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 №2

по дисциплине «Теория информации»

«Побуквенное кодирование текстов»

Выполнил: студент 4 курса

ИВТ, гр. ИП-111

Кузьменок Д.В.

Работу проверил: доцент кафедры ПМиК

Мачикина Елена Павловна

Новосибирск 2025

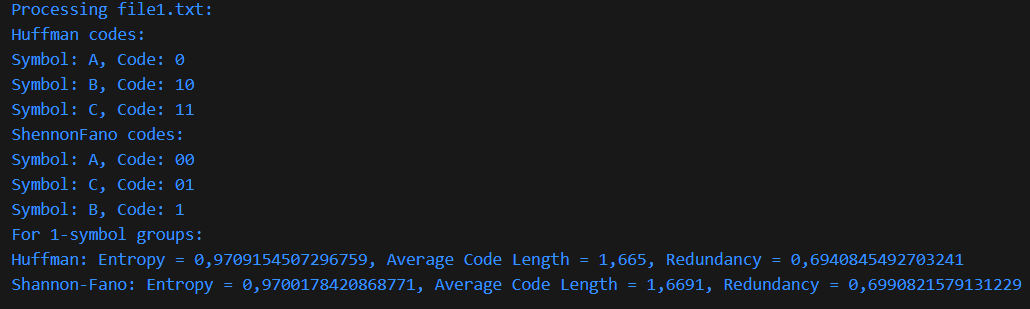
**Цель работы:**

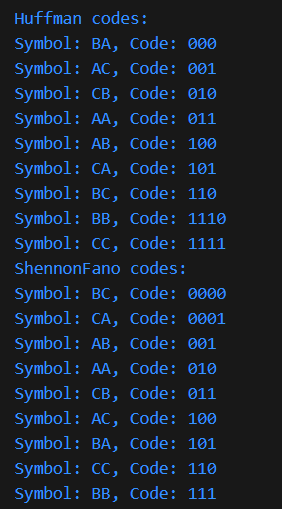
Экспериментальное изучение избыточности сжатия текстового файла.

**Задание:**

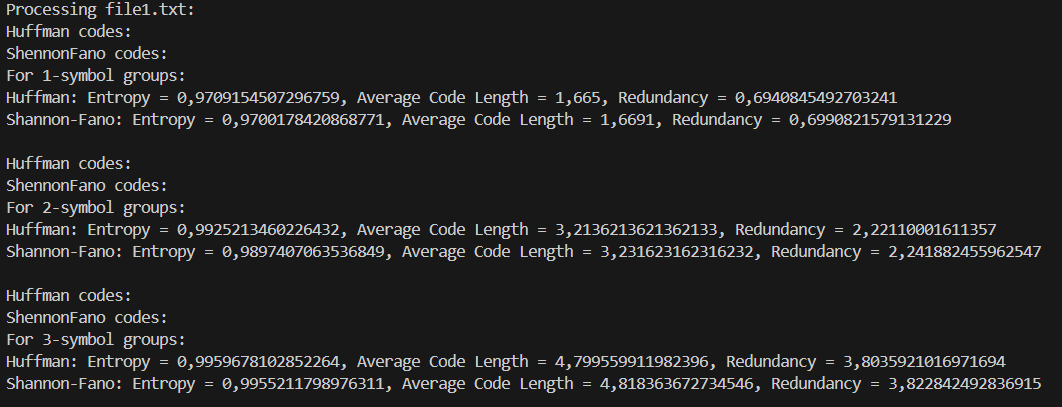
1. Запрограммировать процедуру двоичного кодирования текстового файла побуквенным кодом. В качестве методов сжатия использовать метод Хаффмана и метод Шеннона (или метод Фано). Текстовые файлы использовать те же, что и в практической работе 1.
2. Вычислить среднюю длину кодовых слов и оценить избыточность кодирования для каждого построенного побуквенного кода.
3. После кодирования текстового файла вычислить оценки энтропии файла с закодированным текстом **, , ** (после кодирования последовательность содержит 0 и 1) и заполнить таблицу.

