Министерство образования науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

Тульский государственный университет

Институт прикладной математики и компьютерных наук

Кафедра

Информационная безопасность

**ПРАКТИЧЕСКАЯ РАБОТА №4**

по дисциплине

**ОСНОВЫ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ**

Выполнила:  
ст. гр. 230711, Павлова В.С. \_\_\_\_\_\_\_\_\_\_\_\_

Проверила:

к.т.н., доц. Баранова Е.М. \_\_\_\_\_\_\_\_\_\_\_\_

Тула, 2022 г.

**Практическая работа №4. Основы теории кодирования информационных сообщений**

**Цель работы:** изучить основы теории кодирования информационных сообщений, передаваемых по каналам связи.

**Решение задач, представленных в разделе самостоятельной работы:**

**Задача № 1.** Закодировать методом Шенно-Фано блоки «мы все учились понемногу чему-нибудь и как-нибудь». Каково среднее число символов на знак?

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| блок | мы | все | учились | понемногу | чему | нибудь | и | как | - |
| вероятность | 0,37 | 0,13 | 0,125 | 0,08 | 0,06 | 0,052 | 0,023 | 0,11 | 0,05 |

**Решение:** Из определения вероятности меры количества информации по Шеннону-Фано следует, что сообщение наиболее информативно, если вероятности появления его символов одинаковы. Чтобы повысить эффективность кода (сделать его наиболее оптимальным), следует разбить блоки на классы и подклассы, стремясь получить вероятности для отдельных групп.

Составим таблицу кодов для блоков.

|  |  |  |
| --- | --- | --- |
| Блок | Вероятность | Код |
| Мы | 0,37 | 00 |
| Все | 0,13 | 01 |
| Учились | 0,125 | 100 |
| Понемногу | 0,08 | 1100 |
| Чему | 0,06 | 1101 |
| Нибудь | 0,052 | 11100 |
| И | 0,023 | 11111 |
| Как | 0,11 | 10101 |
| - | 0,05 | 11110 |

Сообщению «мы все учились понемногу чему-нибудь и как-нибудь» соответствует выходная последовательность кодов: 00011001100110111110 11100111111010111100.

Определим среднее число символов на знак как произведение количества символов кода на вероятность появления:

**Ответ:** 00 01 100 1100 1101 11110 11100 11111 10101 11100; 3,11.

**Задача № 2.** Сообщение состоит из последовательности букв А, B и С, вероятности которых не зависят от предыдущего сочетания букв и равны Р(А)=0,7, Р(В)=0,2, Р(С)=0,1. Провести кодирование по алгоритму Шеннона-Фано отдельных букв и двухбуквенных сочетаний. Сравнить коды по их эффективности и избыточности.

**Решение:** Составим коды для отдельных букв.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Комбинации букв | Вероятность | Символы кода | | Код |
| A | 0,7 | 0 |  | 0 |
| B | 0,2 | 1 | 0 | 10 |
| C | 0,1 |  | 1 | 11 |

Составим коды для двухбуквенных сочетаний.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Комбинации букв | Вероятность | Символы кода |  |  |  |  |  | Код |
| AA | 0,49 | 0 |  |  |  |  |  | 0 |
| AB | 0,14 | 1 | 0 | 0 |  |  |  | 100 |
| BA | 0,14 |  |  | 1 |  |  |  | 101 |
| AC | 0,07 |  | 1 | 0 | 0 |  |  | 1100 |
| CA | 0,07 |  |  |  | 1 |  |  | 1101 |
| BB | 0,04 |  |  | 1 | 0 |  |  | 1110 |
| BC | 0,02 |  |  |  | 1 | 0 |  | 11110 |
| CB | 0,02 |  |  |  |  | 1 | 0 | 111110 |
| CC | 0,01 |  |  |  |  |  | 1 | 111111 |

Рассчитаем энтропию источника:

Среднее число двоичных символов кода, приходящихся на один символ:

Эффективность кода:

Избыточность кода:

Появление двухбуквенных сочетаний:

Рассчитаем энтропию источника:

Среднее число двоичных символов кода, приходящихся на один символ:

Эффективность кода:

Избыточность кода:

Таким образом, можно сделать вывод, что второй вариант будет эффективнее.

