|  |  |
| --- | --- |
| 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-02-01оп-22 Маслов Владислав Андреевич    подпись, дата |
| Череповец, 2023 | |

Цель: научиться кодировать и декодировать по методу Фано и Хаффмана.

Задачи для самостоятельного разбора

1. Проведите кодирование по методу Фано алфавита из четырех букв, вероятности которых равны 0,4; 0,3; 0,2 и 0,1.

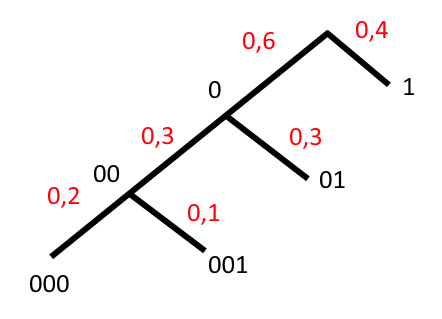


Рис.1. Кодирование алфавита из 4 букв по методу Фано.

|  |  |  |  |
| --- | --- | --- | --- |
| A | Б | В | Г |
| 0,4 | 0,3 | 0,2 | 0,1 |

Табл.1. Сообщения и их вероятности.

|  |  |  |  |
| --- | --- | --- | --- |
| A | Б | В | Г |
| 1 | 01 | 001 | 000 |

Табл.2. Алфавит в результате кодирования.

Цена кодирования: l=Li\*Pi=1\*0,4+2\*0,2+3\*0,2+3\*0,1=1,7

1. Алфавит содержит 7 букв, которые встречаются с вероятностями 0,4; 0,2; 0,1; 0,1; 0,1; 0,05; 0,05. Осуществите кодирование по методу Фано.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Сообщение | A | Б | В | Г | Д | Е | Ж |
| Вероятность, P | 0,4 | 0,2 | 0,1 | 0,1 | 0,1 | 0,05 | 0,05 |

Табл.3. Кодирование алфавита из 7 букв по методу Фано.

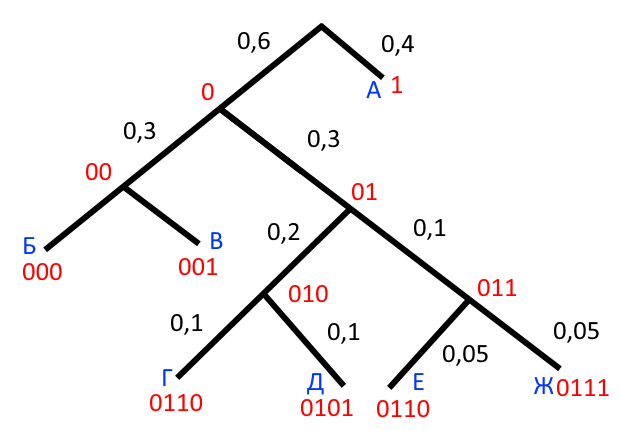


Рис.2. Кодирование алфавита из 7 букв по методу Фано.

Цена кодирования: l=1\*0,4+3\*0,2+3\*0,1+4\*(0,1\*2+0,05\*2) = 2,5

1. Алфавит состоит из двух букв, A и B, встречающихся с вероятностями P(A) = 0,8 и P(B) = 0,2. Примените метод Фано к кодированию всевозможных двухбуквенных и трехбуквенных комбинаций.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Слово | P | Разбиение на подгруппы | | | Код |
| АА | 0,64 | I |  |  | 0 |
| АБ | 0,16 | II | I |  | 10 |
| БА | 0,16 | II | I | 110 |
| ББ | 0,04 | II | 111 |

Табл.4. Кодирование двухбуквенных комбинаций методом Фано

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Слово | P | Разбиение на подгруппы | | | | | Код |
| ААА | 0,512 | I |  |  |  |  | 0 |
| ААБ | 0,128 | II (0,488) | I  (0,256) | I |  |  | 100 |
| АБА | 0,128 | II |  |  | 101 |
| БАА | 0,128 | II (0,232) | I |  |  | 110 |
| АББ | 0,032 | II (0,104) | I (0,064) | I | 11100 |
| БАБ | 0,032 | II | 11101 |
| ББА | 0,032 | II  (0,04) | I | 11110 |
| БББ | 0,008 | II | 11111 |

