

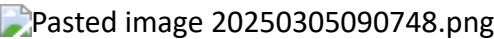
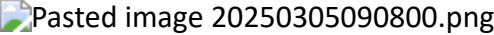
Одно из применений сбалансированных бинарных деревьев (как строили на рк ака деревья сжатия) - поиск
Как строить деревья поиска? 2 подхода:

- Самые частые данные - выше. Данные встречаются чаще -> их чаще будут искать
- Более часто искомые данные - выше. По очевидным причинам: чаще ищут - быстрее выдавать

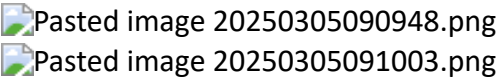
Статистические методы сжатия данных

Код	Условия оптимальности	Схема
Шеннона-Фано, 1948	Асимптотическое достижение оптимальности при увеличении числа сообщений (больше исходник - лучше сжатие)	var-var, block-var
Хаффмана, 1952	1) Каждое сообщение - уникальное кодовое слово 2) Сжатый текст = конкатенация кодовых слов	var-var, block-var
Универсальные коды, 1975		
Арифметические коды, 1963		

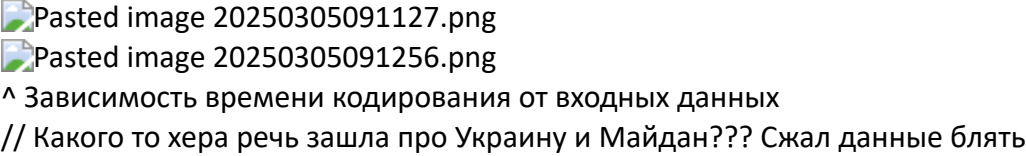
Коды Шеннона-Фано

- длина кода символа обратно пропорциональна частоте встречаемости символа



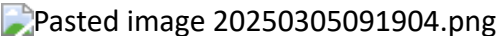
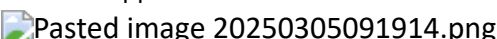
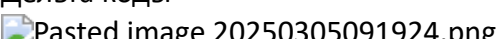
Статические коды Хаффмана



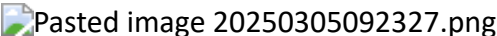
Сравнение Хаффмана и Шеннона-Фано



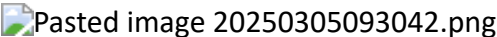
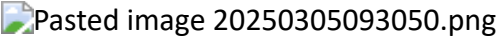
Универсальные коды Элиаса

Два типа - гамма и дельта коды

Гамма коды

Дельта коды

Средняя длина $S = c_1 (H + c_2)$
Особо не понадобятся судя по всему так что ебал учить это всё лол

Коды Фибоначчи


Такой большой разрыв позволяет быстро набирать вес и получить достаточную избыточность
Позволяет прикол - если разница чисел закодированных слов достаточно мала (условно около 100), то два слова можно считать практически одинаковыми

Арифметические коды

Тоже прикалываются с вероятностями охуеть


Прикол судя по всему в том, что наиболее вероятные символы вероятнее попадутся. Весь метод - проекция встречаемости на отрезок [0, 1]

Применение всей этой ебалы

Коды Элиаса - позволяют быстро кодировать/декодировать на ходу благодаря отсутствию особой привязки к вероятностям, пусть они и сравнительно неэффективны
В кодах Элиаса таблица соответствия составляется до кодирования, то есть её не надо передавать. Они интересны когда:

- Заранее неизвестна последовательность
- Имеется возможность составить таблицу кодирования до прохода

Арифметические коды - каждому символу можно подобрать несколько способов кодирования. Интересны когда:

- Символы в сообщениях появляются очень неравномерно
При реализации сжатия на ПЛИСах/МК лучше выбирать более простые алгоритмы