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 достиженеие 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
  - Pasted image 20250305090748%20darkmode.png
  - Pasted image 20250305090800%20darkmode.png

#### Статические коды Хаффмана

- Pasted image 20250305090948%20darkmode.png
- Pasted image 20250305091003%20darkmode.png

## Сравнение Хаффмана и Шеннона-Фано

- Pasted image 20250305091127%20darkmode.png
- Pasted image 20250305091256%20darkmode.png
- ^ Зависимость времени кодирования от входных данных
- // Какого то хера речь зашла про Украину и Майдан??? Сжал данные блять

# Универсальные коды Элиаса

Два типа - гамма и дельта коды

Pasted image 20250305091904%20darkmode.png

Гамма коды

Pasted image 20250305091914%20darkmode.png

Дельта коды

Pasted image 20250305091924%20darkmode.png

Средняя длина  $S=c_1\left(H+c_2\right)$ 

Особо не понадобятся судя по всему так что ебал учить это всё лол

# Коды Фибоначчи

Pasted image 20250305092327%20darkmode.png

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

- Pasted image 20250305093042%20darkmode.png
- Pasted image 20250305093050%20darkmode.png

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