Табл.5. Кодирование трехбуквенных комбинаций методом Фано

1. Проведите кодирование по методу Хаффмана трехбуквенных слов из предыдущей задачи.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Слово | P | | P1 | | | P2 | | | P3 | | | P4 | | | P5 | | | P6 | |
| ААА | 0,512 |  | 0,512 |  | 0,512 | |  | 0,512 | |  | 0,512 | |  | 0,512 | |  | 0,512 | | 0 |
| ААБ | 0,128 |  | 0,128 |  | 0,128 | |  | 0,128 | |  | 0,232 | | 11 | 0,256 | | 10 | 0,488 | | 1 |
| АБА | 0,128 |  | 0,128 |  | 0,128 | |  | 0,128 | |  | 0,128 | | 100 | 0,232 | | 11 |  | |  |
| БАА | 0,128 |  | 0,128 |  | 0,128 | |  | 0,128 | | 110 | 0,128 | | 101 |  | |  |  | |  |
| АББ | 0,032 |  | 0,04 | 1111 | 0,064 | | 1110 | 0,104 | | 111 |  | |  |  | |  |  | |  |
| БАБ | 0,032 |  | 0,032 | 11100 | 0,04 | | 1111 |  | |  |  | |  |  | |  |  | |  |
| ББА | 0,032 | 11110 | 0,032 | 11101 |  | |  |  | |  |  | |  |  | |  |  | |  |
| БББ | 0,008 | 11111 |  |  |  | |  |  | |  |  | |  |  | |  |  | |  |

