

# Хеширование

$K \text{ или } \in \mathbb{Z}_+$

## Прямая адресация

Множество = массив булевых значений размера  $N$ , где  $N$  - это макс. значение ключа

Операции  $\mathcal{O}(1)$

Размер -  $\mathcal{O}(N)$

$N \sim 2^{40} \Rightarrow \text{fail}$

## Хеш-функция

$$h: K \rightarrow \{0, 1, \dots, m-1\}$$

Хранить значение  $k$  в ячейке с  $\mathcal{N} h(k)$ .

Память  $\mathcal{O}(m)$

$L = \frac{n}{m}$  - коэффициент заполнения

$n$  - # ключей

$m$  - размер образа  $h$ .

→ Сложность операции  $\mathcal{O}(L+1)$ .

## Разрешение коллизий:

$$k_1, k_2, k_1 \neq k_2 \quad h(k_1) = h(k_2)$$

- метод цепочек

В  $\mathcal{N}$  ячейке хранить относительн. список операций  $\mathcal{O}(\text{длина цепочки})$

- открытая адресация

$$h: K \times \{0, \dots, m-1\} \rightarrow \{0, \dots, m-1\}$$

Ищем значение в ячейках с  $\mathcal{N} h(k, i)$

$$i = 0, \dots, m-1$$

$$h(k, 0), h(k, 1), \dots, h(k, m-1)$$

$$h(k, i) = (h'(k) + i) \bmod m \quad \text{линейный поиск}$$

$$h' : K \rightarrow \{0, \dots, m-1\}$$

$$\begin{array}{c} h(k, 0) \\ \hline \text{|||X|||} \\ \hline h(k, 0) \quad h(k, 1) \end{array}$$

NB: при удавлении файла запись сразу помечается "deleted".

Двойное хеширование:

$$h(k, i) = h_1(k) + i \cdot h_2(k) \bmod m$$

$$h_2(k) \neq 0, \text{НОД}(h_2(k), m) = 1$$

Можно выбрать  $m$  - простым.

Плотность равномерного хеширования.

-  $h(k)$  - случайная величина, равномерно на  $\{0, \dots, m-1\}$

-  $h(k_1)$  и  $h(k_2)$  - независимы

УТВ: Вероятность коллизии  $k_1$  и  $k_2 = \frac{1}{m}$

$$P[h(k_1) = h(k_2)] = \frac{1}{m}$$

УТВ: Доля элементов в таблице в среднем  $\Theta(1)$

$\Delta$   $X_{kl}$  - случайная величина совб.  $h(k) = l$ .

$$E[X_{kl}] = \frac{1}{m} = P[h(k) \neq l] \cdot 0 + P[h(k) = l] \cdot 1$$

$$E[\text{доля элементов в ячейке } \neq l] = E\left[\sum_{i=1}^n X_{k_i, l}\right] =$$

$$= \sum_{i=1}^n E[X_{k:i}] = n \cdot \frac{1}{m} = 2 \triangleleft$$

УТВ: Кол-во операций при не успешном поиске  $\Theta(1 + 2)$

УТВ: Кол-во операций при успешном поиске  $\Theta(1 + 2)$

$\triangleright \exists k_1 \dots k_n$  - ключи в порядке добавления  
 $\exists X_{i,j}$  - случай. вел.  $[h(k_i) = h(k_j)]$

$$E \left[ \frac{1}{n} \sum_{i=1}^n \sum_{j=i+1}^n X_{i,j} \right] = \quad k_{10} \rightarrow k_5 \rightarrow k_3$$

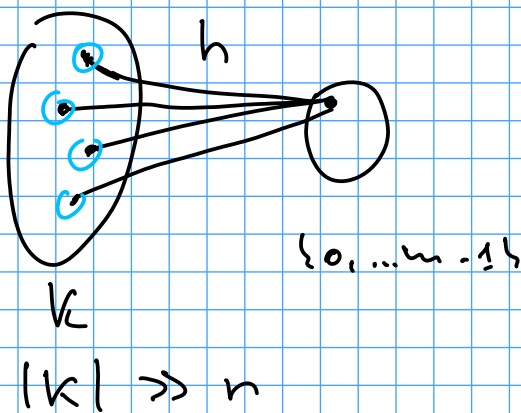
поиск ключа  $k_i$   
 усреднение по ключам

$$E[X_{i,i}] = 1, E[X_{i,j}] = \frac{1}{m}, i \neq j$$

$$= \frac{1}{n} \sum_{i=1}^n \left( 1 + \sum_{j=i+1}^n E[X_{i,j}] \right) = 1 + \frac{1}{n} \sum_{i=1}^n \sum_{j=i+1}^n \frac{1}{m} =$$

$$= 1 + \frac{1}{n \cdot m} \sum_{i=1}^n (n-i) = 1 + \frac{1}{n \cdot m} \left( n^2 - \frac{n(n+1)}{2} \right) =$$

$$= 1 + \frac{n}{m} - \frac{n+1}{2m} = \Theta \left( 1 + \frac{n}{m} \right) \triangleleft$$



$$\frac{N}{m} \sim 2 = \frac{n}{m}$$

$\Downarrow$

можно обойтись без хэши-р-ии.  
 (с.ч.  $n \sim N$ )

Канне Шубанот хем-орин

- берүүтү исне уу интервала  $[0, 1)$

$$h(x) = \lceil m \cdot x \rceil$$

- үенче исне

$$h(k) = k \bmod m$$

- иснего баарлыгы исне.

$$h(a_0, a_1, a_2, a_3 \dots a_{n-1}) = a_0 + a_1 x + a_2 x^2 + \dots + a_{n-1} x^{n-1} \bmod m$$

$$a_0 x^{n-1} + a_1 x^{n-2} + \dots + a_{n-1} \bmod m$$

