

Краевое государственное бюджетное
профессиональное образовательное учреждение
«Хабаровский машиностроительный техникум»

Комплект
контрольно-оценочных средств по учебной дисциплине
ОП.12. Основы теории информации

программы подготовки специалистов среднего звена (ППССЗ)
по специальности 09.02.06 «Сетевое и системное администрирование»
базовой подготовки

г. Хабаровск
2018 г.

Комплект контрольно-оценочных средств разработан на основе Федерального государственного образовательного стандарта среднего профессионального образования по специальности СПО 09.02.06 «Сетевое и системное администрирование» базовой подготовки программы учебной дисциплины «ОП.12. Основы теории информации»

Организация-разработчик:

Краевое государственное бюджетное профессиональное образовательное учреждение «Хабаровский машиностроительный техникум» (КГБ ПОУ ХМТ)

Разработчик:

Богданова Т.С., преподаватель первой квалификационной категории КГБ ПОУ ХМТ

Рассмотрены и одобрены

на заседании цикловой
«Математического
естественнонаучного цикла»

Протокол № _____

от «__» _____ 2018 г.

Председатель ЦК

_____ Т.А. Новикова

комиссии
и

Утверждаю

Заместитель директора по УР
_____ И.Н. Пухляр

«__» _____ 2018 г.

1. Паспорт комплекта контрольно-оценочных средств

Контрольно-оценочные средства (КОС) предназначены для оценки результатов освоения учебной дисциплины «ОП.12. Основы теории информации»

КОС разработаны в соответствии с:

примерной программой учебной дисциплины «Основы теории информации»;

программой учебной дисциплины «ОП.12. Основы теории информации»

В результате освоения учебной дисциплины «ОП.12. Основы теории информации» обучающийся должен обладать предусмотренными ФГОС специальности 09.02.06 «Сетевое и системное администрирование» следующими умениями, знаниями, общими и профессиональными компетенциями:

Умения

У.1. Применять закон аддитивности информации.

У.2. Применять теорему Котельникова.

У.3. Использовать формулу Шеннона.

Знания

З.1. Виды и формы представления информации.

З.2. Методы и средства определения количества информации.

З.3. Принципы кодирования и декодирования информации.

З.4. Способы передачи цифровой информации.

З.5. Методы повышения помехозащищенности передачи и приема данных, основы теории сжатия данных.

З.6. Методы криптографической защиты информации.

З.7. Способы генерации ключей.

Общие компетенции

ОК.01. Выбирать способы решения задач профессиональной деятельности, применительно к различным контекстам.

ОК.02. Осуществлять поиск, анализ и интерпретацию информации, необходимой для выполнения задач профессиональной деятельности.

ОК.04. Работать в коллективе и команде, эффективно взаимодействовать с коллегами, руководством, клиентами.

ОК.05. Осуществлять устную и письменную коммуникацию на государственном языке с учетом особенностей социального и культурного контекста.

ОК.09. Использовать Основы теории информации в профессиональной деятельности.

ОК.10. Пользоваться профессиональной документацией на государственном и иностранном языках.

Профессиональные компетенции

ПК.1.3. Обеспечивать защиту информации в сети с использованием программно-аппаратных средств.

КОС включают контрольные материалы для проведения текущего контроля и промежуточной аттестации в форме *экзамена*.

2. Результаты освоения учебной дисциплины, подлежащие проверке

В результате аттестации по учебной дисциплине осуществляется комплексная проверка следующих умений и знаний, а также динамика формирования общих компетенций:

Таблица 1

Результаты обучения: умения, знания, общие и профессиональные компетенции	Показатели оценки результата	Форма контроля и оценивания
Уметь:		
У.1. Применять закон аддитивности информации.	Применять закон аддитивности информации	текущий контроль, рубежный контроль промежуточная аттестации
У.2. Применять теорему Котельникова.	Применять теорему Котельникова	
У.3. Использовать формулу Шеннона.	Использовать формулу Шеннона	
Знать:		
3.1. Виды и формы представления информации.	Виды и формы представления информации	текущий контроль, рубежный контроль промежуточная аттестации
3.2. Методы и средства определения количества информации.	Методы и средства определения количества информации	
3.3. Принципы кодирования и декодирования информации.	Принципы кодирования и декодирования информации	
3.4. Способы передачи цифровой информации.	Способы передачи цифровой информации	
3.5. Методы повышения помехозащищенности передачи и приема данных, основы теории сжатия данных.	Методы повышения помехозащищенности передачи и приема данных, основы теории сжатия данных	
3.6. Методы криптографической защиты информации.	Методы криптографической защиты информации	
3.7. Способы генерации ключей.	Способы генерации ключей	
Общие компетенции:		
ОК.01. Выбирать способы решения задач профессиональной деятельности, применительно к различным контекстам	Умения: распознавать задачу и/или проблему в профессиональном и/или социальном контексте; анализировать задачу и/или проблему и выделять её составные части; определять этапы решения задачи; выявлять и эффективно искать информацию, необходимую для решения задачи и/или проблемы; составить план действия; определить необходимые ресурсы; владеть актуальными методами работы в профессиональной и смежных сферах; реализовать составленный план; оценивать результат и последствия своих действий (самостоятельно или с помощью наставника). Знания: актуальный профессиональный и социальный	текущий контроль, рубежный контроль промежуточная аттестации

	<p>контекст, в котором приходится работать и жить; основные источники информации и ресурсы для решения задач и проблем в профессиональном и/или социальном контексте; алгоритмы выполнения работ в профессиональной и смежных областях; методы работы в профессиональной и смежных сферах; структуру плана для решения задач; порядок оценки результатов решения задач профессиональной деятельности</p>	
<p>ОК.02. Осуществлять поиск, анализ и интерпретацию информации, необходимой для выполнения задач профессиональной деятельности</p>	<p>Умения: определять задачи для поиска информации; определять необходимые источники информации; планировать процесс поиска; структурировать получаемую информацию; выделять наиболее значимое в перечне информации; оценивать практическую значимость результатов поиска; оформлять результаты поиска.</p> <p>Знания: номенклатура информационных источников применяемых в профессиональной деятельности; приемы структурирования информации; формат оформления результатов поиска информации</p>	
<p>ОК.04. Работать в коллективе и команде, эффективно взаимодействовать с коллегами, руководством, клиентами.</p>	<p>Умения: организовывать работу коллектива и команды; взаимодействовать с коллегами, руководством, клиентами в ходе профессиональной деятельности.</p> <p>Знания: психологические основы деятельности коллектива, психологические особенности личности; основы проектной деятельности.</p>	
<p>ОК.05. Осуществлять устную и письменную коммуникацию на государственном языке с учетом особенностей социального и культурного контекста</p>	<p>Умения: грамотно излагать свои мысли и оформлять документы по профессиональной тематике на государственном языке, проявлять толерантность в рабочем коллективе.</p> <p>Знания: особенности социального и культурного контекста; правила оформления документов и построения устных сообщений.</p>	
<p>ОК.09. Использовать Основы теории информации в профессиональной деятельности</p>	<p>Умения: применять средства информационных технологий для решения профессиональных задач; использовать современное программное обеспечение.</p> <p>Знания: современные средства и устройства информатизации; порядок их применения и программное обеспечение в профессиональной деятельности</p>	
<p>ОК.10. Пользоваться профессиональной</p>	<p>Умения: понимать тексты на базовые профессиональные темы</p>	

документацией на государственном и иностранном языке	Знания: правила построения простых и сложных предложений на профессиональные темы; основные общеупотребительные глаголы (бытовая и профессиональная лексика); лексический минимум, относящийся к описанию предметов, средств и процессов профессиональной деятельности; особенности произношения; правила чтения текстов профессиональной направленности.	
Профессиональные компетенции:		
ПК.1.3. Обеспечивать защиту информации в сети с использованием программно-аппаратных средств.	Умения: Настраивать стек протоколов TCP/IP и использовать встроенные утилиты операционной системы для диагностики работоспособности сети. Использовать программно-аппаратные средства технического контроля. Знания: Требования к компьютерным сетям. Требования к сетевой безопасности. Элементы теории массового обслуживания. Основные понятия теории графов. Основные проблемы синтеза графов атак. Системы топологического анализа защищенности компьютерной сети. Архитектуру сканера безопасности.	текущий контроль, рубежный контроль промежуточная аттестации

3. Оценка освоения учебной дисциплины:

3.1. Формы и методы оценивания

Предметом оценки служат умения и знания, предусмотренные ФГОС по учебной дисциплине «ОП.12. Основы теории информации», направленные на формирование общих и профессиональных компетенций.

3.2. Типовые задания для оценки освоения учебной дисциплины

Типовые задания для оценки знаний (**текущий контроль**)

Тема: Единицы измерения информации

1. Чему равен 1 байт?

А . 10бит

Б. 8 бит

В. 1024 бит

Г. 8 кб

2. Во сколько раз 1 Мбайт больше 1 Кбайта?

А . 1000

Б. 1024

В. 100

Г. 124

3. Сколько байт в 1 Кбайте?

А. 8

Б. 1024

В. равны

Г. 10

4. Расположите в порядке возрастания:

101 бит

1000 байт

1 кб

10 мб

2 гб

5. Сколько бит в 10 байтах?

А. 80

Б. 10

В. 800

Г. 100

6. Расположи в порядке убывания:

0,5 гб

20 мб

18 кб

1000 байт

7. Наименьшая единица информации - это:

Бит

Байт

Мб

Кб

8. 64 бита -это:

А. 8 байт

Б. 8 кб

В. 2 байт

Г. 10 мб

9. 128 бит - это:

А. 1/4 килобита

Б. 1/8 килобита

В. 1/2 килобита

Г. 1/10 килобита

10. В какой строке единицы измерения информации представлены по возрастанию?

А. Гигабайт, мегабайт, килобайт, бит, терабайт

Б. Бит, байт, килобайт, мегабайт, гигабайт, терабайт

В. Бит, байт, мегабайт, килобайт, гигабайт

Г. Байт, бит, килобайт, мегабайт, гигабайт, эксбайт

11. Установите соответствие

Емкость файла	20 кб
Dvd диск	17 гб
Жесткий диск	2 терабайта

12. У Васи есть файл размером 1058 байт. Сможет ли Вася уместить его на флэшку объемом 2 Гб, если свободного места осталось 3 Мб?

- А. Нет, на флэшке мало свободного места
- Б. Да, свободного места хватит, чтобы разместить файл и еще останется
- В. Нет, на флэшку объемом 2 Гбайта данный файл не поместится
- Г. Да, объема флэшки хватит, чтобы разместить файл

13. Сможет Вася отправить файл по электронной почте объемом 73428992 байт, если к письму можно прикрепить файл объемом не более 10 Мб?

- А. Да, но он больше не сможет прикрепить ни один файл
- Б. Нет, объем файла на много превышает 10 Мбайт
- В. Да и еще сможет присоединить какой-нибудь небольшой по объему файл

14. Васе надо опрарвить 3 файла на конкурс, каждый соответственно объемом 256 мб, 550 мб, 1058 байт. Сможет ли Вася уместить данные файлы на пустой флэшке объемом 1073741824 байт?

- А. Нет, он сможет уместить только один файл, самый большой
- Б. Нет, он сможет уместить только файл размером 256 Мбайт
- В. Вася сможет уместить на флэшке все три файла
- Г. Вася сможет уместить все три файла на флэшке и еще останется место

15. Установите соответствие между носителем информации и его объемом

CD-R	700 мб
DVD - R	4,7 гб
USB Flash	4 гб, 8 гб, 16 гб, 32 гб, 64 гб, 128 гб
дискета	1,44 мб
HDD	500 гб, 2 ТБ

16. Какую математическую операцию нужно применить, чтобы перевести из большей единицы измерения информации в меньшую?

- А. Сложение
- Б. Умножение
- В. Деление
- Г. Вычитание

Время на подготовку и выполнение: 30 мин.

За правильный ответ на вопросы или верное решение задачи выставляется положительная оценка – 1 балл.

За не правильный ответ на вопросы или неверное решение задачи выставляется – 0 баллов.

Шкала оценки образовательных достижений

Процент результативности (правильных ответов)	Оценка уровня подготовки	
	балл (отметка)	вербальный аналог
90 ÷ 100	5	отлично
80 ÷ 89	4	хорошо
70 ÷ 79	3	удовлетворительно
менее 70	2	неудовлетворительно

Тема: «Передача информации»

1. Какое устройство системы передачи информации обеспечивает эффективность ее передачи?
 А. модулятор;
 Б. кодер источника;
 В. кодер канала.
2. Какое устройство системы передачи информации обеспечивает достоверность ее передачи?
 А. кодер канала;
 Б. кодер источника;
 В. модулятор.
3. Что является информационной характеристикой только канала связи?
 А. скорость передачи информации;
 Б. пропускная способность.
4. Определить пропускную способность дискретного канала связи без шума, по которому передается 10 сигн./сек. Алфавит сообщений источника состоит из 16 букв.

Время на подготовку и выполнение: 15 мин.

За правильный ответ на вопросы или верное решение задачи выставляется положительная оценка – 1 балл.

За не правильный ответ на вопросы или неверное решение задачи выставляется – 0 баллов.

Шкала оценки образовательных достижений

Процент результативности (правильных ответов)	Оценка уровня подготовки	
	балл (отметка)	вербальный аналог
90 ÷ 100	5	отлично
80 ÷ 89	4	хорошо
70 ÷ 79	3	удовлетворительно
менее 70	2	неудовлетворительно

Расчетное задание

Вариант 1-10

1.1. Определить количество информации (по Хартли), содержащееся в системе, информационная емкость которой характеризуется десятичным числом Q .

№	1	2	3	4	5	6	7	8	9	10
Q	500	1000	750	1250	250	1500	650	900	1100	1600

1.2. Определить пропускную способность канала связи, по которому передаются сигналы S_i . Помехи в канале определяются матрицей условных вероятностей $P(S_j / S_i)$. За секунду может быть передано $N = 10$ сигналов.

1. $\begin{pmatrix} 0,2 & 0,8 & 0 \\ 0 & 0,2 & 0,8 \\ 0,8 & 0 & 0,2 \end{pmatrix}$ 2. $\begin{pmatrix} 0,4 & 0,3 & 0,3 \\ 0,3 & 0,4 & 0,3 \\ 0,3 & 0,3 & 0,4 \end{pmatrix}$ 3. $\begin{pmatrix} 0,7 & 0,3 & 0 \\ 0 & 0,7 & 0,3 \\ 0,3 & 0 & 0,7 \end{pmatrix}$
4. $\begin{pmatrix} 0,2 & 0,4 & 0,4 \\ 0,4 & 0,2 & 0,4 \\ 0,4 & 0,4 & 0,2 \end{pmatrix}$ 5. $\begin{pmatrix} 0,4 & 0,6 & 0 \\ 0 & 0,4 & 0,6 \\ 0,6 & 0 & 0,4 \end{pmatrix}$ 6. $\begin{pmatrix} 0,6 & 0,2 & 0,2 \\ 0,2 & 0,6 & 0,2 \\ 0,2 & 0,2 & 0,6 \end{pmatrix}$
7. $\begin{pmatrix} 1/3 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{pmatrix}$ 8. $\begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,1 & 0,8 & 0,1 \\ 0,1 & 0,1 & 0,8 \end{pmatrix}$ 9. $\begin{pmatrix} 0,4 & 0,4 & 0,1 & 0,1 \\ 0,1 & 0,1 & 0,4 & 0,4 \end{pmatrix}$
10. $\begin{pmatrix} 0,3 & 0,35 & 0,35 \\ 0,35 & 0,3 & 0,35 \\ 0,35 & 0,35 & 0,3 \end{pmatrix}$

Время на подготовку и выполнение: 30 мин.

За верное решение задачи выставляется положительная оценка. За неверное решение задачи предлагается новый вариант задания.

Расчетное задание

1. Источник сообщений вырабатывает ансамбль символов. Символы в последовательности статистически независимы. Вычислить энтропию источника и определить избыточность.

Вариант 1	Вариант 2	Вариант 3
Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 0,4 & 0,2 & 0,15 & 0,1 & 0,1 & 0,05 \end{pmatrix}$	Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 0,2 & 0,3 & 0,4 & 0,1 \end{pmatrix}$	Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0,3 & 0,1 & 0,25 & 0,15 & 0,2 \end{pmatrix}$
Таблица 2 $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0,6 & 0,4 \\ 0 & 0,4 & 0,6 \end{pmatrix}$	Таблица 2 $\begin{pmatrix} 0,1 & 0,2 & 0,3 & 0,4 \\ 0,4 & 0,3 & 0,2 & 0,1 \\ 0,3 & 0,4 & 0,1 & 0,2 \end{pmatrix}$	Таблица 2 $\begin{pmatrix} 0,1 & 0,3 & 0,6 \\ 0,2 & 0,4 & 0,4 \\ 0,3 & 0,6 & 0,1 \end{pmatrix}$
Таблица 3 $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,2 & 0,2 & 0,6 \end{pmatrix}$	Таблица 3. $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,5 & 0,25 & 0,25 \end{pmatrix}$	Таблица 3 $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,25 & 0,25 & 0,5 \end{pmatrix}$

Вариант 4	Вариант 5	Вариант 6
Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0,1 & 0,2 & 0,2 & 0,25 & 0,25 \end{pmatrix}$	Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 0,1 & 0,6 & 0,05 & 0,1 & 0,05 & 0,1 \end{pmatrix}$	Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0,3 & 0,15 & 0,15 & 0,15 & 0,25 \end{pmatrix}$
Таблица 2 $\begin{pmatrix} 0 & 1 & 0 \\ p & 0 & 1-p \\ 0 & \gamma & 1-\gamma \end{pmatrix}$	Таблица 2 $\begin{pmatrix} 0,5 & 0 & 0,5 & 0 \\ 0,1 & 0,2 & 0,3 & 0,4 \\ 0 & 0,5 & 0 & 0,5 \end{pmatrix}$	Таблица 2 $\begin{pmatrix} 0,3 & 0,2 & 0,5 \\ 0,2 & 0,3 & 0,5 \\ 0,5 & 0,2 & 0,3 \end{pmatrix}$
Таблица 3 $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,2 & 0,2 & 0,6 \end{pmatrix}$	Таблица 3 $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,4 & 0,4 & 0,2 \end{pmatrix}$	Таблица 3 $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,15 & 0,3 & 0,55 \end{pmatrix}$
Вариант 7	Вариант 8	Вариант 9
Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 0,25 & 0,15 & 0,25 & 0,35 \end{pmatrix}$	Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0,1 & 0,2 & 0,15 & 0,25 & 0,05 & 0,1 & 0,15 \end{pmatrix}$	Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 0,2 & 0,3 & 0,4 & 0,1 \end{pmatrix}$
Таблица 2 $\begin{pmatrix} 0,8 & 0,1 & 0,1 & 0 \\ 0,1 & 0,2 & 0,3 & 0,4 \\ 0,1 & 0,1 & 0 & 0,8 \end{pmatrix}$	Таблица 2 $\begin{pmatrix} 1/3 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/3 & 1/3 \\ 1 & 0 & 0 & 0 \end{pmatrix}$	Таблица 2 $\begin{pmatrix} 0,1 & 0,2 & 0,3 & 0,4 \\ 0,4 & 0,3 & 0,2 & 0,1 \\ 0,3 & 0,4 & 0,1 & 0,2 \end{pmatrix}$
Таблица 3 $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,9 & 0,05 & 0,05 \end{pmatrix}$	Таблица 3 $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,8 & 0,1 & 0,1 \end{pmatrix}$	Таблица 3 $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,5 & 0,25 & 0,25 \end{pmatrix}$

2. Найти число значений m равномерно распределенной случайной величины V , при которой ее энтропия будет равна энтропии случайной величины X .

Вариант 1	Вариант 2
Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 0,4 & 0,2 & 0,15 & 0,1 & 0,1 & 0,05 \end{pmatrix}$	Таблица 1 $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 0,2 & 0,3 & 0,4 & 0,1 \end{pmatrix}$
Таблица 2 $\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}$	Таблица 2 $\begin{pmatrix} 0,5 & 0,1 & 0,2 & 0,2 \\ 0,3 & 0,4 & 0,15 & 0,15 \\ 0,2 & 0,8 & 0 & 0 \end{pmatrix}$
Таблица 3 $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,1 & 0,2 & 0,7 \end{pmatrix}$	Таблица 3 $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,2 & 0,3 & 0,5 \end{pmatrix}$

<p style="text-align: center;">Вариант 3</p> <p>Таблица 1</p> $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0,3 & 0,1 & 0,25 & 0,15 & 0,2 \end{pmatrix}$ <p>Таблица 2</p> $\begin{pmatrix} P & \frac{1-P}{2} & \frac{1-P}{2} \\ \frac{1-P}{2} & P & \frac{1-P}{2} \\ \frac{1-P}{2} & \frac{1-P}{2} & P \end{pmatrix}$ <p>Таблица 3</p> $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,3 & 0,35 & 0,35 \end{pmatrix}$	<p style="text-align: center;">Вариант 4</p> <p>Таблица 1</p> $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0,1 & 0,2 & 0,2 & 0,25 & 0,25 \end{pmatrix}$ <p>Таблица 2</p> $\begin{pmatrix} 0,3 & 0,6 & 0,1 \\ 0,4 & 0,4 & 0,2 \\ 0,4 & 0,3 & 0,3 \end{pmatrix}$ <p>Таблица 3</p> $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,6 & 0,2 & 0,2 \end{pmatrix}$
<p style="text-align: center;">Вариант 5</p> <p>Таблица 1</p> $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 0,1 & 0,6 & 0,05 & 0,1 & 0,05 & 0,1 \end{pmatrix}$ <p>Таблица 2</p> $\begin{pmatrix} 0,5 & 0,1 & 0,2 & 0,2 \\ 0,1 & 0,2 & 0,3 & 0,4 \\ 0,3 & 0,4 & 0,15 & 0,15 \end{pmatrix}$ <p>Таблица 3</p> $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,3 & 0,3 & 0,4 \end{pmatrix}$	<p style="text-align: center;">Вариант 6</p> <p>Таблица 1</p> $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0,3 & 0,15 & 0,15 & 0,15 & 0,25 \end{pmatrix}$ <p>Таблица 2</p> $\begin{pmatrix} 0,4 & 0,3 & 0,3 \\ 0,2 & 0,3 & 0,5 \\ 0,4 & 0,4 & 0,2 \end{pmatrix}$ <p>Таблица 3</p> $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,25 & 0,35 & 0,4 \end{pmatrix}$
<p style="text-align: center;">Вариант 7</p> <p>Таблица 1</p> $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 0,25 & 0,15 & 0,25 & 0,35 \end{pmatrix}$ <p>Таблица 2</p> $\begin{pmatrix} 0,3 & 0,4 & 0,1 & 0,2 \\ 0,5 & 0,1 & 0,2 & 0,2 \\ 0,1 & 0,1 & 0 & 0,8 \end{pmatrix}$ <p>Таблица 3</p> $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,1 & 0,35 & 0,55 \end{pmatrix}$	<p style="text-align: center;">Вариант 8</p> <p>Таблица 1</p> $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0,1 & 0,2 & 0,15 & 0,25 & 0,05 & 0,1 & 0,15 \end{pmatrix}$ <p>Таблица 2</p> $\begin{pmatrix} 0,4 & 0,3 & 0,3 \\ 0,3 & 0,6 & 0,1 \\ 0,5 & 0,25 & 0,25 \end{pmatrix}$ <p>Таблица 3</p> $\begin{pmatrix} a_1 & a_2 & a_3 \\ 0,15 & 0,25 & 0,6 \end{pmatrix}$

Время на подготовку и выполнение: 30 мин.

За верное решение задачи выставляется положительная оценка. За неверное решение задачи предлагается новый вариант задания.

Тема: «Кодирование информации»

1. Что происходит с длиной сообщения при эффективном кодировании?

- А. увеличивается;
 Б. остается прежней;
 В. уменьшается.
2. Как изменяется эффективность кода при увеличении длины блока при блоковом кодировании?
 А. не убывает;
 Б. не изменяется;
 В. не возрастает.
3. Закодировать сообщение 100110 кодом с проверкой четности.
 А. 1001100;
 Б. 10011011;
 В. 1001101.
4. Закодировать число 13 кодом Хэмминга (4,7).
 А. 1010101;
 Б. 1110101;
 В. 1011101.
5. Исправить ошибку в кодовом слове 1010111 (код Хэмминга (4,7)) и найти передаваемое десятичное число.
 А. 15;
 Б. 13;
 В. 9.

Время на подготовку и выполнение: 15 мин.

За правильный ответ на вопросы или верное решение задачи выставляется положительная оценка – 1 балл.

За не правильный ответ на вопросы или неверное решение задачи выставляется – 0 баллов.

Шкала оценки образовательных достижений

Процент результативности (правильных ответов)	Оценка уровня подготовки	
	балл (отметка)	вербальный аналог
90 ÷ 100	5	отлично
80 ÷ 89	4	хорошо
70 ÷ 79	3	удовлетворительно
менее 70	2	неудовлетворительно

Расчетное задание

Вариант 1-10

Алфавит передаваемых сообщений состоит из независимых букв S_i . Вероятности появления каждой буквы в сообщении заданы. Определить и сравнить эффективность кодирования сообщений методом Хаффмана при побуквенном кодировании и при кодировании блоками по две буквы.

№	$p(S_i)$	№	$p(S_i)$
1	(0,6;0,2;0,08;0,12)	6	(0,7;0,2;0,06;0,04)
2	(0,7;0,1;0,07;0,13)	7	(0,6;0,3;0,08;0,02)
3	(0,8;0,1;0,07;0,03)	8	(0,5;0,2;0,11;0,19)
4	(0,5;0,3;0,04;0,16)	9	(0,5;0,4;0,08;0,02)
5	(0,6;0,2;0,05;0,15)	10	(0,7;0,2;0,06;0,04)

Время на подготовку и выполнение: 30 мин.

За верное решение задачи выставляется положительная оценка – 5 баллов.

За неверное решение задачи выставляется – 0 баллов.

Расчетное задание

1. Сжать наиболее рациональным способом следующие информационные массивы:

3	7	0	0	0	0	0	0	1	9	2
3	7	0	0	0	0	0	0	1	9	3
3	7	0	0	0	0	0	0	1	9	4
3	7	0	0	0	0	0	0	1	9	5
3	7	0	0	0	0	0	0	1	9	6

1	3	7	0	0	0	0	0	0	1	9	2
2	3	7	0	0	0	0	0	0	1	9	3
3	3	7	0	0	0	0	0	0	1	9	4
4	3	7	0	0	0	0	0	0	1	9	5
5	3	7	0	0	0	0	0	0	1	9	6

1	3	7	0	0	0	0	0	1	9	2	7	3	7	0	0	0	0	0
2	3	7	0	0	0	0	0	1	9	3	8	3	7	0	0	0	0	0
3	3	7	0	0	0	0	0	1	9	4	9	3	7	0	0	0	0	0
4	3	7	0	0	0	0	0	1	9	5	1	3	7	0	0	0	0	0
5	3	7	0	0	0	0	0	1	9	6	2	3	7	0	0	0	0	0

2. Определить адрес числа 315 в памяти ИПС, где расположено 600 чисел, максимальное из которых 9800.

3. Слово «пролетариат» закодировано в коде Бодо (приложение 3). Сжать это слово, применив метод поразрядного сдвига (длина сжатого слова специально не оговаривается).

4. Чему равно максимальное число разрядов кодов, сжатых методом поразрядного сдвига, если допустимое количество адресов ЗУ равно 2048.

5. Определить необходимый объем ЗУ для размещения словаря на 500 слов, представленных в семизначном двоичном коде, если допустимая длина кодируемого слова $L_{\text{макс}} = 16$.

6. Восстановить исходный массив чисел по следующему ниже сжатому массиву:

2	4	6	8	1	3	5	7
p	7	p	p	2	p	p	p
1	p						

7. Сжать проведенные ниже массив, используя знак раздела и знак конца строки К:

6	3	1	8	1	2	7
---	---	---	---	---	---	---

6	3	1	8	1	8	6
2	1	1	8	1	2	4
3	1	1	8	1	2	9
4	1	1	8	1	2	9
4	1	1	8	1	2	9
5	1	1	8	1	2	9

8. Показать процесс восстановления исходного массива по следующему сжатому массиву:

1	6	4	3	6	1	8	1	1	3	2	5	7	7
у	2	z	x	x	1	4	у	1	z	x	x	6	у
2	у	x	0	z	x	x	9						

9. В памяти ЭВМ требуется разместить 1000 чисел, максимальное из которых 439. Найти оптимальное значение объема памяти для размещения в ней этих чисел, сжатых методом Г.В.Лавинского.

10. На произвольном примере показать процесс определения местонахождения в памяти ЭВМ чисел, сжатых методом Г.В. Лавинского.

Время на подготовку и выполнение: 60 мин.

За правильный ответ на вопросы или верное решение задачи выставляется положительная оценка – 1 балл.

За не правильный ответ на вопросы или неверное решение задачи выставляется – 0 баллов.

Практическое занятие №1 Способы хранения, обработки и передачи информации

Практическое занятие № 2 Измерение количества информации. Носители информации.

Практическое занятие № 3 Скорость передачи информации

Практическое занятие № 4 Поиск энтропии случайных величин

Практическое занятие № 5 Применение теоремы отчетов

Практическое занятие № 6 Выполнение расчетов по теореме отчетов. Определение пропускной способности дискретного канала

Практическое занятие № 7 Расчет вероятностей

Практическое занятие № 8 Составление закона распределения вероятностей.

Практическое занятие № 9 Системные требования алгоритмов сжатия. Алгоритмы сжатия данных неизвестного формата

Практическое занятие № 10 Практическое применение различных алгоритмов сжатия

Лабораторная работа №1 Сравнение и анализ архиваторов

Практическое занятие № 11 Итоговое практическое занятие

Практическое занятие № 12 Практическое применение алгоритмов кодирования

Практическое занятие № 12 Кодирование информации

Практическое занятие № 13 Кодирование информации

Практическое занятие № 14 Декодирование информации

Практическое занятие № 15 Практическое применение криптографии

Практическое занятие №1

Тема программы: **Формальное представление знаний. Виды информации.**

Тема: **Способы хранения, обработки и передачи информации.**

Цель: научиться сохранять, обрабатывать и передавать данные при помощи технических средств информации.

Время выполнения: 2 часа

Оборудование: ПК, сканер, фотоаппарат, микрофон, диск.

Раздаточный материал: инструкционно – технологическая карта для сканирования фотографии.

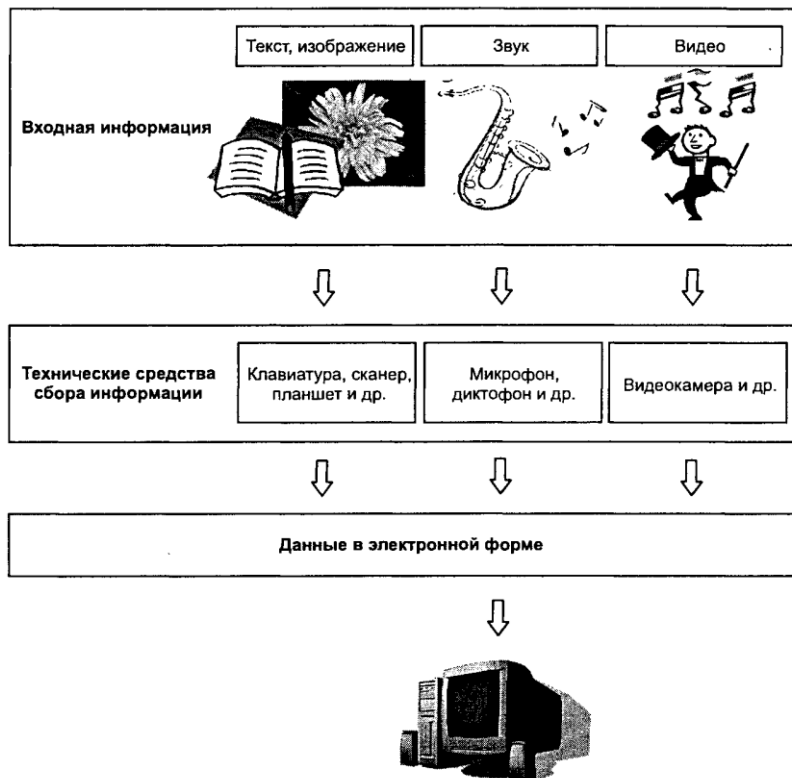
Программное обеспечение: операционная система, программа для работы с видеоинформацией.

Теоретические основы

1. Технологии сбора и хранения информации

Сбор предполагает получение максимально выверенной исходной информации и является одним из самых ответственных этапов в работе с информацией, поскольку от цели сбора и методов последующей обработки полностью зависит конечный результат работы всей информационной системы.

Технология сбора подразумевает использование определенных методов сбора информации и технических средств, выбираемых в зависимости от вида информации и применяемых методов ее сбора. На заключительном этапе сбора, когда информация преобразуется в данные, т. е. в информацию, представленную в формализованном виде, пригодном для компьютерной обработки, осуществляется ее ввод в систему. Для сбора данных необходимо сначала определить технические средства, позволяющие осуществлять сбор быстро и высококачественно и поддерживающие операции ввода информации и представления данных в электронной форме.



2. Технологический процесс обработки информации

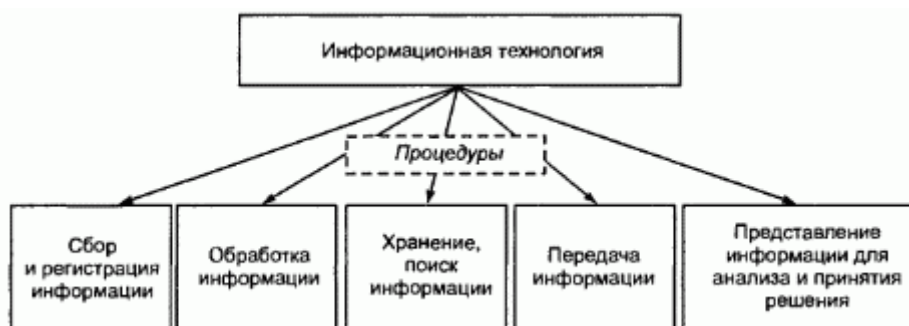
Технологический процесс обработки информации — есть строго определенная последовательность взаимосвязанных процедур, выполняемых для преобразования

первичной информации с момента ее возникновения до получения требуемого результата.

Технологический процесс призван автоматизировать обработку исходной информации за счет привлечения технических средств базовой информационной технологии, сократить финансовые и трудовые затраты, обеспечить высокую степень достоверности результатной информации. Для конкретной задачи той или иной предметной области технологический процесс обработки информации разрабатывается индивидуально. Совокупность процедур зависит от следующих факторов:

- характер и сложность решаемой задачи;
- алгоритм преобразования информации;
- используемые технические средства;
- сроки обработки данных;
- используемые системы контроля;
- число пользователей и т. д.

В общем случае технологический процесс обработки информации включает процедуры.



3. Способы обработки информации

Современные информационные технологии позволяют обрабатывать информацию централизованным и децентрализованным (т. е. распределенным) способами.

Централизованный способ предполагает сосредоточение данных в информационно-вычислительном центре, выполняющем все основные действия технологического процесса обработки информации. Основное достоинство централизованного способа — сравнительная дешевизна обработки больших объемов информации за счет повышения загрузки вычислительных средств.

Децентрализованный способ характеризуется рассредоточением информационно-вычислительных ресурсов и распределением технологического процесса обработки информации по местам возникновения и потребления информации. Достоинством децентрализованного способа является повышение оперативности обработки информации и решения поставленных задач за счет автоматизации деятельности на конкретных рабочих местах, применения надежных средств передачи информации, организации сбора первичных документов и ввода исходных данных в местах их возникновения [6].

Децентрализованный способ обработки информации может быть реализован автономным или сетевым методом. При автономной обработке информации передача документов и данных на электронных носителях осуществляется по почте либо курьером, а при сетевой — через современные каналы связи.

На практике применяют смешанный способ обработки информации, для которого

характерны признаки двух способов одновременно (централизованный с частичной децентрализацией или децентрализованный с частичной централизацией).

4. Режимы обработки информации на компьютере

Вычислительные средства участвуют в процессе обработки информации в двух основных режимах: пакетном или диалоговом.

В случае, когда технология обработки информации на компьютере представляет собой заранее определенную последовательность операций, не требующую вмешательства человека, и диалог с пользователем отсутствует, информация обрабатывается в так называемом пакетном режиме. Суть его состоит в том, что программы обработки данных последовательно выполняются под управлением операционной системы как совокупность (пакет) заданий. Операционная система обеспечивает ввод данных, вызов требуемых программ, включение необходимых внешних устройств, координацию и управление технологическим процессом обработки информации.

Задачи, решаемые в пакетном режиме, характеризуются следующими свойствами:

- алгоритм решения задачи формализован, вмешательства пользователя не требуется;
- наличие большого объема входных и выходных данных, в основном хранящихся на устройствах хранения информации (например, жестких дисках компьютеров);
- расчет выполняется для большинства записей входных файлов;
- длительное время решения задачи — как правило, обусловлено большими объемами обрабатываемых данных;
- регламентность — задачи решаются с заданной периодичностью.

Пакетный режим возник первым и широко использовался с середины XX в., когда обработка информации на ЭВМ осуществлялась в специально создаваемых вычислительных центрах. Заказчики подготавливали исходные данные (обычно на перфокартах или перфолентах) и отправляли их в вычислительный центр, где данные обрабатывались и результаты обработки возвращались заказчику. С развитием персональных ЭВМ (начиная с 80-х гг. прошлого века) обработка данных стала осуществляться, в основном, непосредственно потребителями, поэтому в настоящее время пакетный режим используется достаточно редко. Сегодня более распространен диалоговый режим, когда необходимо непосредственное взаимодействие пользователя с компьютером и на каждое свое действие пользователь получает немедленные ответные действия компьютера. Диалоговый режим позволяет пользователю интерактивно управлять порядком обработки информации и получать результатные данные в виде необходимых документов либо файлов.

5. Технологии передачи и представления информации

Информационные процессы невозможны без средств передачи и представления информации, поскольку зачастую информация требуется в месте, территориально удаленном от источника ее возникновения, и должна быть представлена в виде символов, образов и сигналов, пригодных для восприятия потребителем.

