

Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Сибирский государственный университет телекоммуникаций и информатики»
(СибГУТИ)

Институт информатики и вычислительной техники

09.03.01 "Информатика и вычислительная техника"
профиль "Программное обеспечение средств
вычислительной техники и автоматизированных
систем"

Практическая работа №3

по дисциплине «Теория информации»

«Блочное кодирование»

Выполнил: студент 4 курса
ИВТ, гр. ИП-111
Кузьменок Д.В.

Работу проверил: доцент кафедры ПМиК
Мачикина Елена Павловна

Новосибирск 2025

Цель работы:

Экспериментальное изучение свойств блочного кодирования.

Задание:

1. Для выполнения работы необходим сгенерированный файл с неравномерным распределением из практической работы 1.
При блочном кодировании входная последовательность разбивается на блоки равной длины, которые кодируются целиком. Поскольку вероятностное распределение символов в файле известно, то и вероятности блоков могут быть вычислены и использованы для построения кода.
2. Закодировать файл блочным методом кодирования (можно использовать любой метод кодирования), размер блока $n = 1, 2, 3, 4$.
Вычислить избыточность кодирования на символ входной последовательности для каждого размера блока.
3. После тестирования программы необходимо заполнить таблицу и проанализировать полученные результаты, сравнить с теоретическими оценками.

Результаты работы

```
Энтропия файла 1:
Оценка 1: 1,58494238365313
Оценка 2: 1,5845081162634331
Оценка 3: 1,5829342858058595
Энтропия файла 2:
Оценка 1: 1,4810942368117777
Оценка 2: 1,4804149316879474
Оценка 3: 1,4795132467268444
Энтропия файла 3:
Оценка 1: 4,361135988572305
Оценка 2: 3,9361564407913097
Оценка 3: 3,37716111991902
Средняя длина кодового слова в file1.txt для блока размером 1: 1,665
Избыточность в file1.txt для блока размером 1: 0,05051137326667069
Средняя длина кодового слова в file1.txt для блока размером 2: 3,2114
Избыточность в file1.txt для блока размером 2: 0,013475771803255841
Средняя длина кодового слова в file1.txt для блока размером 3: 4,788178817881788
Избыточность в file1.txt для блока размером 3: 0,008392615007480819
Средняя длина кодового слова в file1.txt для блока размером 4: 6,3492
Избыточность в file1.txt для блока размером 4: 0,0055515493323111224
Средняя длина кодового слова в file2.txt для блока размером 1: 1,4936
Избыточность в file2.txt для блока размером 1: 0,008443597225212507
Средняя длина кодового слова в file2.txt для блока размером 2: 2,9872
Избыточность в file2.txt для блока размером 2: 0,00900723369112631
Средняя длина кодового слова в file2.txt для блока размером 3: 4,465946594659466
Избыточность в file2.txt для блока размером 3: 0,006275383370582566
Средняя длина кодового слова в file2.txt для блока размером 4: 5,9292
Избыточность в file2.txt для блока размером 4: 0,004892941743167648
Средняя длина кодового слова в file3.txt для блока размером 1: 4,3833710773212555
Избыточность в file3.txt для блока размером 1: 0,005098462604059146
Средняя длина кодового слова в file3.txt для блока размером 2: 7,9001294219154445
Избыточность в file3.txt для блока размером 2: 0,003533464783634488
Средняя длина кодового слова в file3.txt для блока размером 3: 10,186347460368813
Избыточность в file3.txt для блока размером 3: 0,005632103735646217
Средняя длина кодового слова в file3.txt для блока размером 4: 10,763589301121657
Избыточность в file3.txt для блока размером 4: 0,004522148107201787
```

Полученные результаты для трех файлов

	Длина блока n=1	Длина блока n=2	Длина блока n=3	Длина блока n=4
Оценка избыточности кодирования на один символ входной последовательности	0,008443597225212507	0,00900723369112631	0,006275383370582566	0,004892941743167648

Теоретическая оценка:

Длина блока 1:

$$H = -(0.5 * \log_2 0.5 + 0.3 * \log_2 0.3 + 0.2 * \log_2 0.2) = 1.485$$

Кодовые слова (кодирование Хаффмана): A = 0, B = 10, C = 11

$$L_{cp} = 0.5 * 1 + 0.3 * 2 + 0.2 * 2 = 1.5$$

$$r = L_{cp} - H = 0.015$$

Длина блока 2:

$$H = -(0.25 * \log_2 0.25 + 0.15 * \log_2 0.15 + 0.1 * \log_2 0.1 + 0.15 * \log_2 0.15 + 0.09 * \log_2 0.09 + 0.06 * \log_2 0.06 + 0.1 * \log_2 0.1 + 0.06 * \log_2 0.06 + 0.04 * \log_2 0.04) = 2.971$$

$$L_{cp} = 2,9872$$

$$r = L_{cp} - H = 0.0162$$

Длина блока 3:

$$H = 4.46$$

$$L_{cp} = 4,466$$

$$r = L_{cp} - H = 0.006$$

Длина блока 4:

$$H = 5,924$$

$$L_{cp} = 5,929$$

$$r = L_{cp} - H = 0.005$$

Вывод:

В ходе исследования блочного кодирования благодаря алгоритму Хаффмана было замечено, что с увеличением размера блока, избыточность закодированного сообщения обычно уменьшается, поскольку алгоритм эффективнее использует зависимости между символами.