Табл.6. Кодирование трехбуквенных слов методом Хаффмана

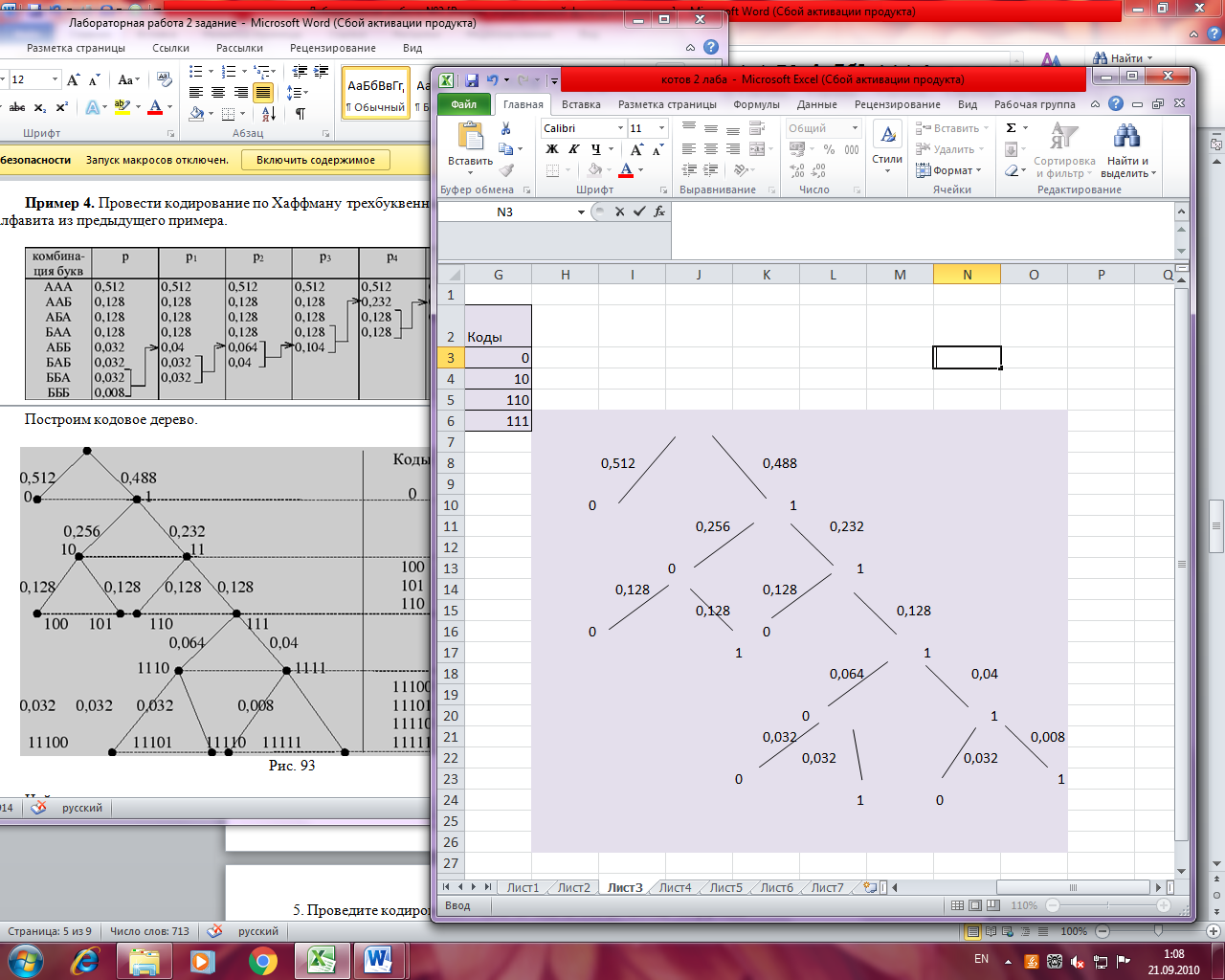


Рис.3. Дерево Хаффмана

Полученные коды:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Слово | Код |  | Слово | Код |
| ААА ААБ БАА  АБА | 1  100  110  101 |  | АББ БАБ ББА БББ | 11100  11101  11110  11111 |

Табл.7. Полученные коды в результате кодирования методом Хаффмана

1. Проведите кодирование 7 букв из задачи 302 по методу Хаффмана.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Сооб. | P | P1 | P2 | P3 | P4 | P5 | Коды |
| 1 | 0,4 | 0,4 | 0,4 | 0,4 | 0,4 | 0,6 (0) | 1 |
| 2 | 0,2 | 0,2 | 0,2 | 0,2 | 0,4 (00) | 0,4 (1) | 01 |
| 3 | 0,1 | 0,1 | 0,2 (000) | 0,2 (000) | 0,2 (01) |  | 0010 |
| 4 | 0,1 | 0,1 | 0,1 (0010) | 0,2 (001) |  |  | 0011 |
| 5 | 0,1 | 0,1 (0000) | 0,1 (0011) |  |  |  | 0000 |
| 6 | 0,05 (00010) | 0,1 (0001) |  |  |  |  | 00010 |
| 7 | 0,05 (00011) |  |  |  |  |  | 00011 |

Табл.6. Кодирование алфавита из 7 букв по методу Фано.

1. Проведите кодирование по методам Фано и Хаффмана пяти букв, равновероятно встречающихся.

Метод Фано:



Табл.8. Кодирование методом Фано пяти букв.

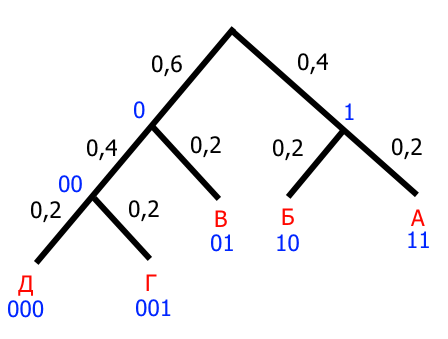


Рис.4. Дерево

Метод Хаффмана:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | P1 | P2 | P3 | P4 | Код |
| 1 | 0,2 | 0,4 | 0,4 | 0,6 0 | 11 |
| 2 | 0,2 | 0,2 | 0,4 00 | 0,4 1 | 10 |
| 3 | 0,2 | 0,2 000 | 0,2 01 |  | 01 |
| 4 | 0,2 10 | 0,2 001 |  |  | 001 |
| 5 | 0,2 11 |  |  |  | 000 |

Табл.9. Кодирование методом Хаффмана пяти букв.

