий по различным формам текущего контроля:

- Основные свойства информации.
- Вычисление энтропии конечного вероятностного источника.
- Построение кодового дерева по заданному множеству кодовых слов.
- Проверка существования префиксного кода с заданными длинами кодовых слов с использованием неравенства Крафта.
- Проверка однозначности декодирования кода с помощью метода Маркова.
- Кодирование двоичным кодом Фано множества сообщений (для различного числа сообщений и частот их появления).
- Определение средней длины слова для заданного кода и частоты появления сообщений источника.
- Кодирование методом Хаффмена заданного конечного источника.
- Нахождение точек шара радиуса r в пространстве B^n (для различных значений r и n).
- Вычисление числа точек в сфере радиуса r в пространстве B^n (для различных значений r и n).
- Вычисление числа точек в шаре радиуса r в пространстве B^n (для различных значений r и n).
- Определение кодового расстояния Хемминга для заданного кодового множества.
- Доказательство неравенства треугольник для расстояния Хемминга.
- Построение линейного кода по заданной проверочной матрице.

- По заданной проверочной матрице найти порождающую матрицу.
- По заданной проверочной матрице произвести разбиение пространства на классы, выбрать лидеров и декодировать заданное сообщение (обнаружить и исправить, если возможно, ошибки).
- Построить конечный автомат (табличным, графическим и матричным способом)
- Преобразовать автомат Мили в автомат Мура
- Преобразовать автомат Мура в автомат Мили
- Используя матричный метод анализа сетей Петри, решить задачу о достижимости (покрываемости) заданной разметки сети.
- Проверка возможности тупикового состояния заданной сети Петри.
- Проверка достижимости заданной разметки сети Петри матричным методом.
- Преобразовать автомат Мили в сеть Петри
- Написать запрос на языке реляционной алгебры

Домашнее задание по дисциплине «Теоретические основы информатики»

Написать программу, читающую побайтно заданный файл и подсчитывающую число появлений каждого из 256 возможных знаков. Можно использовать программу (макрос inByte) на языке VBA для Excel (содержится в файле hw_tits2008.xls). Исследовать с помощью разработанной программы файлы различных типов (.jpeg, .gif, .bmp, .txt, .doc, .xls, .exe).

Для каждого исследуемого в работе файла сделать следующее:

- построить диаграмму, показывающую число каждого из 256 байт в исследуемом файле.
- рассматривая знаки (байты) файла как сообщения, а частоты их появления как вероятности, представить файл как вероятностный источник сообщений. Вычислить энтропию этого источника.

Постараться объяснить наблюдаемые на диаграммах особенности. Основываясь на построенных диаграммах и вычисленных значениях энтропии, указать, какие из рассматриваемых файлов могут быть сжаты в большей степени и почему. Ответы на данные вопросы сформулировать в виде выводов.

В дополнение к описанным действиям разработанная программа должна выполнять аналогичные расчеты только для байтов, соответствующих символам кириллицы и пробелу. Кодовую таблицу CP-1251 для кириллицы можно найти в Internet, например, по адресу <http://dll.botik.ru/educ/clerk/Library/Method/kod-tabl.ru.html>. Байты, соответствующие латинскому алфавиту и специальные знаки должны игнорироваться. При подсчете не различать прописные и строчные буквы. Программа должна вычислять частоты встречаемости букв кириллицы и энтропию текста, содержащегося в файле. При вычислении энтропии не учитывать знаки, отличные от символов кириллицы и пробела. Допускается выполнение данной части работы с использованием Excel. Распечатать частоты появления букв кириллицы и пробела в тексте.

Используя данные о частотах встречаемости букв русского языка (см. <http://www.lg--web.chat.ru/texts.html> или <http://www.statsoft.ru/home/portal/exchange/textanalysis.htm>) дешифровать зашифрованный текст. При шифровании все знаки, кроме букв и пробела, пропускались,

прописные буквы заменялись на строчные. Затем каждый из 33 знаков (32 буквы и пробел) заменялся на другой знак в соответствии с некоторой перестановкой.

Результаты работы представить в виде распечатанного отчета. В электронном виде представить анализируемые файлы и дешифрованный текст, исходные тексты и исполняемые модули программ (если таковые имеются).