Современные средства связи способны передавать информацию в любой форме: телефонные, телевизионные, телеграфные сообщения, массивы данных, печатные материалы, фотографии и т. д. В соответствии со спецификой передаваемых

сообщений организуется канал передачи информации — совокупность технических средств, обеспечивающих передачу сигналов от источника к потребителю.

Основная характеристика канала передачи — скорость передачи информации, а ее предельно допустимое значение называют емкостью канала, которая ограничивается шириной полосы канала и шумом.

Канал связи соединяет передатчик и приемник с помощью линии связи, которая может быть проводной, кабельной, радио, микроволновой, оптической или спутниковой. Примерами линий связи являются телефонные и вычислительные сети, сети телевизионного и радиовещания, мобильной связи, спутниковые технологии передачи данных.

В современных цифровых системах связи функции передатчика и приемника выполняет модем. Основное достоинство передачи информации в цифровой форме заключается в возможности использования кодированных сигналов, обеспечения защиты информации и наилучшего способа приема.

Для представления переданной или хранимой информации потребителю используются процессы воспроизведения и отображения.

Воспроизведение информации — это процесс, при котором ранее записанная на носителе информация считывается устройством воспроизведения.

Отображение информации — есть представление информации, т. е. генерация сигналов на основе исходных данных, а также правил и алгоритмов их преобразования в форме, приемлемой для непосредственного восприятия человеком.

Потребителем информации наиболее часто выступает человек, и для принятия решений ему необходимы результаты обработки информации. Тем не менее человек не способен ощутить машинное представление информации, а может воспринимать ее лишь органами чувств (органами зрения, слуха, осязания, обоняния и т. д.), поэтому для организации взаимодействия человека с информационными моделями объектов информационная система должна быть наделена специальными средствами отображения данных.

Поскольку зрение используется для восприятия информации наиболее активно, то средства отображения в современных ИС должны представлять информацию в лучшей форме для визуального наблюдения. Заметим, что мультимедиа-системы позволяют также представлять информацию в форме аудио- и видеосигналов, однако для управленческих информационных систем наиболее характерно отображение информации в текстовой и графической форме, что осуществляется за счет использования мониторов и печатающих устройств (например, принтеров, плоттеров).

Прежде чем, например, на мониторе, появится информация в доступной для человека форме, компьютером будет автоматически осуществлена следующая последовательность операций:

- преобразование данных, представленных в машинной форме, в вид, приемлемый для экранного отображения;
- согласование формы представления данных с параметрами монитора;
- воспроизведение в соответствии с возможностями воспроизводящего устройства (т. е. в данном примере — монитора).

Порядок выполнения работы

1. Создание досье группы. Заранее заготовить материал: фотографии, текст.
2. Сфотографировать своих однокурсников.
3. Включить компьютер.
4. Создать общую папку на сервере.
5. Сканировать фотографии и сохранить в общую папку.
6. Включить текстовый редактор. Создать титульный лист с общей фотографией и названием группы: специальность и год.
7. Оформить каждый лист на одного человека. Записать данные: дата рождения, номер школы, хобби.
8. Сохранить данные на жесткий диск в свою папку под именем досье группы.

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Что такое сбор информации и каково его предназначение?
2. Что понимается под технологией сбора информации?
3. Чем отличаются понятия «информация» и «данные»?
4. Назовите основные требования к сбору данных и к хранимым данным.
5. Перечислите основные средства сбора текстовой, графической, звуковой и видеoinформации. Какие еще средства сбора информации вам известны?
6. Какие еще методы сбора данных вам известны?
7. В чем заключается процедура хранения информации?
8. Перечислите основные требования к структурам хранения.
9. Что такое база данных?
10. В чем различие между базой и банком данных?
11. Что такое резервное копирование и для чего оно осуществляется?
12. Что такое архивное копирование и в чем его отличие от резервного копирования?
13. Что такое базовая информационная технология?
14. В чем заключается различие между централизованным и децентрализованным способами обработки информации?
15. Какие режимы обработки информации вам известны?

Инструкционно-технологическая карта для сканирования фотографий

Перед началом работы положите фотографию на стекло сканера, лицом вниз. Для того чтобы отсканировать фотографию воспользуйтесь «Мастером работы со сканером» или используйте программное обеспечение поставляемое производителем вашего сканера. Мы рассмотрим пример сканирования фотографии посредством «Мастера работы со сканером».

Для запуска мастера войдите в меню Пуск → Настройки → Панель управления → Сканеры и камеры → Выберите модель вашего сканера и нажмите дважды. В

появившемся окне (см. рис.1) выберите «Особые параметры» и нажмите кнопку «Настроить»

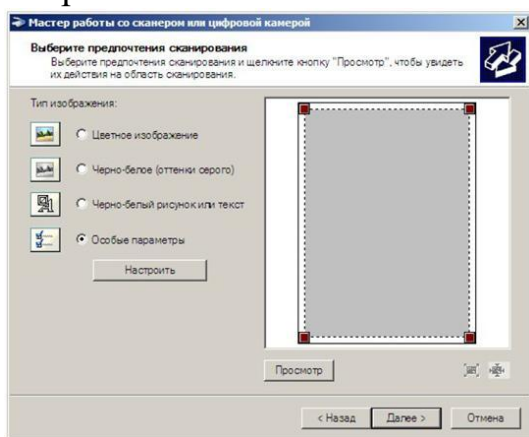


Рисунок 1

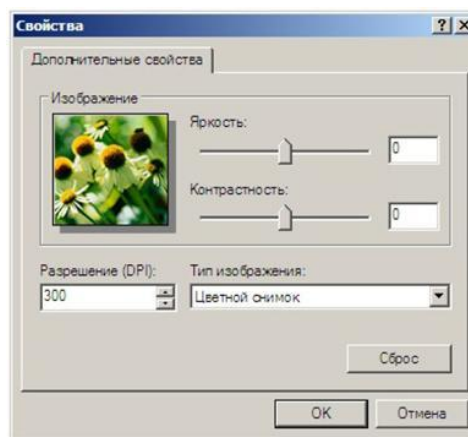


Рисунок 2

В появившемся окне (см. рис. 2) выберите «Разрешение (DPI)» - 300, «Тип изображения» - Цветной снимок и нажмите кнопку ОК, вы попадете в предыдущее меню в котором необходимо нажать кнопку «Просмотр» (см. рис. 1) У вас на экране появится предварительный просмотр изображения с вашего сканера (см.рис. 3)

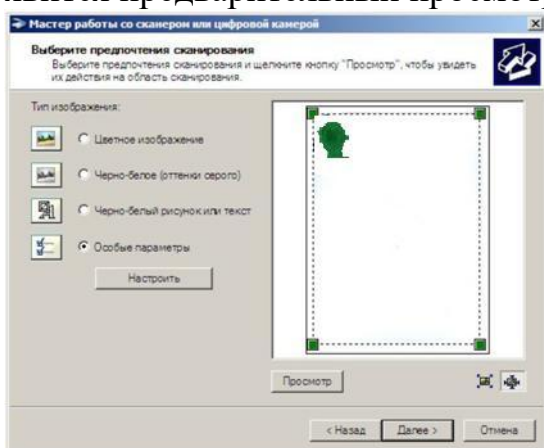


Рисунок 3

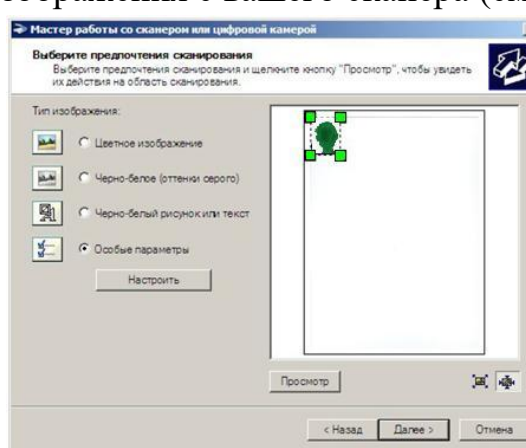


Рисунок 4

Для того чтобы отсканировать фотографию правильно необходимо выбрать область сканирования (выделена пунктирной линией) как показано на рис. 4, в противном случае сканер кроме фотографии туриста отсканирует также и всю плоскость стекла, в результате чего в визе туриста вместо фотографии будет напечатан белый лист, такая виза будет считаться недействительной!

После того как вы выбрали область сканирования нажмите кнопку «Далее», в появившемся окне укажите название файла, выберите формат файла JPG и укажите папку для сохранения фотографии, после чего нажмите кнопку «Далее».

Практическое занятие № 2

Тема программы: **Способы измерения информации.**

Тема: **Измерение количества информации. Носители информации.**

Цель: научиться измерять и вычислять информацию, а также и работать с носителями информации.

Время выполнения: 2 часа

Оборудование: ПК.

Раздаточный материал: дидактический материал

Программное обеспечение: операционная система, текстовый редактор.

Теоретические основы

1 байт = 8 битов.

Существуют еще более крупные единицы измерения информации.

Переход к более крупным единицам измерения информации (килобайт, мегабайт, терабайт, петабайт, эксабайт). Байт наиболее удобная единица измерения информационного объема сообщения, состоящего из последовательности символов компьютерного алфавита. Однако она мала при подсчете емкости информационных носителей. По аналогии с физическими единицами измерения (например, 1 килограмм = 1000 грамм) подбираем по таблице целых степеней двойки значение близкое к тысяче. Это значение равно 1024. Поэтому 1 килобайт = 1024 байт = 2^{10} байт, 1 мегабайт = 1024 килобайт = 2^{10} килобайт и т. д.

1 Кбайт (один килобайт) = 1024 байт;

1 Мбайт (один мегабайт) = 1024 Кбайт;

1 Гбайт (один гигабайт) = 1024 Мбайт.

Что представляют единицы измерения:

5 бит – буква в клетке кроссворда.

1 байт – символ, введенный с клавиатуры.

6 байт – средний размер слова, в тексте на русском языке.

50 байт – строка текста.

2 Кбайта – страница машинописного текста.

100 Кбайт – фотография в низком разрешении

1 Мбайт – небольшая художественная книга.

100 Мбайт – метровая книга с полками.

1 Гбайт – прочитывает человек за всю жизнь.

3 Гбайт – час качественной видеозаписи.

Для сохранения информации используют носители информации: флэш-накопители, флэш-карты, диски разных форматов, съемные жесткие диски.

Порядок выполнения работы

Изучение материала и выполнение заданий на компьютере.

Содержательный (вероятностный) подход к определению количества информации

Если заключённые в каком-то сообщении сведения являются для человека новыми, понятными, пополняют его знания, т.е. приводят к уменьшению неопределённости знаний, то сообщение содержит информацию.

1 бит – количество информации, которое содержится в сообщении, которое уменьшает неопределённость знаний в 2 раза.

Пример1. При бросании монеты возможны 2 события (случая) – монета упадёт орлом или решкой, причём оба события равновероятны (при большом количестве бросаний количество случаев падения монеты орлом и решкой одинаковы). После получения сообщения о результате падения монеты неопределённость знаний уменьшилась в 2 раза, и, поэтому, количество информации, полученное при этом равно 1 бит.

Содержательный (вероятностный) подход является субъективным, т.к. одну и ту же информацию разные люди могут оценивать по разному. Для одного человека сведения в сообщении могут быть важными и понятными, для другого бесполезными, непонятными или вредными.

Единицы измерения информации. Перевод единиц измерения.

1 бит – количество информации, которое содержится в сообщении, которое уменьшает неопределённость знаний в 2 раза.

1 бит – наименьшая единица информации. Более крупные единицы – байт, килобайт, мегабайт, гигабайт.

Система единиц измерения информации:

1 байт = 8 бит

1 Кбайт = 2¹⁰ байт = 1024 байт;

1 Мбайт = 2¹⁰ Кбайт = 1024 Кбайт = 2²⁰ байт;

1 Гбайт = 2¹⁰ Мбайт = 1024 Мбайт = 2³⁰ байт

Информационный объём носителей информации:

Дискета – 1,44 Мбайт; компакт-диск ≈ 700 Мбайт; DVD-диск – до 17 Гбайт (стандарт – 4,7 Гбайт); жёсткий диск – от 20 Гбайт до 80 Гбайт и более (стандарт 80 Гбайт); Flash-память – 256 Мбайт – 2 Гбайт.

Примеры перевода единиц:

1) 5 байт = 5 * 8 бит = 40 бит;

2) 24 бита = 24 : 8 байта = 3 байта;

3) 4 Кбайт = 4 * 1024 байт = 4096 байт;

4) 16384 бита = 16384 : 8 байт = 2048 байт;

1) 2048 байт : 1024 = 2 Кбайта.

Вычисление количества информации для равновероятных событий.

Если события равновероятны, то количество информации можно рассчитать по формуле:

$$N = 2^I,$$

где N – число возможных событий,

I – количество информации в битах.

Формула была предложена американским инженером Р. Хартли в 1928 г.

Задача 1. В коробке 32 карандаша, все карандаши разного цвета. Наугад вытащили красный. Какое количество информации при этом было получено?

Решение.

Так как вытаскивание карандаша любого цвета из имеющихся в коробке 32 карандашей является равновероятным, то число возможных событий равно 32.

$$N = 32, I = ?$$

$$N = 2^I, 32 = 2^5, I = 5 \text{ бит.}$$

Ответ: 5 бит.

Задачи на перевод единиц измерения информации

***Задача 1.* В школьной библиотеке 16 стеллажей с книгами, на каждом – по 8 полок. Ученику сообщили, что нужный учебник находится на 2-ой полке 4-го стеллажа. Какое количество информации получил ученик?

Решение.

1) Число стеллажей (случаев) – 16.

$$N_1 = 16, N_1 = 2^{I_1}, 16 = 2^4, I_1 = 4 \text{ бита.}$$

2) Число полок на каждом стеллаже (случаев) – 8,

$$N_2 = 8, N_2 = 2^{I_2}, 8 = 2^3, I_2 = 3 \text{ бит.}$$

3) $I = I_1 + I_2, I = 4 \text{ бита} + 3 \text{ бита} = 7 \text{ бит.}$

Ответ: 7 бит.

***Задача 3.** Загадывают число в диапазоне от 1 до 200. Какое наименьшее количество вопросов надо задать, чтобы наверняка отгадать число. На вопросы можно отвечать только «Да» или «Нет».

Решение.

Правильная стратегия состоит в том, чтобы количество вариантов каждый раз уменьшалось вдвое.

Например, загадано число 152.

1 вопрос: Число >100? Да.

2 вопрос: Число < 150? Нет.

3 вопрос: Число > 175? Нет. и т.д.

Количество событий в каждом варианте будет одинаково, и их отгадывание равновероятно. $N = 2I$, $200 = 2I$, $7 < I < 8$. Т.к. количество вопросов нецелым числом быть не может, то необходимо задать не более 8 вопросов.

Ответ: 8 вопросов

****Вычисление количества информации для событий с различными вероятностями.**

Существует множество ситуаций, когда возможные события имеют различные вероятности реализации. Рассмотрим примеры таких событий.

1. В коробке 20 карандашей, из них 15 красных и 5 чёрных. Вероятность вытащить наугад красный карандаш больше, чем чёрный.

2. При случайном падении бутерброда вероятность падения его маслом вниз (более тяжёлой стороной) больше, чем маслом вверх.

3. В пруду живут 8000 карасей, 2000 щук и 40000 пескарей. Самая большая вероятность для рыбака – поймать в этом пруду пескаря, на втором месте – карася, на третьем – щуку.

Количество информации в сообщении о некотором событии зависит от его вероятности. Чем меньше вероятность события, тем больше информации оно несёт. $P = K/N$, где K – количество случаев реализации одного из исходов события, N – общее число возможных исходов одного из событий $I = -\log_2(1/p)$, где I – количество информации, p – вероятность события

Задача. В коробке 50 шаров, из них 40 белых и 10 чёрных. Определить количество информации в сообщении о вытаскивании наугад белого шара и чёрного шара.

Решение.

Вероятность вытаскивания белого шара

$$P_1 = 40/50 = 0,8$$

Вероятность вытаскивания чёрного шара $P_2 = 10/50 = 0,2$

Количество информации о вытаскивании белого шара $I_1 = \log_2(1/0,8) = \log_2 1,25 = \log 1,25 / \log 2 \approx 0,32$ бит

Количество информации о вытаскивании чёрного шара

$$I_2 = \log_2(1/0,2) = \log_2 5 = \log 5 / \log 2 \approx 2,32$$
 бит

Ответ: 0,32 бит, 2,32 бит

Что такое логарифм?

Логарифмом числа a по основанию b называется показатель степени, в которую надо возвести число a , чтобы получить число b .

$$a^{\log_a b} = b, a > 0, b > 0, a \neq 1$$

Вычисление логарифмов чисел по основанию 2 с помощью электронного калькулятора

$\log_2 6 = \log 6 / \log 2$, где $\log 6$ и $\log 2$ – десятичные логарифмы

Программа вычисления логарифма числа 6 по основанию 2 ($\log_2 6$) с помощью инженерного калькулятора: 6, log, /, 2, log, =

Количество информации в случае различных вероятностей событий определяется по формуле:

Формула Шеннона: (американский учёный, 1948 г.)

$$I = - \sum_{i=1}^N p_i \log_2 p_i$$

где P_i – вероятность i -го события, N – количество возможных событий

Задача. В озере живут караси и окуни. Подсчитано, что карасей 1500, а окуней – 500. Сколько информации содержится в

сообщениях о том, что рыбак поймал карася, окуня, поймал рыбу?

Решение.

События поимки карася или окуня не являются равновероятными, так как окуней в озере меньше, чем карасей.

Общее количество карасей и окуней в пруду $1500 + 500 = 2000$.

Вероятность попадания на удочку карася

$p_1 = 1500/2000 = 0,75$, окуня $p_2 = 500/2000 = 0,25$.

$I_1 = \log_2(1/p_1)$, $I_2 = \log_2(1/p_2)$, где I_1 и I_2 – вероятности поймать карася и окуня соответственно.

$I_1 = \log_2(1 / 0,75) \approx 0,43$ бит, $I_2 = \log_2(1 / 0,25) \approx 2$ бит – количество информации в сообщении поймать карася и поймать окуня соответственно.

Количество информации в сообщении поймать рыбу (карася или окуня) рассчитывается по формуле Шеннона

$$I = - p_1 \log_2 p_1 - p_2 \log_2 p_2$$

$$I = - 0,75 * \log_2 0,75 - 0,25 * \log_2 0,25 = - 0,75 * (\log 0,75 / \log 2) - 0,25 * (\log 0,25 / \log 2) = 0,604 \text{ бит} \approx 0,6 \text{ бит.}$$

Ответ: в сообщении содержится 0,6 бит информации

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Какое количество информации несет в себе жесткий диск емкостью 4 терабайта, если производитель рассчитывает 1000 за 1024?

2. Чем отличается вероятностный подход к измерению информации от алфавитного?

3. Какие единицы измерения информации используют для флэш-накопителей?

Практическое занятие № 3

Тема программы: **Способы измерения информации.**

Тема: **Скорость передачи информации.**

Цель: научиться измерять и вычислять скорость передачи информации.

Время выполнения: 2 часа

Оборудование: ПК.

Раздаточный материал: дидактический материал

Программное обеспечение: операционная система, текстовый редактор.

Теоретические основы

Скорость передачи информации определяется количеством элементов двоичной информации, передаваемых за 1 с. В синхронной передаче данных по коммутируемым каналам рекомендуется использовать скорости из следующего ряда: 600, 1200, 2400, 4800, 9600 бит/с. Для асинхронной передачи по коммутируемым каналам рекомендуется скорость 300 бит/с.

В синхронной передаче данных по арендованным каналам рекомендуется использовать скорости из следующих рядов:

а) предпочтительные: 600, 1200, 2400, 4800, 9600, 14400 бит/с;

б) дополнительные: 3000, 6000, 7200, 12000 бит/с;

в) допустимого диапазона: $N \times 600$ бит/с ($1 \leq N \leq 24$)

Следует отличать скорость передачи информации от модуляционной (линейной) скорости, измеряемой в Бодах (количество элементов модулированного сигнала, передаваемого за 1 с). Для простых видов модуляции скорость передачи информации совпадает с модуляционной скоростью.

При синхронной передаче скорость должна отличаться от номинального значения не более, чем на 0,01%, а при асинхронной – не более, чем на 2,5%.

Оба компьютера, как правило, могут одновременно обмениваться информацией в обе стороны. Этот режим работы называется полным дуплексным.

Дуплексный режим передачи данных – режим, при котором передача данных осуществляется одновременно в обоих направлениях.

В отличие от дуплексного режима передачи данных, полудуплексный подразумевает передачу в каждый момент времени только в одном направлении.

Кроме собственно модуляции и демодуляции сигналов модемы могут выполнять сжатие и декомпрессию пересылаемой информации, а также заниматься поиском и исправлением ошибок, возникнувших в процессе передачи данных по линиям связи.

Одной из основных характеристик модема является скорость модуляции (modulation speed), которая определяет физическую скорость передачи данных без учета исправления ошибок и сжатия данных. Единицей измерения этого параметра является количество бит в секунду (бит/с), называемое бодом.

Любой канал связи имеет ограниченную пропускную способность (скорость передачи информации), это число ограничивается свойствами аппаратуры и самой линии (кабеля).

Объем переданной информации вычисляется по формуле $Q=q \cdot t$, где q – пропускная способность канала (в битах в секунду), а t – время передачи

Пример 1. Скорость передачи данных через ADSL-соединение равна 128000 бит/с. Через данное соединение передают файл размером 625 кбайт. Определить

время передачи файла в секундах.

Решение:

1) выделим в заданных больших числах степени двойки и переведем размер файла в биты, чтобы «согласовать» единиц измерения:

$$128000 \text{ бит/с} = 128 \cdot 1000 \text{ бит/с} = 27 \cdot 125 \cdot 8 \text{ бит/с} = 27 \cdot 53 \cdot 23 \text{ бит/с} = 210 \cdot 53 \text{ бит/с}$$

$$625 \text{ кбайт} = 54 \text{ кбайт} = 54 \cdot 213 \text{ бит.}$$

2) чтобы найти время передачи в секундах, нужно разделить размер файла на скорость передачи:

$$t = (54 \cdot 213) \text{ бит} / 210 \cdot 53 \text{ бит/с} = 40 \text{ с.}$$

Ответ: 40 с .

Пример 2. Скорость передачи данных через ADSL-соединение равна 512000 бит/с. Передача файла через это соединение заняла 1 минуту. Определить размер файла в килобайтах.

Решение:

1) выделим в заданных больших числах степени двойки; переведем время в секунды (чтобы «согласовать» единицы измерения), а скорость передачи – в кбайты/с, поскольку ответ нужно получить в кбайтах:

$$1 \text{ мин} = 60 \text{ с} = 4 \cdot 15 \text{ с} = 22 \cdot 15 \text{ с}$$

$$512000 \text{ бит/с} = 512 \cdot 1000 \text{ бит/с} = 29 \cdot 125 \cdot 8 \text{ бит/с} = 29 \cdot 53 \cdot 23 \text{ бит/с} = 212 \cdot 53 \text{ бит/с} = 29 \cdot 53 \text{ бит/с} = (29 \cdot 53) / 210 \text{ кбайт/с} = (53 / 2) \text{ кбайт/с}$$

2) чтобы найти объем файла, нужно умножить время передачи на скорость передачи:

$$Q = q \cdot t = 22 \cdot 15 \text{ с} \cdot (53 / 2) \text{ кбайт/с} = 3750 \text{ кбайт}$$

Ответ: 3750 кбайт.

Пример 3. С помощью модема установлена связь с другим компьютером со скоростью соединения 19200, с коррекцией ошибок и сжатием данных.

а) Можно ли при таком соединении файл размером 2,6 килобайт передать за 1 секунду? Обоснуйте свой ответ.

б) Всегда ли при таком соединении файл размером 2,3 килобайт будет передаваться за 1 секунду? Обоснуйте свой ответ.

в) Можно ли при таком соединении оценить время передачи файла размером 4 Мб? Если можно, то каким образом?

Решение:

а) Для начала узнаем, какое количество килобайт мы можем передать за 1 секунду: $19200 / 1024 / 8 = 2,3$ (Кбайт). Следовательно, если бы не было сжатия информации, то данный файл за одну секунду при данной скорости соединения было бы невозможно передать. Но сжатие есть, $2.6 / 2.3 < 4$, следовательно, передача возможна.

б) Нет не всегда, так как скорость соединения это максимально возможная скорость передачи данных при этом соединении. Реальная скорость может быть меньше.

в) Можно указать минимальное время передачи этого файла: $4 \cdot 1024 \cdot 1024 / 4 / 19200$, около 55 с (столько времени будет передаваться файл на указанной скорости с максимальной компрессией). Максимальное же время передачи оценить вообще говоря нельзя, так как в любой момент может произойти обрыв связи.

Практические задания

Задание 1. Решите задачу о передаче информации с помощью модема.

Вариант 1	Скорость передачи данных через ADSL-соединение равна 512000 бит/с. Через данное соединение передают файл размером 1500 Кб. Определите время передачи файла в секундах.
Вариант 2	Скорость передачи данных через ADSL-соединение равна 1024000 бит/с. Через данное соединение передают файл размером 2500 Кб. Определите время передачи файла в секундах.
Вариант 3	Скорость передачи данных через ADSL-соединение равна 1024000 бит/с. Передача файла через данное соединение заняла 5 секунд. Определите размер файла в килобайтах.
Вариант 4	Скорость передачи данных через ADSL-соединение равна 512000 бит/с. Передача файла через данное соединение заняла 8 секунд. Определите размер файла в килобайтах.

Задание 2. Решите задачу о передаче графической информации.

Вариант 1	Определите скорость работы модема, если за 256 с он может передать растровое изображение размером 640x480 пикселей. На каждый пиксель приходится 3 байта.
Вариант 2	Сколько секунд потребуется модему, передающему информацию со скоростью 56 000 бит/с, чтобы передать цветное растровое изображение размером 640 x 480 пикселей, при условии, что цвет каждого пикселя кодируется тремя байтами?
Вариант 3	Определите скорость работы модема, если за 132 с он может передать растровое изображение размером 640x480 пикселей. На каждый пиксель приходится 3 байта.
Вариант 4	Сколько секунд потребуется модему, передающему информацию со скоростью 28800 бит/с, чтобы передать цветное растровое изображение размером 640 x 480 пикселей, при условии, что цвет каждого пикселя кодируется тремя байтами?

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Дайте определение пропускной способности (передача информации)?
2. В чем измеряется пропускная способность?
3. Зависит ли скорость передачи информации от объема передаваемой информации?

Практическое занятие № 4

Тема программы: **Способы измерения информации.**

Тема: Поиск энтропии случайных величин.

Цель: научиться вычислять энтропию случайной величины.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, калькулятор, текстовый редактор.

Теоретические основы

Энтропия в теории информации — мера хаотичности информации, неопределённость появления какого-либо символа первичного алфавита. При отсутствии информационных потерь численно равна количеству информации на символ передаваемого сообщения.

Так, возьмём, например, последовательность символов, составляющих какое-либо предложение на русском языке. Каждый символ появляется с разной частотой, следовательно, неопределённость появления для некоторых символов больше, чем для других. Если же учесть, что некоторые сочетания символов встречаются очень редко, то неопределённость ещё более уменьшается (в этом случае говорят об энтропии n -ого порядка. Концепции информации и энтропии имеют глубокие связи друг с другом, но, несмотря на это, разработка теорий в статистической механике и теории информации заняла много лет, чтобы сделать их соответствующими друг другу.

Энтропия независимых случайных событий x с n возможными состояниями (от 1 до n) рассчитывается по формуле:

$$H(x) = - \sum_{i=1}^n p(i) \log_2 p(i)$$

Эта величина также называется *средней энтропией сообщения*.

Величина $\log_2 \frac{1}{p(i)}$ называется *частной энтропией*, характеризующей только i -е состояние.

Таким образом, энтропия события x является суммой с противоположным знаком всех произведений относительных частот появления события i , умноженных на их же двоичные логарифмы (основание 2 выбрано только для удобства работы с информацией, представленной в двоичной форме). Это определение для дискретных случайных событий можно расширить для функции распределения вероятностей.

Шеннон вывел это определение энтропии из следующих предположений:

- мера должна быть непрерывной; т. е. изменение значения величины вероятности на малую величину должно вызывать малое результирующее изменение энтропии;
- в случае, когда все варианты (буквы в приведенном примере) равновероятны, увеличение количества вариантов (букв) должно всегда увеличивать полную энтропию;
- должна быть возможность сделать выбор (в нашем примере букв) в два шага, в которых энтропия конечного результата должна будет являться суммой энтропий промежуточных результатов.

Шеннон показал, что любое определение энтропии, удовлетворяющее этим предположениям, должно быть в форме:

$$-K \sum_{i=1}^n p(i) \log_2 p(i)$$

где K — константа (и в действительности нужна только для выбора единиц измерения).

Шеннон определил, что измерение энтропии ($H = -p_1 \log_2 p_1 - \dots - p_n \log_2 p_n$), применяемое к источнику информации, может определить требования к минимальной пропускной способности канала, требуемой для надежной передачи информации в виде закодированных двоичных чисел. Для вывода формулы Шеннона необходимо вычислить математическое ожидания «количества информации», содержащегося в цифре из источника информации. Мера энтропии Шеннона выражает неуверенность реализации случайной переменной. Таким образом, энтропия является разницей между информацией, содержащейся в сообщении, и той частью информации, которая точно известна (или хорошо предсказуема) в сообщении. Примером этого является избыточность языка — имеются явные статистические закономерности в появлении букв, пар последовательных букв, троек и т.д.

В общем случае **b-арная энтропия** (где b равно 2,3,...) источника $\mathcal{S} = (S,P)$ с исходным алфавитом $S = \{a_1, \dots, a_n\}$ и дискретным распределением вероятности $P = \{p_1, \dots, p_n\}$ где p_i является вероятностью a_i ($p_i = p(a_i)$) определяется формулой:

$$H_b(\mathcal{S}) = - \sum_{i=1}^n p_i \log_b p_i$$

Определение энтропии Шеннона очень связано с понятием термодинамической энтропии. Больцман и Гиббс проделали большую работу по статистической термодинамике, которая способствовала принятию слова «энтропия» в информационную теорию. Существует связь между понятиями энтропии в термодинамике и теории информации. Например, демон Максвелла также противопоставляет термодинамическую энтропию информации, и получение какого-либо количества информации равно потерянной энтропии.

СВОЙСТВА ЭНТРОПИИ

1. Энтропия является вещественной и неотрицательной величиной.
2. Энтропия – величина ограниченная.
3. Энтропия обращается в нуль лишь в том случае, если вероятность одного из состояний равна единице; тогда вероятности всех остальных состояний, естественно, равны нулю. Это положение соответствует случаю, когда состояние источника полностью определено.
4. Энтропия максимальна, когда все состояния источника равновероятны.
5. Энтропия источника и с двумя состояниями u_1 и u_2 изменяется от нуля до единицы, достигая максимума при равенстве их вероятностей:
 $p(u_1) = p(u_2) = 1 - p = 0,5$.
6. Энтропия объединения нескольких статистически независимых источников информации равна сумме энтропии исходных источников.
7. Энтропия характеризует среднюю неопределенность выбора одного состояния из ансамбля. При ее определении используют только вероятности состояний, полностью игнорируя их содержательную сторону. Поэтому энтропия не может

служить средством решения любых задач, связанных с неопределенностью.

8. Энтропия как мера неопределенности согласуется с экспериментальными данными, полученными при изучении психологических реакций человека, в частности реакции выбора. Установлено, что время безошибочной реакции на последовательность беспорядочно чередующихся равновероятных раздражителей (например, зажигающихся лампочек) растет с увеличением их числа так же, как энтропия. Это время характеризует неопределенность выбора одного раздражителя. Замена равновероятных раздражителей неравновероятными приводит к снижению среднего времени реакции ровно настолько, насколько уменьшается энтропия.

Дифференциальной энтропией случайной величины X называется величина:

$$H_D(x) = H(x) - H(y) = - \int_{-\infty}^{+\infty} p_x(x) * \log_2 d * p_x(x) dx$$

Если произвести квантование случайных величин $X_1, X_2 \dots X_n$ по уровню с числом уровней квантования равным m , то возможное число реализаций длительностью T_n станет конечным и равным $M = m^n$.

Каждая из реализаций $C_1, C_2, \dots, C_i, \dots, C_m$ будет иметь определенную вероятность появления в эксперименте по наблюдению реализаций. Тогда неопределенность (энтропия) и количество информации в реализации (в среднем по всем реализациям) определяются равенством

$$H_n = - \sum_{i=1}^M P(C_i) \log(P(C_i))$$

$$H = \frac{H_n}{n} = - \frac{1}{n} \sum_{i=1}^M P(C_i) \log(P(C_i))$$

Энтропия и количество информации на одну степень свободы (на одну выборку) равны

$$H = \frac{H_n}{n} = - \frac{1}{n} \sum_{i=1}^M P(C_i) \log(P(C_i))$$

Избыточность показывает, какая доля максимально возможной при заданном объеме алфавита неопределенности не используется источником.

$$\mu = (H_{\max} - H_u) / H_{\max},$$

Где H_u – энтропия рассматриваемого источника, H_{\max} – максимально возможное значение его энтропии, которое может быть достигнуто подбором распределения и ликвидацией взаимозависимости элементов алфавита. Так, для дискретного источника с M элементами

$$H_{\max} = \log M$$

Выполнение расчетных задач

Задача №1

Доказать, что $H(x, y) \leq H(x) + H(y)$.

Решение:

По определению,
$$H(x, y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log p(x_i, y_j).$$

Для статистически зависимых событий $p(x_i, y_j) = p(x_i) p(y_j / x_i)$.

$$\begin{aligned}
H(x, y) &= -\sum_{i=1}^n \sum_{j=1}^m p(x_i) p(y_j / x_i) \log [p(x_i) p(y_j / x_i)] = \\
&= -\sum_{i=1}^n \sum_{j=1}^m p(x_i) p(y_j / x_i) [\log p(x_i) + \log p(y_j / x_i)] = \\
&= -\sum_{i=1}^n p(x_i) \log p(x_i) \sum_{j=1}^m p(y_j / x_i) + \sum_{i=1}^n p(x_i) \left(-\sum_{j=1}^m p(y_j / x_i) \log p(y_j / x_i) \right) = \\
&= -\sum_{i=1}^n p(x_i) \log p(x_i) + \sum_{i=1}^n p(x_i) H(x, y / x_i) = H(x) + H(y / x).
\end{aligned}$$

$H(y / x_i)$ - это частная энтропия Y при условии, что известно состояние $X=x_i$.

Наличие информации о состоянии X не может увеличить неопределенность состояния Y , но может уменьшить его в случае зависимости Y от X . Значит, условная энтропия $H(y / x_i)$ не больше безусловной энтропии $H(y)$, то есть $H(y / x_i) \leq H(y)$. Тогда средняя условная энтропия

$$H(y / x) = \sum_{i=1}^n p(x_i) H(y / x_i) \leq \sum_{i=1}^n p(x_i) H(y) = H(y) \sum_{i=1}^n p(x_i) = H(y),$$

то есть $H(y / x) \leq H(y)$.

Значит, $H(x, y) = H(x) + H(y / x) \leq H(x) + H(y)$.

Задача №2

Показать, что для регулярной марковской цепи энтропия $H(x)^{(r)}$ за r шагов равняется энтропии за один шаг, умноженной на число шагов r .

Решение:

Регулярная цепь Маркова полностью характеризуется матрицей переходных

вероятностей $P = \begin{pmatrix} p_{11} & \dots & p_{1m} \\ \cdot & \cdot & \cdot \\ p_{m1} & & p_{mm} \end{pmatrix}$ и предельным стационарным распределением

вероятностей состояний (p_1, p_2, \dots, p_m) .

В стационарном режиме энтропия за один шаг не зависит от номера шага и равна

$$H(X)^{(1)} = \sum_{k=1}^m p_k H_k(x),$$

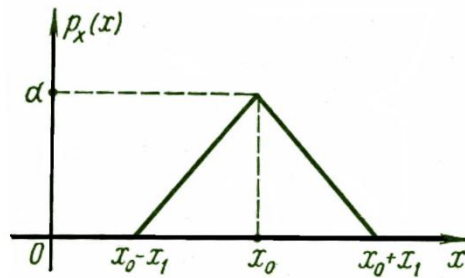
p_k - стационарная вероятность k -го состояния,

$H_k(x) = -\sum_{i=1}^m p_{ki} \log p_{ki}$ - энтропия в k -м состоянии.

Энтропия за r шагов равна сумме энтропий за каждый шаг. Так как энтропия за каждый шаг одинакова, то сумма энтропий равна $H(X)^{(r)} = r \cdot H(X)^{(1)}$.

Задача №3

Вычислить энтропию $H_d(x)$ распределения $p_x(x)$ изображенного на рисунке. Построить график зависимости $H_d(x)$ в функции параметра d .



Ответ: **4.36.** Энтропия треугольного распределения равна $-\ln d + 1/2$.

Решение:

Из условия нормировки $S = d \cdot x_1 = 1 \Rightarrow x_1 = \frac{1}{d}$.