1. Осуществите кодирование двухбуквенных комбинаций четырех букв из задачи 1.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| А | Б | В | Г |  |  | |  | |  | |  | |  | |  |  |  |
| 0,4 | 0,3 | 0,2 | 0,1 |  |  | |  | |  | |  | |  | |  |  |  |
|  |  |  |  |  |  | |  | |  | |  | |  | |  |  |  |
| Слово | P | Разбиение на подгруппы | | | | | | | | | | Код | |
| АА | 0,16 | I | I | I | |  | |  | |  | | 000 | |
| АБ | 0,12 | II | | 001 | |
| БА | 0,12 | II | I | | 010 | |
| ББ | 0,09 | II | | 011 | |
| АВ | 0,08 | II | I | I | | I | | 1000 | |
| ВА | 0,08 | II | | 1001 | |
| БВ | 0,06 | II | | I | | 1010 | |
| ВБ | 0,06 | II | | 1011 | |
| АГ | 0,04 | II | I | | I | | 1100 | |
| ВВ | 0,04 | II | | I | | 11010 | |
| ГА | 0,04 | II | | 11011 | |
| БГ | 0,03 | II | | I | | I | | 11100 | |
| ГБ | 0,03 | II | | 11101 | |
| ВГ | 0,02 | II | | I | | 11110 | |
| ГВ | 0,02 | II | | I | | 111110 | |
| ГГ | 0,01 | II | | 111111 | |

1. Проведите кодирование всевозможных четырехбуквенных слов из задачи 3.

|  |  |
| --- | --- |
| Символ | P |
| А | 0,8 |
| Б | 0,2 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Слово | P | Разбиение на подгруппы | | | | | | | | Код |
| АААА | 0,4096 | I | I |  |  |  |  |  |  | 00 |
| АААБ | 0,1024 | II | 01 |
| ААБА | 0,1024 | II | I | I | 100 |
| АБАА | 0,1024 | II | 101 |
| БААА | 0,1024 | II | I | I | 1100 |
| ААББ | 0,0256 | II | 1101 |
| АБАБ | 0,0256 | II | I | I | 11100 |
| АББА | 0,0256 | II | I | 111010 |
| БААБ | 0,0256 | II | 111011 |
| БАБА | 0,0256 | II | I | I | 111100 |
| ББАА | 0,0256 | II | 111101 |
| АБББ | 0,0064 | II | I | I | 1111100 |
| БАББ | 0,0064 | II | 1111101 |
| ББАБ | 0,0064 | II | I | 1111110 |
| БББА | 0,0064 | II | I | 11111110 |
| ББББ | 0,0016 | II | 11111111 |

1. Сравните эффективность кодов Фано и Хаффмана при кодировании алфавита из десяти букв, которые встречаются с вероятностями 0,3; 0,2; 0,1; 0,1; 0,1; 0,05; 0,05; 0,04; 0,03; 0,03.

|  |  |
| --- | --- |
| Буква | P |
| А | 0,3 |
| Б | 0,2 |
| В | 0,1 |
| Г | 0,1 |
| Д | 0,1 |
| Е | 0,05 |
| Ж | 0,05 |
| З | 0,04 |
| И | 0,03 |
| К | 0,03 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ФАНО | | | | | | | |
| Буква | P | Разбиение на группы | | | | | Код |
| А | 0,3 | I | I |  |  |  | 00 |
| Б | 0,2 | II | 01 |
| В | 0,1 | II | I | I | 100 |
| Г | 0,1 | II | 101 |
| Д | 0,1 | II | I | I | 1100 |
| Е | 0,05 | II | 1101 |
| Ж | 0,05 | II | I | I | 11100 |
| З | 0,04 | II | 11101 |
| И | 0,03 | II | I | 11110 |
| К | 0,03 | II | 11111 |

Цена кодирования: 2,95

2\*0,3 + 2\*0,2 + 3\*0,1 + 3\*0,1 + 4\*0,1 + 4\*0,05 + 5\*0,05 + 5\*0,04 + 5\*0,03 + 5\*0,03 = 2,95

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ХАФФМАН | | | | | | | | | | | | | | | | | | |
|  | P | | P1 | | P2 | | P3 | | P4 | | P5 | | P6 | | P7 | | P8 | |
| А | 0,3 | 00 | 0,3 | 00 | 0,3 | 00 | 0,3 | 00 | 0,3 | 00 | 0,3 | 00 | 0,3 | 00 | 0,4 | 1 | 0,6 | 0 |
| Б | 0,2 | 10 | 0,2 | 10 | 0,2 | 10 | 0,2 | 10 | 0,2 | 10 | 0,2 | 10 | 0,3 | 01 | 0,3 | 00 | 0,4 | 1 |
| В | 0,1 | 110 | 0,1 | 110 | 0,1 | 110 | 0,11 | 011 | 0,19 | 010 | 0,2 | 11 | 0,2 | 10 | 0,3 | 01 |
| Г | 0,1 | 111 | 0,1 | 111 | 0,1 | 111 | 0,1 | 110 | 0,11 | 011 | 0,19 | 010 | 0,2 | 11 |
| Д | 0,1 | 0100 | 0,1 | 0100 | 0,1 | 0100 | 0,1 | 111 | 0,1 | 110 | 0,11 | 011 |
| Е | 0,05 | 0111 | 0,06 | 0110 | 0,09 | 0101 | 0,1 | 0100 | 0,1 | 111 |
| Ж | 0,05 | 01010 | 0,05 | 0111 | 0,06 | 0110 | 0,09 | 0101 |
| З | 0,04 | 01011 | 0,05 | 01010 | 0,05 | 0111 |
| И | 0,03 | 01100 | 0,04 | 01011 |
| К | 0,03 | 01101 |