**Результаты работы**

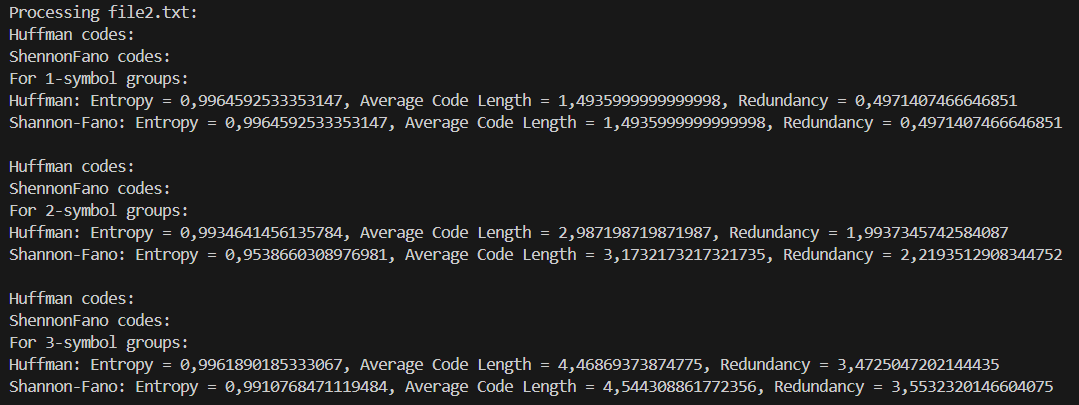




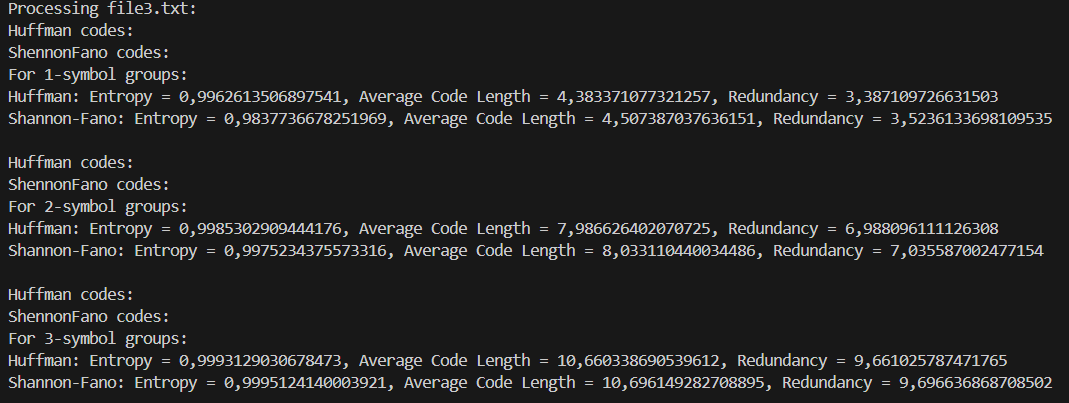
Генерация кодов для сиволов



Результаты для первого файла



Результаты для второго файла



Результаты для третьего файла

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Метод кодирования | Название текста | Оценка  избыточности кодирования |  |  |  |
| Код Хаффмана | Три символа с одинаковыми вероятностями | 0,6940845492703241 | 0,9709154507296759 | 0,9925213460226432 | 0,9959678102852264 |
| Код Фано | Три символа с одинаковыми вероятностями | 0,6990821579131229 | 0,9700178420868771 | 0,9897407063536849 | 0,9955211798976311 |
| Код Хаффмана | Три символа с разными вероятностями (0.5, 0.2, 0.3) | 0,4971407466646851 | 0,9964592533353147 | 0,9934641456135784 | 0,9961890185333067 |
| Код Фано | Три символа с разными вероятностями (0.5, 0.2, 0.3) | 0,4971407466646851 | 0,9964592533353147 | 0,9934641456135784 | 0,9961890185333067 |
| Код Хаффмана | 1984 – Джордж Оруэлл (английский текст) | 3,387109726631503 | 0,9962613506897541 | 0,9985302909444176 | 0,9993129030678473 |
| Код Фано | 1984 – Джордж Оруэлл (английский текст) | 3,5236133698109535 | 0,9837736678251969 | 0,9975234375573316 | 0,9995124140003921 |

**Вывод:**

При кодированиях были получены префиксные код, в которых используется избыточность сообщения (коды более частых символов состоят из коротких последовательностей, а коды более редких символов – из более длинных).

Можно увидеть, что данные методы кодирования обладают высокой избыточностью. Энтропия полученных последовательностей близка к единице, что говорит о том, что на один символ приходится один бит информации. При том, при выборе пар или троек символов энтропия почти не меняется. Это говорит о том, что символы в получившихся кодах равновероятны, что подтверждает эффективность кодирования.