

ФЕДЕРАЛЬНОЕ АГЕНТСТВО СВЯЗИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ
БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ»

Теория информации

Практическая работа №4
«Блочное кодирование»

Выполнил:
студент гр. ИП-911
Мироненко К.А.

Проверила:
доцент кафедры ПМиК
Мачикина Е.П.

Новосибирск
2022-2023 уч.год

Результат работы

```
Windows PowerShell
PS C:\Users\User\Desktop\oh shit, here we go again\8 семестр\Information_theory\labs\lab4> python .\lab4.py
Текст: 1984 - George Orwell. Русский язык. Часть первая
Энтропия оригинального текста: 4.383726570658277
Хаффман:
: 0.1578756618 - 110
о: 0.0936678877 - 000
е: 0.0692817479 - 1011
а: 0.0654439080 - 1001
и: 0.0606223384 - 1000
н: 0.0599945636 - 0111
т: 0.0541698487 - 0101
с: 0.0457045963 - 11111
л: 0.0428893175 - 11110
р: 0.0367604230 - 11101
в: 0.0365080187 - 11100
м: 0.0271237558 - 01001
к: 0.0267095538 - 01000
д: 0.0253245661 - 00111
у: 0.0239201626 - 00101
п: 0.0236548145 - 00100
ы: 0.0177071333 - 101010
ь: 0.0153125283 - 101000
я: 0.0150795397 - 011011
г: 0.0145682592 - 011010
з: 0.0145164839 - 011001
б: 0.0140634506 - 011000
ч: 0.0119730251 - 001100
й: 0.0091253867 - 1010110
ж: 0.0084070052 - 1010011
```

```
Windows PowerShell
п: 0.0236548145 - 00100
ы: 0.0177071333 - 101010
ь: 0.0153125283 - 101000
я: 0.0150795397 - 011011
г: 0.0145682592 - 011010
з: 0.0145164839 - 011001
б: 0.0140634506 - 011000
ч: 0.0119730251 - 001100
й: 0.0091253867 - 1010110
ж: 0.0084070052 - 1010011
х: 0.0071514555 - 1010010
ш: 0.0065366245 - 0011010
ю: 0.0043297048 - 10101110
ц: 0.0035142447 - 00110111
щ: 0.0032100651 - 00110110
э: 0.0023298860 - 101011110
ф: 0.0013785159 - 1010111111
1: 0.0001941572 - 1010111110111
ъ: 0.0001617976 - 1010111110101
4: 0.0001294381 - 1010111110100
0: 0.0001294381 - 1010111110011
3: 0.0001229662 - 1010111110010
9: 0.0001164943 - 1010111110001
8: 0.0000906067 - 10101111101101
2: 0.0000647191 - 10101111100001
5: 0.0000517752 - 10101111100000
7: 0.0000517752 - 101011111011001
6: 0.0000323595 - 101011111011000
Средняя длина кодовых слов (L ср.): 4.411975613860233
PS C:\Users\User\Desktop\oh shit, here we go again\8 семестр\Information_theory\labs\lab4>
```

```
Windows PowerShell
PS C:\Users\User\Desktop\oh shit, here we go again\8 семестр\Information_theory\labs\lab4> python .\lab4.py
Текст: 1984 - George Orwell. Русский язык. Часть первая
Энтропия оригинального текста: 3.9946805268364374
Хаффман:
о : 0.0206130189 - 111010
и : 0.0181989981 - 110010
е : 0.0174741447 - 101111
п : 0.0160632693 - 100110
с : 0.0154225507 - 100011
а : 0.0152995845 - 100010
н : 0.0152607531 - 100001
в : 0.0146523940 - 011011
то : 0.0136233610 - 010100
ст : 0.013321252 - 010000
о : 0.0108598574 - 1111011
но : 0.0102644421 - 1110011
на : 0.0100767568 - 1110001
не : 0.0094554539 - 1101010
ь : 0.0094425101 - 1101001
и : 0.0089894767 - 1100010
по : 0.0085817466 - 1011001
ра : 0.0084328928 - 1010110
ни : 0.0083099266 - 1010010
я : 0.0082128480 - 1010000
он : 0.0081804885 - 1001111
ли : 0.0079345561 - 1001010
м : 0.0076433203 - 1000001
ал : 0.0075527137 - 0111111
к : 0.0075074103 - 0111110
```

```
Windows PowerShell
кш: 0.0000064719 - 00111100111110001
бт: 0.0000064719 - 00111100111110000
бг: 0.0000064719 - 00111100111110011
щр: 0.0000064719 - 00111100111110010
шц: 0.0000064719 - 00111100111111101
цт: 0.0000064719 - 00111100111111100
ыу: 0.0000064719 - 00111100111111111
пч: 0.0000064719 - 00111100111111110
тх: 0.0000064719 - 00111100111111001
уо: 0.0000064719 - 00111100111111000
зм: 0.0000064719 - 00111100111111011
юй: 0.0000064719 - 00111100111111010