Цена кодирования: 2,95

2\*0,3 + 2\*0,2 + 3\*0,1 + 3\*0,1 + 4\*0,1 + 4\*0,05 + 5\*0,05 + 5\*0,04 + 5\*0,03 + 5\*0,03 = 2,95

В данном случае кодирования равно эффективны.

1. Сравните эффективность двоичного кода Фано и кода Хаффмана при кодировании алфавита из 16 букв, которые встречаются с вероятностями 0,25; 0,2; 0,1; 0,1; 0,05; 0,04; 0,04; 0,04; 0,03; 0,03; 0,03; 0,03; 0,02; 0,02; 0,01; 0,01.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ФАНО | | | | | | | | |
| Буква | P | Разбиение на группы | | | | | | Код |
| X1 | 0,25 | I | I |  |  |  |  | 00 |
| X2 | 0,2 | II | I | 010 |
| X3 | 0,1 | II | 011 |
| X4 | 0,1 | II | I | I | I | 1000 |
| X5 | 0,05 | II | 1001 |
| X6 | 0,04 | II | I | 1010 |
| X7 | 0,04 | II | 1011 |
| X8 | 0,04 | II | I | I | 1100 |
| X9 | 0,03 | II | I | 11010 |
| X10 | 0,03 | II | 11011 |
| X11 | 0,03 | II | I | I | 11100 |
| X12 | 0,03 | II | 11101 |
| X13 | 0,02 | II | I | I | 111100 |
| X14 | 0,02 | II | 111101 |
| X15 | 0,01 | II | I | 111110 |
| X16 | 0,01 | II | 111111 |

|  |  |
| --- | --- |
| Буква | P |
| X1 | 0,25 |
| X2 | 0,2 |
| X3 | 0,1 |
| X4 | 0,1 |
| X5 | 0,05 |
| X6 | 0,04 |
| X7 | 0,04 |
| X8 | 0,04 |
| X9 | 0,03 |
| X10 | 0,03 |
| X11 | 0,03 |
| X12 | 0,03 |
| X13 | 0,02 |
| X14 | 0,02 |
| X15 | 0,01 |
| X16 | 0,01 |

Цена кодирования: 3,44

2\*0,25 + 3\*0,2 + 3\*0,1 + 4\*0,1 + 4\*0,05 + 4\*0,04 + 4\*0,04 + 4\*0,04 + 5\*0,03 + 5\*0,03 + 5\*0,03 + 5\*0,03 + 6\*0,02 + 6\*0,02 + 6\*0,01 + 6\*0,01 = 3,44

