

Пояснительная записка

Беспалова Анастасия

Июнь 2020

1 Описание формата `zmh`

За первый проход формируется префиксный код в соответствии с частотой встречаемых символов. Т.е. чем чаще встречается символ — тем короче для него код. Самый длинный код для наименее вероятных символов.

За второй проход выполняется кодирование файла.

Полученный файл формата `zmh` — это файл, состоящий из двух строк: первая строка содержит в себе таблицу соответствия код-символ, вторая строка содержит закодированный сжатый текст.

Декодирование происходит быстро (за линейное время), так как обратное соответствие представляется в коде программы в виде бинарного дерева с листьями-символами.

На самом деле, размер получаемого `zmh` файла больше, чем размер исходного, так как 0 и 1 записаны в виде символов. Потенциально можно перевести данные тексты в бинарный вид, тогда архиватор будет действительно сжимать. Проще всего это сделать, если в отдельном файле хранить таблицу соответствия, а в отдельном — закодированный текст в бинарном виде.