Плотность данного распределения

$$p_X(x) = \begin{cases} d + \frac{d(x-x_0)}{x_1}, & x_0 - x_1 \leq x \leq x_0 \\ d - \frac{d(x-x_0)}{x_1}, & x_0 \leq x \leq x_0 + x_1 \end{cases}$$

Дифференциальная энтропия

$$H_d(x) = - \int_{-\infty}^{\infty} p_X(x) \log p_X(x) dx = - \left(\int_{x_0-x_1}^{x_0} p_X(x) \ln p_X(x) dx + \int_{x_0}^{x_0+x_1} p_X(x) \ln p_X(x) dx \right).$$

$$\int u \ln u du = \ln u \cdot \frac{u^2}{2} - \int (\ln u)' \frac{u^2}{2} du = -\frac{u^2}{4} + \frac{1}{2} u^2 \ln u + C = \frac{1}{2} u^2 \left(\ln u - \frac{1}{2} \right) + C.$$

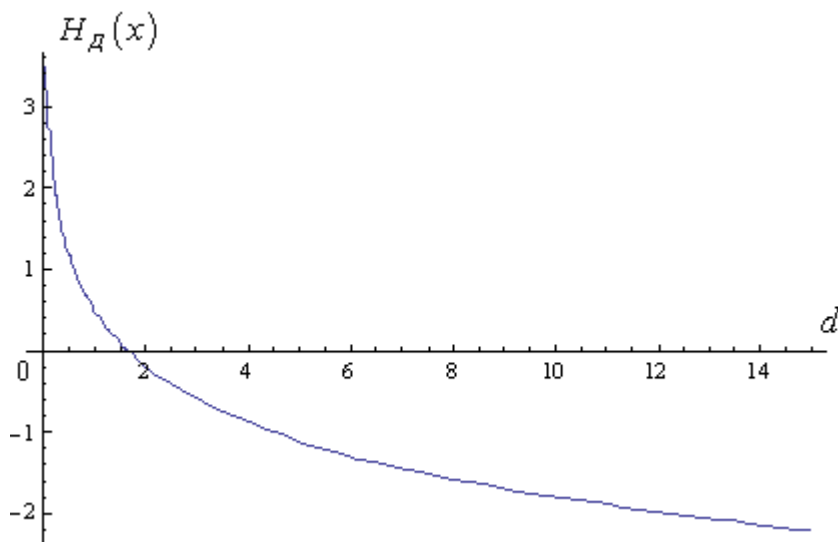
$$\int_{x_0-x_1}^{x_0} \left[d + \frac{d}{x_1}(x-x_0) \right] \ln \left[d + \frac{d}{x_1}(x-x_0) \right] dx =$$

$$= \frac{x_1}{d} \left(\frac{1}{2} \left[d + \frac{d}{x_1}(x-x_0) \right]^2 \left(\ln \left[d + \frac{d}{x_1}(x-x_0) \right] - \frac{1}{2} \right) \right) \Big|_{x_0-x_1}^{x_0} = \frac{x_1}{d} \cdot \frac{1}{2} d^2 \left(\ln d - \frac{1}{2} \right).$$

$$\int_{x_0}^{x_0+x_1} \left[d - \frac{d}{x_1}(x-x_0) \right] \ln \left[d - \frac{d}{x_1}(x-x_0) \right] dx =$$

$$= -\frac{x_1}{d} \left(\frac{1}{2} \left[d - \frac{d}{x_1}(x-x_0) \right]^2 \left(\ln \left[d - \frac{d}{x_1}(x-x_0) \right] - \frac{1}{2} \right) \right) \Big|_{x_0}^{x_0+x_1} = \frac{x_1}{d} \cdot \frac{1}{2} d^2 \left(\ln d - \frac{1}{2} \right).$$

$$H_d(x) = -2 \cdot \frac{x_1}{d} \cdot \frac{1}{2} d^2 \left(\ln d - \frac{1}{2} \right) = -\frac{x_1}{d} d^2 \left(\ln d - \frac{1}{2} \right) = -\frac{1}{d^2} d^2 \left(\ln d - \frac{1}{2} \right) = -\ln d + \frac{1}{2}.$$



Задача 4

В результате полной дезорганизации управления m самолетов летят произвольными курсами. Управление восстановлено, и все самолеты взяли общий курс со среднеквадратической ошибкой отклонения от курса $\sigma=3^\circ$. Найти изменение энтропии, считая, что в первом случае имело место равномерное распределение вероятностей углов, а во втором случае – нормальное.

Ответ: 4.86 бита

Решение.

Начальное распределение вероятностей углов курсов самолетов равномерное в интервале от $a = 0$ до $b = 360^\circ = 2\pi \text{ рад}$ с плотностью вероятности

$$p_{1X}(x) = \frac{1}{b-a}, \quad a \leq x \leq b.$$

Дифференциальная энтропия этого распределения

$$H_{1D}(x) = -\int_a^b \frac{1}{b-a} \log_2 \frac{1}{b-a} dx = -\log_2 \frac{1}{b-a} = \log_2(b-a) = \log_2 2\pi = 2,65 \text{ бит.}$$

Конечное распределение вероятностей углов курсов самолетов нормальное с параметрами $a = 0$, $\sigma = 3^\circ = \frac{\pi}{60} \text{ рад}$ и плотностью вероятности

$$p_{2X}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)}.$$

Дифференциальная энтропия этого распределения

$$\begin{aligned} H_{2D}(x) &= -\int_{-\infty}^{\infty} p_{2X}(x) \log_2 \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)} \right] dx = \\ &= -\log_2 \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} p_{2X}(x) dx + \frac{\log_2 e}{2\sigma^2} \int_{-\infty}^{\infty} x^2 p_{2X}(x) dx = \\ &= -\log_2 \frac{1}{\sigma\sqrt{2\pi}} + \frac{\log_2 e}{2\sigma^2} \cdot \sigma^2 = \log_2 \sigma\sqrt{2\pi} + \frac{\log_2 e}{2} = \log_2 \frac{\pi\sqrt{2\pi e}}{60} = -2,21 \text{ бит.} \end{aligned}$$

Изменение энтропии $\Delta H(x) = H_{2D}(x) - H_{1D}(x) = -2,21 - 2,65 = -4,86 \text{ бит.}$

Энтропия уменьшилась на 4,86 бит.

Задача 5

Измерительное устройство вырабатывает временные интервалы, распределенные случайным образом в пределах от 100 до 500 мс. Как изменится энтропия случайной величины при изменении точности измерения с 1 мс до 1 мкс?

Ответ: Энтропия увеличивается примерно на 10 бит

Решение.

При точности 1 мс дискретная случайная величина X – результат измерения – может равновероятно принимать одно из $n = \frac{500-100}{1} = 400$ значений. Энтропия равна $H_1(x) = \log_2 n$.

При точности 1 мкс дискретная случайная величина X – результат измерения – может равновероятно принимать одно из $m = \frac{500-100}{10^{-3}} = 400 \cdot 10^3 = 1000n$ значений.

Энтропия равна $H_2(x) = \log_2 m$.

Изменение энтропии

$\Delta H(x) = H_2(x) - H_1(x) = \log_2 m - \log_2 n = \log_2 1000n - \log_2 n = \log_2 1000 \approx \log_2 1024 = 10$ бит.

Энтропия увеличилась примерно на 10 бит.

Задача 6

Записать отношения между энтропиями:

$H(x)$, $H(y)$, $H(x|y)$, $H(y|x)$, $H(x,y)$, $H(x|y_j)$, $H(y|x_i)$

Решение.

Связь между энтропией совместного распределения и полными энтропиями и средними условными энтропиями

$$H(x, y) = H(x) + H(y/x) = H(y) + H(x/y).$$

Если x и y независимы, то $H(y/x) = H(y)$, $H(x/y) = H(x)$ и энтропия совместного распределения $H(x, y) = H(x) + H(y)$ максимальна, так как

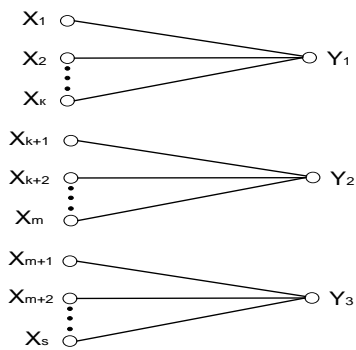
$$H(y/x) \leq H(y), \quad H(x/y) \leq H(x).$$

Связь между частными и средними условными энтропиями

$$H(y/x) = \sum_{i=1}^n p(x_i) H(y/x_i), \quad H(x/y) = \sum_{j=1}^m p(y_j) H(x/y_j).$$

Задача 7

На рисунке представлена диаграмма канала со слабым разрешением. Определить количество информации, передаваемое по каналу.



Ответ: $I(x,y)=H(x)$

Решение:

Количество переданной по каналу информации равно $I = H - H_{аносм.}$.

$H = H(y)$ – энтропия полученных элементов y до отправления элемента x ;

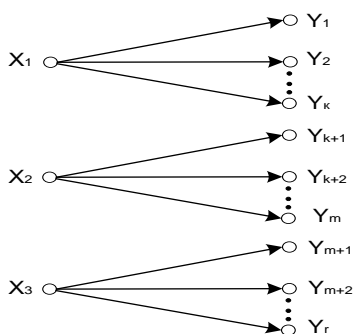
$H_{аносм.} = H(y/x)$ - энтропия полученных элементов y после отправления элемента x .

Для данного канала со слабым разрешением $H(y/x) = 0$, так как полученный элемент y однозначно определяется по отправленному элементу x .

Получаем $I = H(y)$.

Задача 2

На рисунке представлена диаграмма канала с неоднозначностью. Определить количество информации, передаваемое по каналу.



ответ: $I(x,y)=H(x)$

Решение:

Количество переданной по каналу информации равно $I = H - H_{аносм.}$.

$H = H(x)$ – энтропия отправленных элементов x до получения элемента y ;

$H_{аносм.} = H(x/y)$ - энтропия отправленных элементов x после получения элемента y .

Для данного канала с неоднозначностью $H(x/y) = 0$, так как полученный элемент y однозначно определяет отправленный элемент x .

Получаем $I = H(x)$.

Задачи по вычислению энтропии

1. Найдите энтропию для числа белых шаров при извлечении двух шаров из урны,

содержащей два белых и один черный шар.

2. Найдите энтропию для числа козырных карт при извлечении двух карт из колоды в 36 карт.

3. Какую степень неопределенности содержит опыт угадывания суммы очков на извлеченной кости из полного набора домино?

4. Найдите энтропию для числа тузов при извлечении трех карт из карт с картинками.

5. Найдите дифференциальную энтропию для равномерного распределения.

6. Найдите дифференциальную энтропию для показательного закона распределения, если известно, что случайная величина x принимает значение меньше единицы с вероятностью 0,5.

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Как определяется энтропия дискретных случайных величин?
2. Приведите примеры энтропий для классических законов распределения.

Практическое занятие № 5

Тема программы: **Теорема отчетов.**

Тема: **Применение теоремы отчетов.**

Цель: Изучение возможности синтеза сигналов по дискретным отсчетам в соответствии с теоремой Котельникова.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, калькулятор, текстовый редактор.

Теоретические основы

Теорема Котельникова

В 1933 году В.А. Котельниковым доказана теорема отсчетов [6, 32], имеющая важное значение в теории связи: непрерывный сигнал $s(t)$ с ограниченным спектром можно точно восстановить (интерполировать) по его отсчетам $s(k\Delta t)$, взятым через интервалы $\Delta t = \frac{1}{(2F)}$, где F – верхняя частота спектра сигнала.

В соответствии с этой теоремой сигнал $s(t)$ можно представить рядом Котельникова

$$s(t) = \sum_{k=-\infty}^{\infty} s\left(\frac{k}{2F}\right) \frac{\sin 2\pi F \left[t - \frac{k}{2F} \right]}{2\pi F \left[t - \frac{k}{2F} \right]} \quad (1.21)$$

Таким образом, сигнал $s(t)$, можно абсолютно точно представить с помощью последовательности отсчетов $s\left(\frac{k}{2F}\right)$, заданных в дискретных точках $\frac{k}{2F}$ (рис.1.16).

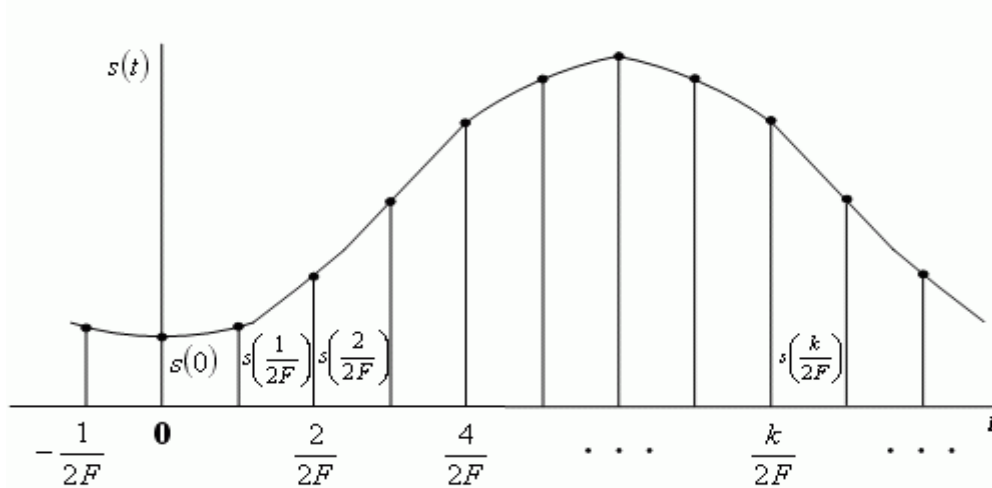


Рис. 1.16. Сигнал и его отсчеты

Функции

$$\psi(t) = \frac{\sin 2\pi F \left[t - \frac{k}{2F} \right]}{2\pi F \left[t - \frac{k}{2F} \right]} \quad (1.22)$$

образуют ортогональный базис в пространстве сигналов, характеризующихся ограниченным спектром:

$$\Phi(f) = 0 \text{ при } |f| > F. \quad (1.23)$$

Обычно для реальных сигналов можно указать диапазон частот, в пределах которого сосредоточена основная часть его энергии и которым определяется ширина спектра сигнала. В ряде случаев спектр сознательно сокращают. Это обусловлено тем, что аппаратура и линия связи должны иметь минимальную полосу частот. Сокращение спектра выполняют, исходя из допустимых искажений сигнала. Например, при телефонной связи хорошая разборчивость речи и узнаваемость абонента обеспечиваются при передаче сигналов в полосе частот $\Delta F = 0,3 \dots 3,4 \text{ [кГц]}$. Увеличение ΔF приводит к неоправданному усложнению аппаратуры и повышению затрат. Для передачи телевизионного изображения при стандарте в 625 строк полоса частот, занимаемая сигналом, составляет около 6 МГц.

Из вышесказанного следует, что процессы с ограниченными спектрами могут служить адекватными математическими моделями многих реальных сигналов.

Функция вида $\frac{\sin 2\pi F \left[t - \frac{k}{2F} \right]}{2\pi F \left[t - \frac{k}{2F} \right]}$ называется функцией отсчетов (рис.1.17).

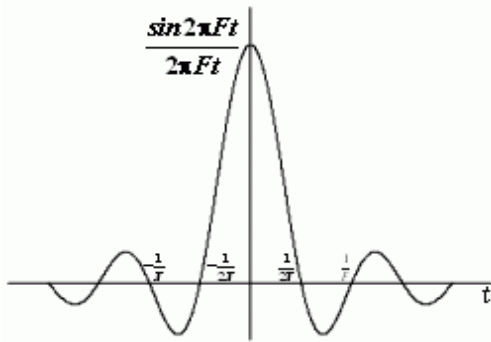


Рис. 1.17. Функция отсчетов

Она характеризуется следующими свойствами. Если $k = 0$, функция отсчетов имеет максимальное значение при $t = 0$, а в моменты

времени $t = \frac{i}{2F}$ ($i = 1, 2, \dots$) она обращается в нуль; ширина главного лепестка функции отсчетов на

нулевом уровне равна $\frac{1}{F}$, поэтому минимальная длительность импульса, который может существовать на выходе линейной системы с

полосой пропускания F , равна $\frac{1}{F}$; функции отсчетов ортогональны на бесконечном интервале времени.

На основании теоремы Котельникова может быть предложен следующий способ дискретной передачи непрерывных сигналов:

Для передачи непрерывного сигнала $s(t)$ по каналу связи с полосой пропускания F определим мгновенные значения сигнала $s(t)$ в дискретные моменты времени $t_k = \frac{k}{2F}$, ($k = 0, 1, 2, \dots$). После этого передадим эти значения по каналу связи каким-либо из возможных способов и восстановим на приемной стороне переданные отсчеты. Для преобразования потока импульсных отсчетов в непрерывную функцию пропустим их через идеальный ФНЧ с граничной частотой F .

Можно показать, что энергия сигнала находится по формуле [6, 32]:

$$E = \int_{-\infty}^{\infty} s^2(t) dt = \frac{1}{2F} \sum_{k=-\infty}^{\infty} s^2\left(\frac{k}{2F}\right) \quad (1.24)$$

Для сигнала, ограниченного во времени, выражение (1.24) преобразуется к виду:

$$E = \int_1^{2FT} s^2(t) dt = \frac{1}{2F} \sum_{k=1}^{2FT} s^2\left(\frac{k}{2F}\right) \quad (1.25)$$

Выражение (1.25) широко применяется в теории помехоустойчивого приема сигналов, но является приближенным, т.к. сигналы не могут быть одновременно ограничены по частоте и времени.

Практическое задание

1. Изобразить сигналы, синтезируемые в лабораторной работе:
 - а) синусоидальный сигнал частотой 5 кГц;
 - б) видеоимпульсы прямоугольной формы длительностью 0,25; 0,5; 1,0 мс;
 - в) видеоимпульсы пилообразной формы длительностью 0,5 мс; 1,0 мс.
2. Рассчитать и построить идеальные выборочные сигналы для сигналов, указанных в п. 1а, 1б, 1в, при $f_{\text{выб}} = 5, 10, 20, 40$ кГц.

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;

- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Сформулируйте теорему Котельникова для сигналов с ограниченным спектром.
2. Объясните погрешности синтезирования реальных сигналов по дискретным отсчетам.

Практическое занятие № 6

Тема программы: Теорема отчетов.

Тема: Выполнение расчетов по теореме отчетов. Определение пропускной способности дискретного канала.

Цель: научиться выполнять расчеты по теореме отчетов и определять пропускную способность дискретного канала.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, калькулятор, текстовый редактор.

Теоретические основы

Пусть на вход аналогово-цифрового преобразователя поступает гармонический сигнал с частотой f (период $T = 1/f$). частоты исходного сигнала

Проведем дискретизацию входного аналогового сигнала с периодом дискретизации T_d меньшим половины периода входного сигнала T (рисунок 1).

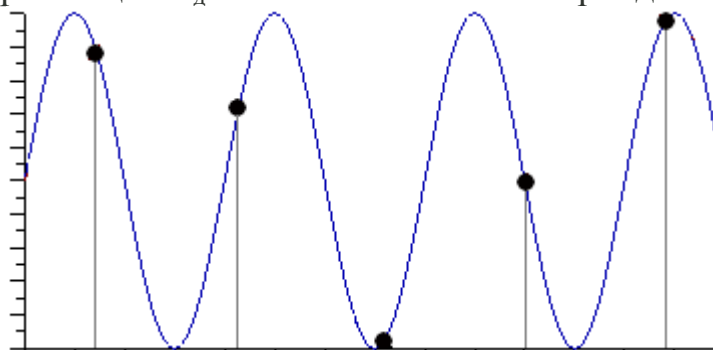


Рисунок 1

Очевидно, что дискретные отсчеты сигнала однозначно не отображают форму исходного сигнала, в частности по получившимся точкам можно построить гармонический сигнал с периодом $T_{искаж.}$, отличающимся от периода исходного сигнала T . Период $T_{искаж.}$ больше периода исходного сигнала T , соответственно частота меньше, частоты исходного сигнала f (рисунок 2).

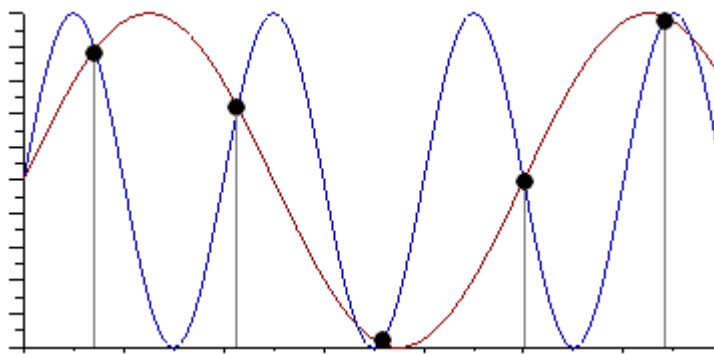


Рисунок 2

Данный эффект называется стробоскопическим эффектом или алиасингом. Он заключается в появлении ложной низкочастотной составляющей при дискретизации сигнала с частотой меньшей удвоенной частоты исходного сигнала (или с периодом большим половины периода исходного сигнала), отсутствующей в исходном сигнале.

Пример 2

Уменьшим период дискретизации до половины периода исходного аналогового сигнала (частоту дискретизации увеличим до удвоенной частоты исходного сигнала). В данной ситуации возникает неопределенность начальной фазы и амплитуды сигнала, при этом частота исходного сигнала не искажается. В крайнем случае мы можем получить отсчеты сигнала равные нулю (рисунок 3).

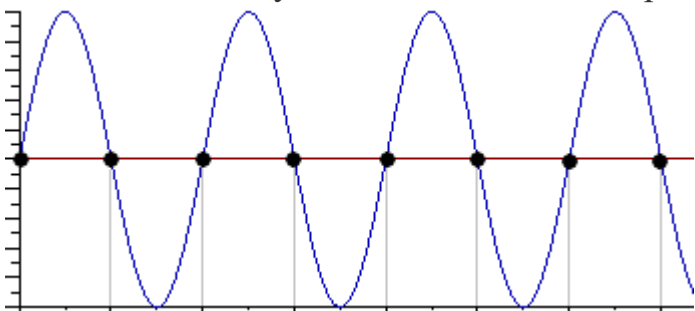


Рисунок 3

Пример 3

Продолжим уменьшение периода дискретизации. Если период дискретизации меньше половины периода исходного сигнала, то очевидно, что через получившиеся после оцифровки точки можно построить только один гармонический сигнал, соответствующий исходному, без искажения начальной фазы, амплитуды и частоты (рисунок 4). Данное утверждение теоретически обосновано и мы его примем без доказательства.

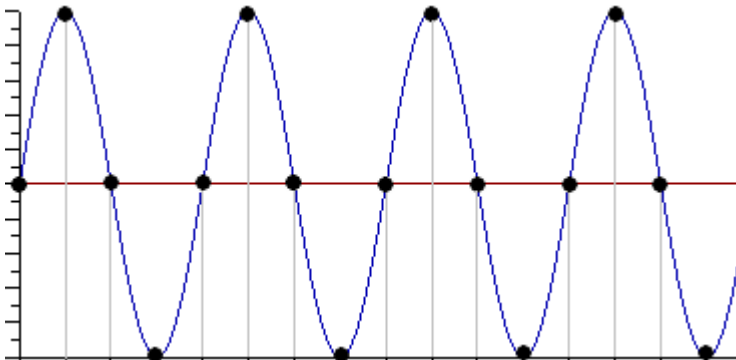


Рисунок 4

Таким образом, для адекватного восстановления гармонического сигнала по дискретным отсчетам, частота дискретизации должна быть не меньше половины частоты сигнала. Частота равная половине частоты дискретизации называется частотой Найквиста $f_N = f_d/2$.

Данное утверждение можно обобщить следующим образом:

Аналоговый сигнал с ограниченным спектром может быть восстановлен однозначно и без искажений по своим дискретным отсчетам, взятым с частотой большей удвоенной максимальной частоты в своем спектре.

$$f_d > 2 \cdot F_{max} \quad (1)$$

Данное утверждение известно как **теорема Котельникова** (в западной литературе **теорема Найквиста-Шеннона**) или теорема отсчетов. В различных источниках в формулировке данной теоремы могут быть различия, основным из которых является знак сравнения в формуле 1: $f_d \geq 2 \cdot F_{max}$ или $f_d > 2 \cdot F_{max}$. Мы придерживаемся формулировки со знаком **строго больше**, так как при частоте оцифровки равной максимальной частоте в спектре возникают неоднозначности начальной фазы и амплитуды.

На практике аналоговый сигнал, как правило, оцифровывают с частотой в несколько раз превышающей удвоенную частоту в спектре сигнала, хотя существуют методики оцифровки сигнала с нарушением теоремы отсчетов.

Пропускная способность непрерывного канала

Пусть сигнал $y(t)$ на выходе канала представляет собой сумму полезного сигнала $x(t)$ и шума $n(t)$, т.е. $y(t) = x(t) + n(t)$, причем $x(t)$ и $n(t)$ статистически независимы. Допустим, что канал имеет ограниченную полосу пропускания шириной $\Delta F_{\text{нх}}$. Тогда в соответствии с теоремой Котельникова (см. п. 1.5) функции $y(t)$, $x(t)$ и $n(t)$ можно представить совокупностями отсчетов y_i , x_i , и n_i , $i = 1, 2, \dots, L$, где $L = 2\Delta F_{\text{нх}}T$. При этом статистические свойства сигнала $x(t)$ можно описать многомерной ПРВ $w(x_1, x_2, \dots, x_L) = w(x)$, а свойства шума – ПРВ $w(n_1, n_2, \dots, n_L) = w(n)$.

Пропускная способность непрерывного канала определяется следующим образом:

$$C = \lim_{T \rightarrow \infty} \frac{1}{T} \max_{w(x)} I(X, Y)$$

где $I(X, Y)$ – количество информации о какой-либо реализации сигнала $x(t)$ длительности T , которое в среднем содержит реализация сигнала $y(t)$ той же длительности T , а максимум ищется по всем возможным распределениям $w(x)$.

Когда сигнал на входе канала имеет нормальное распределение и отсчеты независимы величина $h(X)$ максимизируется [6]. Поэтому пропускная способность гауссовского канала с дискретным временем, рассчитанная на единицу времени, с учетом (4.16) может быть записана в виде

$$C = V_{\text{нх}} \cdot I(Y, X) = \frac{V_{\text{нх}}}{2} \log_2 \left(\frac{\sigma_c^2 + \sigma^2}{\sigma^2} \right) = \frac{V_{\text{нх}}}{2} \log_2 (1 + h^2) \quad (4.17)$$

Полученное выражение показывает, что пропускная способность гауссовского

канала с дискретным временем определяется числом импульсов, передаваемых в секунду, и отношением сигнал/шум (h^2).

С учетом взаимосвязи скорости передачи информации и полосы частот непрерывного канала от (4.17) можно перейти к формуле Шеннона, которая устанавливает связь пропускной способности гауссовского канала с полосой пропускания непрерывного канала и отношением мощности сигнала к мощности помехи:

$$C = \Delta F_{\text{НХ}} \log_2(1 + h^2). \quad (4.18)$$

График отношения $\frac{C}{\Delta F_{\text{НХ}}} = \log_2(1 + h^2)$ изображен на рис. 4.6. Заметим, что при малом отношении $h^2 \ll 1$

$$C \cong \Delta F_{\text{НХ}} \cdot 1,442 \cdot h^2,$$

а пропускная способность канала связи прямо пропорциональна этому отношению.

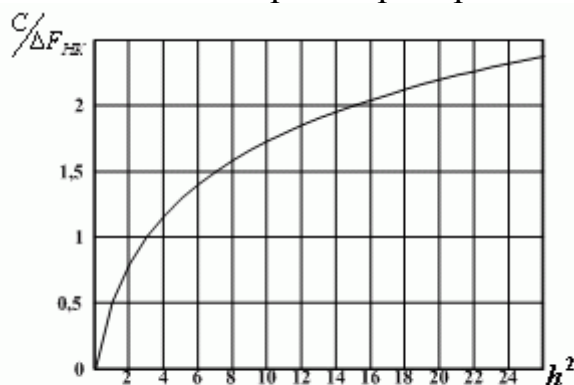


Рис. 4.6. Пропускная способность непрерывного канала

При большом отношении $h^2 \gg 1$ в (4.18) можно пренебречь единицей и считать, что

$$\frac{C}{\Delta F_{\text{НХ}}} \approx \log_2(h^2),$$

т.е. зависимость пропускной способности непрерывного канала от отношения сигнал/шум логарифмическая.

Пропускная способность канала, как предельное значение скорости безошибочной передачи информации, является одной из основных характеристик любого канала.

Определим пропускную способность стандартного канала тональной частоты, имеющего границы эффективно передаваемых частот 0,3...3,4 кГц, среднюю мощность сигнала на выходе 56 мкВт при средней мощности помехи 69000 пВт.

Согласно (4.18), при заданных параметрах

$$C_{\text{НХ}} = 3,1 \cdot 10^3 \cdot \log_2\left(\frac{56 \cdot 10^{-6}}{69 \cdot 10^{-12}}\right) = 3,0 \cdot 10^4 \quad [\text{бит/с}].$$

Для непрерывных каналов справедлива теорема Шеннона, согласно которой сообщения дискретного источника могут быть закодированы и переданы по непрерывному каналу так, что вероятность ошибочного декодирования принятого сигнала $\nu(\epsilon)$ будет меньше наперед заданной положительной величины $P_{\text{ош}}^*$, если производительность источника $H'(X)$ меньше пропускной способности C непрерывного канала.

Для типовых непрерывных каналов многоканальной связи основные технические характеристики и пропускная способность, вычисленная по формуле Шеннона (4.18), при отношении сигнал/шум 20 дБ, приведены в табл. 4.4.

Зная пропускную способность канала и информационные характеристики сообщений (табл. 4.5), можно определить, какие сообщения (первичные сигналы) можно передавать по заданному каналу.

Таблица 4.4. Характеристики типовых каналов многоканальной связи

Наименование канала	Границы передаваемых частот, Гц	Пропускная способность, бит/с
Тональной частоты	300...3400	$20,64 \cdot 10^3$
Предгрупповой широкополосный	$12,3 \cdot 10^3 \dots 23,4 \cdot 10^3$	$73,91 \cdot 10^3$
Первичный широкополосный	$60,6 \cdot 10^3 \dots 107,7 \cdot 10^3$	$313,6 \cdot 10^3$
Вторичный широкополосный	$312,3 \cdot 10^3 \dots 551,4 \cdot 10^3$	$1,59 \cdot 10^6$
Третичный широкополосный	$812,3 \cdot 10^3 \dots 2043,7 \cdot 10^3$	$8,2 \cdot 10^6$

Таблица 4.5. Производительность источников сообщений

Вид сообщения	Характер сообщения	Параметры АЦП		Производительность, бит/с
		f_d , Гц	$N = \log_2 L$	
Телеграфные, 50 Бод	дискретные	–	–	30...50
Телефонные	непрерывные	$8 \cdot 10^3$	8	$64 \cdot 10^3$
Звукового вещания: первого класса высшего класса	непрерывные	$24 \cdot 10^3$	13	$240 \cdot 10^3$
	непрерывные	$32 \cdot 10^3$		$416 \cdot 10^3$
Факсимильные, 120 строк/с: полутонные штриховые	непрерывные	$2,93 \cdot 10^3$	4	$11,72 \cdot 10^3$
	дискретные	–	–	$2,93 \cdot 10^3$
Передача данных, 2400 Бод	дискретные	–	–	$2,4 \cdot 10^3$
Телевизионные	непрерывные	$13 \cdot 10^6$	16	$208 \cdot 10^6$

Например, первичный сигнал телевизионного вещания имеет $H'(X) = 208 \cdot 10^6 \text{ бит/с}$ (табл. 4.5) и поэтому не может быть передан ни по одному из типовых непрерывных или цифровых каналов без потери качества. Следовательно, для передачи сигнала телевизионного вещания требуется создание специальных каналов с более высокой пропускной способностью или снижение скорости цифрового потока.

Выполнение расчетных заданий

1. Число символов алфавита $m = 4$. Вероятности появления символов равны соответственно $p_1 = 0,15$; $p_2 = 0,4$; $p_3 = 0,25$; $p_4 = 0,2$. Длительности символов $t_1 = 3\text{с}$; $t_2 = 2\text{с}$; $t_3 = 5\text{с}$, $t_4 = 6\text{с}$. Чему равна скорость передачи сообщений, составленных из таких символов?
2. Сообщения составлены из пяти качественных признаков ($m = 5$). Длительность элементарной посылки $t = 20\text{мс}$. Определить, чему равна скорость передачи

сигналов и информации.

3. Определить пропускную способность бинарного канала связи, способного передавать 100 символов 0 или 1 в единицу времени, причем каждый из символов искажается (заменяется противоположным) с вероятностью $p = 0,01$.
4. Имеются источник информации с энтропией в единицу времени $H(X) = 100$ дв.ед. и два канала связи; каждый из них может передавать в единицу времени 70 двоичных знаков (0 или 1); каждый двоичный знак заменяется противоположным с вероятностью $p = 0,1$. Требуется выяснить, достаточна ли пропускная способность этих каналов для передачи информации, поставляемой источником.
5. Чему равна пропускная способность симметричного канала, если источник вырабатывает сигналы со скоростью 2 знака в секунду, закодированные кодом с основанием $m = 10$, а вероятность ложного приема $p = 0,3$?
6. Сообщения составлены из алфавита $X = (x_1, x_2, x_3)$. Вероятности появления символов алфавита 0,7; 0,2; 0,1 соответственно. Помехи в канале связи заданы следующей канальной матрицей:

$$P(Y/X) = \begin{vmatrix} 0,98 & 0,01 & 0,01 \\ 0,1 & 0,75 & 0,15 \\ 0,2 & 0,3 & 0,5 \end{vmatrix}$$

Определить скорость передачи информации, если время передачи одного символа $t_1 = 0,02$ с.

7. Чему равна пропускная способность канала связи, описанного канальной матрицей:

$$P(A,B) = \begin{vmatrix} 0,1 & 0 & 0 \\ 0,1 & 0,3 & 0 \\ 0 & 0,1 & 0,4 \end{vmatrix}$$

если известно, что на выходе источника сообщений символы вырабатываются со скоростью 100 знаков в секунду?

8. Определить максимально возможную скорость передачи информации по радиотехническому каналу связи пункта управления с телеуправляемой ракетой, если полоса пропускания канала связи равна 3 МГц, а минимальное отношение сигнал-шум по мощности в процессе наведения ракеты на цель равно 3.
9. Определить полосу пропускания канала передачи телевизионного черно-белого изображения с 5×10^5 элементами, 25 кадрами в секунду и 8 равновероятными градациями яркости для отношения $P/N = 15$ при условии, что изображение может принимать наиболее хаотичный вид «белого шума».

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;

— вывод о проделанной работе.

Контрольные вопросы

1. Что такое пропускная способность канала передачи информации? Чем отличается пропускная способность от скорости передачи информации по каналу связи?
2. Чем отличается информационная скорость передачи от технической, и в каких единицах эти скорости измеряются?
3. Как изменяется пропускная способность дискретного канала связи при воздействии на канал помех.
4. Сформулируйте основную теорему Шеннона о кодировании для канала без помех.
5. Сформулируйте и поясните теорему Шеннона о кодировании для канала с помехами.
6. Приведите выражение пропускной способности для дискретного канала без помех и с помехами.
7. Сформулируйте и поясните теорему отсчетов (Котельникова)
8. Какие параметры влияют на объем сигнала.
9. От чего зависит пропускная способность непрерывного канала связи.
10. Назовите условия согласования источников информации с пропускной способностью непрерывных каналов связи.
11. Какова скорость отображения информации приемным устройством отображения информации.

Практическое занятие № 7

Тема программы: **Смысл энтропии Шеннона.**

Тема: **Расчет вероятностей.**

Цель: научиться вычислять вероятности событий (появление символов в сообщении) и рассчитывать энтропию.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, калькулятор, текстовый редактор.

Теоретические основы

Количество информации по Хартли и Шеннону

Понятие количество информации отождествляется с понятием информация. Эти два понятия являются синонимами. Мера информации должна монотонно возрастать с увеличением длительности сообщения (сигнала), которую естественно измерять числом символов в дискретном сообщении и временем передачи в непрерывном случае. Кроме того, на содержание количества информации должны влиять и статистические характеристики, так как сигнал должен рассматриваться как случайный процесс.

При этом наложено ряд ограничений:

1. Рассматриваются только дискретные сообщения.
2. Множество различных сообщений конечно.
3. Символы, составляющие сообщения равновероятны и независимы.

Хартли впервые предложил в качестве меры количества информации принять логарифм числа возможных последовательностей символов.

$$I = \log m^k = \log N \quad (1)$$

К. Шеннон попытался снять те ограничения, которые наложил Хартли. На самом деле в рассмотренном выше случае равной вероятности и независимости символов при любом k все возможные сообщения оказываются также равновероятными, вероятность каждого из таких сообщений равна $P = 1/N$. Тогда количество информации можно выразить через вероятности появления сообщений $I = -\log P$.

В силу статистической независимости символов, вероятность сообщения длиной в k символов равна

$$P = \prod_{i=1}^k p_i$$

Если i -й символ повторяется в данном сообщении k_i раз, то

$$P = \prod_{i=1}^m p_i^{k_i}$$

так как при повторении i символа k_i раз k уменьшается до m . Из теории вероятностей известно, что, при достаточно длинных сообщениях (большое число

символов k) $k_i \approx k \cdot p_i$ и тогда вероятность сообщений будет равняться

$$P = \prod_{i=1}^m p_i^{k p_i}$$

Тогда окончательно получим

$$I = -\log P = -k \sum_{i=1}^m p_i \log p_i \quad (2)$$

Данное выражение называется формулой Шеннона для определения количества информации.

Формула Шеннона для количества информации на отдельный символ сообщения совпадает с энтропией. Тогда количество информации сообщения состоящего из k символов будет равняться $I = k \cdot H$

Количество информации, как мера снятой неопределенности

При передаче сообщений, о какой либо системе происходит уменьшение неопределенности. Если о системе все известно, то нет смысла посылать сообщение. Количество информации измеряют уменьшением энтропии.

Количество информации, приобретаемое при полном выяснении состояния некоторой физической системы, равно энтропии этой системы:

$$I = -\sum_{i=1}^n p_i \log p_i$$

Количество информации I – есть осредненное значение логарифма вероятности состояния. Тогда каждое отдельное слагаемое $-\log p_i$ необходимо рассматривать как частную информацию, получаемую от отдельного сообщения, то есть

$$I_i = -\log p_i$$

Избыточность информации

Если бы сообщения передавались с помощью равновероятных букв алфавита и между собой статистически независимых, то энтропия таких сообщений была бы максимальной. На самом деле реальные сообщения строятся из не равновероятных букв алфавита с наличием статистических связей между буквами. Поэтому энтропия реальных сообщений $-H_p$, оказывается много меньше оптимальных сообщений $-H_0$. Допустим, нужно передать сообщение, содержащее количество информации, равное I . Источнику, обладающему энтропией на букву, равной H_p , придется затратить

некоторое число n_p , то есть

$$I = n_p H_p$$

Если энтропия источника была бы H_0 , то пришлось бы затратить меньше букв на передачу этого же количества информации

$$I = n_0 H_0 \quad n_0 = \frac{I}{H_0} < n_p$$

Таким образом, часть букв $n_p - n_0$ являются как бы лишними, избыточными. Мера удлинения реальных сообщений по сравнению с оптимально закодированными и представляет собой избыточность D .

$$D = 1 - \frac{H_p}{H_0} = 1 - \frac{n_0}{n_p} = \frac{n_p - n_0}{n_p} \quad (3)$$

Но наличие избыточности нельзя рассматривать как признак несовершенства источника сообщений. Наличие избыточности способствует повышению помехоустойчивости сообщений. Высокая избыточность естественных языков обеспечивает надежное общение между людьми.