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Хаффман | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|  | | P | | | | P1 | | | | P2 | | | | | P3 | | | | P4 | | | | | | P5 | | | | | P6 | |
| X1 | | 0,25 | | 01 | | 0,25 | | 01 | | 0,25 | | 01 | | | 0,25 | | 01 | | 0,25 | | | | 01 | | 0,25 | | 01 | | | 0,25 | 01 |
| X2 | | 0,2 | | 11 | | 0,2 | | 11 | | 0,2 | | 11 | | | 0,2 | | 11 | | 0,2 | | | | 11 | | 0,2 | | 11 | | | 0,2 | 11 |
| X3 | | 0,1 | | 101 | | 0,1 | | 101 | | 0,1 | | 101 | | | 0,1 | | 101 | | 0,1 | | | | 101 | | 0,1 | | 101 | | | 0,1 | 101 |
| X4 | | 0,1 | | 0000 | | 0,1 | | 0000 | | 0,1 | | 0000 | | | 0,1 | | 0000 | | 0,1 | | | | 0000 | | 0,1 | | 0000 | | | 0,1 | 0000 |
| X5 | | 0,05 | | 1001 | | 0,05 | | 1001 | | 0,05 | | 1001 | | | 0,05 | | 1001 | | 0,06 | | | | 1000 | | 0,07 | | 0011 | | | 0,08 | 0010 |
| X6 | | 0,04 | | 00011 | | 0,04 | | 00011 | | 0,04 | | 00011 | | | 0,05 | | 00010 | | 0,05 | | | | 1001 | | 0,06 | | 1000 | | | 0,07 | 0011 |
| X7 | | 0,04 | | 00100 | | 0,04 | | 00100 | | 0,04 | | 00100 | | | 0,04 | | 00011 | | 0,05 | | | | 00010 | | 0,05 | | 1001 | | | 0,06 | 1000 |
| X8 | | 0,04 | | 00101 | | 0,04 | | 00101 | | 0,04 | | 00101 | | | 0,04 | | 00100 | | 0,04 | | | | 00011 | | 0,05 | | 00010 | | | 0,05 | 1001 |
| X9 | | 0,03 | | 00111 | | 0,03 | | 00111 | | 0,04 | | 00110 | | | 0,04 | | 00101 | | 0,04 | | | | 00100 | | 0,04 | | 00011 | | | 0,05 | 00010 |
| X10 | | 0,03 | | 10000 | | 0,03 | | 10000 | | 0,03 | | 00111 | | | 0,04 | | 00110 | | 0,04 | | | | 00101 | | 0,04 | | 00100 | | | 0,04 | 00011 |
| X11 | | 0,03 | | 10001 | | 0,03 | | 10001 | | 0,03 | | 10000 | | | 0,03 | | 00111 | | 0,04 | | | | 00110 | | 0,04 | | 00101 | | |  |  |
| X12 | | 0,03 | | 000100 | | 0,03 | | 000100 | | 0,03 | | 10001 | | | 0,03 | | 10000 | | 0,03 | | | | 00111 | |  | |  | | |  |  |
| X13 | | 0,02 | | 100101 | | 0,02 | | 100101 | | 0,03 | | 000100 | | | 0,03 | | 10001 | |  | | | |  | |  | |  | | |  |  |
| X14 | | 0,02 | | 001100 | | 0,02 | | 001100 | | 0,02 | | 000101 | | |  | |  | |  | | | |  | |  | |  | | |  |  |
| X15 | | 0,01 | | 0011010 | | 0,02 | | 001101 | |  | |  | | |  | |  | |  | | | |  | |  | |  | | |  |  |
| X16 | | 0,01 | | 0011011 | |  | |  | |  | |  | | |  | |  | |  | | | |  | |  | |  | | |  |  |
| P7 | | | P8 | | | | P9 | | | | P10 | | | P11 | | | | P12 | | | P13 | | | | | P14 | | |
| 0,25 | 01 | | 0,25 | | 01 | | 0,25 | | 01 | | 0,25 | | 01 | 0,25 | | 01 | | 0,34 | | 00 | | 0,41 | | 1 | | 0,59 | | 0 |
| 0,2 | 11 | | 0,2 | | 11 | | 0,2 | | 11 | | 0,2 | | 11 | 0,21 | | 10 | | 0,25 | | 01 | | 0,34 | | 00 | | 0,41 | | 1 |
| 0,1 | 101 | | 0,11 | | 100 | | 0,15 | | 001 | | 0,19 | | 000 | 0,2 | | 11 | | 0,21 | | 10 | | 0,25 | | 01 | |  | |  |
| 0,1 | 0000 | | 0,1 | | 101 | | 0,11 | | 100 | | 0,15 | | 001 | 0,19 | | 000 | | 0,2 | | 11 | |  | |  | |  | |  |
| 0,09 | 0001 | | 0,1 | | 0000 | | 0,1 | | 101 | | 0,11 | | 100 | 0,15 | | 001 | |  | |  | |  | |  | |  | |  |
| 0,08 | 0010 | | 0,09 | | 0001 | | 0,1 | | 0000 | | 0,1 | | 101 |  | |  | |  | |  | |  | |  | |  | |  |
| 0,07 | 0011 | | 0,08 | | 0010 | | 0,09 | | 0001 | |  | |  |  | |  | |  | |  | |  | |  | |  | |  |
| 0,06 | 1000 | | 0,07 | | 0011 | |  | |  | |  | |  |  | |  | |  | |  | |  | |  | |  | |  |
| 0,05 | 1001 | |  | |  | |  | |  | |  | |  |  | |  | |  | |  | |  | |  | |  | |  |