йа: 0.0000064719 - 00111100111100101
нш: 0.0000064719 - 00111100111100100
юк: 0.0000064719 - 00111100111100111
оа: 0.0000064719 - 00111100111100110
хэ: 0.0000064719 - 00111100111100001
хъ: 0.0000064719 - 00111100111100000
жл: 0.0000064719 - 00111100111100011
нр: 0.0000064719 - 00111100111100010
фс: 0.0000064719 - 00111100111101101
ез: 0.0000064719 - 00111100111101100
юи: 0.0000064719 - 00111100111101111
23: 0.0000064719 - 00111100111101110
33: 0.0000064719 - 00111100111101001
юо: 0.0000064719 - 00111100111101000
вг: 0.0000064719 - 00111100111101011
а: 0.0000064719 - 00111100111101010
Средняя длина кодовых слов (L ср.): 8.015144258772644
PS C:\Users\User\Desktop\oh shit, here we go again\8 семестр\Information_theory\labs\lab4>
```

```
Windows PowerShell
PS C:\Users\User\Desktop\oh shit, here we go again\8 семестр\Information_theory\labs\lab4> python .\lab4.py
Текст: 1984 - George Orwell. Русский язык. Часть первая
Энтропия оригинального текста: 3.651445161424523
Хаффман:
по: 0.0070737927 - 0110101
не: 0.0065560402 - 0100001
пр: 0.0053004906 - 11110001
и : 0.0051904682 - 11101100
то : 0.0051322210 - 11100111
на: 0.0051063334 - 11100110
ть : 0.0046727157 - 11001100
на : 0.0045303338 - 10111111
в : 0.0044656148 - 10111100
но : 0.0044008957 - 10110110
не : 0.0040643566 - 10011110
он : 0.0039737500 - 10010111
ли : 0.0039608061 - 10010101
го : 0.0039478623 - 10010011
он: 0.0036436828 - 01110101
сто: 0.0034236380 - 01011100
ом : 0.0030547394 - 00100011
ст: 0.0029835484 - 00011010
бы: 0.0029576608 - 00011001
за: 0.0028864698 - 00001010
про: 0.0028282227 - 00000100
ой : 0.0028217508 - 00000011
ся : 0.0028152789 - 00000001
что: 0.0026470093 - 111100001
ми : 0.0025887622 - 111010011
```

```
Windows PowerShell
330: 0.0000064719 - 01001110001111101
сню: 0.0000064719 - 01001110001111100
зус: 0.0000064719 - 01001110001111111
чьс: 0.0000064719 - 01001110001111110
яд: 0.0000064719 - 01001110001111001
чр: 0.0000064719 - 01001110001111000
уац: 0.0000064719 - 01001110001111011
лня: 0.0000064719 - 01001110001111010
изж: 0.0000064719 - 01001110001100101
зжо: 0.0000064719 - 01001110001100100
жог: 0.0000064719 - 01001110001100111
ьюо: 0.0000064719 - 01001110001100110
юн: 0.0000064719 - 01001110001100001
туе: 0.0000064719 - 01001110001100000
лбе: 0.0000064719 - 01001110001100011
лза: 0.0000064719 - 01001110001100010
пощ: 0.0000064719 - 01001110001101101
лту: 0.0000064719 - 01001110001101100
з э: 0.0000064719 - 01001110001101111
тпл: 0.0000064719 - 01001110001101110
люе: 0.0000064719 - 01001110001101001
юеш: 0.0000064719 - 01001110001101000
вг: 0.0000064719 - 01001110001101011
вгл: 0.0000064719 - 01001110001101010
ирс: 0.0000064719 - 01001110000010101
рсв: 0.0000064719 - 01001110000010100
ла: 0.0000064719 - 01001110000010111
а: 0.0000064719 - 01001110000010110
Средняя длина кодовых слов (L ср.): 10.982797675291012
PS C:\Users\User\Desktop\oh shit, here we go again\8 семестр\Information_theory\labs\lab4>
```

Результаты

Метод кодирования	Название текста	Избыточность кодирования (Длина блока $n=1$)	Избыточность Кодирования (Длина блока $n=2$)	Избыточность кодирования (Длина блока $n=3$)
Код Хаффмана	1984 — George Orwell. Русский язык. Часть первая	0.028249043201955537	0.012891602549884418	0.009487397005814469

Энтропия оригинального текста ($n=1$): 4.383726570658277

Средняя длина кодовых слов ($n=1$): 4.411975613860233

Энтропия оригинального текста ($n=2$): 3.9946805268364374

Средняя длина кодовых слов ($n=2$): 8.015144258772644

Энтропия оригинального текста ($n=3$): 3.651445161424523

Средняя длина кодовых слов ($n=3$): 10.982797675291012