**Ответ:** эффективность кода N1=0,89, N2=0,991; избыточность кода R1=0,11, R2=0,009.

**Задача № 3.** Закодировать сообщение методом Шеннона-Фано «Теория информацииКодированияМодуляции».

**Решение:** составим алфавит кода и для каждого символа этого алфавита и определим вес (количество повторений символа в сообщении):

Т-1, Е-1, О-5, Р-3, И-8, Я-3, Н-2, Ф-1, М-2, А-2, Ц-2, К-1, Д-2, В-1, У-1, Л-1.

Запишем алфавит по убыванию веса символа:

И-8, О-5, Р-3, Я-3, Н-2, М-2, А-2, Ц-2, Д-2, Т-1, Е-1, Ф-1, К-1, В-1, У-1, Л-1.

Создадим таблицу кодов для символов:

|  |  |  |
| --- | --- | --- |
| Буква | Частота | Код |
| И | 8 | 00 |
| О | 5 | 010 |
| Р | 3 | 0110 |
| Я | 3 | 0111 |
| Н | 2 | 10000 |
| М | 2 | 10001 |
| А | 2 | 1001 |
| Ц | 2 | 1010 |
| Д | 2 | 1011 |
| Т | 1 | 11000 |
| Е | 1 | 11001 |
| Ф | 1 | 11010 |
| К | 1 | 11011 |
| В | 1 | 1110 |
| У | 1 | 11100 |
| Л | 1 | 11111 |

Используя полученную таблицу кодов, закодируем сообщение «Теория информацииКодированияМодуляции»:

11000 11001 010 0110 00 0111 00 10000 11010 010 0110 10001 1001 1010 00 00 11011 010 1011 00 0110 010 1110 1001 10000 00 0111 10001 010 1011 11100 11111 0111 1010 00 00.

**Ответ:** 11000 11001 010 0110 00 0111 00 10000 11010 010 0110 10001 1001 1010 00 00 11011 010 1011 00 0110 010 1110 1001 10000 00 0111 10001 010 1011 11100 11111 0111 1010 00 00.

**Задача № 4.** Построить код Шеннона-Фано для системы из семи букв: A, B, C, D, E, F, G, вероятности появления которых соответственно 0,1, 0,2, 0,05, 0,3, 0,05, 0,15, 0,15. Определить среднее количество разрядов на одну букву. Декодировать этим кодом последовательность:

10011101001000111101110101111000

**Решение.** Составим таблицу с алфавитом символов и их вероятностями:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| A | B | C | D | E | F | G |
| 0,1 | 0,2 | 0,05 | 0,3 | 0,05 | 0,15 | 0,15 |

Теперь составим таблицу кодов для символов:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Символ | Вероятность | Символы кода | | | | Код |
| D | 0,3 | 0 | 0 | | | 00 |
| B | 0,2 | 1 | | | 01 |
| F | 0,15 | 1 | 0 | 0 | | 100 |
| G | 0,15 | 1 | | 101 |
| A | 0,1 | 1 | 0 | | 110 |
| C | 0,05 | 1 | 0 | 1110 |
| E | 0,05 | 1 | 1111 |

Определим среднее количество разрядов на одну букву: 0,3\*2+0,2\*2+0,15\*3\*2+0,1\*3+0,05\*4\*2 = 2,6.

Последовательность кодов 100.1110.100.100.01.1110.1110.101.1110.00 соответствует сообщению «FCFFBCCGCD».

**Ответ:** FCFFBCCGCD.

**Задача № 5.** Провести эффективное кодирование ансамбля из восьми знаков (m=8), используя метод Шеннона-Фано.

**Решение.** Пусть имеется последовательность символов «qwwertty». Составим таблицу с алфавитом символов и их вероятностями:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| q | w | e | r | t | y |
| 0,125 | 0,25 | 0,125 | 0,125 | 0,25 | 0,125 |

Теперь составим таблицу кодов для символов:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Буква | Вероятности | Символы кода | | | Код |
| W | 0,25 | 0 | 0 |  | 00 |
| T | 0,25 | 1 | 01 |
| Q | 0,125 | 1 | 0 | 0 | 100 |
| E | 0,125 | 1 | 101 |
| R | 0,125 | 1 | 0 | 110 |
| Y | 0,125 | 1 | 111 |

Средняя длина кодового слова равна 0,25\*2\*2+0,125\*3\*4 = 2,5.

По формуле Шеннона , значит данный код можно считать оптимальным. Таким образом сообщению «qwwertty» соответствует выходная последовательность кодов 10000001011100101111.