Цена кодирования: 3,41

2\*0,25 + 2\*0,2 + 3\*0,1 + 4\*0,1 + 4\*0,05 + 5\*0,04 + 5\*0,04 + 5\*0,04 + 5\*0,03 + 5\*0,03 + 5\*0,03 + 6\*0,03 + 6\*0,02 + 6\*0,02 + 7\*0,01 + 7\*0,01 = 3,41

Цена кодирования алфавита по Хаффману меньше, чем по Фано, то есть для этого алфавита кодировать по Хаффману эффективнее.

Заключение

В ходе лабораторной работы были закодированы и декодированы различные сообщения по методу Фано и Хаффмана. Было произведено кодирование по методу Фано двухбуквенных комбинаций, кодирование по Хаффману трехбуквенных комбинаций, а также построено тринарное дерево Хаффмана для кодирования трехбуквенных комбинаций. Задачи, поставленные в ходе выполнения лабораторной работы, выполнены в полной мере.

Вопросы для самоконтроля

1. Как определяется среднее число элементарных сигналов, приходящихся на одну букву алфавита?

Среднее число элементарных сигналов, приходящихся на одну букву алфавита, может быть определено с использованием понятия энтропии информации. Энтропия информации представляет собой меру неопределенности или информационного содержания и выражается в единицах бит.

Для определения среднего числа элементарных сигналов на одну букву используется следующая формула: N = 1 / H, где N - среднее число элементарных сигналов, приходящихся на одну букву, H - энтропия информации.

Энтропия информации для заданного алфавита может быть вычислена по формуле:

H = - Σ (P(i) \* log2(P(i))), где P(i) - вероятность появления i-й буквы в тексте или сообщении.

1. Префиксные коды – это такие коды, в которых ни одна более короткая комбинация не является началом более длинной комбинации (удовлетворяют условию Фано).
2. Сколько требуется двоичных знаков для записи кодированного сообщения?

Количество двоичных знаков, необходимых для записи, зависит от длины сообщения и используемого метода кодирования.

1. На чем основано построение кода Фано?

Построение кода Фано основано на принципе оптимального кодирования, при котором избыточность кода минимальна. Код Фано является префиксным кодом.

1. Сжатие алфавита – это?

Сжатие кода – процесс уменьшения размера данных путем замены одного или нескольких символов в алфавите на более короткие символы или их последовательности.

Сжатие алфавита – это процесс уменьшения количества символов или букв, используемых для представления информации, с целью уменьшить объем данных или сообщения. Вместо использования полного алфавита, сжатие алфавита заменяет некоторые символы или комбинации символов на более короткие представления или коды, что позволяет сократить количество бит, необходимых для хранения или передачи информации. Это позволяет экономить место на диске, ускорять передачу данных и уменьшать потребление пропускной способности сети.

Сжатие алфавита – метод Хаффмана. Для того, чтобы закодировать сообщения по Хаффману, предварительно преобразуется таблица, задающая вероятности сообщений. Исходные данные записываются в столбец, две последние (наименьшие) вероятности в котором складываются, а полученная сумма становится новым элементом таблицы, занимающим соответствующее место в списке убывающих по величине вероятностей. Эта процедура продолжается до тех пор, пока в столбце не останутся всего два элемента

1. Какой код самый выгодный?

Самым выгодным кодом считается код, который обеспечивает минимальную среднюю длину кодового слова.

1. Основная теорема о кодировании.

Основная теорема о кодировании (теорема Шеннона) утверждает, что средняя длина кодового слова в оптимальном префиксном коде для источника информации с заданным алфавитом и вероятностями символов близка к энтропии этого источника. Энтропия является мерой неопределенности и определяет нижнюю границу для средней длины кодового слова.

1. Энтропия конкретных типов сообщений.

Энтропия конкретного типа сообщения – мера неопределённости сообщения, выраженная в количестве битов, необходимых для кодирования сообщения с минимальной погрешностью.

Энтропия конкретного типа сообщений зависит от вероятностей появления различных символов или элементов в этом сообщении. Чем более равномерное распределение вероятностей символов, тем выше энтропия и, соответственно, больше информации требуется для представления каждого символа.