|  |  |
| --- | --- |
| Череповецкий государственный университет  Кафедра «Математического и программного обеспечения ЭВМ» | |
| ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ  по дисциплине «Теория информации»  КОДИРОВАНИЕ СООБЩЕНИЙ МЕТОДОМ RLE | |
|  | Принял:  преподаватель Е.Н. Руденко    подпись, дата  Выполнил:  студент гр. 1ПИб-02-01оп-22 Маслов Владислав Андреевич    подпись, дата |
| Череповец, 2023 | |

Реферат

Предметом исследования являются формулы RLE для кодирования сообщений

Цельработы**–**научиться кодировать методом RLE

Входе работы проводились теоретические исследования кодирования, а также графическое моделирование кодирования

В результате работы численно получены значения вероятностей и кодировок для различных сообщений.

Ход работы

1. Выполнить вручную кодирование сообщения “Цена 2598888666611 коп” методом RLE. С помощью таблицы CP-1251 перевести символы заданной фразы в десятичные числа, а затем десятичные числа перевести в двоичные. Выполнить сжатие информации, вычислить контрольные суммы и коэффициент сжатия.

|  |  |  |
| --- | --- | --- |
| 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 |

Табл.1.

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

1. Выполнить сжатие информации методом Шеннона-Фано. Построить кодовое дерево и определить коэффициент сжатия методом Шеннона-Фано.

Нахождение вероятности появления символа в сообщении представлено в соответствии с табл.2:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |

Табл.2. Нахождение вероятности появления символа в сообщении.

Построение кодового дерева представлено в соответствии с рис.1.

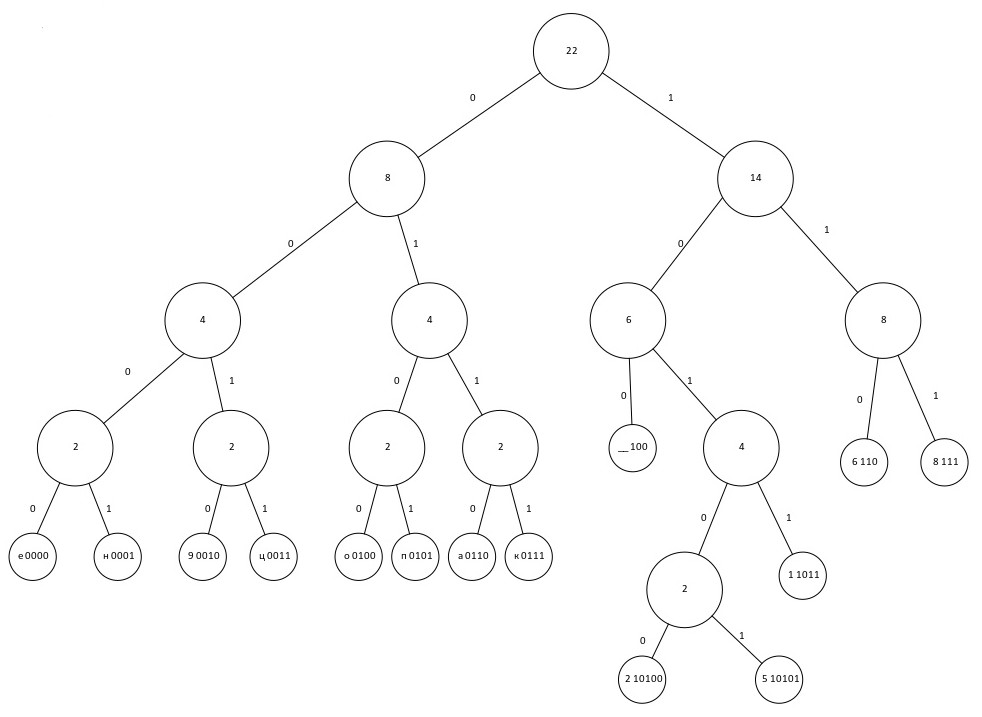


Рис.1. Кодовое дерево

|  |  |
| --- | --- |
| Символ | Двоичный код |
| 8 | 111 |
| 6 | 110 |
| \_ | 100 |
| 1 | 1011 |
| ц | 0011 |
| е | 0000 |
| н | 0001 |
| а | 0110 |
| 2 | 10100 |
| 5 | 10101 |
| 9 | 0010 |
| к | 0111 |
| о | 0100 |
| п | 0101 |

Табл.3.