Частотные характеристики текстовых сообщений

Важными характеристиками текста являются повторяемость букв, пар букв (биграмм) и вообще m -ок (m -грамм), сочетаемость букв друг с другом, чередование гласных и согласных и некоторые другие. Замечательно, что эти характеристики являются достаточно устойчивыми.

Идея состоит в подсчете чисел вхождений каждой n^m возможных m -грамм в достаточно длинных открытых текстах $T = t_1 t_2 \dots t_l$, составленных из букв алфавита $\{a_1, a_2, \dots, a_n\}$. При этом просматриваются подряд идущие m -граммы текста

$$t_1 t_2 \dots t_m, t_2 t_3 \dots t_{m+1}, \dots, t_{l-m+1} t_{l-m+2} \dots t_l.$$

Если $\mathcal{A}(a_{i1} a_{i2} \dots a_{im})$ — число появлений m -граммы $a_{i1} a_{i2} \dots a_{im}$ в тексте T , а L общее число подсчитанных m -грамм, то опыт показывает, что при достаточно больших L частоты

$$\frac{\mathcal{A}(a_{i1} a_{i2} \dots a_{im})}{L}$$

для данной m -граммы мало отличаются друг от друга.

В силу этого, относительную частоту считают приближением вероятности $P(a_{i1} a_{i2} \dots a_{im})$ появления данной m -граммы в случайно выбранном месте текста (такой подход принят при статистическом определении вероятности).

Для русского языка частоты (в порядке убывания) знаков алфавита, в котором отождествлены $Е$ с $Ё$, $Ь$ с $Ъ$, а также имеется знак пробела (-) между словами, приведены в таблице 1.

Таблица 1

- 0.175	О 0.090	Е, Ё 0.072	А 0.062
И 0.062	Т 0.053	Н 0.053	С 0.045
Р 0.040	В 0.038	Л 0.035	К 0.028
М 0.026	Д 0.025	П 0.023	У 0.021
Я 0.018	Ы 0.016	З 0.016	Ь, Ъ 0.014
Б 0.014	Г 0.013	Ч 0.012	Й 0.010
Х 0.009	Ж 0.007	Ю 0.006	Ш 0.006
Ц 0.004	Щ 0.003	Э 0.003	Ф 0.002

Некоторая разница значений частот в приводимых в различных источниках

таблицах объясняется тем, что частоты существенно зависят не только от длины текста, но и от его характера.

Устойчивыми являются также частотные характеристики биграмм, триграмм и четырехграмм осмысленных текстов.

Порядок выполнения работы

Определить количество информации (по Хартли), содержащееся в заданном сообщении, при условии, что значениями являются буквы кириллицы.

«Фамилия Имя Отчество» завершил ежегодный съезд эрудированных школьников, мечтающих глубоко проникнуть в тайны физических явлений и химических реакций

Построить таблицу распределения частот символов, характерные для заданного сообщения. Производится так называемая частотная селекция, текст сообщения анализируется как поток символов и высчитывается частота встречаемости каждого символа. Сравнить с имеющимися данными в табл 1.

На основании полученных данных определить среднее и полное количество информации, содержащееся в заданном сообщении. Оценить избыточность сообщения.

1. Построить таблицу распределения частот символов, характерных для заданного сообщения путём деления количества определённого символа в данном сообщении на общее число символов

По формуле

$$\sum_{i=1}^m p_i \log p_i$$

$H =$ вычислил энтропию сообщения

2. Далее по формуле Шеннона для определения кол-ва информации

$I = -\log P = -k \sum_{i=1}^m p_i \log p_i$ вычислил кол-во информации в передаваемом сообщении

3. Вычислил избыточность D по формуле

$$D = 1 - \frac{H_p}{H_o} = 1 - \frac{n_o}{n_p} = \frac{n_p - n_o}{n_p}$$

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Дать определение понятие энтропия?
2. Что означает вероятностный способ измерения информации?
3. Что означает статическое определение вероятности?
4. Запишите уравнение Хартли?
5. Какие основные разработки внес в основу теории информации Шеннон?

Расчет вероятностей Тема программы: Смысл энтропии Шеннона.

Тема: Составление закона распределения вероятностей.

Цель: научиться составлять законы распределения вероятностей.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, калькулятор.

Теоретические основы

Дискретной называют случайную величину, возможные значения которой есть отдельные изолированные числа, которые эта величина принимает с определенными ненулевыми вероятностями. Число возможных значений может быть конечным или бесконечным (счетным).

Законом распределения дискретной случайной величины называют перечень её возможных значений и соответствующих им вероятностей. Закон распределения может быть задан одним из следующих способов.

1. Таблицей

x	x ₁	x ₂	...	x _n
p	p ₁	p ₂	...	p _n

где $\sum_{i=1}^n p_i = 1$.

2. Аналитически $P(X = x_i) = \varphi(x_i)$. Например:

а) *биномиальное распределение*

$$P(X = k) = C_n^k p^k q^{n-k}, 0 < p < 1, k = 0, 1, 2, \dots, n;$$

б) *распределение Пуассона*

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \lambda > 0, k = 0, 1, 2, \dots$$

3. С помощью *функции распределения* $F(x)$, определяющей для каждого значения x вероятность того, что случайная величина X примет значение, меньшее x , т. е. $F(x) = P(X < x)$.

Свойства $F(x)$:

1) $0 \leq F(x) \leq 1$;

2) $F(x_2) \geq F(x_1)$, если $x_2 > x_1$;

3) $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$.

4. Закон распределения может быть задан графически - *многоугольником распределения* (см. пример 1).

Числовые характеристики дискретных случайных величин

Математическое ожидание $M(X) = \sum_i x_i p_i$;

Дисперсия $D(X) = M[X - M(X)]^2$ или $D(X) = M(X^2) - [M(X)]^2$;

Среднее квадратическое отклонение $\sigma(X) = \sqrt{D(X)}$.

Для биномиального распределения $M(X) = np$, $D(X) = npq$. Для распределения Пуассона $M(X) = \lambda$, $D(X) = \lambda$.

Пример 1.

Устройство состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0,1. Составить закон распределения числа отказавших элементов в одном опыте, построить многоугольник распределения. Найти функцию распределения $F(x)$ и построить её график. Найти $M(X)$, $D(X)$, $\sigma(X)$.

Решение: Дискретная случайная величина X (число отказавших элементов в одном опыте) имеет следующие возможные значения: $x_1=0$ (ни один из элементов устройства не отказал), $x_2=1$ (отказал один элемент), $x_3=2$ (отказало два элемента) и $x_4=3$ (отказали три элемента).

Отказы элементов независимы один от другого, вероятности отказа каждого элемента равны между собой, поэтому применима формула Бернулли. Учитывая, что, по условию, $n=3$, $p=0,1$ (следовательно, $q=1-0,1=0,9$), получим:

$$P_3(0)=q^3=0,9^3=0,729; \quad P_3(1)=C_3^1 p q^2 = 3 \cdot 0,1 \cdot 0,9^2 = 0,243;$$

$$P_3(2)=C_3^2 p^2 q = 3 \cdot 0,1^2 \cdot 0,9 = 0,027; \quad P_3(3)=p^3 = 0,1^3 = 0,001.$$

Контроль: $\sum_{i=1}^4 p_i = 1; \quad 0,729+0,243+0,027+0,001=1.$

Искомый биномиальный закон распределения X :

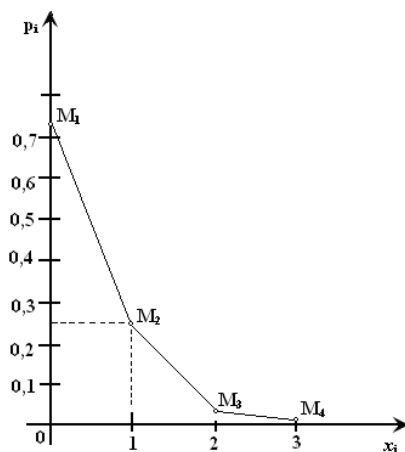
X	0	1	2	3
p	0,729	0,243	0,027	0,001

Для построения

многоугольника распределения строим прямоугольную систему координат. По оси абсцисс откладываем возможные значения x_i , а по оси ординат – соответствующие им вероятности p_i . Построим точки $M_1(0;0,729)$, $M_2(1;0,243)$, $M_3(2;0,027)$, $M_4(3;0,001)$. Соединив эти точки отрезками прямых, получаем искомый многоугольник распределения (Рис.1).

Рис.1

Найдем функцию
 Для $x \leq 0$ имеем
 для $0 < x \leq 1$ имеем
 для $1 < x \leq 2$



распределения $F(x)=P(X < x)$.
 $F(x)=P(X < 0)=0$;
 $F(x)=P(X < 1)=P(X=0)=0,729$;

$$F(x)=P(X < 2)=P(X=0)+P(X=1)=0,729+0,243=0,972;$$

$$\text{для } 2 < x \leq 3 \quad F(x)=P(X < 3)=P(X=0)+P(X=1)+P(X=2)=0,972+0,027=0,999;$$

для $x > 3$ будет $F(x)=1$, т. к. событие достоверно.

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ 0,729 & \text{при } 0 < x \leq 1, \\ 0,972 & \text{при } 1 < x \leq 2, \\ 0,999 & \text{при } 2 < x \leq 3, \\ 1 & \text{при } x > 3. \end{cases}$$

График этой функции приведен на Рис. 2.

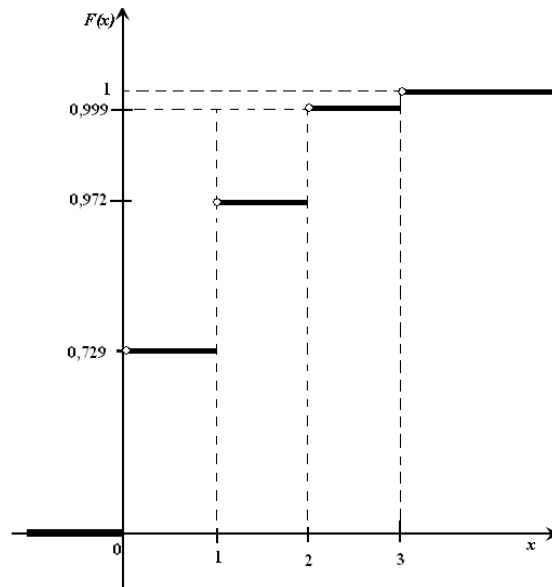


Рис. 2

Для биномиального распределения $M(X) = np = 3 \cdot 0,1 = 0,3$; $D(X) = npq = 3 \cdot 0,1 \cdot 0,9 = 0,27$; $\sigma(X) = \sqrt{D(X)} = \sqrt{0,27} \approx 0,52$.

Пример 2.

В партии из 10 деталей имеется 8 стандартных. Наудачу отобраны две детали. Составить закон распределения случайной величины X – числа стандартных деталей среди отобранных. Найти $M(X)$, $D(X)$.

Решение: Случайная величина X – число стандартных деталей среди отобранных деталей – имеет следующие возможные значения: $x_1=0$; $x_2=1$; $x_3=2$. Найдем вероятности возможных значений X по формуле (пример 2)

$$P(X = k) = \frac{C_n^k \cdot C_{N-n}^{m-k}}{C_N^m} \quad (N - \text{число деталей в партии, } n - \text{число стандартных деталей в}$$

партии, m – число отобранных деталей, k – число стандартных деталей среди отобранных),

находим:

$$P(X = 0) = \frac{C_8^0 \cdot C_2^2}{C_{10}^2} = \frac{1}{10 \cdot 9 / (1 \cdot 2)} = \frac{1}{45};$$

$$P(X = 1) = \frac{C_8^1 \cdot C_2^1}{C_{10}^2} = \frac{8 \cdot 2}{45} = \frac{16}{45};$$

$$P(X = 2) = \frac{C_8^2 \cdot C_2^0}{C_{10}^2} = \frac{8 \cdot 7 / (1 \cdot 2)}{45} = \frac{28}{45}.$$

Составим искомый закон распределения:

X	0	1	2
-----	---	---	---

p	$\frac{1}{45}$	$\frac{16}{45}$	$\frac{28}{45}$
-----	----------------	-----------------	-----------------

Контроль: $\frac{1}{45} + \frac{16}{45} + \frac{28}{45} = 1$.

$$M(X) = \sum_{i=1}^3 x_i p_i = 0 \cdot \frac{1}{45} + 1 \cdot \frac{16}{45} + 2 \cdot \frac{28}{45} = \frac{72}{45} = \frac{8}{5}$$

$$D(X) = M(X^2) - [M(X)]^2; M(X^2) = \sum_{i=1}^3 x_i^2 p_i = 0^2 \cdot \frac{1}{45} + 1 \cdot \frac{16}{45} + 2^2 \cdot \frac{28}{45} = \frac{128}{45};$$

$$D(X) = \frac{128}{45} - \left(\frac{8}{5}\right)^2 = \frac{128}{45} - \frac{64}{25} = \frac{64}{225}$$

Пример 3.

В устройстве независимо друг от друга выходят из строя три элемента. Вероятность выхода из строя первого элемента – 0,3, второго – 0,2, третьего – 0,4. Составить закон распределения случайной величины X – числа вышедших из строя элементов.

Решение: случайная величина X имеет следующие возможные значения: $x_1=0$, $x_2=1$, $x_3=2$, $x_4=3$. $p_1=0,3, q_1=1-p_1=0,7$, $p_2=0,2, q_2=1-p_2=0,8$, $p_3=0,4, q_3=1-p_3=0,6$.

$P(X=k)$ вычисляем по следующим формулам (см. пример 4)

$$P(X=0) = q_1 \cdot q_2 \cdot q_3 = 0,7 \cdot 0,8 \cdot 0,9 = 0,336;$$

$$P(X=1) = p_1 \cdot q_2 \cdot q_3 + q_1 \cdot p_2 \cdot q_3 + q_1 \cdot q_2 \cdot p_3 = 0,3 \cdot 0,8 \cdot 0,6 + 0,7 \cdot 0,2 \cdot 0,6 + 0,7 \cdot 0,8 \cdot 0,4 = 0,144 + 0,084 + 0,224 = 0,452;$$

$$P(X=2) = p_1 \cdot p_2 \cdot q_3 + p_1 \cdot q_2 \cdot p_3 + q_1 \cdot p_2 \cdot p_3 = 0,3 \cdot 0,2 \cdot 0,6 + 0,3 \cdot 0,8 \cdot 0,4 + 0,7 \cdot 0,2 \cdot 0,4 = 0,118;$$

$$P(X=3) = p_1 \cdot p_2 \cdot p_3 = 0,3 \cdot 0,2 \cdot 0,4 = 0,024.$$

Контроль: $0,336 + 0,452 + 0,118 + 0,024 = 1$.

Искомый закон распределения:

X	0	1	2	3
p	0,336	0,452	0,118	0,024

Пример 4.

Среднее число заказов такси, поступающих на диспетчерский пункт в одну минуту, равно двум. Составить закон распределения случайной величины X – числа заказов, поступающих за 4 минуты. Найти $M(X)$, $D(X)$.

Решение: Поток заказов на такси можно считать *простейшим*, т. е. обладающим стационарностью, «отсутствием последствия» и ординарностью. *Интенсивность потока* (среднее число событий появляющихся в единицу времени) $\lambda=2$. Вероятность появления k событий простейшего потока за время $t=4$

определяется формулой Пуассона $P_i(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$, для данной задачи

$P_4(k) = \frac{8^k e^{-8t}}{k!}$. Совокупность возможных значений X есть счетное множество, т.е.

$x_1=0, x_2=1, \dots, x_k=k+1, \dots$; тогда закон распределения случайной величины X – числа заказов, поступающих за 4 минуты принимает вид:

X	0	1	2	...	k	...
p	$e^{-\lambda t}$	$\frac{(\lambda t) e^{-\lambda t}}{1!}$	$\frac{(\lambda t)^2 e^{-\lambda t}}{2!}$...	$\frac{(\lambda t)^k e^{-\lambda t}}{k!}$...

или

X	0	1	2	...	k	...
p	e^{-8}	$\frac{8 e^8}{1!}$	$\frac{8^2 e^{-8}}{2!}$...	$\frac{8^k e^{-8}}{k!}$...

Воспользовавшись таблицей 3 приложения, окончательно получим:

X	0	1	2	...	k	...
p	0,00035	0,002684	0,010735	...	$\frac{8^k e^{-8}}{k!}$...

Наивероятнейшее число заказов такси за 4 минуты можно определить по получившемуся закону распределения (значения x , при которых p максимально): $k'_0=7, k''_0=8$. Для простейшего потока событий: математическое ожидание $M(X) = \lambda t = 8$, дисперсия $D(X) = \lambda t = 8$.

Пример 5.

Даны законы распределения независимых случайных величин X и Y . Составить закон распределения случайной величины $Z=X+2Y$. Найти $M(Z), D(Z)$.

X	-3	0	1
p	0,1	0,03	0,06

Y	1	3	6
p	0,2	0,5	0,3

Решение: Закон распределения $V=2Y$ получается из распределения Y путем умножения всех значений y_i на 2. Получаем:

V	2	6	12
p	0,2	0,5	0,3

Для составления закона распределения случайной величины Z вычислим все ее возможные значения по формуле $z_k = x_i + v_j$, $k = 1, 2, \dots, 9$, $i, j = 1, 2, 3$.

Соответствующие данным значениям z_k вероятности p_k можно вычислить по формуле умножения вероятностей $P_k = P(Z = z_k) = P(X = x_i) \cdot P(V = v_j)$, т. к. события $X = x_i$ и $V = v_j$ - независимы (исходим из независимости случайных величин X и Y) и наступают совместно (событие $\{Z = z_k\} = \{\text{совместное наступление событий } X = x_i \text{ и } V = v_j\}$). Тогда распределение Z принимает вид

Z	-1	3	9	2	6	12	3	7	13
p	0,02	0,05	0,03	0,06	0,15	0,09	0,12	0,3	0,18

Рассмотрим значения $z_2 = z_7 = 3$. События $Z = z_2$ и $Z = z_7$ несовместны, поэтому вероятность наступления хотя бы одного из этих событий вычисляется по правилу сложения вероятностей

$$P(Z = z_2 \cup Z = z_7) = P(Z = z_2) + P(Z = z_7) = 0,05 + 0,12 = 0,17$$

Искомый закон распределения случайной величины Z получается после размещения z_k по возрастанию.

Z	-1	2	3	6	7	9	12	13
p	0,02	0,06	0,17	0,15	0,3	0,03	0,09	0,18

Математическое ожидание $M(Z)$ и дисперсию $D(Z)$ можно найти по формулам:

$$M(Z) = \sum_{k=1}^8 z_k p_k; \quad D(Z) = M(Z^2) - [M(Z)]^2, \quad \text{где } M(Z^2) = \sum_{k=1}^8 z_k^2 p_k.$$

Рассмотрим другой способ.

$M(Z)$ и $D(Z)$ можно найти через $M(X)$, $M(Y)$, $D(X)$, $D(Y)$.

$$M(X) = \sum_{i=1}^3 x_i p_i = -3 \cdot 0,1 + 0 \cdot 0,3 + 1 \cdot 0,6 = 0,3$$

$$M(Y) = \sum_{j=1}^3 y_j p_j = 0 \cdot 0,2 + 3 \cdot 0,5 + 6 \cdot 0,3 = 3,3$$

$$D(X) = \sum_{i=1}^3 x_i^2 p_i - (M(X))^2 = (-3)^2 \cdot 0,1 + 0^2 \cdot 0,3 + 1^2 \cdot 0,6 - (0,3)^2 = 1,41$$

$$D(Y) = \sum_{j=1}^3 y_j^2 p_j - (M(Y))^2 = 0^2 \cdot 0,2 + 3^2 \cdot 0,5 + 6^2 \cdot 0,3 - (3,3)^2 = 4,41$$

$$M(Z) = M(X + 2Y) = M(X) + M(2Y) = M(X) + 2M(Y) = 0,3 + 2 \cdot 3,3 = 6,9,$$

т. к. математическое ожидание суммы равно сумме математических ожиданий слагаемых; постоянный множитель можно вынести за знак математического ожидания.

$$D(Z) = D(X + 2Y) = D(X) + D(2Y) = D(X) + 4D(Y) = 1,41 + 4 \cdot 4,41 = 19,05,$$

т. к. дисперсия суммы независимых случайных величин равна сумме дисперсий слагаемых; постоянный множитель можно вынести за знак дисперсии, возведя его в квадрат.

Пример 6.

Стрелок ведет стрельбу с вероятностью попадания в цель 0,8 при каждом выстреле. Стрельба ведется до первого попадания, но делается не более 3 выстрелов. Составить закон распределения случайной величины X , если: а) X – число промахов; б) X – число попаданий; в) X – число произведенных выстрелов.

Решение: Вероятность попадания $p=0,8$; вероятность промаха $q=1-p=0,2$.

а) Случайная величина X – число промахов при трех выстрелах – имеет следующие возможные значения: $x_1 = 0$; $x_2 = 1$; $x_3 = 2$; $x_4 = 3$.

Событие $X=0$ равносильно попаданию с первой попытки, следовательно, $P(X=0)=p=0,8$.

Событие $X=1$ равносильно попаданию со второй попытки, т. е. совместному наступлению двух событий: промаха и попадания; следовательно, $P(X=1)=q \cdot p=0,2 \cdot 0,8=0,16$.

Событие $X=2$ равносильно попаданию с третьей попытки, т. е. $P(X=2)=q \cdot q \cdot p=0,2 \cdot 0,2 \cdot 0,8=0,032$.

Событие $X=3$ означает отсутствие попаданий, $P(X=3)=q \cdot q \cdot q=0,2^3=0,008$.

Искомый закон распределения X :

X	0	1	2	3
p	0,8	0,16	0,032	0,008

б) Случайная величина X – число попаданий – имеет следующие возможные значения: $x_1 = 0$ (допущено три промаха); $x_2 = 1$ (произошло попадание с первой, второй или третьей попытки).

Тогда $P(X=0)=q^3=0,2^3=0,008$;

$$P(X=1)=p+q \cdot p+q \cdot q \cdot p=0,8+0,16+0,032=0,992$$

$$\text{или } P(X=1)=1-P(X=0)=1-0,008=0,992.$$

Искомый закон распределения X :

X	0	1
P	0,008	0,992

в) Случайная величина X – число произведенных выстрелов – имеет следующие возможные значения: $x_1 = 1$; $x_2 = 2$; $x_3 = 3$.

Событие $X=1$ равносильно попаданию с первой попытки, т. е. $P(X=1)=p=0,8$.

Событие $X=2$ равносильно попаданию со второй попытки, т. е. $P(X=2)=q \cdot p=0,16$.

Событие $X=3$ означает, что либо произошло попадание с третьей попытки, либо было три промаха. Тогда $P(X=3)=q \cdot q \cdot p+q \cdot q \cdot q=0,032+0,008=0,04$.

Искомый закон распределения X :

X	1	2	3
P	0,8	0,16	0,04

Задачи

Вариант 1. Производятся последовательные независимые испытания приборов на надёжность. Каждый следующий прибор испытывается лишь в том случае, если предыдущий оказался надежным. Построить закон распределения случайного числа испытанных приборов, если вероятность выдержать испытание для каждого из них равна 0,9. Найти математическое ожидание числа испытанных приборов. Найти функцию распределения $F(x)$ и построить ее график; найти $M(X)$,

$\sigma(X)$; построить многоугольник распределения.

Вариант 2. Известно, что в партии из 20 телефонных аппаратов 5 недействующих. Случайным образом из этой партии взято 4 аппарата. Построить закон распределения случайной величины X – числа недействующих аппаратов из отобранных. Найти дисперсию этой случайной величины. В каких единицах она измеряется? Построить график функции распределения $F(x)$ случайной величины X , многоугольник распределения.

Вариант 3. Сырье на завод привозят от трех независимо работающих поставщиков. Вероятность своевременного прибытия сырья от первого поставщика равна 0,4, от второго – 0,7, от третьего – 0,6. Найти математическое ожидание $M(X)$, дисперсию $D(X)$ числа своевременных поставок сырья. Найти функцию распределения и построить ее график.

Вариант 4. Завод получает сырье на автомашинах от трех независимо работающих поставщиков. Вероятность прибытия автомашины от первого поставщика равна 0,2, от второго – 0,3 и от третьего – 0,1. Составить распределение числа прибывших автомашин. Найти математическое ожидание и дисперсию полученной величины. Построить график функции распределения $F(x)$.

Вариант 5. Вероятность изготовления бракованной детали $p=0,1$. Изготовлено 4 детали. X – случайное число бракованных деталей. Построить закон распределения случайной величины X , найти ее математическое ожидание и дисперсию. Построить график функции распределения, многоугольник распределения.

Вариант 6. Среднее число заявок, поступающих на предприятие бытового обслуживания за 1 час, равно 2. Составить закон распределения случайной величины X – числа заявок, поступивших за 3 часа. Найти $M(X)$, $D(X)$ и наиболее вероятное число заявок за 3 часа.

Вариант 7. В среднем в магазин заходит 3 человека в минуту. Составить закон распределения случайной величины X – числа зашедших в магазин человек за 2 минуты. Построить многоугольник распределения. Найти $M(X)$, $D(X)$.

Вариант 8. Даны законы распределения независимых случайных величин

X	-3	0	1
P	0,1	0,3	0,6

Y	0	3	6
p	0,2	0,5	0,3

Составить законы распределения

случайных величин:

а) XY ; б) $X+Y$. Найти $M(X+Y)$, $D(X+Y)$. Справедливо ли равенство $M(X) \cdot M(Y) = M(X \cdot Y)$?

Вариант 9. Команда состоит из двух стрелков. Числа очков, выбиваемых каждым из них при одном выстреле, являются случайными величинами X_1 и X_2 , которые характеризуются следующими законами распределения:

X_1	3	4	5
P	0,3	0,4	0,3

и

X_2	2	3	4	5
P	0,2	0,1	0,2	0,5

Результаты

стрельбы

одного стрелка не влияют на результат стрельбы другого. Составить закон

распределения числа очков, выбиваемых командой, если стрелки сделают по одному выстрелу. Убедиться в справедливости равенства $D(X_1+X_2)=D(X_1)+D(X_2)$.

Вариант 10. Производятся выстрелы из орудия с вероятностью попадания в цель 0,9 при каждом выстреле. Стрельба ведётся до первого попадания, но делается не более 4 выстрелов. Составить закон распределения случайной величины X , если: а) X – число произведенных выстрелов; б) X – число промахов; в) X – число попаданий. Найдите математическое ожидание всех найденных случайных величин.

Отчет

Отчет должен содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Какие значения не может принимать вероятность?
2. Чему равна вероятность достоверного события? Невозможного?
3. Дайте определение закону распределения дискретной случайной величины.
4. Дайте определение математическому ожиданию?

Практическое занятие № 9

Тема программы: **Сжатие информации.**

Тема: **Системные требования алгоритмов сжатия. Алгоритмы сжатия данных неизвестного формата.**

Цель: научиться сжимать информацию с помощью метода Хаффмана и метода RLE.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, калькулятор, текстовый редактор.

Теоретические основы

Сжатие данных (англ. *data compression*) — алгоритмическое преобразование данных, производимое с целью уменьшения их объёма. Применяется для более рационального использования устройств хранения и передачи данных. Синонимы — упаковка данных, компрессия, сжимающее кодирование, кодирование источника. Обратная процедура называется восстановлением данных (распаковкой, декомпрессией). Сжатие основано на устранении избыточности, содержащейся в исходных данных. Простейшим примером избыточности является повторение в тексте фрагментов (например, слов естественного или машинного языка). Подобная избыточность обычно устраняется заменой повторяющейся последовательности ссылкой на уже закодированный фрагмент с указанием его длины. Другой вид избыточности связан с тем, что некоторые значения в сжимаемых данных встречаются чаще других. Сокращение объёма данных достигается за счёт замены часто встречающихся данных короткими кодовыми словами, а редких — длинными (энтропийное кодирование). Сжатие

данных, не обладающих свойством избыточности (например, случайный сигнал или белый шум, зашифрованные сообщения), принципиально невозможно без потерь. В основе любого способа сжатия лежит модель источника данных, или, точнее, модель избыточности. Иными словами, для сжатия данных используются некоторые априорные сведения о том, какого рода данные сжимаются. Не обладая такими сведениями об источнике, невозможно сделать никаких предположений о преобразовании, которое позволило бы уменьшить объём сообщения. Модель избыточности может быть статической, неизменной для всего сжимаемого сообщения, либо строиться или параметризоваться на этапе сжатия (и восстановления). Методы, позволяющие на основе входных данных изменять модель избыточности информации, называются адаптивными. Неадаптивными являются обычно узкоспециализированные алгоритмы, применяемые для работы с данными, обладающими хорошо определёнными и неизменными характеристиками. Подавляющая часть достаточно универсальных алгоритмов являются в той или иной мере адаптивными.

Все методы сжатия данных делятся на два основных класса:

- Сжатие без потерь
- Сжатие с потерями

При использовании сжатия без потерь возможно полное восстановление исходных данных, сжатие с потерями позволяет восстановить данные с искажениями, обычно несущественными с точки зрения дальнейшего использования восстановленных данных. Сжатие без потерь обычно используется для передачи и хранения текстовых данных, компьютерных программ, реже — для сокращения объёма аудио- и видеоданных, цифровых фотографий и т. п., в случаях, когда искажения недопустимы или нежелательны. Сжатие с потерями, обладающее значительно большей, чем сжатие без потерь, эффективностью, обычно применяется для сокращения объёма аудио- и видеоданных и цифровых фотографий в тех случаях, когда такое сокращение является приоритетным, а полное соответствие исходных и восстановленных данных не требуется.

Системные требования алгоритмов

Различные алгоритмы могут требовать различного количества ресурсов вычислительной системы, на которых они реализованы:

- оперативной памяти (под промежуточные данные);
- постоянной памяти (под код программы и константы);
- процессорного времени.

В целом, эти требования зависят от сложности и «интеллектуальности» алгоритма. Общая тенденция такова: чем эффективнее и универсальнее алгоритм, тем большие требования к вычислительным ресурсам он предъявляет. Тем не менее, в специфических случаях простые и компактные алгоритмы могут работать не хуже сложных и универсальных. Системные требования определяют их потребительские качества: чем менее требователен алгоритм, тем на более простой, а следовательно, компактной, надёжной и дешёвой системе он может быть реализован.

Так как алгоритмы сжатия и восстановления работают в паре, имеет значение соотношение системных требований к ним. Нередко можно усложнив один алгоритм значительно упростить другой. Таким образом, возможны три варианта:

Алгоритм сжатия требует больших вычислительных ресурсов, нежели

алгоритм восстановления.

Это наиболее распространённое соотношение, характерное для случаев, когда однократно сжатые данные будут использоваться многократно. В качестве примера можно привести цифровые аудио- и видеопроигрыватели.

Алгоритмы сжатия и восстановления требуют приблизительно равных вычислительных ресурсов.

Наиболее приемлемый вариант для линий связи, когда сжатие и восстановление происходит однократно на двух её концах (например, в цифровой телефонии).

Алгоритм сжатия существенно менее требователен, чем алгоритм восстановления.

Такая ситуация характерна для случаев, когда процедура сжатия реализуется простым, часто портативным устройством, для которого объём доступных ресурсов весьма критичен, например, космический аппарат или большая распределённая сеть датчиков. Это могут быть также данные, распаковка которых требуется в очень малом проценте случаев, например запись камер видеонаблюдения.

Алгоритмы сжатия данных неизвестного формата

Имеется два основных подхода к сжатию данных неизвестного формата.

- На каждом шаге алгоритма сжатия очередной сжимаемый символ либо помещается в выходной буфер сжимающего кодера как есть (со специальным флагом, помечающим, что он не был сжат), либо группа из нескольких сжимаемых символов заменяется ссылкой на совпадающую с ней группу из уже закодированных символов. Поскольку восстановление сжатых таким образом данных выполняется очень быстро, такой подход часто используется для создания самораспаковывающихся программ.
- Для каждой сжимаемой последовательности символов однократно либо в каждый момент времени собирается статистика её встречаемости в кодируемых данных. На основе этой статистики вычисляется вероятность значения очередного кодируемого символа (либо последовательности символов). После этого применяется та или иная разновидность энтропийного кодирования, например, арифметическое кодирование или кодирование Хаффмана, для представления часто встречающихся последовательностей короткими кодовыми словами, а редко встречающихся — более длинными.

Код Хаффмана

Определение 1: Пусть $A=\{a_1, a_2, \dots, a_n\}$ - алфавит из n различных символов, $W=\{w_1, w_2, \dots, w_n\}$ - соответствующий ему набор положительных целых весов. Тогда набор бинарных кодов $C=\{c_1, c_2, \dots, c_n\}$, такой что:

- (c_i не является префиксом для c_j , при
1) $i \neq j$
- ($\sum_{i=1}^n w_i$ минимальна ($|c_i|$ длина кода c_i)
2)

называется *минимально-избыточным префиксным кодом* или иначе *кодом Хаффмана*.

Замечания:

1. Свойство (1) называется *свойством префиксности*. Оно позволяет однозначно декодировать коды переменной длины.

2. Сумму в свойстве (2) можно трактовать как размер закодированных данных в битах. На практике это очень удобно, т.к. позволяет оценить степень сжатия не прибегая непосредственно к кодированию.

3. В дальнейшем, чтобы избежать недоразумений, под кодом будем понимать битовую строку определенной длины, а под минимально-избыточным кодом или кодом Хаффмана - множество кодов (битовых строк), соответствующих определенным символам и обладающих определенными свойствами.

Известно, что любому бинарному префиксному коду соответствует определенное бинарное дерево.

Определение 2: Бинарное дерево, соответствующее коду Хаффмана, будем называть *деревом Хаффмана*.

Задача построения кода Хаффмана равносильна задаче построения соответствующего ему дерева. Приведем общую схему построения дерева Хаффмана:

1. Составим список кодируемых символов (при этом будем рассматривать каждый символ как одноэлементное бинарное дерево, вес которого равен весу символа).

2. Из списка выберем 2 узла с наименьшим весом.

3. Сформируем новый узел и присоединим к нему, в качестве дочерних, два узла выбранных из списка. При этом вес сформированного узла положим равным сумме весов дочерних узлов.

4. Добавим сформированный узел к списку.

5. Если в списке больше одного узла, то повторить 2-5.

Приведем пример: построим дерево Хаффмана для сообщения $S = \text{"А Н F В Н С Е Н Е Н С Е А Н D С Е Е Н Н Н С Н Н Н D E G H G G E H С Н Н"}$.

Для начала введем несколько обозначений:

1. Символы кодируемого алфавита будем выделять жирным шрифтом: **A, B, C**.

2. Веса узлов будем обозначать нижними индексами: **A₅, B₃, C₇**.

3. Составные узлы будем заключать в скобки: $((A_5 + B_3)_8 + C_7)_{15}$.

Итак, в нашем случае $A = \{A, B, C, D, E, F, G, H\}$, $W = \{2, 1, 5, 2, 7, 1, 3, 15\}$.

1. **A₂ B₁ C₅ D₂ E₇ F₁ G₃ H₁₅**

2. **A₂ C₅ D₂ E₇ G₃ H₁₅ (F₁+B₁)₂**

3. **C₅ E₇ G₃ H₁₅ (F₁+B₁)₂ (A₂+D₂)₄**

4. **C₅ E₇ H₁₅ (A₂+D₂)₄ ((F₁+B₁)₂+G₃)₅**

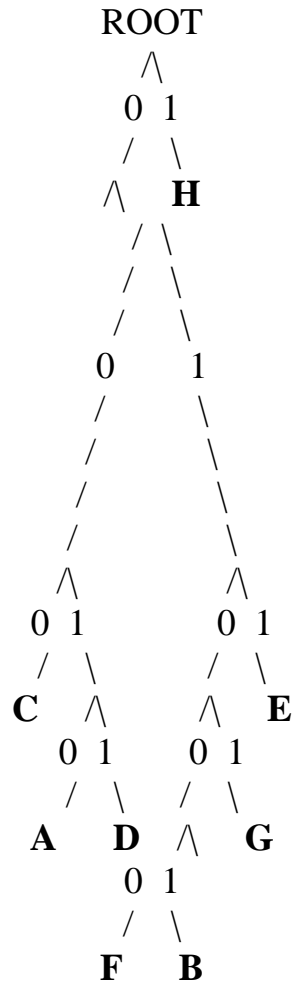
5. **E₇ H₁₅ ((F₁+B₁)₂+G₃)₅ (C₅+(A₂+D₂)₄)₉**

6. **H₁₅ (C₅+(A₂+D₂)₄)₉ (((F₁+B₁)₂+G₃)₅+E₇)₁₂**

7. **H₁₅ ((C₅+(A₂+D₂)₄)₉+(((F₁+B₁)₂+G₃)₅+E₇)₁₂)₂₁**

8. **(((C₅+(A₂+D₂)₄)₉+(((F₁+B₁)₂+G₃)₅+E₇)₁₂)₂₁+H₁₅)₃₆**

В списке, как и требовалось, остался всего один узел. Дерево Хаффмана построено. Теперь запишем его в более привычном для нас виде.



Листовые узлы дерева Хаффмана соответствуют символам кодируемого алфавита. Глубина листовых узлов равна длине кода соответствующих символов.

