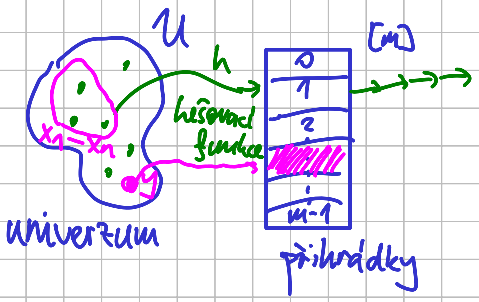


Df: Systém $\mathcal{H}$ funkcí z $\mathcal{U}$ do $[m]$ je c-univerzální pro $c>0$

$\equiv \forall x,y \in \mathcal{U},\ x \neq y:\ \Pr_{h\in\mathcal{H}}[h(x)=h(y)] \le \frac{c}{m}$.

Víme: $\exists$ 1-univerzální systém

Lemma: Nechť $\mathcal{H}$ je c-univerzální systém fcí z $\mathcal{U}$ do $[m]$, $x_1 \ldots x_n \in \mathcal{U}$ navzájem různé, $y \in \mathcal{U}$.

Potom $\mathbb{E}_{h\in\mathcal{H}}[\underbrace{\#i : h(x_i)=h(y)}_{Y}] \le c \cdot \frac{n}{m} + 1$.

obsazení přihrádky, kam padlo y

Dk: Předpokládejme, že $\forall i\ y \neq x_i$ ... případ $y = x_i$ zařídí $+1$ na konci.

Zavedeme indikátory $I_1 \ldots I_n$:

$$I_i := \begin{cases} 0 \\ 1 & \text{pro } h(x_i)=h(y) \end{cases}$$

- $Y = \sum_i I_i$
- $\mathbb{E}[Y] = \sum_i \mathbb{E}[I_i]$ — linearita $\mathbb{E}$
- $\mathbb{E}[I_i] = \Pr[h(x_i)=h(y)] \le \frac{c}{m}$ — def. $\mathbb{E}$; c-univerzalita $\mathcal{H}$

$\mathbb{E}[Y] \le c \cdot \frac{n}{m}$

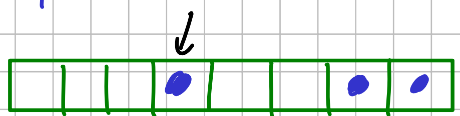
Důsledek: $\mathbb{E}$ časové složitosti operací Find, Insert, Delete je $O(\frac{n}{m})$.

umíme udržet $O(1)$ přehešováváním

Otevřená adresace

Df: Každému $x \in \mathcal{U}$ přiřadíme vyhledávací posloupnost $h(x,0), h(x,1), \ldots, h(x,m-1)$ což je permutace na $[m]$.

Operace:



Insert(x)

Find(x) — zastaví se o prázdnou př.

Delete(x) — prvek nahradíme pomníčkem, časem strukturu přebudujeme

Příklady

lineární přidávání

$h(x,i) = (f(x) + i) \bmod m$ — f: heš. fce

Ukáže se $O(\frac{1}{(1-\alpha)^2})$

0	1	2	3	4	5	6	7	8	9
		42		14	75	36	24	95	17

← degradace

Vkládáme: 75 36 14 42 24 95 17

dvojité hešování

$h(x,i) = (f(x) + i \cdot g(x)) \bmod m$

2 heš. fce; prvočíslo

Ukáže se: $O(\frac{1}{1-\alpha})$

Věta: Pokud vyhl. posloupnosti jsou nezávislé plně náhodné permutace, pak $\mathbb{E}[\#$ přihrádek navštívených při neúspěšném Findu$] \le \frac{1}{1-\alpha}$, kde $\alpha = \frac{n}{m}$.

Dk věty: Nechť $y \in \mathcal{U}$ hledáme, $h_1 \ldots h_m$ je jeho vyhl. posl.