Кодирование исходного сообщения 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

1. Выполнить сжатие информации методом Хаффмана. Построить кодовое дерево и определить коэффициент сжатия методом Хаффмана.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Вероятность появления | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 | 0,182 | 000 | 0,182 | 000 | | 0,182 | | 000 | | | 0,182 | | | 000 | | | | 0,182 | | | | 000 | | | 0,182 | 000 | |
| 6 | 0,182 | 001 | 0,182 | 001 | | 0,182 | | 001 | | | 0,182 | | | 001 | | | | 0,182 | | | | 001 | | | 0,182 | 001 | |
| \_ | 0,091 | 011 | 0,091 | 011 | | 0,091 | | 011 | | | 0,091 | | | 011 | | | | 0,091 | | | | 011 | | | 0,091 | 011 | |
| 1 | 0,091 | 110 | 0,091 | 110 | | 0,091 | | 110 | | | 0,091 | | | 110 | | | | 0,091 | | | | 110 | | | 0,091 | 110 | |
| ц | 0,045 | 01010 | 0,091 | 111 | | 0,091 | | 111 | | | 0,091 | | | 111 | | | | 0,091 | | | | 111 | | | 0,091 | 111 | |
| е | 0,045 | 01011 | 0,045 | 01010 | | 0,091 | | 0100 | | | 0,091 | | | 0100 | | | | 0,091 | | | | 0100 | | | 0,091 | 0100 | |
| н | 0,045 | 1010 | 0,045 | 01011 | | 0,045 | | 01010 | | | 0,091 | | | 100 | | | | 0,091 | | | | 100 | | | 0,091 | 0101 | |
| а | 0,045 | 1011 | 0,045 | 1010 | | 0,045 | | 01011 | | | 0,045 | | | 01010 | | | | 0,091 | | | | 101 | | | 0,091 | 100 | |
| 2 | 0,045 | 1000 | 0,045 | 1011 | | 0,045 | | 1010 | | | 0,045 | | | 01011 | | | | 0,045 | | | | 01010 | | | 0,091 | 101 | |
| 5 | 0,045 | 1001 | 0,045 | 1000 | | 0,045 | | 1011 | | | 0,045 | | | 1010 | | | | 0,045 | | | | 01011 | | |  | |
| 9 | 0,045 | 01000 | 0,045 | 1001 | | 0,045 | | 1000 | | | 0,045 | | | 1011 | | | |  | | |  | | |
| к | 0,045 | 01001 | 0,045 | 01000 | | 0,045 | | 1001 | | |  | | | |  | | | |  | | | |
| о | 0,045 | 1110 | 0,045 | 01001 | |  | | |  | | |  | | | |  | | | |
| п | 0,045 | 1111 |  | |  | |  | | |  | | |  | | | |

Табл.4. Сжатие информации методом Хаффмана

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 8 | 0,182 | 000 | 0,182 | | 000 | | 0,182 | | 000 | | 0,273 | | 01 | 0,364 | 1 | 0,364 | 1 | 0,636 | 0 | 1 |
| 6 | 0,182 | 001 | 0,182 | | 001 | | 0,182 | | 001 | | 0,182 | | 000 | 0,273 | 01 | 0,364 | 00 | 0,364 | 1 |
| \_ | 0,182 | 10 | 0,182 | | 10 | | 0,182 | | 10 | | 0,182 | | 001 | 0,182 | 000 | 0,273 | 01 |
| 1 | 0,091 | 011 | 0,182 | | 010 | | 0,182 | | 11 | | 0,182 | | 10 | 0,182 | 001 |
| ц | 0,091 | 110 | 0,091 | | 011 | | 0,182 | | 010 | | 0,182 | | 11 |
| е | 0,091 | 111 | 0,091 | | 110 | | 0,091 | | 011 | |  |
| н | 0,091 | 0100 | 0,091 | | 111 | |  |  | |
| а | 0,091 | 0101 |  |  | |

Табл.4. Сжатие информации методом Хаффмана - продолжение

Результат сжатия информации методом Хаффмана представлен в соответствии с табл.5:

|  |  |
| --- | --- |
| 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 |

Табл.5. Результат сжатия методом Хаффмана

Кодирование исходного сообщения “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

1. Архивирование файлов

Сжатие различных документов при помощи стандартного архиватора (WinZip, WinRar, 7-Zip и т.п.) представлено в соответствии с табл.6.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Документ | Расширение | Объем файла до архивации, Кбайт | Объем файла после архивации, Кбайт | Коэффициент сжатия |
| Текст | .doc | 159 | 156 | 0,0189 (1,9%) |
| Фотография | .jpg | 151,284 | 151,138 | 0,0015 (0,2%) |
| Рисунок | .bmp | 22300 | 3,63 | 0,9998 (100%) |
| Видео | .avi | 4200 | 4070 | 0,031 (3%) |
| Звук | .mp3 | 25,5 | 20,9 | 0,1804 (18,4%) |

Табл.6. Сжатие различных документов при помощи архиватора

Заключение

В ходе лабораторной работы были изучены такие методы сжатия информации, как метод RLE, дерево Шеннона-Фано и дерево Хаффмана. Также получен опыт использования архиватора 7-Zip для сжатия медиа-файлов.