Путь от корня дерева к листовому узлу можно представить в виде битовой строки, в которой "0" соответствует выбору левого поддерева, а "1" - правого. Используя этот механизм, мы без труда можем присвоить коды всем символам кодируемого алфавита. Выпишем, к примеру, коды для всех символов в нашем примере:

A=0010_{bin} **C**=000_{bin} **E**=011_{bin} **G**=0101_{bin}
B=01001_{bin} **D**=0011_{bin} **F**=01000_{bin} **H**=1_{bin}

Теперь у нас есть все необходимое для того чтобы закодировать сообщение S. Достаточно просто заменить каждый символ соответствующим ему кодом:

S'="0010 1 01000 01001 1 000 011 1 011 1 000 011 0010 1 0011 000 011 011 1 1 1 000 1 1 1 0011 011 0101 1 0101 0101 011 1 000 1 1".

Оценим теперь степень сжатия. В исходном сообщении S было 36 символов, на

каждый из которых отводилось по $\lceil \log_2 |A| \rceil = 3$ бита (здесь и далее будем понимать квадратные скобки $\lceil \cdot \rceil$ как целую часть, округленную в положительную сторону, т.е. $\lceil 3,018 \rceil = 4$). Таким образом, размер S равен $36 \cdot 3 = 108$ бит

Размер закодированного сообщения S' можно получить воспользовавшись замечанием 2 к определению 1, или непосредственно, подсчитав количество бит в S' . И в том и другом случае мы получим 89 бит.

Итак, нам удалось сжать 108 в 89 бит.

Теперь декодируем сообщение S' . Начиная с корня дерева будем двигаться вниз, выбирая левое поддерево, если очередной бит в потоке равен "0", и правое - если "1". Дойдя до листового узла мы декодируем соответствующий ему символ.

Ясно, что следуя этому алгоритму мы в точности получим исходное сообщение S .

Метод RLE.

Наиболее известный простой подход и алгоритм сжатия информации обратимым путем - это кодирование серий последовательностей (Run Length Encoding - RLE). Суть методов данного подхода состоит в замене цепочек или серий повторяющихся байтов или их последовательностей на один кодирующий байт и счетчик числа их повторений. Проблема всех аналогичных методов заключается лишь в определении способа, при помощи которого распаковывающий алгоритм мог бы отличить в результирующем потоке байтов закодированную серию от других - не закодированных последовательностей байтов. Решение проблемы достигается обычно простановкой меток в начале закодированных цепочек. Такими метками могут быть, например, характерные значения битов в первом байте закодированной серии, значения первого байта закодированной серии и т.п. Данные методы, как правило, достаточно эффективны для сжатия растровых графических изображений (BMP, PCX, TIF, GIF), т.к. последние содержат достаточно много длинных серий повторяющихся последовательностей байтов. Недостатком метода RLE является достаточно низкая степень сжатия или стоимость кодирования файлов с малым числом серий и, что еще хуже - с малым числом повторяющихся байтов в сериях.

Практическое задание

1. Сжатие методом Хаффмана

«КАКАЯ ЗИМА ЗОЛОТАЯ!

КАК БУДТО ИЗ ДЕТСКИХ ВРЕМЕН...

НЕ НАДО НИ СОЛНЦА, НИ МАЯ –

ПУСТЬ ДЛИТСЯ ТОРЖЕСТВЕННЫЙ СОН.

ПУСТЬ Я В ЭТОМ СНЕ ПОЗАБУДУ

КОГДА-ТО МАНИВШИЙ ОГОНЬ,

И ЛЕТО ПРЕДАМ, КАК ИУДА,

ЗА ТРИДЦАТЬ СНЕЖИНОК В ЛАДОНЬ.

ЗАТЕМ, ЧТО И Я ХОЛОДЕЮ,

ТЕПЛО УЖЕ СТРАШНО ПРИНЯТЬ:

Я СЛИШКОМ ДАВНО НЕ УМЕЮ

НИ ТЛЕТЬ, НИ ГОРЕТЬ, НИ СЖИГАТЬ...

ВСЕ ЧАЩЕ, ВСЕ ДОЛЬШЕ НЕМЕЮ:

К ЗИМЕ УЖЕ ДЕЛО, К ЗИМЕ...
И ТОЛЬКО ТОГО ОТОГРЕЮ,
КОМУ ХОЛОДНЕЕ, ЧЕМ МНЕ»

2. С помощью сжатия по методу RLE.

1 последовательность:

SSSSOOOEEERROOOAAAYYYYYDDDDDOEUUUUUWWWWJJJORRUUUUUUUU
UUXXXKNNNNNNMMMMMMGGGLLLLLLLLJJJ

2 последовательность:

FFFFFFFKKKKSSSSUURERRRRRRRRRPPPPPPDDDDKKKKKGLDDDD
DDDDKGGGGMGMMMM

Отчет

Отчет должен содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Запишите какие Вы знаете методы сжатия информации?
2. Перечислите основные программы применяемые для сжатия информации?

Практическое занятие № 10

Тема программы: Сжатие информации.

Тема: Практическое применение различных алгоритмов сжатия

Цель: Целью лабораторной работы является получение навыков работы с архиваторами RAR, ARJ и ZIP, и ознакомление с основными алгоритмами сжатия информации.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, программы архиваторы: RAR, ARJ и ZIP

Теоретические основы

При эксплуатации персональных компьютеров по самым различным причинам возможны порча или потеря информации на магнитных дисках. Это может произойти из-за физической порчи магнитного диска, неправильной корректировки или случайного уничтожения файлов, разрушения информации компьютерным вирусом и т.д. Для того чтобы уменьшить потери в таких ситуациях, следует иметь архивные копии используемых файлов и систематически обновлять копии изменяемых файлов. Для хранения архивов данных можно использовать внешние запоминающие устройства большой емкости, которые дают возможность легко скопировать жесткий диск (например, магнитооптика, стримеры, "Арвид" и др.) Однако при этом резервные копии занимают столько

же места, сколько занимают исходные файлы, и для копирования нужных файлов может потребоваться много дискет. Более удобно для создания архивных копий использовать специально разработанные программы архивации файлов, которые сжимают информацию. При архивировании степень сжатия файлов сильно зависит от их формата. Некоторые форматы данных (графические, Page Maker и др.) имеют упакованные разновидности, при этом сжатие производится создающей исходный файл программой, однако лучшие архиваторы способны поджать и их. Совсем другая картина наблюдается при архивации текстовых файлов. Текстовые файлы обычно сжимаются на 50-70%, а программы на 20-30%. Принцип работы архиваторов основан на поиске в файле "избыточной" информации и последующем ее кодировании с целью получения минимального объема. Самым известным методом архивации файлов является сжатие последовательностей одинаковых символов. Например, внутри вашего файла находятся последовательности байтов, которые часто повторяются. Вместо того чтобы хранить каждый байт, фиксируется количество повторяющихся символов и их позиция. Для наглядности приведем следующий пример: Упаковываемый файл занимает 15 байт и состоит из следующей последовательности символов: **VBBBBLLLLLAAAAA**

В шестнадцатиричной системе :

42 42 42 42 42 4 С 4 С 4 С 4 С 4 С 41 41 41 41 41

Архиватор может представить этот файл в следующем шестнадцатиричном виде : 01 05 42 06 05 4 С 0А 05 41

Эти последовательности можно интерпретировать следующим образом: с первой позиции 5 раз повторяется знак В, с шестой позиции 5 раз повторяется знак L и с позиции 11 5 раз повторяется знак А. Согласитесь, очень простая демонстрация алгоритма архивации. Очевидно, что для хранения файла в его последней форме требуется лишь 9 байт - меньше на 6 байт. Описанный метод является простым и очень эффективным способом сжатия файлов. Однако он не обеспечивает большой экономии объема, если обрабатываемый текст содержит небольшое количество последовательностей повторяющихся символов.

Существуют два основных способа проведения сжатия:

- статистический
- словарный.

Лучшие статистические методы применяют арифметическое кодирование, лучшие словарные - метод Зива -Лемпела. В статистическом сжатии каждому символу присваивается код, основанный на вероятности его появления в тексте. Высоко вероятные символы получают короткие коды, и наоборот. Такой способ сжатия называют оптимальным префиксным кодом. Для его построения используют алгоритмы Хаффмана или Шеннона- Фано. Например, анализируя любой английский текст, можно установить, что буква Е встречается гораздо чаще, чем Z, а X и Q относятся к наименее встречающимся. Таким образом, используя специальную таблицу соответствия, можно закодировать каждую букву Е меньшим числом бит, используя более длинный код для более редких букв, тогда как в обычных кодировках любому символу соответствует битовая последовательность фиксированной длины (как правило, кратной байту). В словарном методе группы последовательных символов или "фраз" заменяются

кодом. Замененная фраза может быть найдена в некотором " словаре". Популярными архиваторы ARJ, RAR работают на основе алгоритма Лемпела - Зива . Сущность алгоритмов Зива и Лемпела состоит в том , что фразы заменяются указателем на то место, где они в тексте уже ранее появлялись . Это семейство алгоритмов обозначается как LZ- сжатие . Такой метод быстро приспосабливается к структуре текста и может кодировать короткие функциональные слова, т.к. они очень часто в нем появляются. Новые слова и фразы могут также формироваться из частей ранее встреченных слов. Декодирование сжатого текста осуществляется напрямую - происходит простая замена указателя готовой фразой из словаря, на которую тот указывает. На практике LZ-метод добивается хорошего сжатия, его важным свойством является очень быстрая работа декодировщика . Одной из форм такого указателя является пара (m,l) , которая заменяет фразу из l символов, начинающуюся со смещения m во входном потоке. Например, указатель $(7,2)$ адресует 7-ой и 8-ой символы исходной строки . Используя это обозначение, строка " abbaabbbabab" будет закодирована как " abba(1,3)(3,2)(8,3) ". Заметим , что несмотря на рекурсию в последнем указателе, производимое кодирование не будет двусмысленным . Распространено не верное представление, что за понятием LZ- метода стоит единственный алгоритм . Из- за большого числа вариантов этого метода лучшее описание можно осуществить только через его растущую семью , где каждый член отражает свое решение разработчика . Эти версии отличаются друг от друга в двух главных факторах : есть ли предел обратного хода указателя, и на какие подстроки из этого множества он может ссылаться.

Продвижение указателя в ранее просмотренную часть текста может быть неограниченным (расширяющееся окно) или ограничено окном постоянного размера из N предшествующих символов, где N обычно составляет несколько тысяч. Выбранные подстроки также могут быть неограниченным или ограниченным множеством фраз, выбранных согласно некоторому замыслу . Каждая комбинация этих условий является компромиссом между скоростью выполнения, объемом требуемой ОП и качеством сжатия.

Расширяющееся окно предлагает лучшее сжатие за счет организации доступа к большому количеству подстрок . Но по мере роста окна , кодировщик может замедлить свою работу из - за возрастания времени поиска соответствующих подстрок, а сжатие может ухудшиться из - за увеличения размеров указателей. Если памяти для окна будет не хватать, произойдет сброс

процесса, что также ухудшит сжатие до поры нового увеличения окна .

Окно постоянного размера лишено этих проблем , но содержит меньше подстрок, доступных указателю. Ограничение множества доступных подстрок размерами фиксированного окна уменьшает размер указателей и убыстряет кодирование.

К основным функциям архиваторов относятся :

- архивация указанных файлов или всего текущего каталога ;
- извлечение отдельных или всех файлов из архива ;
- просмотр содержимого архивного файла;
- проверка целостности архивов;

- восстановление поврежденных архивов;
- ведение многотомных архивов;
- вывод файлов из архива на экран или на печать ;
- парольная защита архива .

Архиватор ARJ не имеет графического интерфейса , и вся работа с ним осуществляется с командной строки . Формат команд имеет следующий вид :
 arj <команда > [- <спецификация 1> [- <спецификация 2>]...] <имя архива>
 [<имя файла >...]

В Подробную информацию о списке команд архиватора можно получить, набрав

командной строке: arj /?

Рассмотрим наиболее популярные команды архиватора:

Для архивации файлов: arj a <имя архива> <имя файла 1> <имя файла 2>

Для извлечения файлов из архива: arj e <имя архива> <имя файла 1> <имя файла 2>

Для просмотра содержимого архивного файла: arj l <имя архива>

Для проверки целостности архива: arj t <имя архива>

Для восстановления испорченного архива : arj -jr <имя архива>

Для создания многотомного архива: arj a -v< размер тома> <имя архива>
 <имя файла 1> <имя файла 2>

Для вывода файла из архива на экран: arj p <имя архива> <имя файла >

Для создания архива с паролем: arj a -g< пароль> <имя архива> <имя файла > или arj a -g? <имя архива> <имя файла > в последнем случае пароль будет запрошен отдельной строкой.

Для создания самораспаковывающихся архивов: arj a -je <имя архива> <имя файла 1> <имя файла 2>

Архиватор RAR имеет версии , как для Dos, Win 3.XX так и для Windows 95/98. Последние версии WinRar имеют графический интерфейс и работа с ними очень проста и понятна. Данный архиватор позволяет создавать как архивы *.rar так и архивы *.zip

К достоинствам данного архиватора можно отнести:

- графический интерфейс;
- высокую степень сжатия, даже мультимедийных файлов;
- возможность оценить размер архива, не производя архивирование.
- большую вероятность восстановления поврежденных архивов.

Практическое задание

- 1.Найдите на компьютере не менее 5-ти текстовых файлов (расширение .txt)
- 2.Произведите их сжатие архиватором RAR в обычный и SFX- архив.
- 3.Зафиксируйте размер файла до сжатия и после него.
- 4.Вычислите коэффициент сжатия (отношение размера исходного файла к размеру сжатого файла)
- 5.Повторите пункты 1-4 для графических файлов (расширение .bmp)
- 6.Повторите пункты 1-4 для графических файлов (расширение .jpg)
- 7.Повторите пункты 1-4 для звуковых файлов (расширение .wav)
- 8.Сведите полученные результаты в таблицу. Сделайте выводы о том, какие

файлы сжимаются лучше.

9. Напишите отчет о проделанной работе.

Содержание отчета

Отчет должен содержать следующие разделы:

Ответы на контрольные вопросы.

Результаты сжатия файлов в виде таблицы.

Выводы о проделанной работе.

Контрольные вопросы

1. Зачем нужно архивировать информацию ?
2. На чем основана работа архиваторов. По какому принципу они сжимают информацию.
3. Каковы функции архиваторов.
4. Чем отличаются SFХ – архивы.

Лабораторная работа №1

Тема программы: **Сжатие информации.**

Тема: **Сравнение и анализ архиваторов**

Цель: формирование практических навыков и умений архивирования и сжатия файлов.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, программы архиваторы.

Теоретические основы

Характерной особенностью большинства типов данных, с которыми традиционно работают пользователи, является определенная избыточность. Степень избыточности зависит от типа данных.

При обработке информации избыточность также играет важную роль. Так, например, при преобразовании или селекции информации избыточность используют для повышения ее качества (репрезентативности, актуальности, адекватности и т.п.). Однако, когда речь заходит не об обработке, а о хранении готовых документов или их передаче, то избыточность можно уменьшить, что дает эффект сжатия данных.

Если методы сжатия информации применяют к готовым документам, то нередко термин *сжатие данных* подменяют термином *архивация данных*, а программные средства, выполняющие эти операции, называют *архиваторами*.

Степень сжатия файлов характеризуется коэффициентом K_c , определяемым как отношение объема сжатого файла V_c к объему исходного файла V , выраженное в процентах: $K_o = (V_c / V * 100)$. Степень сжатия зависит от используемой программы, метода сжатия и типа исходного файла. Наиболее хорошо сжимаются файлы графических образов, текстовые файлы и файлы данных, для которых степень сжатия может достигать 5-40%, меньше сжимаются файлы исполняемых программ и загрузочных модулей – 60-90%. Почти не сжимаются архивные файлы.

Объекты сжатия

В зависимости от того, в каком объекте размещены данные, подвергаемые сжатию, различают:

- уплотнение (архивацию) файлов;
- уплотнение (архивацию) папок;
- уплотнение дисков.

Уплотнение файлов применяют для уменьшения их размеров при подготовке к передаче по каналам электронных сетей или к транспортировке на внешнем носителе малой емкости, например на гибком диске.

Уплотнение папок используют как средство архивации данных перед длительным хранением, в частности, при резервном копировании.

Уплотнение дисков служит целям повышения эффективности использования их рабочего пространства и, как правило, применяется к дискам, имеющим недостаточную емкость.

Несмотря на изобилие алгоритмов сжатия данных, теоретически есть только три способа уменьшения их избыточности. Это либо изменение содержания данных, либо изменение их структуры, либо и то и другое вместе.

Если при сжатии данных происходит изменение их содержания, метод сжатия необратим и при восстановлении данных из сжатого файла не происходит полного восстановления исходной последовательности. Такие методы называют также *методами сжатия с регулируемой потерей информации*. Они применимы только для тех типов данных, для которых формальная утрата части содержания не приводит к значительному снижению потребительских свойств. В первую очередь, это относится к мультимедийным данным: видеорядам, музыкальным записям, звукозаписям и рисункам. Методы сжатия с потерей информации обычно обеспечивают гораздо более высокую степень сжатия, чем обратимые методы, но их нельзя применять к текстовым документам, базам данных и, тем более, к программному коду. Характерными форматами сжатия с потерей информации являются: JPG для графических данных; .MPG для видеоданных; .MP3 для звуковых данных.

Если при сжатии данных происходит только изменение их структуры, то метод сжатия обратим. Из результирующего кода можно восстановить исходный массив путем применения обратного метода. Обратимые методы применяют для сжатия любых типов данных. Характерными форматами сжатия без потери информации являются: .GIF, .TIF, .PCX и многие другие для графических данных; .AVI для видеоданных; .ZIP, .ARJ, .RAR, .LZH, .LH, .CAB и многие другие для любых типов данных.

Архиваторы

Современные программные средства для создания и обслуживания архивов отличаются большим объемом функциональных возможностей, многие из которых выходят за рамки простого сжатия данных и эффективно дополняют стандартные средства операционной системы. В этом смысле современные средства архивации данных называют *диспетчерами архивов*.

К базовым функциям, которые выполняют современные диспетчеры архивов, относятся: извлечение файлов из архивов, создание новых архивов, добавление файлов в имеющийся архив, создание самораспаковывающихся архивов, создание распределенных архивов на носителях малой емкости, тестирование целостности структуры архивов, полное или частичное восстановление поврежденных архивов, защита архивов от просмотра и несанкционированной модификации.

К дополнительным функциям диспетчеров архивов относятся сервисные функции, делающие работу более удобной. Они часто реализуются внешним

подключением дополнительных служебных программ и обеспечивают:

- просмотр файлов различных форматов без извлечения их из архива;
- поиск файлов и данных внутри архивов;
- установку программ из архивов без предварительной распаковки;
- проверку отсутствия компьютерных вирусов в архиве до его распаковки;
- криптографическую защиту архивной информации;
- декодирование сообщений электронной почты;
- «прозрачное» уплотнение исполнимых файлов .EXE и .DLL;
- создание самораспаковывающихся многотомных архивов;
- выбор или настройку коэффициента сжатия информации.

Структура окон WinRAR и WinZip типична для приложений Windows. Вид панели инструментов WinRAR приведен на рис. 1.

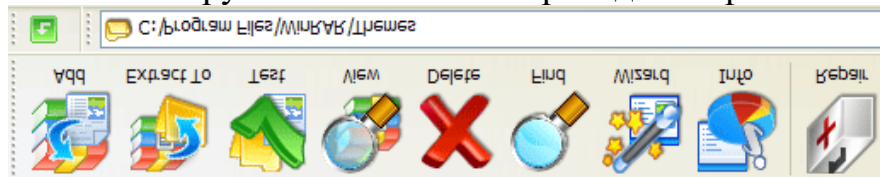


Рис. 1. Панель инструментов WinRAR

Самораспаковывающиеся архивы

В тех случаях, когда архивация производится для передачи документа потребителю, следует предусмотреть наличие у него программного средства, необходимого для извлечения исходных данных из уплотненного архива. Если таких средств у потребителя нет – создают самораспаковывающиеся архивы. Самораспаковывающийся архив готовится на базе обычного архива путем присоединения к нему небольшого программного модуля. Сам архив получает расширение .EXE, характерное для исполняемых файлов (рис. 2). Потребитель сможет выполнить его запуск как программы, после чего распаковка архива произойдет на его компьютере автоматически.

Распределенные архивы

В тех случаях, когда предполагается передача большого архива на носителях малой емкости, например на гибких дисках, возможно распределение одного архива в виде малых фрагментов на нескольких носителях. Некоторые диспетчеры (например, WinZip) выполняют разбиение сразу на гибкие диски, а некоторые (например, WinRAR) позволяют выполнить предварительное разбиение архива на фрагменты заданного размера на жестком диске. Впоследствии их можно перенести на внешние носители путем копирования.

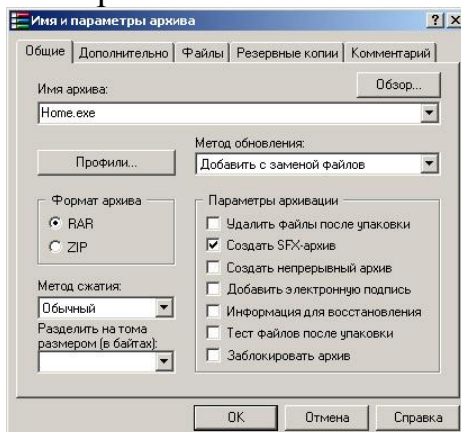


Рис. 2. Определение параметров архива

При создании распределенных архивов диспетчер WinZip обладает особенностью: каждый том несет файлы с одинаковыми именами. В результате этого нет возможности установить номера томов, хранящихся на каждом из гибких дисков, по названию файла. Поэтому каждый диск следует маркировать пометками на наклейке, а при создании распределенного архива следует быть внимательнее, чтобы не перепутать последовательность немаркированных томов.

В случае необходимости узнать номер тома можно не по названию файла, а по метке на диске, хотя эта операция не слишком удобна. Для этого следует открыть окно «Мой компьютер», выбрать значок дисковод, щелкнуть на нем правой кнопкой мыши и выбрать в контекстном меню пункт «Свойства». В диалоговом окне «Свойства: Диск ...» на вкладке «Общие» можно узнать номер тома распределенного архива в поле «Метка тома».

Консольная версия WinRAR

Консольная версия WinRAR поддерживает архивы только в формате RAR, у которых обычно расширение ".rar". ZIP и прочие форматы не поддерживаются. Пользователи Windows могут установить GUI-версию RAR – WinRAR, которая обрабатывает и архивы других типов.

Некоторые отличительные особенности RAR:

- оригинальный высокоэффективный алгоритм сжатия данных;
- специальные алгоритмы сжатия, оптимизированные для текстовых, аудио- и графических данных, а также для 32- и 64-битовых исполняемых файлов архитектуры Intel;
- лучшая, чем у аналогичных продуктов, степень сжатия при использовании режима "непрерывного" (solid) архивирования;
- электронная подпись (только в зарегистрированной версии);
- самораспаковывающиеся (SFX) архивы и тома;
- восстановление физически поврежденных архивов;
- блокировка, шифрование, задание порядка архивирования файлов;
- сохранение прав доступа к файлам, меток тома и др.

Следует отметить, что при создании томов RAR в FAT или FAT32 WinRAR автоматически ограничивает максимальный объем тома до 4 ГБ минус 1 байт, так как эти файловые системы не поддерживают файлы объемом больше 4 ГБ.

Работа с WinRAR из консоли

Синтаксис командной строки WinRAR

Формат вызова:

```
WinRAR <команда> [ -<ключи>... ] <архив> [<@файлы-списки...>]  
[<файлы...>] [ <путь_для_извлечения\> ]
```

Для создания и управления архивами служат параметры командной строки (команды и ключи). *Команда* – это строка (или одна буква), указывающая, что WinRAR должен выполнить соответствующее действие. *Ключи* модифицируют действие команды. Остальные параметры – это имена архива и файлов, которые будут добавлены или извлечены из архива.

Файлы-списки – это обычные текстовые файлы, содержащие имена файлов для обработки. Каждое имя файла должно быть указано на отдельной строке и

начинаться с первой позиции строки. В файл-список допускается помещать комментарии, признак начала комментария – символы //. Например, для архивирования файлов *.txt из каталога c:\1kurs\doc, файлов *.bmp из каталога c:\1kurs\image и всех файлов из каталога c:\evm\misc можно создать backup.lst, содержащий следующие строки:

```
c:\1kurs\doc\*.txt //резервная копия текстов
c:\1kurs\image\*.bmp //резервная копия рисунков
c:\1kurs\misc
```

После этого для архивирования достаточно будет выполнить команду:

```
winrar a backup @backup.lst
```

Если требуется прочитать имена файлов с устройства stdin (стандартный ввод), то после символа "@" не указывайте имя файла (просто @).

В одной командной строке разрешается указывать как обычные имена или группы файлов для обработки, так и файлы-списки. Если не указаны ни файлы, ни файлы-списки, то подразумевается шаблон *.* (т.е. WinRAR обрабатывает все файлы).

Команды:

a – добавляет указанные файлы к архиву;
m – переносит указанные файлы и подкаталоги в архив;
d – удаляет указанные файлы из архива;
x – извлекает указанные файлы из архива с восстановлением структуры подкаталогов;

e – извлекает указанные файлы из архива в текущий подкаталог;

v – просмотр содержимого архива;

u – добавляет те файлы к архиву, которых в нем нет;

c – добавляет комментарии к архиву;

k – защита данных от модификации.

Ключи:

-? – выводит экран помощи;

-r – сохраняет структуру подкаталогов;

-o+ – при распаковке разрешает перезаписывать существующие файлы;

-o- – при распаковке не разрешает перезаписывать существующие файлы;

-x<name> – все файлы, с соответствующими name именами, будут исключены из обработки (можно использовать шаблоны);

-x@<list> – задает файл, в котором содержатся имена файлов, исключаемых из обработки;

-v<size> – создание архивных томов;

-p<password> – назначить пароль.

Примеры команд

1). Добавить комментарий к архиву:

```
rar c distrib.rar
```

Комментарии отображаются во время обработки архива. Длина комментария не должна превышать 62000 байт.

2). Добавить комментарий из файла: `rar c -zinfo.txt dummy`

3). Записать комментарий архива в указанный файл:

```
rar cw oldarch comment.txt
```

4). Выполнить регистрозависимый поиск строки "first level" в файлах

*.txt, находящихся в архивах *.rar на диске c:.

```
rar "ic=first level" -r c:\*.rar *.txt
```

Поддерживаются следующие необязательные параметры:

i – не различать прописные и строчные буквы (по умолчанию);

c – различать прописные и строчные буквы;

h – поиск в шестнадцатеричном режиме;

t – использовать таблицы символов ANSI, Unicode и др.

Если ни один параметр не указан, вместо синтаксиса i=<строка> можно использовать более простую команду i<строка>. Модификатор 't' допускается применять вместе с другими параметрами.

5). Найти шестнадцатеричную строку f0e0aeaeab2d83e3a9 в архивах RAR, расположенных в каталоге e:\texts

```
rar ih=f0e0aeaeab2d83e3a9 -r e:\texts
```

б). Добавить к пути назначения имя архива

```
rar x -ad *.rar data\
```

Эта опция может пригодиться при распаковке группы архивов. По умолчанию RAR извлекает файлы из всех архивов в одну и ту же папку, если же указать этот ключ, то файлы из каждого архива будут распакованы в отдельные папки (в данном случае в папке 'data').

Работа с архиватором WinRAR

Получение справки о программе

Для получения справочной информации выберите команду ? Содержание. В окне Справка выберите на вкладке Содержание раздел WinRAR Interface, подраздел WinRAR menus, как показано на рис. 3.

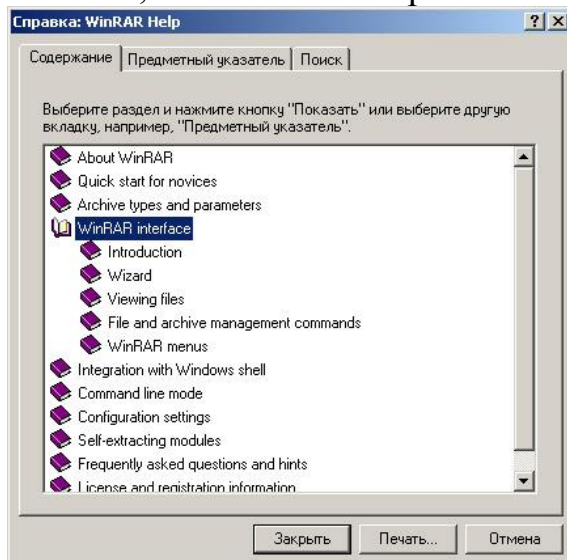


Рис. 3. Окно справки WinRAR

После запуска архиватора WinRAR на экране будет раскрыто окно, приведенное на рис. 4.



Рис. 4. Окно WinRAR в режиме операций с файлами

Окно архиватора WinRAR, в отличие от окна, WinZip, имеет средства навигации по дискам и папкам компьютера: поле списка для выбора дисков и папок, кнопку для перехода на верхний уровень в иерархии папок.

Для выбора нужного диска используйте окно списка дисков. Для выхода в родительский каталог щелкните ярлык папки с именем «..». Для открытия нужной папки щелкните ярлык с названием папки.

При проведении процессов архивации (разархивации) с группой файлов, имена которых задаются шаблонами, применяются следующие действия. Для выделения группы файлов выберите в меню *File* команду *Select group* или щелкните кнопку *Серый плюс* и задайте в окне выбора маску «*0*.**», как показано на рис. 5. Щелкнув кнопку «ОК», завершите создание маски для выбора группы файлов.

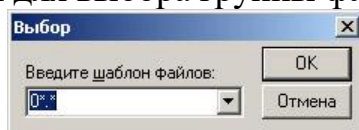


Рис. 5 Выделение группы файлов в архиве

Для создания архива из нескольких файлов, выделите нужные файлы и щелкните кнопку «Добавить» (Add) на панели инструментов (рис. 6).

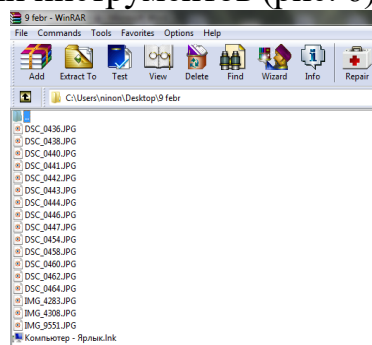


Рис. 6. Добавление выбранных файлов в архив

Для удаления из архива файла необходимо открыть архив в окне архиватора WinRAR, указать удаляемый файл и щелкнуть кнопку «Удалить» на панели инструментов или выбрать последовательность команд: *Команды-Удалить файлы*. Подтвердить удаление можно, нажав кнопку «Да» в окне подтверждения *Удаление* (рис. 7).



Рис. 7. Окно WinRAR в режиме «удаление файла из архива»

Изменение настроек программы WinRAR

Для изменения настроек выберите команду *Параметры-Установки*, после чего на экране развернется окно настройки параметров WinRAR. Выбирая различные вкладки окна *Параметры* для получения подсказки по параметрам настройки, используйте всплывающую подсказку. Задайте следующие параметры настройки WinRAR:

- на вкладке *Архивация* щелкните кнопку «Создать по умолчанию» для создания опций архивирования по умолчанию, в открывшемся после этого окне *Установить параметры сжатия по умолчанию* включите опции *Создать SFX-архив*, в списке *Размер тома* выберите стандартный размер тома сменного носителя. Щелкнув кнопку «ОК», закройте окно *Установить параметры сжатия по умолчанию* (рис. 8). Можно отредактировать значение размера тома в списке *Размер тома*, задав его величину вручную;

- на вкладке *Интеграция* включите все флажки в поле *Связать WinRAR с* и щелкните кнопку «ОК» для применения внесенных изменений. Проверьте действие измененных параметров, выделив несколько файлов и щелкнув кнопку «Добавить» на панели инструментов. После этого откроется окно *Имя и параметры архива*, в поле *Имя архива* которого выводится имя с расширением .exe (как было установлено, по умолчанию создается SFX-архив), в поле *Размер тома* отображается значение заданного по умолчанию размера тома. Щелкнув клавишу *Esc*, отмените архивацию.

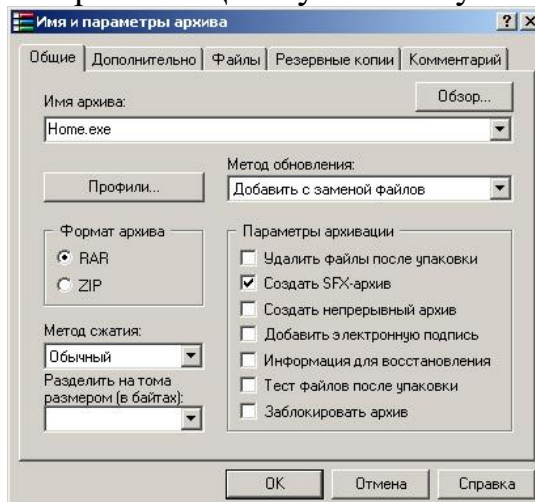


Рис. 8. Определение параметров архива

Создание многотомных архивов

Для создания многотомного архива файлов необходимо открыть окно архиватора, выбрать в поле списка дисков и папок папку, подлежащую архивации, выделить все файлы и щелкните кнопку «Добавить» на панели инструментов.

В окне *Имя и параметры архива* выбрать вкладку *Общие*. Далее в поле *Имя архива* задайте имя архива (например, Archive2.rar), выберите вариант формата архива RAR, в поле *Volume size* (Размер тома) задайте размер тома архива (например, 1.44).

Внимание: при выполнении лабораторной работы размер тома определите в несколько раз меньше суммарного объема файлов, включаемых в архив, чтобы в процессе архивации было создано нескольких томов.

Щелкнув кнопку «ОК», запустите операцию упаковки файлов в архив. По окончании архивации в текущем каталоге появится несколько файлов с именем созданного архива, с расширениями, отличающимися нумерацией, например: Archive2.rar, Archive2.r00, Archive2.r01, Archive2.r02 и т.п., где файл с расширением .rar – первый том архива, файлы с расширением .r00, .r01, .r02 и т.п. – файлы следующих томов архива.

Создание защищенных архивов

Для создания архивов, доступ к которым защищен паролем, выберите в меню *Файл* команду *Пароль*, в окне *Ввод пароля* по умолчанию в поле *Введите пароль* введите значение пароля и повторите ввод пароля в поле *Повторите пароль для проверки*. Щелкнув кнопку «ОК», завершите определение пароля. После этого в данном сеансе работы архиватора доступ ко всем создающимся архивам будет закрываться заданным паролем (рис. 9).

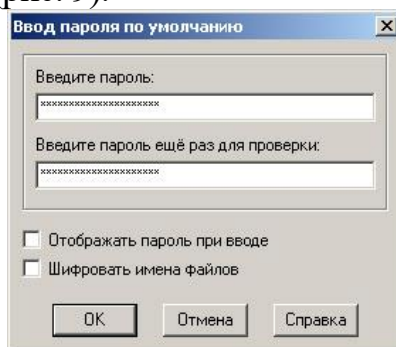


Рис. 9. Задание пароля архива

Внимание: при вводе пароля обратите внимание на включенный регистр символов.

Создайте архив из нескольких файлов в рабочем каталоге.

При извлечении файлов из защищенного паролем архива откроется окно *Ввод пароля*. Введите в поле *Введите пароль для зашифрованного файла* любое сочетание символов – неправильный пароль и щелкните кнопку «ОК». Если пароль неправильный, то раскроется окно сообщений, в котором будет выведено сообщение: *Ошибка CRC* в зашифрованном файле (неправильный пароль). Щелкнув кнопку «Закреть», закройте окно сообщения. Повторно щелкнув кнопку «Извлечь в» на панели инструментов, в окне *Ввод пароля* введите правильный пароль и щелкните кнопку «ОК». Если пароль был введен правильно, то файл будет распакован из архива.

Создание самораспаковывающегося ZIP-архива

1). Запустите программу WinZip.

2). Выполните команду File/Open Archive (Файл /Открыть архив). Откройте ранее созданный архив .zip.

3). Выполните команду Actions /Make .Exe File (Действия/Создать исполнимый файл) – откроется диалоговое окно WinZip Self-Extractor (Генератор самораспаковывающегося архива).

4). В поле Create Self-Extracting Zip files from (Создать самораспаковывающийся архив из ...) необходимо записать адрес исходного ZIP-файла. Можно воспользоваться кнопкой Browse (Обзор) для поиска нужного файла.

5). В группе Self Extractor Type (Тип самораспаковывающегося архива) включите переключатель, соответствующий операционной системе компьютера, для которого готовится архив.

6). В группе Spanning Support (Поддержка распределенного архива) включите переключатель No spanning (Без распределения) и нажмите кнопку ОК.

Создание самораспаковывающегося распределенного архива

1). Запустите программу WinZip.

2). Выполните команду File/Open Archive (Файл /Открыть архив). Откройте ранее созданный архив .zip.

3). Выполните команду Actions /Make .Exe File (Действия /Создать исполнимый файл) – откроется диалоговое окно WinZip Self-Extractor (Генератор самораспаковывающегося архива).

4). В группе элементов управления Spanning Support (Поддержка распределенного архива) включите переключатель Safe Spanning Method (Защищенный метод распределения) или Old Spanning Method (Обычный метод распределения).

Защищенный метод создает на первом гибком диске два файла: исполнимый файл, выполняющий автоматическую распаковку, и первый том распределенного архива. На последующих дисках создается продолжение распределенного архива. Такой подход повышает уровень безопасности, поскольку даже в том случае, когда исполнимый файл поврежден, например компьютерным вирусом, информация остается в архивном файле. Этот метод применяют для передачи архивных материалов на гибких дисках.

Обычный метод не создает отдельного исполнимого файла и весь архив хранится в одном исполнимом файле, распределенном по нескольким носителям. Данный метод используют для самораспаковывающихся архивов, передаваемых по каналам компьютерных сетей.

5). Откройте диалоговое окно WinZip Self-Extractor (Генератор самораспаковывающегося архива) и установите флажок Erase any existing files on the new disk before continuing (Предварительно стереть все существующие файлы на гибких дисках).