$p_i := \Pr[\text{projdeme aspoň } i \text{ přihrádek}]$

$p_1 = 1 \qquad p_2 = \frac{n}{m} = \alpha$

$p_i = 1 \cdot \frac{n}{m} \cdot \underbrace{\frac{n-1}{m-1}}_{\le \frac{n}{m}} \cdot \underbrace{\frac{n-2}{m-2}}_{\le \frac{n}{m}} \cdots \underbrace{\frac{n-(i-1)}{m-(i-1)}}_{\le \frac{n}{m}} \le \alpha^{i-1}$

$\frac{3}{4} \ge \frac{2}{3}$

$\mathbb{E}[\#\text{ navštívených př.}] = \sum_{i \ge 1} i \cdot \underbrace{\Pr[\text{navštíveno právě } i]}_{(p_i - p_{i+1})}$

$= \sum_{j \ge 1} p_j \underbrace{(j-(j-1))}_{1}$

$\le \sum_{j \ge 1} \alpha^{j-1} = \sum_{j \ge 0} \alpha^j = \frac{1}{1-\alpha}$.

Rozděl a panuj (Divide et impera (?))

Mergesort (třídění sléváním) $\longrightarrow$ Merge $(x_1 - x_a,\ y_1 - y_b)$ v čase $\Theta(a+b)$

Mergesort $(x_1 - x_n)$: Pokud $n \leq 1$: vrátíme vstup

$a_1 - a_{\lfloor n/2 \rfloor} \leftarrow$ Mergesort $(x_1 - x_{\lfloor n/2 \rfloor})$

$b_1 - b_{\lceil n/2 \rceil} \leftarrow$ Mergesort $(x_{\lfloor n/2 \rfloor + 1} - x_n)$

Vrátíme Merge $(a_1 \dots,\ b_1 \dots)$

Analýza složitosti (BÚNO $n = 2^k$)

$$T(n) = 2T(n/2) + n \qquad\qquad M(n) = M(n/2) + n$$

$T(n)$: čas na vstupu délky n

n: má být $\Theta(n)$, ale zvolím vhodné jednotky

$M(n)$: paměť na vstupu délky n

$$T(1) = 1$$

$M(n) = n + \frac{n}{2} + \frac{n}{4} + \dots$ je $\Theta(n)$

Rozepisujeme...

$$T(n) = 2\underbrace{T(n/2)}_{(2 \cdot T(n/4) + n/2)} + n$$

$$T(n) = 4 \cdot \underbrace{T(n/4)}_{2T(n/8) + n/4} + 2n$$

$$T(n) = 8T(n/8) + 3n$$

$\vdots$

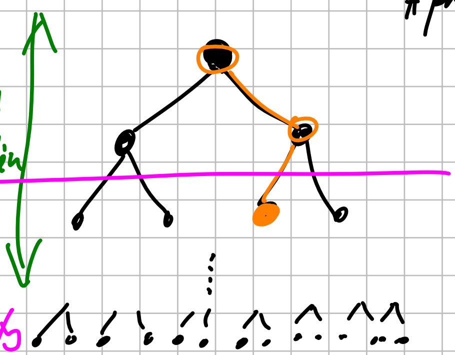
$$T(n) = 2^i \cdot \underbrace{T(n/2^i)}_{T(1) \text{ pro } n/2^i = 1,\ \text{čili } i = \log n} + i \cdot n$$

$$T(n) = n \cdot \underbrace{T(1)}_{1} + \log n \cdot n \ \dots\ \Theta(n \log n)$$

Strom rekurze

log n úrovní

i



listy

# problémů	velikost problému	paměť na úrovni	čas na úrovni
1	n	n	n
2	$\frac{n}{2}$	$\frac{n}{2}$	n
4	$\frac{n}{4}$	$\frac{n}{4}$	n
2^i	$\frac{n}{2^i}$	$\frac{n}{2^i}$	n
n	1		n

Celkem $n \cdot \log n$

$\Sigma \leq 2n$