

Кодирование и сжатие информации: домашнее задание¹

Задача 1. Указать какие из кодов для букв алфавита $\{a, b, c, d\}$ являются префиксными.

- $\{1, 10, 0, 01\}$ • $\{00, 010, 011, 01\}$ • $\{10, 010, 011, 11\}$ • $\{1, 00, 010, 011\}$

Задача 2. Указать какие из кодов для букв алфавита $\{a, b, c, d\}$ являются разделимыми.

- $\{00, 010, 011, 01\}$ • $\{1, 10, 0, 01\}$ • $\{10, 010, 011, 11\}$ • $\{01, 010, 110, 11\}$

Задача 3. Какова средняя длина кода Шеннона–Фано для источника информации с распределением вероятностей $(0.5, 0.3, 0.1, 0.1)$?

- 1.2 • 1.7 • 0.7 • 1.5

Задача 4. Для сообщения “abcbbbbbacabbacddacdbbaccbbadadadddd abccccbacabbacbbadddbacbbddadadcc bcabbcdabacbbacbbddcbbaccbbdbdadaac<EOF>”, приняв за алфавит $\{a, b, c, d, <space>, <EOF>\}$

1. Определить энтропию
2. Построить коды Шеннона, Шеннона–Фано и Хаффмана для отдельных символов
3. Построить коды Шеннона, Шеннона–Фано и Хаффмана для двухбуквенных блоков символов
4. Закодировать сообщение
5. Определить среднюю длину и избыточность для всех кодов
6. Закодировать и декодировать сообщение, используя адаптивное сжатие по Хаффмену
7. Закодировать и декодировать, используя адаптивное сжатие по Хаффмену (с символом ESC)
8. Закодировать и декодировать сообщение, используя арифметическое кодирование²
9. Закодировать и декодировать сообщение, используя любой из алгоритмов семейства LZ*: LZ77, LZ78, LZW, LZSS², LZMA², LZ4², ...

Задача* 5.³ Реализовать⁴ любые (все) методы и алгоритмы из этого задания на любом языке программирования. Для тестирования можно взять любой реальный файл/изображение, обработав его, например, побайтово (char).

¹Задачи без * являются обязательными

²Не обсуждались на паре, изучить самостоятельно

³К следующей паре будет утверждена система баллов и/или плюшек за выполнение необязательных заданий

⁴На любом нормальном (не эзотерическом) языке программирования; задание должно быть оформлено как pull request в git репозитории (поставить danyaberezun@gmail.com в проверяющие), с соблюдением стилей, автоматическим CI, описанием и т.д.