Hešování III

h. s otevřenou adresací

$h(x,0), h(x,1), h(x,2), \dots$ vyhledávací posl.

h. s lineárním přidáváním: $h(x,i) = h(x) + i \pmod{m}$

dvojité hešování: $h(x,i) = h(x) + i\, g(x)$

...

INSERT … ulož do prvního volného místa ve vyhl. posl.

FIND … po vyhl. posl. dokud nenajdeme x nebo volné místo

DELETE … FIND, označ za smazaný, přehešuj pokud jich je $(\geq m/4)$ (amortizace)

$\Rightarrow$ čas složitost $\approx$ očekávaný #testů (delete amort.)

Přehled

$\varepsilon \leq \frac{\#\text{volných}}{\#\text{obsazených}}$

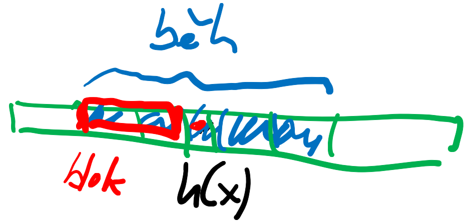
Nechť $m \geq (1+\varepsilon)n$. Potom oček. # testů při FIND při h. s lin. přidáváním je:

- $O(1/\varepsilon^2)$ pro zcela náh. h. fce
- $O(f(\varepsilon))$ pro $(\log n)$-nez. systém
- $O(1/\varepsilon^{13/6})$ pro 5-nez. systém
- $O(1/\varepsilon^2)$ pro tab. hešování !
- $\Omega(\log n)$ pro něj. 4-nez. systém
- $\Omega(\sqrt{n})$ pro něj. 2-nez. systém

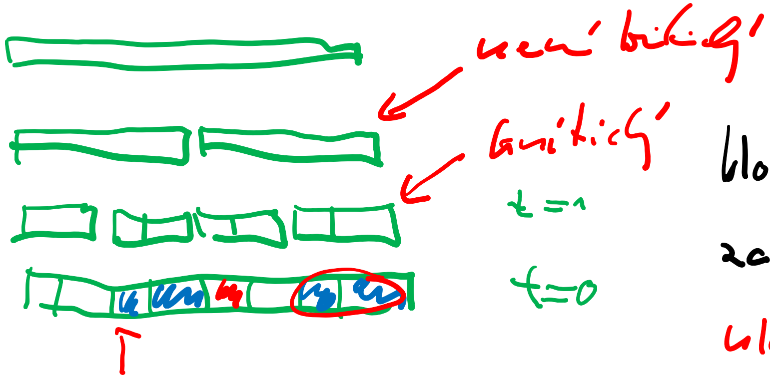
Věta (Černovova nerov.): Nechť $X = X_1 + \dots + X_k$, kde X_i jsou nez. binární n.p., $\mu = \mathbb{E}[X]$, $c > 1$. Pak

$$\Pr[X \geq c\mu] \leq \left(\frac{e^{c-1}}{c^c}\right)^{\mu}.$$

Věta: Nechť $m \geq 3n$ je mocnina 2, $h: \mathcal{U} \to [m]$ zcela náh. fce. Pak očekávaný # testů při $\mathrm{FIND}(x)$ je $O(1)$.



Důkaz: $m = 3n$.



blok vel. 2^t je **kritický** pokud je do něj zahešováno (# uloženo) $\geq \frac{2}{3} 2^t$ prvků.

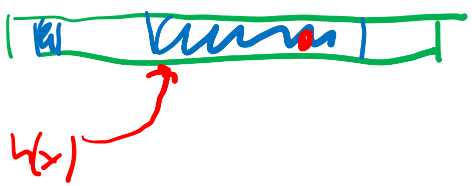
Lemma: Nechť B je blok vel. 2^t. Pak

$$P[B \text{ je kritický}] \leq (e/4)^{2^t/3}$$

Důk: $X = \sum_{i=1}^{n} X_i$ kde $X_i = \begin{cases} 1 & h(x_i) \in B \\ 0 & \text{jinak} \end{cases}$ (indikátory)

$\Rightarrow \mu = \mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = n \frac{2^t}{m} = 2^t/3$ ← zcela náh. fce h

Černovova nerov.: $\Pr[X \geq 2\mu] \leq (e/4)^{2^t/3}$ □



x uložen v běhu $\Rightarrow$ hešován do běhu

Lemma: Nechť je R běh velikosti $\geq 2^{\ell+2}$. Pak alespoň jeden z bloků B_0, B_1, B_2, B_3 vel. 2^ℓ je kritický.

