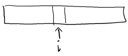
Хеш-таблица с откр. адресацией

T ht_key[M]
D ht_data[M]

$M \geq c \cdot N$, N – max кол.-во ключей, $c > 1$

$c = 2$
$c = 1{,}5$



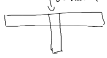
```
int find (T key):  // O(1) в среднем
    i = hash(key)
    while ht_key[i] is not None and ht_key[i] != key:
        i = (i+1) mod M
    return i

assign (T key, D data):
    i = find(key)
    ht_key[i] = key
    ht_data[i] = data

get(T key):
    return ht_data[find(key)]

exists(T key):
    return ht_key[find(key)] is not None
```

T key, data
i = hash(key)

Хеш-таблица с закр. адресацией – оценка времени работы

$\alpha = \dfrac{N}{M}$

i = hash(key)



$hash: A \to [0..M-1]$

Семейство хеш-функций $\mathcal{H} = \{h_i\}_{i \in I}$ $\quad h_i : A \to B$

В начале программы $hash \leftarrow U(\mathcal{H})$

Def. $\mathcal{H}$ наз. универсальным семейством х.-ф. $\overset{def}{\Longleftrightarrow}$ $\forall a_1, a_2 \in A$, $a_1 \neq a_2$ $\quad \Pr\limits_{h_i \leftarrow U(\mathcal{H})}\{h_i(a_1) = h_i(a_2)\} \leq \dfrac{1}{|B|}$

Пример $\mathcal{H} = \{h_{m,x}\}_{x \in \mathbb{Z}_m, m \in \mathbb{P}}$, где $h_{m,x} : \Sigma^* \to \mathbb{Z}_m$

$h_{m,x}(s) = (s_0 \cdot x^{|s|-1} + s_1 \cdot x^{|s|-2} + \ldots + s_{|s|-1} \cdot x^0) \bmod m$

Th. Хеш-таблица с закр. адрес. и унив. сем. х.-ф.: матожидание времени работы любой операции $O(\alpha + 1)$.

Д-во $k_1, k_2, \ldots, k_N$ – все различные ключи в хеш-таблице

Операция с ключом k_j: $E[\text{длина списка под индексом } hash(k_j)] \overset{?}{\leq} O(\alpha+1)$

$$X_{a,b} = \begin{cases} 1, & hash(k_a) = hash(k_b) \\ 0, & \text{иначе} \end{cases}$$

$$(\text{длина списка } hash(k_j)) \leq \sum_{a=1}^{N} X_{a,j}$$

$$E\{\text{ — " — }\} \leq \sum_{a=1}^{N} E[X_{a,j}] = \sum_{a=1}^{N} \Pr[hash(k_a) = hash(k_b)] \leq$$

$$\leq \sum_{\substack{a=1 \\ a \neq j}}^{N} \frac{1}{M} + \Pr\{hash(k_j) = hash(k_j)\} = \frac{N}{M} + 1 = \alpha + 1$$

по унив. $\mathcal{H}$

∎

Полиномиальный хеш: оценка вероятности коллизии

$h_{m,x}(s) = \left(\sum_{i=0}^{|s|-1} x^i \cdot s_{|s|-i-1}\right) \bmod m \qquad m \in \mathbb{P}$

(1) строчки случайны

Утв. $x_0 \in \mathbb{Z}_m$. P – случайный: $P(x) = a_0 x^0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}$, $\deg P < n$

$a_0, a_1, \ldots, a_{n-1}$ – выбир. независ. р/вер. из $\mathbb{Z}_m$

$$\Pr_P[P(x_0) \underset{m}{\equiv} 0] = \frac{1}{m}$$

Д-во. $P(x) = x(a_1 + a_2 x + \ldots + a_{n-1} x^{n-2}) + a_0$

1). сгенер. $a_1, \ldots, a_{n-1}$

$V = x_0 (a_1 + a_2 x_0 + \ldots + a_{n-1} x_0^{n-2})$ – число

2). сгенер. a_0

$P(x_0) \underset{m}{\equiv} 0 \Longleftrightarrow V + a_0 \underset{m}{\equiv} 0 \Leftrightarrow a_0 \underset{m}{\equiv} -V$, т.е. $\Pr\limits_{a_0}[V + a_0 \underset{m}{\equiv} 0] = \dfrac{1}{m}$