6). Далее нажмите кнопку ОК – начнется процесс создания первого тома распределенного архива. По окончании процесса по указанию программы извлеките записанный гибкий диск и вставьте новый.

7). Создав последний том, программа предложит извлечь последний диск и вставить первый для внесения правок в заголовок архива.

Альтернативные архиваторы

Среди альтернативных архиваторов можно выделить 5 программ: *Universal Extractor* – программа, служащая для извлечения данных из архивов практически

любых типов; *7-Zip* – бесплатный файловый архиватор для Windows с высокой степенью сжатия; *PeaZip* – свободный бесплатный архиватор и графическая оболочка для других архиваторов; *IZArc* – бесплатный архиватор для Windows, поддерживающий большое количество форматов; *TUGZip* – простой в использовании архиватор, поддерживающий большое количество форматов. Среди перечисленных архиваторов лидирующие позиции занимает *7-Zip*. По степени сжатия он является лучшим не только среди бесплатных программ, но и подавляющего большинства коммерческих продуктов. *7-Zip* работает со всеми популярными форматами архивов, поддерживает шифрование, умеет создавать самораспаковывающиеся архивы и обладает многими другими удобными функциями. К недостаткам *7-Zip* можно отнести сравнительно малое количество поддерживаемых форматов. Программа *IZArc* умеет распаковывать около 50 типов архивов, включая многие редкие. Также он может архивировать и сохранять файлы в 12 различных форматах и обрабатывать многотомные ZIP-архивы. Мультиформатный архиватор *TUGZip* имеет некоторые специальные возможности, например, восстановление поврежденных архивов ZIP и SQX. *PeaZip* – небольшой, бесплатный архиватор с открытыми кодами, как и *IZArc* поддерживает множество форматов архивов, включая ACE, ARJ, CAB, DMG, ISO, LHA, RAR, и UDF, работает как с 32, так и с 64-битными версиями Windows. *Universal Extractor* нельзя назвать настоящим архиватором, так как сжимать файлы он не умеет, но является наилучшим распаковщиком редких форматов. Огромное количество поддерживаемых форматов делает его лучшим в этом секторе

Интеграция служебных и прикладных программ с ОС

Под интеграцией программного обеспечения понимают возможность совместной работы нескольких различных программ в рамках единой системы управления. Так, например, известным системным средством интеграции является концепция внедрения и связывания объектов и основанный на ней буфер обмена Windows. Другим приемом интеграции, в основе которого лежит изменение свойств программы Проводник и связанного с ней контекстного меню объектов. Для эпизодических работ по архивации и извлечению файлов и папок удобнее использовать систему, хорошо интегрированную в Windows, например, *WinZip*. Для регулярных работ по созданию резервных копий папок и дисков удобнее использовать автономные средства, поскольку для них проще организуется взаимодействие с прочими программами (в частности, со средствами автоматизации). В этих случаях можно рекомендовать, например, программу *WinRAR*.

- 1). Запустите программу «Проводник» (Пуск / Программы / Проводник).
- 2). Скопируйте в созданную папку несколько произвольных файлов.
- 3). Выделите один из файлов и откройте контекстное меню. Обратите внимание на то, что в нем имеются два пункта для создания архива (создание архива с произвольным именем и с именем, соответствующим текущему файлу). Появление этих пунктов связано с наличием в компьютерной системе диспетчера архивов и интеграции *WinZip* с Проводником Windows.

- 4). Выполните команду Add to Zip (Добавить в архив). Далее произойдет автоматический запуск диспетчера архивов *WinZip* и откроется диалоговое окно Add (Добавление в архив).

5). В поле Add to archive (Добавить в архив) ввести название файла создаваемого архива, адрес текущей папки заносится автоматически. Проверив настройку прочих элементов управления, запустите процесс архивации щелчком на командной кнопке Add (Добавить).

6). Перейдите в окно программы Проводник и убедитесь в том, что в папке появился архивный файл test.zip. Щелкните на значке архивного файла правой кнопкой мыши и изучите новые команды контекстного меню, позволяющие выполнить операции с архивным файлом.

7). Выполните команду Create Self-Extractor (Создать самораспаковывающийся архив). В открывшемся диалоговом окне щелкните на командной кнопке «Да» и в последующих диалоговых окнах откажитесь от проверки созданного архива.

8). Закройте открытые окна программы WinZip и в программе Проводник убедитесь в том, что в рабочей папке появился исполняемый файл (.exe).

9). В программе Проводник выполните перетаскивание значка любого файла (или группы файлов) на значок созданного ZIP-архива. При отпускании кнопки мыши в конце перетаскивания происходит автоматическое добавление новых файлов в архив. Если содержимое правой панели Проводника открыто в режиме Таблица, после каждого перетаскивания можно наблюдать увеличение размера файла архива.

Исследование свойств форматов сжатия графических данных

- 1). Откройте графический редактор Paint (Пуск/Программы/Стандартные/ Paint).
- 2). Загрузите в него заранее подготовленный многоцветный рисунок.
- 3). Определите размер рисунка в пикселях (Рисунок/Атрибуты).
- 4). Оцените теоретический размер рисунка в 24-разрядной палитре (3 байта на точку) по формуле:

$S=M \cdot N \cdot 3$, где S – размер файла с рисунком (байт);

M – ширина рисунка (точек);

N – высота рисунка (точек).

5). Сохраните рисунок в папку C:\Temp\Pictures, выбрав имя файла test и назначив тип файла: 24-разрядный рисунок (.BMP).

6). Повторно сохраните рисунок, выбрав то же имя test, но назначив тип файла .GIF. При сохранении произойдет потеря определенной части графической информации.

7). Восстановите рисунок, загрузив его из ранее сохраненного файла Test.bmp.

8). Вновь сохраните его под тем же именем, но выбрав в качестве типа файла формата .JPEG.

9). Запустите программу Проводник.

10). Откройте папку C:\Temp\Pictures в режиме Таблица.

11). Определите размеры файлов Test.bmp, Test.gif и Test.jpg.

12). Определите коэффициент сжатия файлов (R), взяв отношения размеров файлов к теоретической величине, полученной расчетным путем.

Порядок выполнения работы

1) Создать или скопировать на рабочем диске в рабочей директории 5-7 файлов (текстовых, исполняемых, командных, программных).

2) Создать архивы для этих файлов с помощью различных архиваторов, например, WinRar, WinZip и др.

3) Сравнить объемы получившихся файлов, результаты занести в таблицу и

сделать выводы:

Название архиватора	Тип файла	Размер файла	Размер файла после сжатия	Степень сжатия(%)

4) С помощью архиватора (в соответствии с заданием преподавателя) выполнить следующие команды:

- а) добавить в архив заданный файл;
- б) поместить в архив все файлы из текущего каталога, за исключением файлов с заданным расширением;
- в) создать защищенный архив;
- г) создать архивный файл, позволяющий сохранить структуру каталогов;
- д) добавить комментарии к архивам;
- е) извлечь заданный файл из архива.
- ж) создать многотомный архив, указав размер тома – 80 К;
- з) выполнить поиск заданной строки в архивах по различным поисковым признакам.

5) Используя программу архивации, создать на диске, заданном в параметрах, многотомный архив с паролем, заданным в параметрах, поместив в них все файлы из каталога LAB рабочего диска, исключив файлы с расширением EXE.

6) **Просмотреть списки созданных архивов.**

7) Создать командный файл, который с помощью архиватора позволяет расположить файлы в архиве в заданном порядке, просмотреть архив, извлечь файлы из архива в заранее созданный каталог.

8) Создать самораспаковывающиеся RAR- и ZIP-архивы, не поддерживающие распределенные архивы (включить переключатель «Без распределения» в группе Spanning Support – Поддержка распределенного архива).

9) Создать самораспаковывающиеся распределенные архивы RAR- и ZIP-архивы.

10) Используя диспетчер архивов WinZip, выполнить интеграцию служебных и прикладных программ с операционной системой Windows.

11) Исследуйте свойства форматов сжатия графических данных (файлы .bmp, .gif, .jpg). Результаты занесите в таблицу:

Формат файла	Размер файла (Кбайт)	Степень сжатия (%)
24 разрядный .bmp		
.gif		
.jpg		

12) Используя программу, например, Excel, построить диаграммы по результатам, приведенным в таблицах, и сделать выводы.

Содержание отчета

Отчет должен содержать следующие разделы:

Ответы на контрольные вопросы.

Результаты сжатия файлов в виде таблицы.

Выводы о проделанной работе.

Контрольные вопросы

1. Для чего необходимо создавать архив?
2. Поясните основные алгоритмы архивации.
3. Как можно упаковать информацию при хранении на диске?
4. Приведите команды упаковки данных в архив и распаковки данных из архива для архиватора Winrar.exe в консольном режиме.
5. Как создать защищенный архив?
6. Приведите команды упаковки данных в архив Winzip.exe и распаковки данных из архива.
7. Как создать многотомный архив?
8. Укажите расширение имен файлов продолжения архива.
9. Как получить полную справку по всем возможным режимам работы программы-архиватора?
10. Как создать самораспаковывающийся архив?
11. Приведите примеры альтернативных программ архивации.
12. В чем особенность альтернативных программ архивации.
13. Что понимается под интеграцией служебных и прикладных программ с ОС?

Практическое занятие № 11

Тема программы: **Сжатие информации.**

Тема: **Итоговое практическое занятие**

Цель: Закрепление теоретических знаний о методе сжатия сообщений с использованием динамических словарей (алгоритм LZ)

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, программы архиваторы.

Теоретические основы

Сообщения, включающие текст, числовые данные код программ и т.п. не допускают потери информации. Наиболее распространенным средством их сжатия является метод динамических словарей. В частности, в современных программах архиваторов широко используется алгоритм Лемпеля-Зива (LZ), основанный на этом методе. Применение динамических словарей позволяет эффективно сжимать повторяющиеся цепочки знаков, независимо от того, являются ли они однородными. Описание алгоритма LZ приведено в электронном конспекте лекций (тема 8). Для разных типов сообщений – и, соответственно, форматов файлов, - характерны различные вероятности и длины повторяющихся цепочек знаков. В связи с этим эффективность их сжатия может существенно различаться. Например, текстовые данные обычно сжимаются в 2-3 раза, сжатие табличных данных может достигать 8-10 раз, в то же время, изображения в формате jpeg почти не сжимаются LZ-архиватором.

Эта эффективность зависит также от некоторых настраиваемых параметров

алгоритма, в частности, используемой длины словаря. В современных архиваторах такого рода настройка как правила выполняется автоматически – с учетом типа обрабатываемых файлов.

Содержание работы

В настоящей работе исследуется сжатие файлов с помощью популярного архиватора WinRAR, в котором реализован алгоритм LZ.

Исследуется степень сжатия файлов различных типов, а также влияние настраиваемых параметров алгоритма сжатия.

Для этого используются следующие типы файлов:

- текст в формате doc и txt;
- табличные данные в формате htm;
- изображения в форматах bmp и jpeg;
- звуковые файлы формата wav.

Оценка сжатия звуковых файлов и изображений в последствие будет использована для сравнения методов сжатия этих типов данных без потерь информации и с потерями (со снижением качества).

Порядок выполнения работы

Используя программу WinRAR, выполнить исследование степени сжатия файлов разных типов при автоматически выбираемых настройках архиватора (режим сжатия “обычный”). Результаты представить в виде табл.1.

Объяснить различия в степени сжатия различных типов файлов исходя из того, как организованы данные в них. Сформулировать и записать соответствующие выводы.

Таблица 1

Исследование степени сжатия файлов разных типов

Имя и тип файла	Исходный объем (байт)	Объем после сжатия (байт)	Степень сжатия (%)
.txt			
.doc			
.htm			
.wav			
.bmp			
.jpg			

С помощью WinRAR, выполнить исследование степени сжатия файлов в различных режимах сжатия. В частности, использовать следующие параметры:

- режим (“метод”) сжатия: обычный, скоростной, максимальный (для файла, указанного преподавателем);
- объем используемой памяти для сжатия doc-файла (режим Дополнительно-Параметры сжатия-Сжатие текста-Принудительно) – опробовать три различных значения;
- длина словаря для сжатия bmp-файла (режим Дополнительно-Параметры сжатия-Сжатие полноцветной графики-Принудительно) – опробовать три различных значения.

Результаты испытаний поместить в табл..2

Таблица 2

Исследование режимов сжатия файлов

Имя и тип файла	Режим сжатия	Исходный объем (байт)	Степень сжатия (%)

Объяснить зафиксированное влияние параметров сжатия. Сделать вывод об эффективности автоматического выбора параметров сжатия в WinRAR.

Содержание отчета

Отчет должен содержать следующие разделы:

- Ответы на контрольные вопросы.
- Результаты сжатия файлов в виде таблицы.
- Выводы о проделанной работе.

Контрольные вопросы

1. Какого формата лучше сжимается файл?
2. Какие программы применяются для сжатия файлов?
3. Какие методы применяются для сжатия графических файлов?

Практическое занятие № 12

Тема программы: **Арифметическое кодирование.**

Тема: **Практическое применение алгоритмов кодирования**

Цель: Познакомиться с различными кодировками символов, используя текстовые редакторы, выполнить задания в различных текстовых приложениях.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, текстовые редакторы.

Теоретические основы

Правило цифрового представления символов следующее: каждому символу ставится в соответствие некоторое целое число, то есть каждый символ нумеруется.

Пример:

Рассмотрим последовательность строчных букв русского алфавита: а, б, в, г, д, е, ё, ж, з, и, й, к, л, м, н, о, п, р, с, т, у, ф, х, ц, ч, ш, щ, ь, ы, в, э, ю, я. Присвоив каждой букве номер от 0 до 33, получим простейший способ представления символов. Последнее число - 32 в двоичной форме имеет вид 100000, то есть для хранения символа в памяти понадобится 6 бит. Так как с помощью шести бит можно представить число $2^6 - 1 = 63$, то шести бит будет достаточно для представления 64 букв.

Имеются разные стандарты для представления, символов, которые отличаются лишь порядком нумерации символов. Наиболее распространён американский стандартный код для информационного обмена - ASCII [American Standard-Code for Information Interchange] введён в США в 1963г. В 1977 году в несколько модифицированном виде он был принят в качестве всемирного стандарта

Международной организации стандартов [International Standards Organization -. ISO] под названием ISO-646. Согласно этому стандарту каждому символу поставлено в соответствие число от 0 до 255. Символы от 0 до 127 - латинские буквы, цифры и знаки препинания - составляют постоянную часть таблицы. Остальные символы используются для представления национальных алфавитов. Конкретный состав этих символов определяется кодовой страницей. В русской версии ОС Windows95 используется кодовая, страница 866. В ОС Linux для представления русских букв более употребительна кодировка КОИ-8. Недостатки такого способа кодировки национального, алфавита очевидны. Во-первых, невозможно одновременное представление русских и ,например, французских букв. Во-вторых, такая кодировка совершенно непригодна для представления, китайских иероглифов. В 1991 году была создана некоммерческая организация Unicode, в которую входят представители ряда фирм (Borland. IBM, Noyell, Sun и др) и которая занимается развитием и внедрением нового стандарта. Кодировка Unicode использует 16 разрядов ,и может содержать 65536 символов. Это символы большинства народов мира, элементы иероглифов, спецсимволы, 5000 – мест для частного использования, резерв из 30000 мест.

Пример:

ASCII-код символа A= $65_{10} = 41_{16} = 01000111_2$;

Unicode-код символа C= $67_{10} = 0000000001100111_2$

Задания

1. Закодируйте свое имя, фамилию и отчество с помощью одной из таблиц (win-1251, КОИ-8)

2. Раскодируйте ФИО соседа

3. Закодируйте следующие слова, используя таблицы ASCII-кодов:
ИНФОРМАТИЗАЦИЯ, МИКРОПРОЦЕССОР, МОДЕЛИРОВАНИЕ

4. Раскодируйте следующие слова, используя таблицы ASCII-кодов:

88 AD E4 AE E0 AC A0 E2 A8 AA A0

50 72 6F 67 72 61 6D

43 6F 6D 70 75 74 65 72 20 49 42 4D 20 50 43

5. Текстовый редактор Блокнот

Открыть блокнот.

а) Используя клавишу Alt и малую цифровую клавиатуру раскодировать фразу:
145 170 174 224 174 255 170 160 173 168 170 227 171 235;

Технология выполнения задания: При удерживаемой клавише Alt, набрать на малой цифровой клавиатуре указанные цифры. Отпустить клавишу Alt, после чего в тексте появится буква, закодированная набранным кодом.

б) Используя ключ к кодированию, закодировать слово – зима;

Технология выполнения задания: Из предыдущего задания выяснить, каким кодом записана буква а. Учтя, что буквы кодируются в алфавитном порядке, выяснить коды остальных букв.

Что вы заметили при выполнении этого задания во время раскодировки? Запишите свои наблюдения.

6. Текстовый процессор.

Технология выполнения задания: рассмотрим на примере: представить в

различных кодировках слово Кодировка

Решение:

- Создать новый текстовый документ в текстовом редакторе;
- Выбрать – Команда – Вставка – Символ.

В открывшемся окне «Символ» установить из: Юникод (шестн.),

• В наборе символов находим букву **К** и щелкнем на ней левой кнопкой мыши (ЩЛКМ).

• В строке код знака появится код выбранной буквы 041A (незначащие нули тоже записываем).

• У буквы **о** код – 043E и так далее: д – 0434, и – 0438, р – 0440, о – 043E, в – 0432, к – 043A, а – 0430.

• Установить Кириллица (дес.)

• К – 0202, о – 0238, д – 0228, и – 0232, р – 0240, о – 0238, в – 0226, к – 0202, а – 0224.

7. Открыть Текстовый редактор

Используя окно «Вставка символа» выполнить задания: Закодировать слово **Forest**

а) Выбрать шрифт Courier New, кодировку ASCII(дес.) Ответ: **70 111 114 101 115 116**

б) Выбрать шрифт Courier New, кодировку Юникод(шест.) Ответ: **0046 006F 0072 0665 0073 0074**

в) Выбрать шрифт Times New Roman, кодировку Кириллица(дес.) Ответ: **70 111 114 101 115 116**

г) Выбрать шрифт Times New Roman, кодировку ASCII(дес.) Ответ: **70 111 114 101 115 116**

Вывод: _____

Выполнение лабораторной работы оформить в виде таблицы.

8. Буква **Z** имеет десятичный код 90, а **z** – 122. Записать слово «sport» в десятичном коде.

9. С помощью десятичных кодов зашифровано слово «info» 105 110 102 111. Записать последовательность десятичных кодов для этого же слова, но записанного заглавными буквами.

10. Буква **Z** имеет десятичный код 90, а **z** – 122. Записать слово «forma» в десятичном коде.

11. С помощью десятичных кодов зашифровано слово «port» 112 111 114 116. Записать последовательность десятичных кодов для этого же слова, но записанного заглавными буквами. Ответ: **80 79 82 84**

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Какие возможности предоставляет текстовые редакторы по работе с символами?
2. Какие вы знаете алгоритмы кодирования информации?
3. Где применяется алгоритм кодирования информации?

Практическое занятие № 12

Тема программы: **Арифметическое кодирование.**

Тема: **Кодирование информации**

Цель: познакомиться с различными кодировками звуковой информации и с характеристиками звуковых файлов.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, звуковые редакторы.

Теоретические основы

С начала 90-х годов персональные компьютеры получили возможность работать со звуковой информацией. Каждый компьютер, имеющий звуковую плату, может сохранять в виде файлов (файл - это определённое количество информации, хранящееся на диске и имеющее имя) и воспроизводить звуковую информацию. С помощью специальных программных средств (редакторов аудио файлов) открываются широкие возможности по созданию, редактированию и прослушиванию звуковых файлов. Создаются программы распознавания речи, и появляется возможность управления компьютером голосом.

Именно звуковая плата (карта) преобразует аналоговый сигнал в дискретную фонограмму и наоборот, «оцифрованный» звук – в аналоговый (непрерывный) сигнал, который поступает на вход динамика.

Схема записи звука:

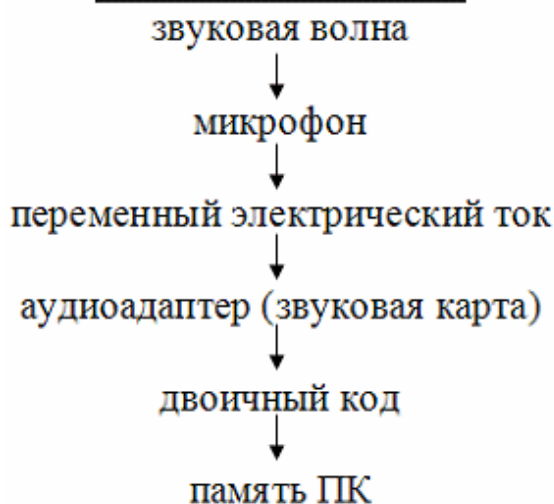
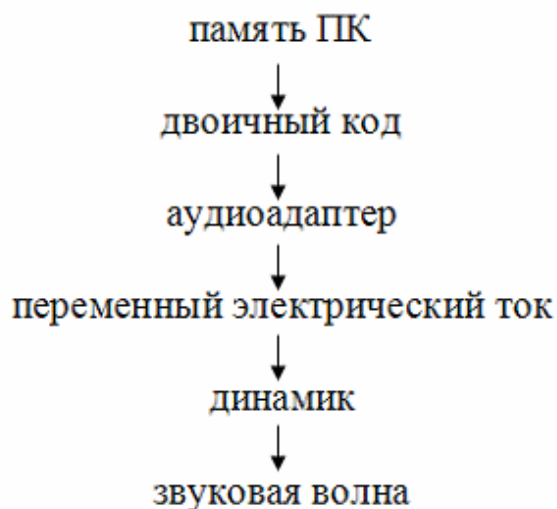


Схема воспроизведения звука:



При двоичном кодировании аналогового звукового сигнала непрерывный сигнал дискретизируется, т.е. заменяется серией его отдельных выборок - отсчётов. Качество двоичного кодирования зависит от двух параметров: количества дискретных уровней сигнала и количества выборок в секунду. Количество выборок или частота дискретизации в аудиоадаптерах бывает различной: 11 кГц, 22 кГц, 44,1 кГц и др. Если количество уровней равно 65536, то на один звуковой сигнал рассчитано 16 бит (216). 16-разрядный аудиоадаптер точнее кодирует и

воспроизводит звук, чем 8-разрядный.

Количество бит, необходимое для кодирования одного уровня звука, называется глубиной звука. Объем моноаудиофайла (в байтах) определяется по формуле:

$$V_{\text{моно}} = \frac{\nu \cdot t \cdot G}{8},$$

где ν - частота дискретизации в Гц,

G – глубина звука в битах, t – время в секундах.

При стереофоническом звучании объем аудиофайла удваивается, при квадрофоническом звучании – учетверяется.

По мере усложнения программ и увеличения их функций, а также появления мультимедиа-приложений, растёт функциональный объем программ и данных. Если в середине 80-х годов обычный объем программ и данных составлял десятки и лишь иногда сотни килобайт, то в середине 90-х годов он стал составлять десятки мегабайт. Соответственно растёт объем оперативной памяти.

Пример решения: Подсчитать, сколько места будет занимать одна минута цифрового звука на жестком диске или любом другом цифровом носителе, записанного с частотой

а) 44.1 кГц;

б) 11 кГц;

в) 22 кГц;

г) 32 кГц

и разрядностью 16 бит.

Решение.

а) Если записывают моносигнал с частотой 44.1 кГц, разрядностью 16 бит (2 байта), то каждую минуту аналого-цифровой преобразователь будет выдавать $441000 \cdot 2 \cdot 60 = 529\,000$ байт (около 5 Мб) данных об амплитуде аналогового сигнала, который в компьютере записываются на жесткий диск.

Если записывают стереосигнал, то 1 058 000 байт (около 10 Мб).

Задания

1. Какой объем памяти требуется для хранения цифрового аудиофайла с записью звука высокого качества при условии, что время звучания составляет 3 минуты?

2. Какой объем данных имеет моноаудиофайл, длительность звучания которого 1 секунда, при среднем качестве звука (16 бит, 24 кГц)?

3. Рассчитайте объем стереоаудиофайла длительностью 20 секунд при 20-битном кодировании и частоте дискретизации 44.1 кГц. Варианты: 44,1 Мб, 4.21 Мб, 3,53 Мб.

4. Оцените информационный объем моноаудиофайла длительностью звучания 20 с, если "глубина" кодирования и частота дискретизации звукового сигнала равны соответственно 8 бит и 8 кГц;

5. Рассчитайте время звучания моноаудиофайла, если при 16-битном кодировании и частоте дискретизации 32 кГц его объем равен 700 Кбайт;

6. Запишите звуковой моноаудиофайл длительностью 20 с, с "глубиной" кодирования 8 бит и частотой дискретизации 8 кГц.

7. Определите качество звука (качество радиотрансляции, среднее качество,

качество аудио-CD) если известно, что объем стереоаудиофайла длительностью звучания в 10 сек. Равен 940 Кбайт;

8. Оцените информационный объем стереоаудиофайла длительностью звучания 30 с, если "глубина" кодирования и частота дискретизации звукового сигнала равны соответственно 8 бит и 8 кГц;

9. Запишите звуковой файл длительностью 30с с "глубиной" кодирования 8бит и частотой дискретизации 8 кГц. Вычислите его объем и сверьтесь с полученным на практике значением.

10. Аналоговый звуковой сигнал был дискретизирован сначала с использованием 256 уровней интенсивности сигнала (качество звучания радиотрансляции), а затем с использованием 65536 уровней интенсивности сигнала (качество звучания аудио-CD). Во сколько раз различаются информационные объемы оцифрованного звука?

11. Оцените информационный объем моноаудиофайла длительностью звучания 1 мин. если "глубина" кодирования и частота дискретизации звукового сигнала равны соответственно:

16 бит и 48 кГц.

12. Запишите звуковой моноаудиофайл длительностью 1 минута с "глубиной" кодирования 16 бит и частотой дискретизации 48 кГц.

13. Подсчитать объем файла с 10 минутной речью записанного с частотой дискретизации 11025 Гц при 4 разрядном кодировании

14. Подсчитать время звучания звукового файла объемом 3.5 Мбайт содержащего стерео запись с частотой дискретизации 44100 Гц, 16-ти разрядном кодировании.

15. Определите количество уровней звукового сигнала при использовании 8-битных звуковых карт. Варианты: 256, 512, 1024, 65 536.

16. Приведите пример:

а) аналогового способа представления звуковой информации;

б) дискретного способа представления звуковой информации.

17. Подготовить презентацию, демонстрирующую возможности звуковых форматов midi, wav, mp3, mod.

18. Перечислите параметры, от которых зависит качество двоичного кодирования звука.

Отчет

Отчет должен быть оформлен в текстовом редакторе и содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. С какими звуковыми форматами вы встречаетесь чаще в повседневной жизни?
2. Дайте определение аудиоадаптеру?
3. Что значит оцифровка звука?

Практическое занятие № 13
Тема программы: **Арифметическое кодирование**
Тема: **Кодирование информации**

Цель: научиться кодировать растровые графические файлы; научиться измерять информационный объем графических файлов.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, графические редакторы.

Теоретические основы

Графическая информация на экране дисплея представляется в виде изображения, которое формируется из точек (пикселей). Вспомните газетную фотографию, и вы увидите, что она тоже состоит из мельчайших точек. Если это только чёрные и белые точки, то каждую из них можно закодировать 1 битом. Но если на фотографии оттенки, то два бита позволяют закодировать 4 оттенка точек: 00 - белый цвет, 01 - светло-серый, 10 - тёмно-серый, 11 - чёрный. Три бита позволяют закодировать 8 оттенков и т.д.

Количество бит, необходимое для кодирования одного оттенка цвета, называется глубиной цвета.

$$K=2^G, \text{ где } K - \text{ количество оттенков, } G - \text{ глубина цвета в битах.}$$

В современных компьютерах разрешающая способность (количество точек на экране), а также количество цветов зависит от видеоадаптера и может изменяться программно.

Цветные изображения могут иметь различные режимы: 16 цветов, 256 цветов, 65536 цветов (high color), 16777216 цветов (true color). На одну точку для режима high color необходимо 16 бит или 2 байта.

Наиболее распространённой разрешающей способностью экрана является разрешение 800 на 600 точек, т.е. 480000 точек. Рассчитаем необходимый для режима high color объём видеопамати: 2 байт * 480000 = 960000 байт.

Для измерения объёма информации используются и более крупные единицы:

$$\begin{aligned} 1 \text{ Кбайт (один килобайт)} &= 2^{10} \text{ байт} = 1024 \text{ байт} \\ 1 \text{ Мбайт (один мегабайт)} &= 2^{20} \text{ байт} = 1048576 \text{ байт} \\ 1 \text{ Гбайт (один гигабайт)} &= 2^{30} \text{ байт} \approx 1 \text{ млрд. байт} \end{aligned}$$

Следовательно, 960000 байт приблизительно равно 937,5 Кбайт. Если человек говорит по восемь часов в день без перерыва, то за 70 лет жизни он наговорит около 10 гигабайт информации (это 5 миллионов страниц - стопка бумаги высотой 500 метров).

Скорость передачи информации - это количество битов, передаваемых в 1 секунду. Скорость передачи 1 бит в 1 секунду называется 1 бод.

$$1 \text{ Кбод} = 1024 \text{ бит/сек}; 1 \text{ Мбод} = 1024 \text{ Кбод}; 1 \text{ Гбод} = 1024 \text{ Мбод}$$

В видеопамяти компьютера хранится битовая карта, являющаяся двоичным кодом изображения, откуда она считывается процессором (не реже 50 раз в секунду) и отображается на экран.

Таблица. Объём видеопамяти в зависимости от типов разрешающей способности компьютеров.

Разрешение	16 цветов	256 цветов	65536 цветов	16777216 цветов
640 x 480	150 Кб	300 Кб	600 Кб	900 Кб
800 x 600	234,4 Кб	468,8 Кб	937,5 Кб	1,4 Мб
1024 x 768	384 Кб	768 Кб	1,5 Мб	2,25 Мб
1280x 1024	640 Кб	1,25 Мб	2,5 Мб	3,75 Мб

Задачи:

1. Известно, что видеопамять компьютера имеет объем 512 Кбайт. Разрешающая способность экрана 640 на 200. Сколько страниц экрана одновременно разместится в видеопамяти при палитре: а) из 8 цветов, б) 16 цветов; в) 256 цветов?

2. Сколько бит требуется, чтобы закодировать информацию о 130 оттенках?

3. Подумайте, как уплотнить информацию о рисунке при его записи в файл, если известно, что: а) в рисунке одновременно содержится только 16 цветовых оттенков из 138 возможных; б) в рисунке присутствуют все 130 оттенков одновременно, но количество точек, закрасенных разными оттенками, сильно различаются.

4. Найдите в сети Интернет информацию на тему «Цветовые модели HSB, RGB, CMYK» и создайте на эту тему презентацию. В ней отобразите положительные и отрицательные стороны каждой цветовой модели, принцип ее функционирования и применение.

5. В приложении «Точечный рисунок» создайте файл размером (по вариантам):

- А) 200*300, (№ по списку 1, 8, 15, 22, 29)
- Б) 590*350, (№ по списку 2, 9, 16, 23, 30)
- В) 478*472, (№ по списку 3, 10, 17, 24, 31)
- Г) 190*367, (№ по списку 4, 11, 18, 25, 32)
- Д) 288*577; (№ по списку 5, 12, 19, 26, 33)
- Е) 100*466, (№ по списку 5, 13, 20, 27, 34)
- Ж) 390*277. (№ по списку 6, 14, 21, 28)

Сохраните его под следующими расширениями:

- монохромный рисунок,
- 16-цветный рисунок,
- 256-цветный рисунок,
- 24-битный рисунок,
- формат JPG.

Используя информацию о размере каждого из полученных файлов, вычислите количество используемых цветов в каждом из файлов, проверьте с

полученным на практике. Объясните, почему формула расчета количества цветов не подходит для формата JPG. Для этого воспользуйтесь информацией из сети Интернет.

6. На бумаге в клетку (или в приложении табличный редактор) нарисуйте произвольный рисунок 10*10 клеток. Закодируйте его двоичным кодом (закрашена клетка – 1, не закрашена - 0). Полученный код отдайте однокласснику для декодирования и получения изображения.

Отчет

Отчет должен содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Влияет разрешение рисунка на объем файла?
2. Чем отличаются цветовые модели?
3. Что означает глубина цвета?

Практическое занятие № 14

Тема программы: **Арифметическое кодирование**

Тема: **Декодирование информации.**

Цель работы: Ознакомление с правилами и получение практических навыков кодирования и декодирования информации.

Время выполнения: 2 часа

Оборудование: ПК.

Программное обеспечение: операционная система, графические редакторы.

Теоретические основы

Кодирование информации это преобразование формы представления информации с целью ее передачи или хранения. Кодирование это представление символов одного алфавита символами другого. Правила, по которым осуществляется кодирование называются кодом. Под словом понимают последовательность символов, количество символов в которой называется длиной кода. Слова так же называют кодовыми комбинациями. Если при кодировании получают комбинации одинаковой длины, то такой код называют равномерным, а длину кодовых комбинаций в этом слове называют значимостью кода. Если кодовые комбинации различной длины, то код называется неравномерным.

Процесс обратный кодированию называется декодированием.

Если в коде ни одна более короткая комбинация не является началом более длинной кодовой комбинации, то код называется префиксным.

- кодирование – это перевод информации с одного языка на другой (запись в другой системе символов, в другом алфавите)
- обычно кодированием называют перевод информации с «человеческого» языка на формальный, например, в двоичный код, а декодированием – обратный переход
- один символ исходного сообщения может заменяться одним символом нового кода или несколькими символами, а может быть и наоборот – несколько символов исходного сообщения заменяются одним символом в новом коде (китайские иероглифы обозначают целые слова и понятия)
- кодирование может быть *равномерное* и *неравномерное*;
при равномерном кодировании все символы кодируются кодами равной длины;
при неравномерном кодировании разные символы могут кодироваться кодами разной длины, это затрудняет декодирование

Пример задания:

Для кодирования букв А, Б, В, Г решили использовать двухразрядные последовательные двоичные числа (от 00 до 11, соответственно). Если таким способом закодировать последовательность символов БАВГ и записать результат шестнадцатеричным кодом, то получится

- 1) $4B_{16}$ 2) 411_{16} 3) $BACD_{16}$ 4) 1023_{16}

Решение:

- 1) из условия коды букв такие: А – 00, Б – 01, В – 10 и Г – 11, код равномерный
- 2) последовательность БАВГ кодируется так: 01 00 10 11 = 1001011
- 3) разобьем такую запись на тетрады справа налево и каждую тетраду переведем в шестнадцатеричную систему (то есть, сначала в десятичную, а потом заменим все числа от 10 до 15 на буквы А, В, С, D, E, F); получаем
 $1001011 = 0100 \ 1011_2 = 4B_{16}$
- 4) правильный ответ – 1.

Возможные ловушки:

- расчет на то, что при переводе тетрад в шестнадцатеричную систему можно забыть заменить большие числа (10–15) на буквы ($1011_2 = 11$, получаем неверный ответ 411_{16})
- может быть дан неверный ответ, в котором нужные цифры поменяли местами (расчет на невнимательность), например, $B4_{16}$
- в ответах дана последовательность, напоминающая исходную (неверный ответ $BACD_{16}$), чтобы сбить случайное угадывание

Пример задания:

Для 5 букв латинского алфавита заданы их двоичные коды (для некоторых букв – из двух бит, для некоторых – из трех). Эти коды представлены в таблице:

A	B	C	D	E
---	---	---	---	---

000	01	100	10	011
-----	----	-----	----	-----

Определить, какой набор букв закодирован двоичной строкой 0110100011000

1) BVCEA 2) BDDEA 3) BDCEA 4) EBAEA

Решение (вариант 1, декодирование с начала):

- 1) здесь используется неравномерное кодирование, при котором декодирование может быть неоднозначным, то есть, заданному коду может соответствовать несколько разных исходных сообщений
- 2) попробуем декодировать с начала цепочки, первой буквой может быть В или Е, эти случаи нужно рассматривать отдельно
- 3) пусть первая буква – Е с кодом 011, тогда остается цепочка 0100011000
 - для кода 0100011000 первой буквой может быть только В с кодом 01, тогда остается 00011000 (начало исходной цепочки – EV?)
 - для кода 00011000 первой буквой может быть только А с кодом 000, тогда остается 11000, а эта цепочка не может быть разложена на заданные коды букв
 - поэтому наше предположение о том, что первая буква – Е, неверно
- 4) пусть первая буква – В с кодом 01, тогда остается цепочка 10100011000
 - для кода 10100011000 первой буквой может быть только D с кодом 10, тогда остается 100011000 (можно полагать, что начало исходной цепочки – VD?)
 - для кода 100011000 первой буквой может быть только С с кодом 100, тогда остается 011000 (начало исходной цепочки – VDC?)

Несмотря на то, что среди ответов есть единственная цепочка, которая начинается с BDC, здесь нельзя останавливаться, потому что «хвост» цепочки может «не сойтись»

- для кода 011000 на первом месте может быть В (код 01) или Е (011); в первом случае «хвост» 1000 нельзя разбить на заданные коды букв, а во втором – остается код 000 (буква А), поэтому исходная цепочка может быть декодирована как BDCEA

5) правильный ответ – 3

Возможные ловушки и проблемы:

- при декодировании неравномерных кодов может быть очень много вариантов, их нужно рассмотреть все; это требует серьезных усилий и можно легко запутаться
- нельзя останавливаться, не закончив декодирование до конца и не убедившись, что все «сходится», на это обычно и рассчитаны неверные ответы

Решение (вариант 2, декодирование с конца):

- 1) для кода 0110100011000 последней буквой может быть только А (код 000), тогда остается цепочка 0110100011
- 2) для 0110100011 последней может быть только буква Е (011), тогда остается цепочка 0110100

- 3) для 0110100 последней может быть только буква С (100), тогда остается цепочка 0110
- 4) для 0110 последней может быть только буква D (10), тогда остается 01 – это код буквы В
- 5) таким образом, получилась цепочка BDCEA
- 6) правильный ответ – 3

Возможные ловушки и проблемы:

- при декодировании неравномерных кодов может быть очень много вариантов (здесь *случайно* получилась единственно возможная цепочка), их нужно рассмотреть все; это требует серьезных усилий и можно легко запутаться
- нельзя останавливаться, не закончив декодирование до конца и не убедившись, что все «сходится», на это обычно и рассчитаны неверные ответы

Решение (вариант 3, кодирование ответов):

- 1) в данном случае самое простое и надежное – просто закодировать все ответы, используя приведенную таблицу кодов, а затем сравнить результаты с заданной цепочкой
- 2) получим
 - 1) EBCEA – 01101100011000 2) BDDEA – 011010011000
 - 3) BDCEA – 0110100011000 4) EBAEA – 01101000011000
- 3) сравнивая эти цепочки с заданной, находим, что правильный ответ – 3.

