

Практическая работа №3

Побуквенное кодирование текстов

Цель работы: Экспериментальное изучение процесса сжатия текстового файла.

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

Результат: программа, тестовые примеры, отчет.

1. Запрограммировать процедуру двоичного кодирования текстового файла. В качестве метода побуквенного кодирования использовать метод Шеннона, или метод Фано, или метод Хаффмана. Текстовые файлы использовать те же, что и в практических работах 1, 2.
2. Проверить, что построенный код для каждого файла является префиксным. Вычислить среднюю длину кодовых слов и оценить избыточность каждого построенного побуквенного кода.
3. После кодирования текстового файла вычислить оценки энтропии выходной последовательности (последовательность содержит 0 и 1), используя частоты отдельных символов, пар символов и троек символов и заполнить таблицу.

Метод кодирования	Название текста	Оценка избыточности кодирования	Оценка энтропии выходной посл-ти (частоты символов)	Оценка энтропии выходной посл-ти (частоты пар символов)	Оценка энтропии выходной посл-ти (частоты троек символов)

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.