|  |  |
| --- | --- |
| Череповецкий государственный университет  Кафедра «Математического и программного обеспечения ЭВМ» | |
| ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ  по дисциплине «Теория информации»  КОДЫ ФАНО И ХАФФМАНА | |
|  | Принял:  преподаватель Е.Н. Руденко    подпись, дата  Выполнил:  студент гр. 1ПИб-02-1оп-22  Митин Юрий Олегович  подпись, дата |
| Череповец, 2023 | |

**Реферат**

Цель работы – познакомиться с методами сжатия информации и применить их на практике.

В ходе работы исследуются три метода сжатия информации: RLE, Шеннона-Фано и Хаффмана.

**Ход работы**

Вариант 14

Задание 1.

Выполнить вручную кодирование сообщения методом RLE. В качестве исходной фразы взять текст из табл. 3.1. С помощью таблицы CP-1251 (см. Приложение 1) перевести символы заданной фразы в десятичные числа, а затем десятичные числа перевести в двоичные. Выполнить сжатие информации, вычислить контрольные суммы и коэффициент сжатия.

 Табл. 3.1.

|  |  |
| --- | --- |
| Вар | Текст |
| 14 | Цена 2598888666611 коп |

Решение:

|  |  |  |
| --- | --- | --- |
| Символ | Десятичные числа | Двоичный код |
| Ц | 199 | 11000111 |
| е | 229 | 11100101 |
| н | 240 | 11110000 |
| а | 224 | 11100000 |
| - | 32 | 00100000 |
| 2 | 50 | 00110010 |
| 5 | 53 | 00110101 |
| 9 | 57 | 00111001 |
| 8 | 56 | 00111000 |
| 6 | 54 | 00110110 |
| 1 | 49 | 00110001 |
| к | 234 | 11101010 |
| о | 239 | 11101111 |
| п | 241 | 11110001 |

Переведём исходное сообщение в двоичный код:

11000111 11100101 11110000 11100000 00100000 00110010 00110101 00111001 00111000 00111000 00111000 00111000 00110110 00110110 00110110 00110110 00110110 00110001 00110001 00100000 11101010 11101111 11110001

Проведём сжатие на основе фиксированной длины кодового слова:

1Ц1е1н1а1 1215194846211 1к1о1п

Переведем это сообщение в двоичный код:

00110001 11000111 00110001 11100101 00110001 11110000 00110001 11100000 00110001 00100000 00110001 00110010 00110001 00110101 00110001 00111001 00110001 00111000 00110001 00110110 00110001 00110001 00110001 11101010 00110001 11101111 00110001 11110001

До сжатия: 184 бит

После сжатия: 224 бит

Кс = 224/184 = 1,22

**Задание 2. Выполнить сжатие информации методом Шеннона-Фано**

Используя фразу из табл. 3.1, построить кодовое дерево и определить коэффициент сжатия методом Шеннона-Фано

|  |  |
| --- | --- |
| Вар | Текст |
| 14 | Цена 2598888666611 коп |

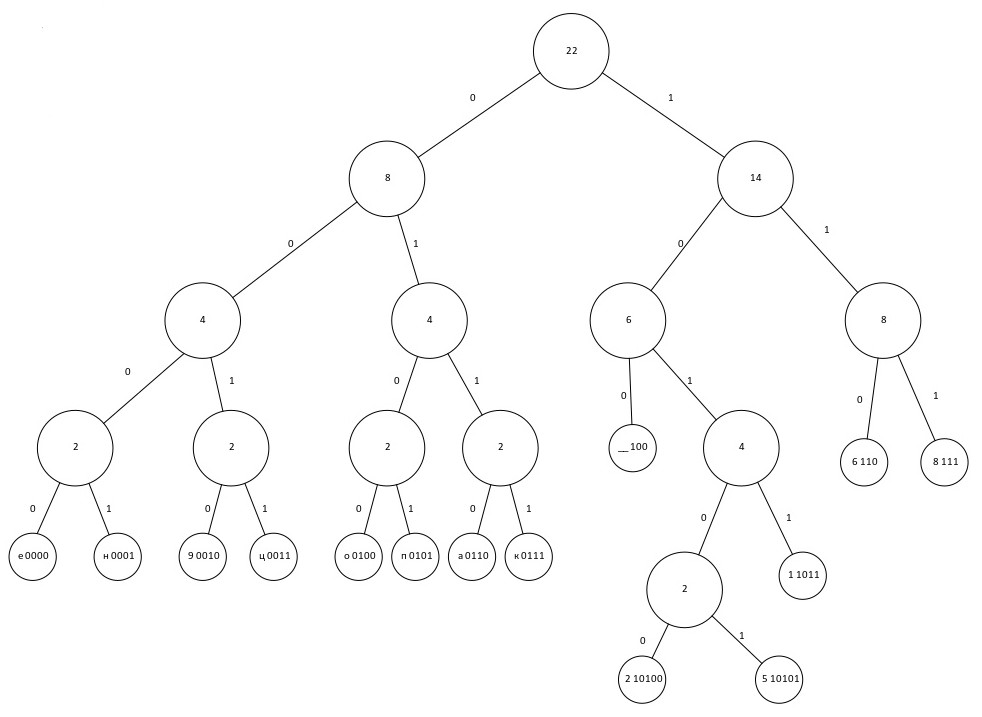
Решение:

Найдём вероятности появления символа в сообщении:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Символ | Частота |  |  |  |  |  |  |  |
| 8 | 4 | 4 | 4 | 4 | 4 | 8 | 14 | 22 |
| 6 | 4 | 4 | 4 | 4 | 4 |
| \_ | 2 | 2 | 2 | 6 | 6 | 6 |
| 1 | 2 | 2 | 4 |
| ц | 1 | 2 |
| е | 1 |
| н | 1 | 2 | 4 | 4 | 8 | 8 | 8 |
| а | 1 |
| 2 | 1 | 2 |
| 5 | 1 |
| 9 | 1 | 2 | 4 | 4 |
| к | 1 |
| о | 1 | 2 |
| п | 1 |
| Строим кодовое дерево. В таблице сложно изобразить, но идём по веткам влево. | | | | | | | | |
| В таблице сложно изобразить, но понятно, что верхнее слагаемое = 1, нижнее = 0 | | | | | | | | |

Для удобства пары одинаковых вероятностей складывал сразу, так как их сумма на следующем слое всегда была бы наименьшей, и ошибки бы не произошло.

Построим кодовое дерево:



Из дерева следует, что

|  |  |
| --- | --- |
| Символ | Двоичный код |
| 8 | 111 |
| 6 | 110 |
| \_ | 100 |
| 1 | 1011 |
| ц | OO11 |
| е | OOOO |
| н | OOO1 |
| а | O11O |
| 2 | 10100 |
| 5 | 10101 |
| 9 | OO1O |
| к | O111 |
| о | O1OO |
| п | O1O1 |

Закодируем исходное сообщение в соответствии с деревом Фано. Закодированное сообщение Цена 2598888666611 коп выглядит следующим образом:

0011 0000 0001 0110 100 10100 10101 0010 111 111 111 111 111 110 110 110 110 1011 1011 100 0111 0100 0101

До сжатия: 184 бит

После сжатия: 83 бит

Кс = 83/184 = 0,45

**Задание 3. Выполнить сжатие информации методом Хаффмана**

Используя фразу из табл. 3.1, построить кодовое дерево и определить коэффициент сжатия методом Хаффмана

|  |  |
| --- | --- |
| Вар | Текст |
| 14 | Цена 2598888666611 коп |

Решение:

Построим таблицу по методу Хаффмана:



Увеличенная таблица есть в Экселе

В результате получаем:

|  |  |
| --- | --- |
| Символ | Код |
| 8 | 000 |
| 6 | 001 |
| \_ | 011 |
| 1 | 110 |
| ц | 01010 |
| е | 01011 |
| н | 1010 |
| а | 1011 |
| 2 | 1000 |
| 5 | 1001 |
| 9 | 01000 |
| к | 01001 |
| о | 1110 |
| п | 1111 |

Закодируем исходное сообщение Цена 2598888666611 коп в соответствии с методом Хаффмана. Закодированное сообщение выглядит следующим образом:

01010 01011 1010 1011 011 1000 1001 01000 000 000 000 000 001 001 001 001 110 110 011 01001 1110 1111

До сжатия: 184 бит

После сжатия: 80 бит

Кс = 80/184 = 0,435

**Задание 4. Архивирование файлов**

С помощью стандартного архиватора (WinZip, WinRar, 7-Zip и т.п.) выполнить сжатие различных документов, тип которых указан в таблице 3.4.1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Документ | Расширение | Объем файла до архивации, Кбайт | Объем файла после архивации, Кбайт | Коэффициент сжатия |
| Текст | .doc | 330,54 | 67,1 | 0,203 |
| Фотография | .jpg | 321,1 | 320,8 | 1,0008 |
| Рисунок | .bmp | 2800 | 42 | 0,015 |
| Видео | .avi | 701,1 | 694 | ~0,99 |
| Звук | .mp3 | 679,6 | 672,8 | ~0,99 |

В качестве текстового документа нужно взять файл, который не содержит рисунков. Число символов должно быть более 3000 знаков.

Фотографии нужно взять на сайте **samara.psati.ru** в соответствии с вариантом (таблица 3.4.2.). Следует взять одну фотографию из указанного раздела.

В качестве рисунка формата BMP нужно взять изображение флага государства в соответствии с таблицей 3.4.3.

Видеоклип желательно снять самостоятельно (с помощью видеокамеры, цифрового фотоаппарата, мобильного телефона, планшетника).

В качестве звукового файла используйте своё любимое музыкальное произведение

**Заключение**

Познакомился с такими методами сжатия информации, как метод RLE, дерево Шеннона-Фано и дерево Хаффмана. Также воспользовался архиватором 7-Zip для сжатия медиа-файлов.