Вопросы для подготовки к экзамену по дисциплине «Теоретические основы информатики»

1. Понятие информации.
2. Основные свойства информации.
3. Информационные процессы в живой природе, обществе и технике: получение (сбор), передача, обработка (преобразование), хранение и использование информации.
4. Информация и сообщения, формы сообщений.
5. Вероятностный подход к определению количества информации, конечный вероятностный источник сообщений. энтропия.
6. Язык как способ представления информации. Понятие формального языка.
7. Кодирование знаков и слов. Условия однозначности декодирования.
8. Префиксный код. Свойства префиксного кода, полный префиксный код. Дерево кода.
9. Метод Маркова
10. Условие существования префиксного кода, неравенство и теорема Крафта.
11. Построение префиксных кодов, код Фано.
12. Средняя длина кода, избыточность кодирования, свойства избыточности префиксного кода.
13. Оптимальное кодирование, свойства оптимальных кодов.
14. Код Хаффмена, сжатие источника и расщепление кода, оптимальность кода Хаффмена.
15. Передача информации, общая схема передачи информации, двоичный симметричный канал, способы борьбы с помехами в канале.
16. Геометрическая интерпретация кодов, расстояние Хэмминга, помехоустойчивое кодирование.
17. Минимальное расстояние кода, коды, обнаруживающие и исправляющие ошибки.
18. Линейные групповые коды, способы задания, проверочная и порождающая матрицы кода, систематический вид кода.
19. Связь минимального расстояния линейного кода с проверочной матрицей.
20. Декодирование линейных кодов, синдром, разбиение пространства B^n на смежные классы, лидеры классов.
21. Защита информации при передаче, основные угрозы и методы защиты от них.
22. Цифровая стеганография.
23. Симметричная, асимметричная и комбинированная криптосистемы.
24. Электронная цифровая подпись и принципы ее использования.
25. Способы построения конечного автомата.
26. Преобразование автоматов.
27. Минимизация числа состояний автомата.
28. Моделирование систем с использованием сетей Петри. Структура сети, разметка сети, функционирование сети.
29. Свойства сетей Петри безопасность, ограниченность, сохранение, достижимость.
30. Матричный метод анализа сетей Петри.
31. Дерево достижимости и его свойства, алгоритм построения дерева, теорема конечности дерева достижимости (без доказательства). Анализ сетей Петри с использованием дерева достижимости.
32. Модели информационного поиска.

33. Реляционная модель данных. Отношения, кортежи, атрибуты, домены.
34. Реляционная алгебра. Поисквые запросы в виде реляционных выражений.
35. .Операции реляционной алгебры (объединение, пересечение, разность, произведение, проекция, селекция, естественное соединение, деление).

V. РЕСУРСЫ

5.1 Основная литература

1. Новиков Ф.А. Дискретная математика для программистов. 3-е изд. – СПб.: Питер 2017.

5.2 Дополнительная литература

1. Иванов, И. В. Теория информационных процессов и систем + доп. Материалы в ЭБС : учеб. пособие для академического бакалавриата / И. В. Иванов. — 3-е изд., перераб. и доп. — М. : Издательство Юрайт, 2018. — 228 с. — (Серия : Бакалавр. Академический курс). — ISBN 978-5-534-05705-8. — Режим доступа : www.biblio-online.ru/book/5D3F8A8A-3340-42EF-977A-EE364CDED004.
2. Черпаков, И. В. Теоретические основы информатики : учебник и практикум для академического бакалавриата / И. В. Черпаков. — М. : Издательство Юрайт, 2018. — 353 с. — (Серия : Бакалавр. Академический курс). — ISBN 978-5-9916-8562-7. — Режим доступа : www.biblio-online.ru/book/A4AE728B-6E4D-4DAD-87CB-932D15097ED7.
3. Информатика: в 2 ч. Ч. 1:учебник / С.В.Назаров и др. – М.: БИНОМ. Лаборатория знаний, 2012.

5.3 Программное обеспечение

Для подготовки практических задач, докладов и выступлений студентами используется современная учебно-лабораторная база, в том числе:

- стандартные пакеты прикладных программ офисного назначения, в том числе:
 - системы подготовки текстов (Microsoft Word);
 - системы электронных таблиц (Microsoft Excel);
 - системы подготовки презентаций (Microsoft PowerPoint);

№ п/п	Наименование	Условия доступа
1.	Microsoft Windows 7 Professional RUS Microsoft Windows 10 Microsoft Windows 8.1 Professional RUS	<i>Из внутренней сети университета (договор)</i>
2.	Microsoft Office Professional Plus 2010	<i>Из внутренней сети университета (договор)</i>

5.4 Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

№ п/п	Наименование	Условия доступа
1.	Электронно-библиотечная система Юрайт (доступна НИУ ВШЭ)	URL: https://biblio-online.ru/

5.5 Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);
- мультимедийный проектор с дистанционным управлением.

Учебные аудитории для лабораторных и самостоятельных занятий по дисциплине оснащены персональными компьютерами, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.

Автор программы: профессор НИУ ВШЭ А.П.Кирсанов