Возможные проблемы:

- сложно сравнивать длинные двоичные последовательности, поскольку они однородны, содержат много одинаковых нулей и единиц

Пример задания:

Для передачи по каналу связи сообщения, состоящего только из букв А, Б, В, Г, решили использовать неравномерный по длине код: А=0, Б=10, В=110. Как нужно закодировать букву Г, чтобы длина кода была минимальной и допускалось однозначное разбиение кодированного сообщения на буквы?

- 1) 1 2) 1110 3) 111 4) 11

Решение (вариант 1, метод подбора):

- 1) рассмотрим все варианты в порядке увеличения длины кода буквы Г
- 2) начнем с Г=1; при этом получается, что сообщение «10» может быть раскодировано двояко: как ГА или Б, поэтому этот вариант не подходит
- 3) следующий по длине вариант – Г=11; в этом случае сообщение «110» может быть раскодировано как ГА или В, поэтому этот вариант тоже не подходит

- 4) третий вариант, $\Gamma=111$, дает однозначное раскодирование во всех сочетаниях букв, поэтому...
- 5) ... правильный ответ – 3.

Возможные проблемы:

- при переборе можно ошибиться и «просмотреть» какой-нибудь вариант

Решение (вариант 2, «умный» метод):

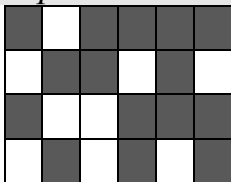
- 1) для того, чтобы сообщение, записанное с помощью неравномерного по длине кода, однозначно раскодировалось, требуется, чтобы никакой код не был началом другого (более длинного) кода; это условие называют *условием Фано*
- 2) как и в первом решении, рассматриваем варианты, начиная с самого короткого кода для буквы Γ ; в нашем случае код $\Gamma=1$ является началом кодов букв Б и В , поэтому условие Фано не выполняется, такой код не подходит
- 3) код $\Gamma=11$ также является началом другого кода (кода буквы В), поэтому это тоже ошибочный вариант
- 4) третий вариант кода, $\Gamma=111$, не является началом никакого уже известного кода; кроме того, ни один уже имеющийся код не является началом кода 111 ; таким образом, условие Фано выполняется
- 5) поэтому правильный ответ – 3.

Возможные проблемы:

- нужно знать условие Фано

Пример задания:

Черно-белое растровое изображение кодируется построчно, начиная с левого верхнего угла и заканчивая в правом нижнем углу. При кодировании 1 обозначает черный цвет, а 0 – белый.



Для компактности результат записали в шестнадцатеричной системе счисления. Выберите правильную запись кода.

- 1) BD9AA5
- 2) BDA9B5
- 3) BDA9D5
- 4) DB9DAB

Решение:

- 1) «вытянем» растровое изображение в цепочку: сначала первая (верхняя) строка, потом – вторая, и т.д.:



2) в этой полоске 24 ячейки, черные заполним единицами, а белые – нулями:

1	0	1	1	1	1	0	1	1	0	1	0	1	0	0	1	1	1	0	1	0	1	0	1
1 строка						2 строка						3 строка						4 строка					

3) поскольку каждая цифра в шестнадцатеричной системе раскладывается ровно в 4 двоичных цифры, разобьем полоску на **тетрады** – группы из четырех ячеек (в данном случае все равно, откуда начинать разбивку, поскольку в полоске целое число тетрад – б):

1	0	1	1	1	1	0	1	1	0	1	0	1	0	0	1	1	1	0	1	0	1	0	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

4) переводя тетрады в шестнадцатеричную систему, получаем последовательно цифры В (11), D(13), А(10), 9, D(13) и 5, то есть, цепочку BDA9D5

5) поэтому правильный ответ – 3.

Возможные проблемы:

- нужно уметь быстро переводить тетрады в шестнадцатеричные цифры (в крайнем случае, это можно сделать через десятичную систему)

Пример задания:

Для передачи чисел по каналу с помехами используется код проверки четности. Каждая его цифра записывается в двоичном представлении, с добавлением ведущих нулей до длины 4, и к получившейся последовательности дописывается сумма её элементов по модулю 2 (например, если передаём **23**, то получим последовательность **0010100110**). Определите, какое число передавалось по каналу в виде **01010100100111100011**?

- 1) 59143 2) 5971 3) 102153 4) 10273

Решение:

- 1) сначала разберемся, как закодированы числа в примере; очевидно, что используется код равномерной длины; поскольку 2 знака кодируются 10 двоичными разрядами (битами), на каждую цифру отводится 5 бит, то есть
 $2 \rightarrow 00101$ и $3 \rightarrow 00110$
- 2) как следует из условия, четыре первых бита в каждой последовательности – это двоичный код цифры, а пятый бит (бит четности) используется для проверки и рассчитывается как «сумма по модулю два», то есть остаток от деления суммы битов на 2; тогда
 $2 = 0010_2$, бит четности $(0 + 0 + 1 + 0) \bmod 2 = 1$
 $3 = 0011_2$, бит четности $(0 + 0 + 1 + 1) \bmod 2 = 0$
- 3) но бит четности нам совсем **не нужен**, важно другое: пятый бит в каждой пятерке **можно отбросить!**
- 4) разобьем заданную последовательность на группы по 5 бит в каждой:

01010, 10010, 01111, 00011.

5) отбросим пятый (последний) бит в каждой группе:

0101, 1001, 0111, 0001.

это и есть двоичные коды передаваемых чисел:

$0101_2 = 5$, $1001_2 = 9$, $0111_2 = 7$, $0001_2 = 1$.

6) таким образом, были переданы числа 5, 9, 7, 1 или число 5971.

7) Ответ: 2.

Отчет

Отчет должен содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Где применяется кодирование информации?
2. Дать определение понятие декодер?
3. Какие существуют методы декодирования?

Практическое занятие № 15

Тема программы: **Стандарты шифрования данных. Криптография**

Тема: **Практическое применение криптографии**

Цель работы: исследование простейших методов криптографической защиты информации.

Время выполнения: 6 часа

Оборудование: ПК.

Программное обеспечение: операционная система, калькулятор.

Часть 1

Теоретические основы

Под **конфиденциальностью** понимают невозможность получения информации из преобразованного массива без знания дополнительной информации (ключа).

Аутентичность информации состоит в подлинности авторства и целостности.

Криптоанализ объединяет математические методы нарушения конфиденциальности и аутентичности информации без знания ключей.

Алфавит - конечное множество используемых для кодирования информации знаков.

Текст - упорядоченный набор из элементов алфавита. В качестве примеров алфавитов можно привести следующие:

алфавит Z_{33-32} - буквы русского алфавита (исключая "ё") и пробел;

алфавит Z_{256} - символы, входящие в стандартные коды ASCII и КОИ-8;

двоичный алфавит - $Z_2 = \{0, 1\}$;

восьмеричный или шестнадцатеричный алфавит

Под **шифром** понимается совокупность обратимых преобразований множества открытых данных на множество зашифрованных данных, заданных алгоритмом криптографического преобразования. В шифре всегда различают два элемента: алгоритм и ключ. Алгоритм позволяет использовать сравнительно короткий ключ для шифрования сколь угодно большого текста.

Криптографическая система, или **шифр** представляет собой семейство T обратимых преобразований открытого текста в зашифрованный. Членам этого семейства можно взаимно однозначно сопоставить число k , называемое **ключом**. Преобразование Tk определяется соответствующим алгоритмом и значением ключа k .

Ключ - конкретное секретное состояние некоторых параметров алгоритма криптографического преобразования данных, обеспечивающее выбор одного варианта из совокупности всевозможных для данного алгоритма. Секретность ключа должна обеспечивать невозможность восстановления исходного текста по зашифрованному.

Пространство ключей K - это набор возможных значений ключа.

Обычно ключ представляет собой последовательный ряд букв алфавита. Следует отличать понятия "ключ" и "пароль". **Пароль** также является секретной последовательностью букв алфавита, однако используется не для шифрования (как ключ), а для аутентификации субъектов.

Электронной (цифровой) подписью называется присоединяемое к тексту его криптографическое преобразование, которое позволяет при получении текста другим пользователем проверить авторство и целостность сообщения.

Зашифрованием данных называется процесс преобразования открытых данных в зашифрованные с помощью шифра, а **расшифрованием** данных - процесс преобразования закрытых данных в открытые с помощью шифра.

Дешифрованием называется процесс преобразования закрытых данных в открытые при неизвестном ключе и, возможно, неизвестном

алгоритме, т.е. методами криптоанализа.

Шифрованием называется процесс зашифрования или расшифрования данных. Также термин шифрование используется как синоним зашифрования. Однако неверно в качестве синонима шифрования использовать термин "кодирование" (а вместо "шифра" - "код"), так как под кодированием обычно понимают представление информации в виде знаков (букв алфавита).

Криптостойкостью называется характеристика шифра, определяющая его стойкость к дешифрованию. Обычно эта характеристика определяется периодом времени, необходимым для дешифрования.

Криптология - это наука о преобразовании информации для обеспечения ее секретности, состоящая из двух ветвей: криптографии и криптоанализа.

Криптоанализ - наука (и практика ее применения) о методах и способах вскрытия шифров.

Криптография - наука о способах преобразования (шифрования) информации с целью ее защиты от незаконных пользователей. Исторически первой задачей криптографии была защита передаваемых текстовых сообщений от несанкционированного ознакомления с их содержанием, известного только отправителю и получателю, все методы шифрования являются лишь развитием этой философской идеи. С усложнением информационных взаимодействий в человеческом обществе возникли и продолжают возникать новые задачи по их защите, некоторые из них были решены в рамках криптографии, что потребовало развития новых подходов и методов.

Шифры простой замены

Система шифрования Цезаря - частный случай шифра простой замены. Метод основан на замене каждой буквы сообщения на другую букву того же алфавита, путем смещения от исходной буквы на K букв.

Известная фраза Юлия Цезаря

VENI VI D I VICI, где

A	B	C	D	E	F	G	H	I	G	K	L	M
←	←	←	←	←	←	←	←	←				
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
					←	←	←	←				

пришел, увидел, победил, зашифрованная с помощью данного метода, преобразуется в

SBKF SFAF SFZF

при смещении на 4 символа влево.

Греческим писателем Полибием за 100 лет до н.э. был изобретен так называемый **полибианский квадрат** размером 5*5, заполненный алфавитом в случайном порядке. Греческий алфавит имеет 24 буквы, а 25-м символом является пробел. Для шифрования на квадрате находили букву текста и

С целью повышения стойкости шифрования можно использовать усовершенствованные варианты таблицы Вижинера. Приведем только некоторые из них:

- во всех (кроме первой) строках таблицы буквы располагаются в произвольном порядке.
- В качестве ключа используется случайность последовательных чисел. Из таблицы Вижинера выбираются десять произвольных строк, которые кодируются натуральными числами от 0 до 10. Эти строки используются в соответствии с чередованием цифр в выбранном ключе.

Известны также и многие другие модификации метода.

Алгоритм перестановки

Этот метод заключается в том, что символы шифруемого текста переставляются по определенным правилам внутри шифруемого блока символов. Рассмотрим некоторые разновидности этого метода, которые могут быть использованы в автоматизированных системах.

Самая простая перестановка — написать исходный текст задом наперед и одновременно разбить шифрограмму на пятерки букв. Например, из фразы
ПУСТЬ БУДЕТ ТАК, КАК МЫ ХОТЕЛИ.

получится такой шифротекст:

ИЛЕТО ХЫМКА ККАТТ ЕДУБЪ ТСУП

В последней группе (пятерке) не хватает одной буквы. Значит, прежде чем зашифровать исходное выражение, следует его дополнить незначащей буквой (например, О) до числа, кратного пяти:

ПУСТЬ-БУДЕТ-ТАККА-КМЫХО-ТЕЛИО.

Тогда шифрограмма, несмотря на столь незначительные изменения, будет выглядеть по-другому:

ОИЛЕТ ОХЫМК АККАТ ТЕДУБ ЪТСУП

Кажется, ничего сложного, но при расшифровке проявляются серьезные неудобства.

Во время Гражданской войны в США в ходу был такой шифр: исходную фразу писали в несколько строк. Например, по пятнадцать букв в каждой (с заполнением последней строки незначащими буквами).

П У С Т Ь Б У Д Е Т Т А К К А
К М Ы Х О Т Е Л И К Л М Н О П

После этого вертикальные столбцы по порядку писали в строку с разбивкой на пятерки букв:

ПКУМС ЫТХЬО БТУЕД ЛЕИТК ТЛАМК НКОАП

Если строки укоротить, а количество строк увеличить, то получится прямоугольник-решетка, в который можно записывать исходный текст. Но тут уже потребуется предварительная договоренность между адресатом и отправителем посланий, поскольку сама решетка может быть различной длины-высоты, записывать к ней можно по строкам, по столбцам, по спирали

туда или по спирали обратно, можно писать и по диагоналям, а для шифрования можно брать тоже различные направления.

Шифры сложной замены

Шифр Гронсфелда состоит в модификации шифра Цезаря числовым ключом. Для этого под буквами сообщения записывают цифры числового ключа. Если ключ короче сообщения, то его запись циклически повторяют. Зашифрованное сообщение получают примерно также, как в шифре Цезаря, но используют не одно жестко заданное смещение а фрагменты ключа.

Пусть в качестве ключа используется группа из трех цифр – 314, тогда сообщение

С О В Е Р Ш Е Н Н О С Е К Р Е Т Н О
3 1 4 3 1 4 3 1 4 3 1 4 3 1 4 3 1 4

Ф П Ё С Ъ З О С С А Х З Л Ф З У С С

В шифрах многоалфавитной замены для шифрования каждого символа исходного сообщения применяется свой шифр простой замены (свой алфавит).

	АБВГДЕЁЖЗИКЛМНОПРСТУФХЧШЩЪЫЬЭЮЯ
А	АБВГДЕЁЖЗИКЛМНОПРСТУФХЧШЩЪЫЬЭЮЯ
Б	АБВГДЕЁЖЗИКЛМНОПРСТУФХЧШЩЪЫЬЭЮЯ
В	Я АБВГДЕЁЖЗИКЛМНОПРСТУФХЧШЩЪЫЬЭЮ
Г	ЮЯ АБВГДЕЁЖЗИКЛМНОПРСТУФХЧШЩЪЫЬЭ
.
Я	ВГДЕЁЖЗИКЛМНОПРСТУФХЧШЩЪЫЬЭЮЯ АБ
_	БВГДЕЁЖЗИКЛМНОПРСТУФХЧШЩЪЫЬЭЮЯ А

Каждая строка в этой таблице соответствует одному шифру замены аналогично шифру Цезаря для алфавита, дополненного пробелом. При шифровании сообщения его выписывают в строку, а под ним ключ. Если ключ оказался короче сообщения, то его циклически повторяют. Зашифрованное сообщение получают, находя символ в колонке таблицы по букве текста и строке, соответствующей букве ключа. Например, используя ключ АГАВА, из сообщения ПРИЕЗЖАЮ ШЕСТОГО получаем следующую шифровку:

ПРИЕЗЖАЮ_ШЕСТОГО
АГАВААГАВААГАВАА
ПОИГЗЖЮЮЮШЕПТНГО

Такая операция соответствует сложению кодов ASCII символов сообщения и ключа по модулю 256.

Практическое задание

Придумайте 3 фразы, каждая минимум из 7 слов. Реализуйте шифрование этой фразы всеми перечисленными видами шифрования.

Часть 2
«Практика криптографической защиты информации»

Введение

Криптография - наука о методах преобразования информации с целью ее защиты от незаконных пользователей.

Стеганография - набор средств и методов сокрытия факта передачи информации.

Некоторые методы стеганографии:

1. В древности голову раба брили, на коже головы писали сообщение и, после отрастания волос, раба отправляли к адресанту.
2. Скрытое письмо между строк: молоком, апельсиновым (или лимонным) соком, другими химическими веществами.
3. "Микроточка". Сообщение с помощью современной технологии записывается на очень маленький носитель ("микроточку"), которая пересылается адресату, например, под обычной маркой.
4. Акrostих - первые буквы слов стихотворения несут информацию:
Добрый удод наелся ягод,
Умный удод наелся на год.
Наелся удод и песни поет.
Ягод наелся удод.
Задание 1: Придумать акrostих, в котором скрыто ваше имя.
5. Например, каждое четвертое слово в посылаемом сообщении несет информацию (остальные слова ничего не значат). Пример: "Тридцать первого августа встреча судебного совета округа состоялась. Подтвердите дату следующего как можно скорее. Участники договорились собраться там же. Борис."

Задание 2: Придумать подобное послание.

Разновидности шифров

1. Шифр замены. Каждая буква заменяется на определенный символ или последовательность символов. Пример: "Пляшущие человечки" Конан Дойля.
2. Шифр перестановки. Буквы в передаваемом сообщении меняются местами в соответствии с определенным правилом. Примеры: МАМА - АМАМ. КРИПТОГРАФИЯ - ИПКРГРТОИЯАФ. (ИП□ КР ГР□ ТО ИЯ□ АФ)
3. Книжный шифр. В зашифрованном тексте каждое слово заменено на пару чисел номер страницы в книге и номер этого слова на странице. (т.е. текст выглядит примерно так: 3-45 45-67 ...).
Ключ - сменный элемент шифра, который применен для шифрования конкретного сообщения.

Шифры перестановки
Маршрутная транспозиция

1. Т - дополнительная буква.

В О С К Р Е

А М Я А Н С

Т Е М А Т И

Я А К С Е Ч

Ш К О Л А Т

Фраза "Воскресная математическая школа" становится: "ЕСИЧТ АЕТНР
КААСЛ ОКМЯС ОМЕАК ШЯТАВ".

Ключ - число 6.

Задание 3:

1. Зашифровать:

а) Французский математик Пьер Ферма по образованию был юрист.

б) Леонардо Пизанского математики знают под именем "сын добряка" или Фибоначчи.

2. Дешифровать (восстановить сообщение, зная ключ) Ключ 8.

Чиной сечем лчгмс хыеоо еайтн ккыйн лтсбч втрйы еоосс ееорс неомв
бадер покп.

Примечание: АБ-дополнительные буквы.

3. Расшифровать (восстановить сообщение, не зная ключа).

Осуз уаан евем исчи тдьм одоа ьльвь рдво быи.

4. Расшифровать:

Етгртуой дкмиуиав цлишлаег врныинис аяоплыдб аанполбр.

2. Ключом является правило расстановки.

В О Е С М А Л А Воесмала срнето катик яачш мяя

С Р Н Е Т О са к.

К А Т И К Или

Я А Ч Ш Воес мала срне тока тикя ачшм

М Е Я еяса к.

С А

К

Задание 4:

1. Расшифровать фразу: Сошки ввнлы охеде нванз бркое еуквс изазх.

2. Расшифровать фразу: Леор тиюд тнет мауа ялее очнм кжхо йчей ооот
лсеч и _пчс днит _киех са _чл илж _а шоо _в рп _уо _к _

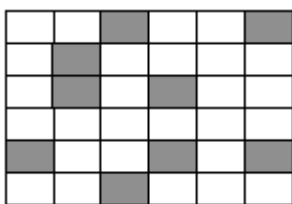
Постолбцовая транспозиция

К	А	Ш	А	Лпше ссис пссл пшао соак ааои ау. Над столбцами записывается ключевое слово, затем в соответствии с порядком букв в алфавите столбцы нумеруются, а затем выписываются подряд: первый столбец, за ним второй и т.д. Ключ здесь - "каша".
3	1	4	2	
Ш	Л	А	С	
А	Ш	А	П	
О	Ш	О	С	
С	Е	И	С	
О	С	А	Л	
А	С	У	Ш	
К	И			

Задание 5:

1. Зашифруйте фразу: Не плюй в колодец: вылетит - не поймаешь.
2. Дешифруйте старинное японское хайку: (ключом будет имя известного японского поэта "Басё") Тйдг адга лвис ыуы лояк пкшр ррув лшсс иеап пнву увет н.
3. Расшифруйте высказывание Козьмы Пруtkова: Ако еаь дне дсц тан жод сск даг рео о.

Шифр "Решетка"



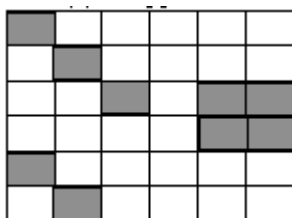
Если хочешь быть красивым поступи в гусары. (Высказывание Козьмы Пруtkова)

Пъеиис влыбым тивхгь укпрос оарчсе ташуыс.

Ключом является решетка

Задание 6:

1. Придумать решетку и зашифровать фразу: "Евклид был древнегреческим математиком"
2. Дешифровать:



- а) Сиекпе тпароо коолко йслтйс тськво нвоски.
 б) Двлпго раимид рувтай гукете гньдот ыоруам.

Шифры замены

Шифр "Британский флаг"



Ответ: флаг.

Задание 7: зашифровать фразу: "ВГГУ".

Каждая буква заменяется несколькими символами

Криптография	11179161915417121932
	K1K7A9K6K9K5A4K7A1Ф1A9Ю2
	И2О2ИО1О4ОА3О2АУ1ИЯ

Задание 8:

- Расшифровать: 1111712 20 111211728 201117112 11517112121228
- Расшифровать: 11212941191517 16122812 1615 176116 111514415
- Расшифровать: 11176191832 5157529 1815291716191832
- Расшифровать: И1А2А5И5О4 Е3И5О4Я1О2 Я1И5О4Е3И5
 И6О1И6О1О5 ИЗЮ1О2И5У8 И1А3И6И3Е3 А2О5А4О3И2
 Е3И1О2А5Е1 Е3О3О3А5О2.

Шифры "Пляшущие человечки"

Основной метод расшифровки подобных шифров - частотный анализ. (+ логические рассуждения).

Таблица частот:

В русском языке в каждой тысяче символов в среднем встречается

пробел	175	р	40	я	18	х	9
о	90	в	38	з	16	ж	7
е, ё	72	л	35	ы	16	ш	6
а	62	к	28	б	14	ю	6

и	62	м	26	ъ, ь	14	ц	4
н	53	д	25	г	13	э	3
т	53	п	23	ч	12	щ	3
с	45	у	21	й	10	ф	2

Чаще всего буквы заменяют другими буквами.

Задание 9: расшифровать текст:

Сзргйзю тсуцьлн Уйиефнлм цкрго, ъхс зов кргнспфхег ф зиецынсм ргзс тзсмхл, ритулрцйзиррс тжжесулхя с тжжси л тсфои ахжс туизфхгелхяфв. Рг сзрсм лк тусжцосн ср тсефхуиьго жцовбьцб ф дсосрнсм зиецынц. Тсуцьлн тзсыио н рим, трцо ии дсосрнц хгн, ъхс хг згоинс цоихиог л фнгкго:
- Рлкнс оихлх. Елзгхя, н зсйзб. Нфхгхл, угкуиылхи туизфхгелхяфв, тсуцьлн Уйиефнлм.

С – 37

И - 21

Г-18

Н – 18

Л - 18

З – 17

Р -16

Ц - 14

Т – 14

Ф – 13

У – 13

Е – 10

М - 7

Й - 5

Я -3

А – 1

Ю – 1

Шифр Цезаря.

В шифре Цезаря каждая буква заменяется на букву, которая идет через 3 после этой: т.е. А=>Г, О=>С, Я=>В.

Примечание: можно делать сдвиг не на три, а на произвольное количество букв.

Задание 10: расшифровать текст

Сзргйзю тсуцьлн Уйиефнлм цкрго, ъхс зов кргнспфхег ф зиецынсм ргзс тзсмхл, ритулрцйзиррс тжжесулхя с тжжси л тсфои ахжс туизфхгелхяфв. Рг сзрсм лк тусжцосн ср тсефхуиьго жцовбьцб ф дсосрнсм зиецынц. Тсуцьлн тзсыио н рим, трцо ии дсосрнц хгн, ъхс хг згоинс цоихиог л фнгкго:

- Рлкнс оихлх. Елзгхя, н зсйзб. Нфхгхл, угкуиылхи туизфхгелхяфв, тсуцълн Уйиефнлм.

Задание 11: зашифровать фразу: "Идет занятие по криптографии".

к	а	ш	а
л	б	щ	б
м	в	ъ	в
н	г	ы	г
о	д	ь	д
п	е	э	е

Ключом является сдвиг.

Метод полосок: берутся полоски, прикладываются и в определенном месте читается слово. Например, слово НГЫГ легко расшифровать. Получается слово КАША.

Задание 12: расшифровать фразу: Схсоябхфв нсныи пюынлрю фоикнл.

Шифр Виженера.

Ключ ВАЗА: /3 1 8 1/

Сдвиг осуществляется не на постоянную величину, а на номер буквы в ключевом слове. КРИПТОГРАФИЯ => НСРРХПЛСГХРА.

Сложность при расшифровке в том, что одинаковые буквы переходят в разные, а разные - в одинаковые => частотный анализ не применим.

Задание 13:

Зашифровать фразу: Математика - царица наук.

Отчет

Отчет должен содержать:

- наименование работы;
- цель работы;
- задание;
- последовательность выполнения работы;
- ответы на контрольные вопросы;
- вывод о проделанной работе.

Контрольные вопросы

1. Где применяется криптография?
2. Какой смысл в Шифре Гронсфельда?
3. С помощью системы шифрования Цезаря зашифровать свое имя?

Шифрование данных – это лишь один из важных элементов системы информационной безопасности, но в отдельности совершенно не достаточный. Система шифрования лишь тогда эффективна, если грамотно настроены системы разграничения доступа, контроля целостности операционной среды, средств обнаружения проникновений, антивирусной и антитроянской защиты и т.д.

Последствия потери данных:

- Отправка серверов или жестких дисков в ремонт;

- Перевозка компьютеров из одного офиса в другой, например, при переезде;
- Утилизация компьютеров, серверов, жестких дисков и лент;
- Хранение магнитных лент в специальной депозитарии (off-site storage);
- Перевозка ленты, например, в депозитарий;
- Кража или потеря жестких дисков или лент. \

Зачем нужно шифрование?

Снижение риска раскрытия конфиденциальных данных из-за так называемого "человеческого фактора", особенно проявляющегося при возникновении экстремальных ситуаций, когда злоумышленники могут получить физический доступ к серверам или к зашифрованным дискам, завладеть администраторским ключом eToken и узнать его PIN-код;

Данные на защищенных дисках всегда хранятся в зашифрованном виде. Поэтому использовать их, даже сделав копию, например, при транспортировке сервера, ремонте, краже или изъятии дисков, невозможно;

Высочайшая надежность - в процессе шифрования реализована защита данных от сбоев, в том числе и в результате сбоев питания компьютера;

Получить доступ к данным и расшифровать их невозможно, даже если под принуждением попробуют заставить это сделать администратора или владельцев. Система уничтожит ключи шифрования дисков по сигналу "тревога", полученному при нажатии "красной кнопки" или от датчиков, обнаруживших несанкционированное проникновение в серверную комнату или открывание серверной стойки.

1. Криптография. Сравнительный анализ алгоритмов симметричного шифрования

Криптография (от др.-греч. κρυπτός — скрытый и γράφω — пишу) — наука о методах обеспечения конфиденциальности (невозможности прочтения информации посторонним) и аутентичности (целостности и подлинности авторства, а также невозможности отказа от авторства) информации.

Изначально криптография изучала методы шифрования информации — обратимого преобразования открытого (исходного) текста на основе секретного алгоритма и/или ключа в зашифрованный текст (шифротекст). Традиционная криптография образует раздел симметричных криптосистем, в которых зашифрование и расшифрование проводится с использованием одного и того же секретного ключа. Помимо этого раздела современная криптография включает в себя асимметричные криптосистемы, системы электронной цифровой подписи (ЭЦП), хеш-функции, управление ключами, получение скрытой информации, квантовую криптографию.

Криптография не занимается: защитой от обмана, подкупа или шантажа законных абонентов, кражи ключей и других угроз информации, возникающих в защищенных системах передачи данных.

Известны два вида шифрования - традиционное (оно же симметрическое) и "открытое шифрование" (асимметрическое). При традиционном шифровании законный пользователь с помощью некоторого конечного автомата (шифратора) преобразует последовательность, называемую открытой информацией, в зашифрованную информацию. Шифратор зависит от параметра (ключа), известного пользователю. Законные пользователи, которым предназначена информация, осуществляют расшифрование информации также с помощью некоторого конечного автомата. В рассматриваемом случае каждый законный пользователь изначально обладает как преобразованием, так и обратным преобразованием, в то время как незаконный пользователь не имеет ключа, т. е. не полностью знает преобразования и. В качестве ключа обычно используется начальное состояние автомата либо его функция перехода.

Симметричные алгоритмы шифрования (или криптография с секретными ключами) основаны на том, что отправитель и получатель информации используют один и тот же ключ. Этот ключ должен храниться в тайне и передаваться способом, исключающим его перехват. Рассмотрим общую схему симметричной, или традиционной, криптографии.



Рисунок 2.1 Общая схема симметричного шифрования

В процессе шифрования используется определенный алгоритм шифрования, на вход которому подаются исходное незашифрованное сообщение, называемое также plaintext, и ключ. Выходом алгоритма является зашифрованное сообщение, называемое также ciphertext. Ключ является значением, не зависящим от шифруемого сообщения. Изменение ключа должно приводить к изменению зашифрованного сообщения.

Зашифрованное сообщение передается получателю. Получатель преобразует зашифрованное сообщение в исходное незашифрованное сообщение с помощью алгоритма дешифрования и того же самого ключа, который использовался при шифровании, или ключа, легко получаемого из ключа шифрования.

Незашифрованное сообщение будем обозначать P или M , от слов plaintext и message. Зашифрованное сообщение будем обозначать C , от слова ciphertext.

Безопасность, обеспечиваемая традиционной криптографией, зависит от нескольких факторов.

Во-первых, криптографический алгоритм должен быть достаточно сильным, чтобы передаваемое зашифрованное сообщение невозможно было расшифровать без ключа, используя только различные статистические закономерности зашифрованного сообщения или какие-либо другие способы его анализа.

Во-вторых, безопасность передаваемого сообщения должна зависеть от секретности ключа, но не от секретности алгоритма. Алгоритм должен быть проанализирован специалистами, чтобы исключить наличие слабых мест, при которых плохо скрыта взаимосвязь между незашифрованными и зашифрованными сообщениями. К тому же при выполнении этого условия производители могут создавать дешевые аппаратные чипы и свободно распространяемые программы, реализующие данный алгоритм шифрования.

В-третьих, алгоритм должен быть таким, чтобы нельзя было узнать ключ, даже зная достаточно много пар (зашифрованное сообщение, незашифрованное сообщение), полученных при шифровании с использованием данного ключа.

Классическим примером таких алгоритмов являются симметричные криптографические алгоритмы, перечисленные ниже:

- Простая подстановка
- Одиночная перестановка по ключу
- Гаммирование

1.1 Простая перестановка

При шифровании простой перестановкой ключевое слово с неповторяющимися символами или цифровой ключ. Число колонок в

таблице задаётся количеством символов в ключе, а число строк может быть фиксировано или может задаваться длиной сообщения. Шифруемый текст записывается последовательными строками под символами ключа. Для заполнения пустых клеток (если объём текста меньше ёмкости таблицы) можно использовать любые символы. Затем текст выписывается колонками в той последовательности, в которой располагаются в алфавите буквы ключа или в порядке следования цифр, если ключ цифровой. В качестве примера рассмотрим шифрование сообщения: «БУДЬТЕ ОСТОРОЖНЫ С ПРЕДСТАВИТЕЛЕМ ФИРМЫ «СПЕКТР». Применим цифровой ключ - 5 1 8 3 7 4 6 2. Выписывая текст по колонкам, получаем абракадабру: УОРТМССВИТЬОДЛСЕНТМЕБТПИРРОЫАФКТЖСЕПДРЕЕЫ.

Расшифрование выполняется в следующем порядке. Подсчитываем число знаков в зашифрованном тексте и делим на число знаков ключа ($41:8=5$ и 1 знак в остатке). Под знаками ключа в соответствующей последовательности записываем вертикально (колонками) символы зашифрованного текста в определенном выше количестве. В каждой колонке по 5 символов, а в одной (первой слева) - 6 символов (5+1 буква в остатке). По строкам таблицы (горизонтально) читаем исходный текст. Выше, в «Истории тайнописи», упоминается шифр называемый в некоторых книгах по криптографии «Считала» (наматывание ленты на жезл). Это не что иное, как перестановка по таблице с простым ключом - 1 2 3 4 ...

1.2 Подстановка

Каждая из 33 букв русского алфавита заменяется на другую букву того же алфавита (моноалфавитная подстановка). Такой шифр (одноалфавитная замена) имеет низкую (временную) стойкость, т. к. зашифрованный (закрытый) текст имеет те же статистические характеристики, что и исходный (открытый) - каждая буква имеет свою частоту появления. Поэтому использовать этот метод целесообразно для шифрования только короткого текста.

Для дополнительной скрытности можно повторно шифровать сообщение, которое уже было зашифровано. Этот способ известен под названием двойная перестановка. Для этого размер второй таблицы подбирают так, чтобы длины ее строк и столбцов были другие, чем в первой таблице. Лучше всего, если они будут взаимно простыми. Кроме того, в первой таблице можно переставлять столбцы, а во второй строки. Наконец, можно заполнять таблицу зигзагом, змейкой, по спирали или каким-то другим способом. Такие способы заполнения таблицы если и не усиливают стойкость шифра, то делают процесс шифрования гораздо более занимательным.

1.3 Гаммирование

Метод гаммирования состоит в том, что символы шифруемого текста

последовательно складываются с символами некоторой специальной последовательности, называемой гаммой.

В потоковых криптосистемах на основе ключа вырабатывается гамма, которая затем накладывается на текст сообщения. Наложение осуществляется посредством сложения по модулю 2 (операции XOR).

Зашифрование производится следующим образом:

$$c_i = m_i \otimes k_i \text{ для } i=1,2,3\dots(1.1)$$

где c_i - знак шифротекста;

m_i - знак открытого текста;

k_i - знак ключевой последовательности;

\otimes - сложение по модулю 2.

Поскольку повторное применение операции XOR восстанавливает первоначальное значение, расшифрование производится повторным наложением гаммы:

$$m_i = c_i \otimes k_i \text{ для } i=1,2,3\dots(1.2)$$

Преобразование текста осуществляется потоком по мере выработки гаммы. Поэтому поточные шифры подходят для шифрования непрерывных потоков данных - голоса, видео и т.д.

Принцип шифрования гаммированием заключается в генерации бесконечного ключа (гаммы шифра) с помощью датчика псевдослучайных чисел (ПСЧ) и наложении полученной гаммы на исходные данные обратимым образом. Процесс расшифрования данных сводится к повторной генерации гаммы шифра при известном ключе и наложении такой гаммы на зашифрованные данные.

Если период гаммы превышает длину всего зашифрованного текста и неизвестна никакая часть исходного текста, то зашифрованный текст можно раскрыть только путем прямого перебора.

1.4 Сравнительный анализ методов шифрования

Сравнительный анализ методов шифрования симметрическим методом изложена в таблице 1.1.

После проведенного анализа было выявлено, что шифрованием методом простой подставки совершенно не эффективные, так как его можно расшифровать после несложного анализа. Алгоритм одиночной перестановки гораздо надежнее подставки, но если ключ используется несколько раз, то его можно проанализировать и взломать. Из трех проанализированных методов, алгоритм методом гаммирования является самым эффективным.

Но у всех методов есть общий недостаток отправитель и получатель должны некоторым тайным образом получить копии секретного ключа и сохранять их в тайне.

Таблица 1.1 Сравнительный анализ алгоритмов симметричного шифрования

	Простая подстановка	Одиночная перестановка по ключу	Гаммирование
Методы взлома	Шифр Цезаря легко вскрывается на основе анализа частот появления букв в шифр тексте	При своей несложности система легко уязвима. Если злоумышленник имеет зашифрованный и соответствующий исходный текст	Метод гаммирования становится бессильным, если злоумышленнику становится известен фрагмент исходного текста и соответствующая ему шифрограмма.
Параметры ключа	Число возможных ключей мало	Число возможных ключей ограничено	Длина гаммы шифрующей должна быть не менее длины защищаемого сообщения
Передача ключа	Отправитель и получатель должны некоторым тайным образом получить копии секретного ключа и сохранить их в тайне.		
Стойкость	Имеет низкую стойкость	Имеет выше стойкость чем подстановка	Исходный текст практически невозможно восстановить без ключа.
Недостатки	Использовать этот метод целесообразно для шифрования только короткого текста	Если ключ используется несколько раз, то его можно проанализировать и взломать.	Полученный зашифрованный текст является достаточно трудным для раскрытия в том случае, если гамма шифра не содержит повторяющихся битовых последовательностей.

2. Практический раздел

2.1 Идентификация пользователя

Идентификация пользователя - распознавание пользователя компьютерной системы на основании ранее заданного описания. Идентификация имеет целью определение полномочий пользователя (права доступа к данным и выбора режима их использования).

Для идентификация пользователя по серийному номеру флеш-карты понадобилось 2 процедуры:

GetDriveType – определяет и возвращает тип носителя;

GetVolumeInformation – определяет информацию о носителе, среди которой содержится серийный номер.