∎

Lm. $\forall m \in \mathbb{P}$, $x_0 \in \mathbb{Z}_m$ $\quad \Pr\limits_{\substack{s,t \\ |s|=n, |t|=n}}[h_{m,x_0}(s) = h_{m,x_0}(t)] = \dfrac{1}{m}$

Д-во

$h_{m,x_0}(s) = S(x_0) \qquad S(z) = s_{|s|-1} + s_{|s|-2} \cdot z + \ldots + s_0 \cdot z^{|s|-1}$ – случ. многочлен

$h_{m,x_0}(t) = T(x_0) \qquad T(z) = \ldots$

$$\Pr_{S,T}[S(x_0) \underset{m}{\equiv} T(x_0)] = \Pr_{S,T}[\underbrace{(S-T)}_{P}(x_0) \underset{m}{\equiv} 0] = \Pr_P[P(x_0) \underset{m}{\equiv} 0] = \frac{1}{m}$$

∎

k сравнений строк. $\Pr\{\text{была} \geq 1 \text{ коллизия}\} = \Pr\{C_1 \vee C_2 \vee \ldots \vee C_k\} \leq$

$$\leq \sum_{i=1}^{k} \Pr\{C_i\} = \frac{k}{m}$$

C_i – событие: на i-м сравнении случилась коллизия

(2) x_0 выбираем случайно из $\mathbb{Z}_m$ $\qquad m \in \mathbb{P}$

Утв. P – многочлен над $\mathbb{Z}_m$, $\deg P \leq n$. Тогда P имеет $\leq n$ корней.

Пример: $m = 2^{64}$
x^{64} 2, 4, 6, 8, 10, …
2^{63} корней

Lm. s, t – строки, $s \neq t$, $|s|, |t| \leq n$ $\quad \Pr\limits_{x_0}[h_{m,x_0}(s) \underset{m}{\equiv} h_{m,x_0}(t)] =$

$$= \Pr_{x_0}[S(x_0) \underset{m}{\equiv} T(x_0)] = \Pr_{x_0}[(S-T)(x_0) \underset{m}{\equiv} 0] = \frac{\overbrace{\#\{x_0 : \overbrace{(S-T)}^{P}(x_0) \underset{m}{\equiv} 0\}}^{\#\text{корней } P}}{\#\{x_0 \in \mathbb{Z}_m\}} \leq$$

S, T – многочлены

$$\leq \frac{\deg(S-T)}{m} \leq \frac{n}{m}$$

k сравнений $\Rightarrow \Pr\{\exists \geq 1 \text{ колл.}\} \leq \dfrac{kn}{m}$.

Полином. хеш и хеш-таблицы

$i = hash(key) \to 0..m-1 \qquad m \sim 10^{18}$

$i = hash(key) \bmod M \qquad M$ – размер хеш-таблицы 10^6

Th. $G_{p \to M} = \{g_{a,b}\}_{a \in \{1,\ldots,p-1\}, b \in \{0,\ldots,p-1\}}$, $p \in \mathbb{P}$ $\qquad g_{a,b} : \mathbb{Z}_p \to \mathbb{Z}_M \qquad p \geq M$

$g_{a,b}(x) = ((a \cdot x + b) \bmod p) \bmod M$

G – универсальное семейство х-ф (т.е. вер. коллизии $\leq \frac{1}{M}$)

Утв. $i = ((a * hash(key) + b) \bmod m) \bmod M$

$g_{a,b} \circ hash \qquad g_{a,b} \leftarrow G_{m \to M}$, $m \in \mathbb{P}$

$\Pr\{\text{коллизий } g_{a,b} \circ hash\} \leq \Pr\{\text{коллизии } hash\} + \Pr\{g_{a,b} \text{ колл.} \mid \text{нет колл. } hash\} \leq$

$\leq \dfrac{len}{m} + \dfrac{1}{M}$

$\dfrac{10^6}{10^{18}} + \dfrac{1}{10^6} = 10^{-12} + 10^{-6}$