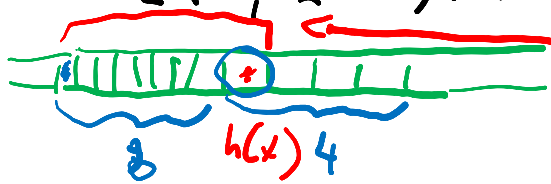
první 4 bloky do kterých zasahuje R

Dk.: $L = R \cap (B_0 \cup B_1 \cup B_2 \cup B_3) \geq 1 + 3 \cdot 2^\ell$ prvků (uložených)

Kdyby žádný z nich nebyl kritický, tak do L zahešováno $< 4 \frac{2}{3} 2^\ell$.

$1 + 3 \cdot 2^\ell \leq$ # uložených v $\leq$ # zahešovaných do $L < 4 \frac{2}{3} 2^\ell$. ↯ $\square$

Lemma: Nechť je R běh velikosti $|R| \in [2^{\ell+2}, 2^{\ell+3})$. Pak alespoň jeden z následujících 12 bloků vel. 2^ℓ je kritický:



Dk.: $4|B| \leq |R| < 8|B| \Rightarrow$ začátek běhu R je v některém z těchto bloků

předch. lemma $\Rightarrow$ některý z 12 bloků je kritický. $\square$

Důsl. Velikost R běhu obsahující $h(x)$. Pak

$$\Pr\left[|R| \in [2^{\ell+2}, 2^{\ell+3})\right] \leq 12 \cdot q^{2^\ell} \quad \text{pro} \quad q = (e/4)^{1/3} \doteq 0.879$$

Dů:

$$-\|- \leq \Pr\left[\begin{array}{l}\text{aspoň 1 z těch 12} \\ \text{bloků je kritický}\end{array}\right] \leq 12 \Pr[\text{blok je kritický}]$$

Booleova ner.

$$\leq 12 \cdot (e/4)^{2^\ell/3} = 12 \cdot q^{2^\ell} \qquad \square$$

1. lemma

Pozorování: $E[X] \leq \sum_{i=1}^{k} \max I_i \cdot \Pr[X \in I_i]$ pro $\operatorname{rng}(X) = I_1 \,\dot\cup \ldots \dot\cup\, I_k$

$(0,4)$, $[2^{\ell+2}, 2^{\ell+3})$, $\ell \geq 0$

$$E[|R|] \leq 3 \cdot \underbrace{\Pr[|R| \leq 3]}_{\leq 1} + \sum_{\ell \geq 0} 2^{\ell+3} \Pr\left[|R| \in [2^{\ell+2}, 2^{\ell+3})\right]$$

$$\leq 3 + \sum_{\ell \geq 0} 2^{\ell+3}\, 12 \cdot q^{2^\ell} \leq 3 + 8 \cdot 12 \sum_{i \geq 0} i\, q^i = O(1). \qquad \square$$

konverguje
$q \in (0,1)$

Rolující hešování

$$h_a(x_0, \dots, x_{d-1}) = \sum_{i=0}^{d-1} x_i a^i$$ kde $d \geq 1$, $a \in \mathbb{Z}_p$, p prvočíslo, $h_a : \mathbb{Z}_p^d \to \mathbb{Z}_p$

Tvr. Systém $Q = \{ h_a \mid a \in \mathbb{Z}_p \}$ je d-univ. pro lib. $d \geq 1$, prvočíslo p.

rolování:

$$---\ , \underbrace{x_0 \dots x_{d-1}}_{x}, x_d, \dots$$

x'

(Rabin–Karp)

$$h(x') = \frac{h_a(x) - x_0}{a} + x_d a^{d-1}$$

← umíme v $O(1)$ čase.