Листинг модуля для идентификация пользователя:

```
procedure TParol.InputClick(Sender: TObject);
var
  SerialNum,dtyp:DWORD;
  a,b:DWORD;
  Buffer,disk :Array[0..255]of char;
  Nomer: cardinal;
begin
  dtyp:=GetDriveType('G:');
  if dtyp <> DRIVE_REMOVABLE then
  begin
    ShowMessage('Диск не обнаружен. Вход не выполнен!');
    exit;
  end;
  GetVolumeInformation(
  'G:',
  Buffer,
  sizeof(Buffer),
  @SerialNum,
  a, b,
  nil, 0);
  nomer:=3830754817;
  if SerialNum = nomer then //сравниваем серийный номер
  begin
    ShowMessage('Вход выполнен!!') ;
    Main.Show();
    Parol.Hide();
  end
  else
  ShowMessage('Вход не выполнен!!');
  end;
```

2.2 Симметричное шифрование

Симметричное шифрование - посторонним лицам может быть известен алгоритм шифрования, но неизвестна небольшая порция секретной информации — ключа, одинакового для отправителя и получателя сообщения.

2.2.1 Метод подстановки

В шифре простой подстановки производится замена каждой буквы сообщения некоторым заранее определенным символом (обычно это также буква). В данном шифре ключом является просто перестановка алфавита (это верно в том случае, если буквы заменяются буквами).

Как можно понять из определения, данный шифр является довольно простым. Перейдем к примеру, показывающему одну из возможных его реализаций.

Программа будет шифровать и дешифровать только русский текст, оставляя неизменным все остальное.

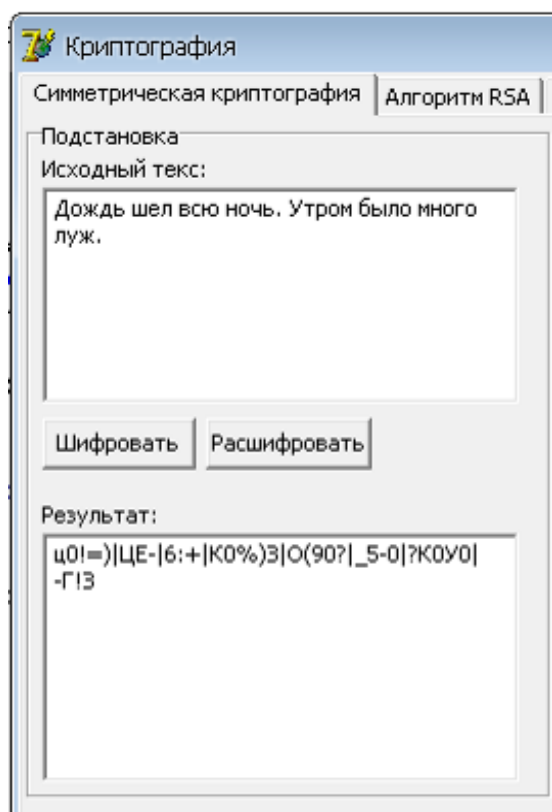


Рисунок 2.1 Экранная форма - Шифрование метом подстановки

Листинг алгоритма шифрования методом подстановки:

```
procedure TMain.Shifrovat1Click(Sender: TObject);  
var  
i, j, q: integer;  
str1, str2: string;  
mas: array [1..256, 1..2] of char;
```

```

begin
mas[1,1]:='1'; mas[1,2]:='я';
mas[2,1]:='2'; mas[2,2]:='ч';
mas[3,1]:='3'; mas[3,2]:='с';
mas[4,1]:='4'; mas[4,2]:='м';
mas[5,1]:='5'; mas[5,2]:='и';
mas[6,1]:='6'; mas[6,2]:='т';
mas[7,1]:='7'; mas[7,2]:='ь';
mas[8,1]:='8'; mas[8,2]:='б';
mas[9,1]:='9'; mas[9,2]:='ю';
mas[10,1]:='0'; mas[10,2]:='.';
mas[11,1]:='-'; mas[11,2]:='Я';
mas[12,1]:='='; mas[12,2]:='Ч';
mas[13,1]:='!'; mas[13,2]:='С';
mas[14,1]:=''''; mas[14,2]:='М';
mas[15,1]:='№'; mas[15,2]:='И';
mas[16,1]:=';'; mas[16,2]:='Т';
mas[17,1]:='%'; mas[17,2]:='Б';
mas[18,1]:=':': mas[18,2]:='Ь';
mas[19,1]:='?'; mas[19,2]:='Ю';
mas[20,1]:='*'; mas[20,2]:='.';
mas[21,1]:='('; mas[21,2]:='Ф';
mas[22,1]:=')'; mas[22,2]:='БІ';
mas[23,1]:='_'; mas[23,2]:='В';
mas[24,1]:='+'; mas[24,2]:='А';
mas[25,1]:='Ў'; mas[25,2]:='П';
mas[26,1]:='Ц'; mas[26,2]:='Р';
mas[27,1]:='У'; mas[27,2]:='О';
mas[28,1]:='К'; mas[28,2]:='Л';
mas[29,1]:='Е'; mas[29,2]:='Д';
mas[30,1]:='Н'; mas[30,2]:='Ж';
mas[31,1]:='Г'; mas[31,2]:='Э';
mas[32,1]:='Ш'; mas[32,2]:='\';
mas[33,1]:='Щ'; mas[33,2]:='д';
mas[34,1]:='З'; mas[34,2]:='э';
mas[35,1]:='Х'; mas[35,2]:='ж';
mas[36,1]:='Ъ'; mas[36,2]:='л';
mas[37,1]:='Ф'; mas[37,2]:='о';
mas[38,1]:='БІ'; mas[38,2]:='р';
mas[39,1]:='В'; mas[39,2]:='п';
mas[40,1]:='А'; mas[40,2]:='а';
mas[41,1]:='П'; mas[41,2]:='в';
mas[42,1]:='Р'; mas[42,2]:='ы';
mas[43,1]:='О'; mas[43,2]:='ф';
mas[44,1]:='Л'; mas[44,2]:='й';

```

mas[45,1]:='Д'; mas[45,2]:='д';
mas[46,1]:='Ж'; mas[46,2]:='ж';
mas[47,1]:='Э'; mas[47,2]:='э';
mas[48,1]:='\'; mas[48,2]:='e';
mas[49,1]:='Я'; mas[49,2]:='я';
mas[50,1]:='Ч'; mas[50,2]:='ч';
mas[51,1]:='С'; mas[51,2]:='с';
mas[52,1]:='М'; mas[52,2]:='м';
mas[53,1]:='И'; mas[53,2]:='и';
mas[54,1]:='Т'; mas[54,2]:='т';
mas[55,1]:='Б'; mas[55,2]:='б';
mas[56,1]:='Ь'; mas[56,2]:='ь';
mas[57,1]:='Ю'; mas[57,2]:='ю';
mas[58,1]:='.'; mas[58,2]:='3';
mas[59,1]:='Й'; mas[59,2]:='й';
mas[60,1]:='Ц'; mas[60,2]:='ц';
mas[61,1]:='У'; mas[61,2]:='у';
mas[62,1]:='К'; mas[62,2]:='к';
mas[63,1]:='Е'; mas[63,2]:='е';
mas[64,1]:='Н'; mas[64,2]:='н';
mas[65,1]:='Р'; mas[65,2]:='р';
mas[66,1]:='Ш'; mas[66,2]:='ш';
mas[67,1]:='Щ'; mas[67,2]:='щ';
mas[68,1]:='З'; mas[68,2]:='з';
mas[69,1]:='Х'; mas[69,2]:='х';
mas[70,1]:='Ъ'; mas[70,2]:='ь';
mas[71,1]:='Ф'; mas[71,2]:='ф';
mas[72,1]:='Ы'; mas[72,2]:='ы';
mas[73,1]:='В'; mas[73,2]:='в';
mas[74,1]:='А'; mas[74,2]:='а';
mas[75,1]:='П'; mas[75,2]:='п';
mas[76,1]:='Р'; mas[76,2]:='р';
mas[77,1]:='О'; mas[77,2]:='о';
mas[78,1]:='Л'; mas[78,2]:='л';
mas[79,1]:='Д'; mas[79,2]:='д';
mas[80,1]:='Ж'; mas[80,2]:='ж';
mas[81,1]:='Э'; mas[81,2]:='э';
mas[82,1]:='\'; mas[82,2]:='№';
mas[83,1]:='Я'; mas[83,2]:='я';
mas[84,1]:='Ч'; mas[84,2]:='%';
mas[85,1]:='С'; mas[85,2]:=':';
mas[86,1]:='М'; mas[86,2]:='?';
mas[87,1]:='И'; mas[87,2]:='*';
mas[88,1]:='Т'; mas[88,2]:='(';
mas[89,1]:='Ь'; mas[89,2]:=')';

```

mas[90,1]:='б'; mas[90,2]:='_';
mas[91,1]:='ю'; mas[91,2]:='+';
mas[92,1]:='.'; mas[92,2]:=',';
mas[93,1]:=','; mas[93,2]:='ё';
mas[94,1]:='ё'; mas[94,2]:=']';
mas[95,1]:=' '; mas[95,2]:='|';
result1.Clear();
for i:=0 to text1.Lines.Count-1 do
begin
str2:="";
str1:= Text1.Lines[i];
for j:=1 to length(text1.Lines[i]) do
for q:=1 to 95 do
if str1[j]=mas[q,1] then
begin
str2:=str2 + mas[q,2];
break;
end;
Result1.Lines.Add(str2);
end;
end;

```

2.2.2 Метод перестановки

При шифровании простой перестановкой ключевое слово с неповторяющимися символами или цифровой ключ. Число колонок в таблице задаётся количеством символов в ключе, а число строк может быть фиксировано или может задаваться длиной сообщения. Шифруемый текст записывается последовательными строками под символами ключа. Для заполнения пустых клеток (если объём текста меньше ёмкости таблицы) можно использовать любые символы. Затем текст выписывается колонками в той последовательности, в которой располагаются в алфавите буквы ключа или в порядке следования цифр, если ключ цифровой.

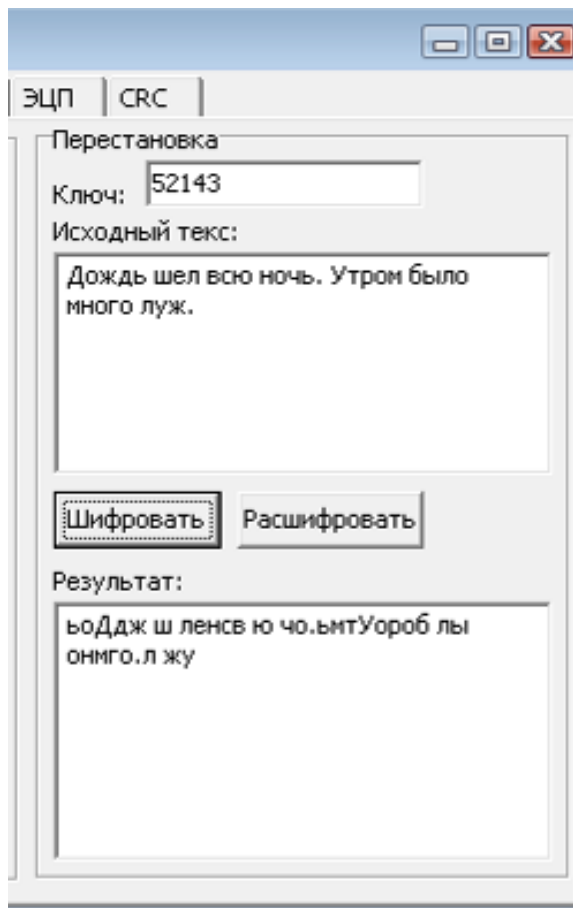


Рисунок 2.2 Экранная форма - Шифрование методом перестановки

Листинг алгоритма шифрования методом перестановки:

```

procedure TMain.Button1Click(Sender: TObject);
var
k: array [1..20] of integer;
str1, str2 : string;
i, j, kol, x, q: integer;
begin
kol:= length(Kluch2.Text);
for i := 1 to kol do
k[i] := StrToInt(Copy(Kluch2.Text, i, 1));
Result2.Clear();
for i := 0 to Text2.Lines.Count - 1 do
begin
str2:="";
str1 := Text2.Lines[i];
x:=length(str1) mod kol;
if x>0 then
for j:=1 to kol-x do
str1:=str1+' ';
for j:=1 to (length(str1) div kol) do

```

```

for q:=1 to kol do
str2 := str2+ copy(str1,k[q]+(kol*(j-1)),1);
Result2.Lines.Add(str2);
end;
end;

```

2.3 Шифрование с открытым ключом

Алгоритм RSA - (буквенная аббревиатура от фамилий Rivest, Shamir и Adleman) — криптографический алгоритм с открытым ключом.

RSA стал первым алгоритмом такого типа, пригодным и для шифрования, и для цифровой подписи. Алгоритм используется в большом числе криптографических приложений.

Листинг алгоритма изложен в приложении.

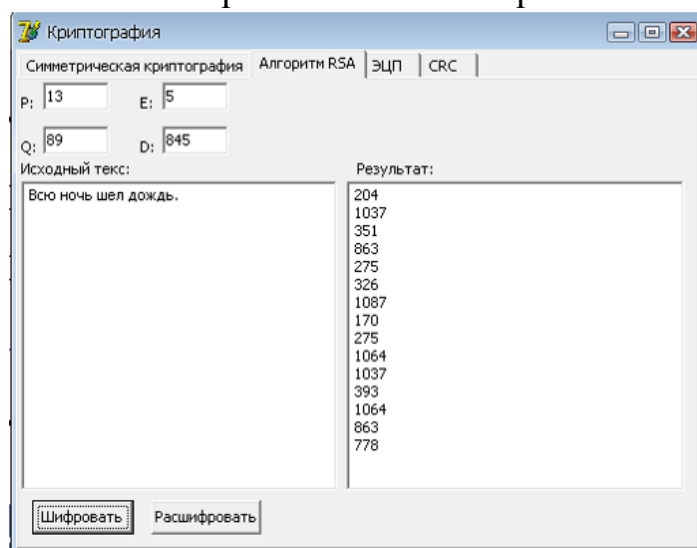


Рисунок 2.3. Экранная форма – шифрование алгоритмом RSA

2.4 Постановка и верификация ЭЦП

Электронная цифровая подпись (ЭП) — реквизит электронного документа, позволяющий установить отсутствие искажения информации в электронном документе с момента формирования ЭП и проверить принадлежность подписи владельцу сертификата ключа ЭП. Значение реквизита получается в результате криптографического преобразования информации с использованием *закрытого ключа ЭП*.

3.3. Типовые задания для оценки освоения учебной дисциплины Типовые задания для оценки знаний (**рубежный контроль**)

Расчетные задания

1. Указать наименьшее количество вопросов, позволяющих всегда угадать день рождения любого человека при ответах: «Да», «Нет».

2. Составить равномерный двоичный код для передачи слов некоторого условного языка, алфавит которого состоит из 20 букв. Чему равен объем информации при передаче семибуквенного слова в этом алфавите?
3. Определить энтропию физической системы В, которая может находиться в одном из 10 состояний. Вероятности состояний системы В:

$$B = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 & b_9 & b_{10} \\ 0,01 & 0,07 & 0,035 & 0,035 & 0,35 & 0,14 & 0,14 & 0,07 & 0,15 & 0,07 \end{bmatrix}$$

4. Определить объем и количество информации в принятом тексте:
 «Товарищ, верь: взойдет она,
 Звезда пленительного счастья,
 Россия воспрянет ото сна...»
5. Длина кода во вторичном алфавите равна 10 символам. Количество информации на символ первичного алфавита равно 2,5 бит/символ. Какое количество информации мы получим, если примем:
- 7 символов вторичного алфавита?
 - 17 символов вторичного алфавита?
6. Методом Шеннона-Фано построить оптимальный код для передачи 100 сообщений при помощи 10 качественных признаков вторичного алфавита.
7. Чему равна общая и частная избыточность некоторого 32-буквенного алфавита, если известно, что его энтропия с учетом неравновероятности букв уменьшается на 0,98 бит/символ, а с учетом взаимозависимости – на 0,4 бит/символ.
8. Какое количество кодовых комбинаций, обнаруживающих одиночную ошибку, можно выбрать из семиразрядного двоичного кода на все сочетания?
9. Какой код имеет полный код Хэмминга для Информационной комбинации 1011?
10. Чему равна скорость передачи информации, если сообщения составлены из русского алфавита, а каждая буква передается за 20 мсек? Взаимозависимость между буквами не учитывается.

За правильный ответ на вопросы или верное решение задачи выставляется положительная оценка – 5 баллов.

За неверное решение задачи предлагается другой вариант.

4. Контрольно-оценочные материалы для промежуточной аттестации по учебной дисциплине

Промежуточная аттестация осуществляется в форме *экзамена*.

I. ПАСПОРТ

Назначение:

КОС предназначен для контроля и оценки результатов освоения учебной дисциплины «ОП.12. Основы теории информации».

Умения

- У.1. Применять закон аддитивности информации.
- У.2. Применять теорему Котельникова.
- У.3. Использовать формулу Шеннона.

Знания

- З.1. Виды и формы представления информации.
- З.2. Методы и средства определения количества информации.
- З.3. Принципы кодирования и декодирования информации.
- З.4. Способы передачи цифровой информации.
- З.5. Методы повышения помехозащищенности передачи и приема данных, основы теории сжатия данных.
- З.6. Методы криптографической защиты информации.
- З.7. Способы генерации ключей.

II. ЗАДАНИЕ ДЛЯ ЭКЗАМЕНУЮЩЕГОСЯ. Вариант № _

Вариант _15

Инструкция для обучающихся

Внимательно прочитайте задание и выполните его.

Время выполнения задания – 1 час 30 мин.

Задание

Комплект разноуровневых заданий

1. Первый уровень состоит из 4 теоретических вопросов
2. Второй уровень состоит из 6 задач

Вариант 1.

Уровень А

1. Что такое информационные процессы (дать определение каждого компонента, привести СВОИ примеры)
2. Формул комбинаторики
3. Матричное кодирование
4. Комбинаторный подход к измерению информации по Колмогорову.

Уровень Б

1. Вычислить $ML_1(\bar{X})$ для блочного кода Хаффмена для X . Длина блока - 2 бита. д.с.в. X берется из последнего примера.

2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.2 & 0.12 \\ 0.3 & 0.13 \\ 0.05 & 0.2 \end{pmatrix}$$

Найти

а) $H_x - ?$

б) $H_{1x}(y) - ?$

3. Устройство состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0,2. Составить закон распределения числа отказавших элементов в одном опыте. Найти $M(X), D(X)$.

4. Имеется (3,4)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%. Вычислить также вероятность ошибочной передачи без использования кода. Сделать аналогичные расчеты для случая, когда вероятность ошибки в десять раз меньше.
5. Закодировать сообщение «ЗЕЛЕНАЯ ЗЕЛЕНЬ» используя алгоритм LZ77(словарь – 12 байт, буфер – 4 байта).
6. Имеется (4,3)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%

Вариант 2.

Уровень А

1. Виды, свойства и формы представления информации
2. Математическое ожидание. Дисперсия.
3. Групповые коды
4. Алгоритмический подход к измерению информации по Колмогорову

Уровень Б

1. Есть три источника сигнала $X(x_1, x_2, x_3)$, $Y(y_1, y_2)$, $Z(z_1, z_2)$. Известно, что $P_x(3p \ 2p \ p)$, $P_y(2q \ q)$, $P_z(4p \ q)$, где $p = \frac{1}{6}, q = \frac{1}{3}$. Найти источник с максимальной энтропией.

2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.4 & 0.05 \\ 0.14 & 0.15 \\ 0.02 & 0.24 \end{pmatrix}$$

Найти а) $H_x - ?$

б) $H_{1,x}(y) - ?$

3. Книга издана тиражом 100 000 экземпляров. Вероятность того, что книга сброшюрована неправильно, равна 0,0001. Найти вероятность того, что тираж содержит ровно пять бракованных книг. . Найти $M(X), D(X)$.
4. Может ли (7,15)-код, минимальное расстояние между кодовыми словами которого 5, быть совершенным?
5. Закодировать сообщение «БЕЛАЯ БЕЛИЗНА» используя алгоритм LZSS(словарь – 12 байт, буфер – 4 байта).
6. Пусть передаваемое слово $a = 10$ кодируется словом $b = 1001$, а строка ошибок – $\varepsilon = 0001$. Какие слова можно раскодировать этим кодом.

Вариант 3.

Уровень А

1. Системы передачи информации. Общая схема передачи информации по линиям связи.
2. Простейшие алгоритмы сжатия информации.
3. Код Хэмминга
4. Комбинаторный подход к измерению информации по Колмогорову

Уровень Б

1. Зашифровать сообщение «мировоззрение людей» ключом «мозг».

2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.1 & 0.03 \\ 0.17 & 0.2 \\ 0.08 & 0.42 \end{pmatrix}$$

Найти

a) $H_x - ?$

b) $H_{1,x}(y) - ?$

3. Вычислить среднее количество бит на единицу сжатого сообщения о значении каждой из д.с.в., из заданных следующими распределениями вероятностей, при сжатии методами Шеннона-Фэнно, Хаффмена и арифметическим.

X	1	4	9	16	25	36
p	0.1	0.2	0.1	0.3	0.1	0.2

4. Пусть двоичный симметричный канал используется для передачи строк из двух бит. Построить таблицу вероятностей приема.
5. Распакуйте сообщение сжатое LZ77(словарь – 12 байт, буфер – 4байта) $\{0,0,'A'\}, \{0,0,'F'\}, \{0,0,'X'\}, \{9,2,'F'\}, \{8,1,'F'\}, \{6,2,'X'\}, \{4,3,'A'\}$
6. Имеется (5,6)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%.

Вариант 4.

Уровень А

1. Формула Хартли. Формула Шеннона.
2. Адаптивные алгоритмы сжатия. Кодирование Хаффмена
3. Понятие о кодах Боуза-Чоудхури-Хоккенгема
4. Вероятностный подход к измерению информации по Колмогорову

Уровень Б

1. Вычислить $ML_1(\bar{X})$ для блочного кода Хаффмена для X . Длина блока - 4 бита. д.с.в. X берется из последнего примера.

2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.2 & 0.12 \\ 0.3 & 0.13 \\ 0.05 & 0.2 \end{pmatrix}$$

Найти

a) $H_x - ?$

b) $H_{1,x}(y) - ?$

3. Устройство состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0,2. Составить закон распределения числа отказавших элементов в одном опыте. Найти $M(X), D(X)$.
4. Имеется (3,4)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%. Вычислить также вероятность ошибочной передачи без использования кода. Сделать аналогичные расчеты для случая, когда вероятность ошибки в десять раз меньше.
5. Закодировать сообщение «ЗЕЛЕНАЯ ЗЕЛЕНЬ» используя алгоритм LZ77(словарь – 12 байт, буфер – 4 байта).

6. Имеется (4,3)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%.

Вариант 5.

Уровень А

1. Формула Хартли. Формула Шеннона.
2. Адаптивные алгоритмы сжатия. Кодирование Хаффмена
3. Понятие о кодах Боуза-Чоудхури-Хоккенгема
4. Вероятностный подход к измерению информации по Колмогорову

Уровень Б

1. Вычислить $ML_1(\bar{X})$ для блочного кода Хаффмена для X . Длина блока - 4 бита. д.с.в. X берется из последнего примера.

2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.2 & 0.12 \\ 0.3 & 0.13 \\ 0.05 & 0.2 \end{pmatrix}$$

Найти

а) $H_x - ?$

б) $H_{1x}(y) - ?$

3. Устройство состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0,2. Составить закон распределения числа отказавших элементов в одном опыте. Найти $M(X), D(X)$.
4. Имеется (3,4)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%. Вычислить также вероятность ошибочной передачи без использования кода. Сделать аналогичные расчеты для случая, когда вероятность ошибки в десять раз меньше.
5. Закодировать сообщение «ЗЕЛЕНАЯ ЗЕЛЕНЬ» используя алгоритм LZ77(словарь – 12 байт, буфер – 4 байта).
6. Имеется (4,3)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%.

Вариант 6.

Уровень А

1. Способы передачи информации. Основные информационные характеристики
2. Подстановочные или словарно-ориентированные алгоритмы сжатия информации.
3. Основы теории защиты информации.
4. Алгоритмический подход к измерению информации по Колмогорову.

Уровень Б

1. Есть три источника сигнала $X(x_1, x_2, x_3)$, $Y(y_1, y_2)$, $Z(z_1, z_2)$. Известно, что $P_x(3p, 2p, p)$, $P_y(2q, q)$, $P_z(4p, q)$, где $p = \frac{1}{6}, q = \frac{1}{3}$. Найти источник с максимальной энтропией.

2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.4 & 0.06 \\ 0.14 & 0.15 \\ 0.02 & 0.23 \end{pmatrix}$$

Найти

а) $H_x - ?$

б) $H_{1x}(y) - ?$

- Книга издана тиражом 100 000 экземпляров. Вероятность того, что книга сброшюрована неправильно, равна 0,0001. Найти вероятность того, что тираж содержит ровно пять бракованных книг. Найти $M(X), D(X)$.
- Может ли (7,15)-код, минимальное расстояние между кодовыми словами которого 5, быть совершенным?
- Закодировать сообщение «БЕЛАЯ БЕЛИЗНА» используя алгоритм LZSS(словарь – 12 байт, буфер – 4 байта).
- Пусть передаваемое слово $a = 10$ кодируется словом $b = 1001$, а строка ошибок – $\varepsilon = 0001$. Какие слова можно раскодировать этим кодом.

Вариант 7.

Уровень А

- Энтропия и её свойства.
- Методы Лемпела-Зива (LZ77 и LZSS).
- Криптография.
- Вероятностный подход к измерению информации по Колмогорову.

Уровень Б

- Закодировать сообщение АССВСВА, используя адаптивный алгоритм Хаффмена с упорядоченным деревом.

$$P_{xy} = \begin{pmatrix} \frac{1}{8} & 0 & \frac{1}{3} & \frac{3}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & 1 & 0 & 0 \\ \frac{5}{8} & 0 & \frac{3}{8} & 0 \end{pmatrix}$$

- причем $p_1 = 0.125; p_2 = 0.0625; p_3 = 0.0625; p_4 = 0.75$.

Найти энтропию источника.

- Вычислить для кодов Хаффмена и Шеннона-Фэно для X . д.с.в. X задается следующим распределением вероятностей:

4. Может ли (8,14)-код, минимальное расстояние между	X	-2	-1			
	p					

расстояние между

кодowymi словами которого 4, быть совершенным?

- Закодировать сообщение «ACCDACDABABCD» используя алгоритм LZW(словарь – ASCII+, 16 фраз).
- Пусть передаваемое слово $a = 11$ кодируется словом $b = 1111$, а строка ошибок – $\varepsilon = 0101$. Какие слова можно раскодировать этим кодом.

Вариант 8.

Уровень А

1. Теорема отсчетов. Физический и информационный смысл.
2. Методы Лемпела-Зива (LZ78 и LZW).
3. Криптосистема без передачи ключей.
4. Комбинаторный подход к измерению информации по Колмогорову

Уровень Б

1. Зашифровать сообщение «мировоззрение людей» ключом «мозг».
2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.1 & 0.03 \\ 0.17 & 0.2 \\ 0.08 & 0.42 \end{pmatrix}$$

Найти

а) $H_x - ?$

б) $H_{1,x}(y) - ?$

3. Вычислить среднее количество бит на единицу сжатого сообщения о значении каждой из д.с.в., из заданных следующими распределениями вероятностей, при сжатии методами Шеннона-Фэнно, Хаффмена и

X	1	4	9	16	25	36
p	0.1	0.2	0.1	0.3	0.1	0.2

- арифметическим.
4. Пусть двоичный симметричный канал используется для передачи строк из двух бит. Построить таблицу вероятностей приема.
5. Распакуйте сообщение сжатое LZ77(словарь – 12 байт, буфер – 4байта) $(0,0,'A'), (0,0,'F'), (0,0,'X'), (9,2,'F'), (8,1,'F'), (6,2,'X'), (4,3,'A')$
6. Имеется (5,6)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%

Вариант 9.

Уровень А

1. Физическая энтропия. Информационная энтропия
2. LZ-алгоритмы распаковки данных.
3. Криптосистема с открытым ключом.
4. Алгоритмический подход к измерению информации по Колмогорову.

Уровень Б

1. Вычислить $ML_1(\bar{X})$ для блочного кода Хаффмена для X . Длина блока - 2 бита. д.с.в. X берется из последнего примера.

2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.2 & 0.12 \\ 0.3 & 0.13 \\ 0.05 & 0.2 \end{pmatrix}$$

Найти

а) $H_x - ?$

б) $H_{1,x}(y) - ?$

3. Устройство состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0,2. Составить

закон распределения числа отказавших элементов в одном опыте. Найти $M(X), D(X)$.

4. Имеется (3,4)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%. Вычислить также вероятность ошибочной передачи без использования кода. Сделать аналогичные расчеты для случая, когда вероятность ошибки в десять раз меньше.
5. Закодировать сообщение «ЗЕЛЕНАЯ ЗЕЛЕНЬ» используя алгоритм LZ77(словарь – 12 байт, буфер – 4 байта).
6. Имеется (4,3)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%.

Вариант 10.

Уровень А

1. Энтропия двух и трех взаимосвязанных источников.
2. Особенности программ-архиваторов.
3. Электронная подпись.
4. Вероятностный подход к измерению информации по Колмогорову.

Уровень Б

1. Закодировать сообщение $DCBCBDB$, используя адаптивный алгоритм Хаффмена с упорядоченным деревом.

$$P_{xy} = \begin{pmatrix} \frac{1}{8} & 0 & \frac{1}{3} & \frac{3}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{2}{2} \\ \frac{5}{8} & 0 & \frac{3}{8} & 0 \end{pmatrix}$$

2. причем $p_1 = 0.0625; p_2 = 0.125; p_3 = 0.75; p_4 = 0.0625$.

Найти энтропию источника.

3. Вычислить HX и $ML(X)$ для кодов Хаффмена и Шеннона-Фэнно для X .

x	1	2	3	4	5
p	7/18	1/6	1/6	1/6	1/9

д.с.в. X задается следующим распределением вероятностей:
4. Имеется (4,6)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%. Вычислить также вероятность ошибочной передачи без использования кода. Сделать аналогичные расчеты для случая, когда вероятность ошибки в десять раз меньше.
5. Закодировать сообщение «МАСЛО МАСЛЕННОЕ» используя алгоритм LZ78(словарь – 16 фраз).
6. Имеется (3,4)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%.

Вариант 11.

Уровень А

1. Математические модели детерминированных периодических сигналов. Их физический смысл .
2. Сжатие информации с потерями (основная теорема о кодировании при наличии помех).
3. Стандарт шифрования данных
4. Комбинаторный подход к измерению информации по Колмогорову.

Уровень Б

1. Есть три источника сигнала $X(x_1, x_2, x_3)$, $Y(y_1, y_2)$, $Z(z_1, z_2)$. Известно, что $P_x(3p \ 2p \ p)$, $P_y(2q \ q)$, $P_z(4p \ q)$, где $p = \frac{1}{6}, q = \frac{1}{3}$. Найти источник с максимальной энтропией.

2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.4 & 0.05 \\ 0.14 & 0.15 \\ 0.02 & 0.24 \end{pmatrix}$$

Найти а) $H_x - ?$

б) $H_{1X}(y) - ?$

3. Книга издана тиражом 100 000 экземпляров. Вероятность того, что книга сброшюрована неправильно, равна 0,0001. Найти вероятность того, что тираж содержит ровно пять бракованных книг. Найти $M(X), D(X)$.
4. Может ли (7,15)-код, минимальное расстояние между кодовыми словами которого 5, быть совершенным?
5. Закодировать сообщение «БЕЛАЯ БЕЛИЗНА» используя алгоритм LZSS(словарь – 12 байт, буфер – 4 байта).
6. Пусть передаваемое слово $a = 10$ кодируется словом $b = 1001$, а строка ошибок – $\varepsilon = 0001$. Какие слова можно декодировать этим кодом.

Вариант 12.

Уровень А

1. Математические модели детерминированных непериодических сигналов. Их физический смысл
2. Помехозащитное кодирование. Двоичный симметричный канал.
3. Компьютерный шрифт. HTML, XML и SGML
4. Алгоритмический подход к измерению информации по Колмогорову.

Уровень Б

1. Закодировать сообщение *АССВСВА*, используя адаптивный алгоритм Хаффмена с упорядоченным деревом.

$$P_{xy} = \begin{pmatrix} \frac{1}{8} & 0 & \frac{1}{3} & \frac{3}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\ \frac{5}{8} & 0 & \frac{3}{8} & 0 \end{pmatrix}$$

2. причем $p_1 = 0.125; p_2 = 0.0625; p_3 = 0.0625; p_4 = 0.75$.

Найти энтропию источника.

3. Вычислить HX и $ML(X)$ для кодов Хаффмена и Шеннона-Фэнно для X . д.с.в.

X задается следующим распределением вероятностей:

X	-2	-1	0	1	2
p	1/3	1/4	1/5	1/6	1/20

4. Может ли (8,14)-код, минимальное расстояние между кодовыми словами которого 4, быть совершенным?
5. Закодировать сообщение «ACCDCCDABBACD» используя алгоритм LZW(словарь – ASCII+, 16 фраз).
6. Пусть передаваемое слово $a = 11$ кодируется словом $b = 1111$, а строка ошибок – $\varepsilon = 0101$. Какие слова можно раскодировать этим кодом.

Вариант 13.

Уровень А

1. Алгебра событий
2. Математическая модель системы связи (коды с исправлением ошибок).
3. Кодировка букв русского алфавита
4. Вероятностный подход к измерению информации по Колмогорову.

Уровень Б

1. Зашифровать сообщение «мировоззрение людей» ключом «мозг».
2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.1 & 0.03 \\ 0.17 & 0.2 \\ 0.08 & 0.42 \end{pmatrix}$$

Найти

а) $H_x - ?$

б) $H_{1X}(y) - ?$

3. Вычислить среднее количество бит на единицу сжатого сообщения о значении каждой из д.с.в., из заданных следующими распределениями вероятностей, при сжатии методами Шеннона-Фэнно, Хаффмена и арифметическим.

X	1	4	9	16	25	36
p	0.1	0.2	0.1	0.3	0.1	0.2

4. Пусть двоичный симметричный канал используется для передачи строк из двух бит. Построить таблицу вероятностей приема.
5. Распакуйте сообщение сжатое LZ77(словарь – 12 байт, буфер – 4байта) $(0,0, 'A'), (0,0, 'F'), (0,0, 'X'), (9,2, 'F'), (8,1, 'F'), (6,2, 'X'), (4,3, 'A')$
6. Имеется (5,6)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%.

Вариант 14.

Уровень А

1. Формула Хартли. Формула Шеннона.
2. Адаптивные алгоритмы сжатия. Кодирование Хаффмена
3. Понятие о кодах Боуза-Чоудхури-Хоккенгема
4. Вероятностный подход к измерению информации по Колмогорову

Уровень Б

1. Вычислить $ML_1(\bar{X})$ для блочного кода Хаффмена для X . Длина блока - 2 бита. д.с.в. X берется из последнего примера.

2. Вычислить

$$P_{xy} = \begin{pmatrix} 0.2 & 0.12 \\ 0.3 & 0.13 \\ 0.05 & 0.2 \end{pmatrix}$$

Найти

а) $H_x - ?$

б) $H_{yx}(y) - ?$

3. Устройство состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в одном опыте равна 0,2. Составить закон распределения числа отказавших элементов в одном опыте. Найти $M(X), D(X)$.
4. Имеется (3,4)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%. Вычислить также вероятность ошибочной передачи без использования кода. Сделать аналогичные расчеты для случая, когда вероятность ошибки в десять раз меньше.
5. Закодировать сообщение «ЗЕЛЕНАЯ ЗЕЛЕНЬ» используя алгоритм LZ77 (словарь – 12 байт, буфер – 4 байта).
6. Имеется (4,3)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%.

Вариант 15.

Уровень А

1. Дискретные и непрерывные сообщения
2. Адаптивное арифметическое кодирование
3. Циклические избыточные коды
4. Комбинаторный подход к измерению информации по Колмогорову.

Уровень Б

1. Закодировать сообщение $DCDDCDB$, используя адаптивный алгоритм Хаффмена с упорядоченным деревом.

$$P_{xy} = \begin{pmatrix} \frac{1}{8} & 0 & \frac{1}{3} & \frac{3}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\ \frac{5}{8} & 0 & \frac{3}{8} & 0 \end{pmatrix}$$

2. причем $p_1 = 0.0625; p_2 = 0.125; p_3 = 0.75; p_4 = 0.0625$.

Найти энтропию источника.

3. Вычислить HX и $ML(X)$ для кодов Хаффмена и Шеннона-Фэно для X . д.с.в. X задается следующим распределением вероятностей:

X	1	2	3	4	5
p	7/18	1/6	1/6	1/6	1/9

4. Имеется (4,6)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%. Вычислить также вероятность ошибочной передачи без использования кода. Сделать аналогичные расчеты для случая, когда вероятность ошибки в десять раз меньше.
5. Закодировать сообщение «МАСЛО МАСЛЕННОЕ» используя алгоритм LZ78(словарь – 16 фраз).
6. Имеется (3,4)-код с проверкой четности. Вычислить вероятность того, что в случае ошибки этот код ее не обнаружит, если вероятность ошибки при передаче каждого бита равна 1%

III. ПАКЕТ ЭКЗАМЕНАТОРА

III а. УСЛОВИЯ

Количество вариантов задания для экзаменуемого – 15

Время выполнения задания – 1 час 30 мин.

Экзаменационная ведомость – единая форма утвержденная учебной частью

III б. КРИТЕРИИ ОЦЕНКИ

1. При выполнении 90 – 100 % от общего объема заданий 1-го уровня выставляется оценка **3 балла**.
2. При выполнении заданий 1-го уровня полностью и 4 ответа на вопросы второго уровня выставляется оценка **4 балла**.
3. При выполнении полностью заданий 1-го и 2-го уровня выставляется оценка **5 баллов**.

Лист согласования

Дополнения и изменения к комплекту КОС на учебный год

Дополнения и изменения к комплекту КОС на _____ учебный год по дисциплине

В комплект КОС внесены следующие изменения:

Дополнения и изменения в комплекте КОС обсуждены на заседании ЦК «Математического и естественнонаучного цикла» от «_____» _____ 20____ г. протокол № _____

Председатель ПК _____ /Т.А. Новикова/