

Федеральное государственное бюджетное образовательное учреждение высшего
образования
«Сибирский государственный университет телекоммуникаций и
информатики»
(СибГУТИ)

Кафедра прикладной математики и кибернетики

Практическая работа № 5

Недвоичное кодирование

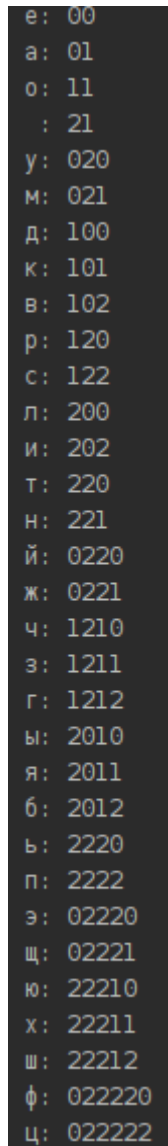
Выполнил:
студент группы ИП-715
Комашко Т.М,

Работу проверила:
доцент кафедры ПМ и К
Мачикина Е. П.

Новосибирск 2020 г.

В качестве не двоичного кодирования было рассмотрено троичное кодирование.

Результаты, полученные при работе программы:



```
е: 00
а: 01
о: 11
: 21
у: 020
м: 021
д: 100
к: 101
в: 102
р: 120
с: 122
л: 200
и: 202
т: 220
н: 221
й: 0220
ж: 0221
ч: 1210
з: 1211
г: 1212
ы: 2010
я: 2011
б: 2012
ь: 2220
п: 2222
э: 02220
щ: 02221
ю: 22210
х: 22211
ш: 22212
ф: 022220
ц: 022222
```

Рисунок 1. Символы с соответствующими им кодами



```
Средняя длина кодового слова: 2.813488449118158
```

Рисунок 2. Средняя длина кода

Средняя энтропия русского языка примерно равна 1,5.

Метод кодирования	Название текста	Оценка энтропии выходной послед-ти (частоты символов)	Оценка энтропии выходной послед-ти (частоты пар символов)	Оценка энтропии выходной послед-ти (частоты троек символов)
Метод Хаффмана	Эрнест Хэммингуэй «По ком звонит колокол» (отрывок на русском)	1.566202039 5798737	1.565138795 1888867	1.563375824 289553

Вывод:

Как и для двоичного для недвоичного кодирования в качестве метода был выбран метод Хаффмана, который позволяет получить оптимальный префиксный код. Оптимальным данный код является по двум причинам:

- 1) Чем больше вероятность появления символа, тем меньше длина его кодового слова;
- 2) Самые длинные коды различаются только в последнем порядке.

При сравнении результатов предыдущей лабораторной работы и данной, можно заметить, что оценки энтропии все так же близки к максимальному ее значению, но разница стала больше. Кодирование все так же является эффективным, но чем больше символов будет использоваться для кодирования, тем больше будет разрыв между оценками энтропии и максимальным ее значением и как следствие будет хуже сжатие.