

Практическая работа №2

Побуквенное кодирование текстов

Цель работы: Экспериментальное изучение избыточности сжатия текстового файла.

Язык программирования: C, C++, C#, Python

Результат: программа, тестовые примеры, отчет.

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 H_1 , H_2 , H_3 (после кодирования последовательность содержит 0 и 1) и заполнить таблицу.

Метод кодирования	Название текста	Оценка избыточности кодирования	H_1	H_2	H_3

Избыточность кодирования определяется как $r = L_{cp} - H$, где H – предельная энтропия текста, L_{cp} – средняя длина кодовых слов.

4. Оформить отчет, загрузить отчет в электронную среду. Отчет обязательно должен содержать заполненную таблицу и анализ полученных результатов.

По желанию в отчет можно включить описание программной реализации.

В отчет не нужно включать содержимое этого файла.

5. Анализ полученных результатов можно оформить в виде ответов на вопросы

1. Какой из реализованных методов побуквенного кодирования более оптимальный? Почему? Обоснуйте ответ, используя полученные результаты значений избыточности

2. Как соотносятся между собой полученные оценки H_1 , H_2 , H_3 для каждого закодированного файла? Равны? не равны? Чем объясняется такое поведение оценок?
3. Можно ли судить о качестве побуквенного кода по закодированной последовательности (по оценкам H_1 , H_2 , H_3)? Обоснуйте свой ответ.