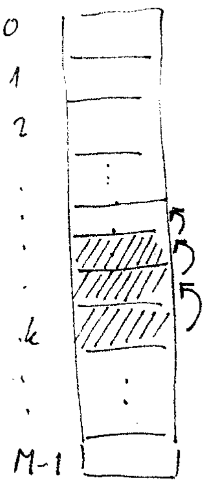
Hašování s lineárním přidáváním (Knuth III) 1962–63



INSERT(x):

$h(x) = k$

když k je obsazené, zkusím $k-1$, $k-2$, atd.

musí zůstat aspoň 1 prázdná adresa

M – velikost tabulky, N – počet vložených prvků

$a_1, a_2, \dots, a_N$ – haš. adresy (původní) vložených prvků

všech M^N posloupností stejně pravděpod.

C_N = očekávaný počet testů při úspěšném vyhledávání

C'_N = — " — neúspěšném — " —

platí: C'_{N-1} = oč. počet testů při insertu N-tého prvku

platí: $C_N = \frac{1}{N} \sum_{k=0}^{N-1} C'_k$ (hledaný prvek byl vložen do tabulky jako 1. nebo 2. nebo … N-tý vše s pravd. $\frac{1}{N}$)

platí:

$$\left.\begin{aligned} C'_N &\approx \frac{1}{2}\left(1 + \left(\frac{1}{1-\alpha}\right)^2\right) \\ C_N &\approx \frac{1}{2}\left(1 + \frac{1}{1-\alpha}\right) \end{aligned}\right\} \text{ budeme dokazovat}$$

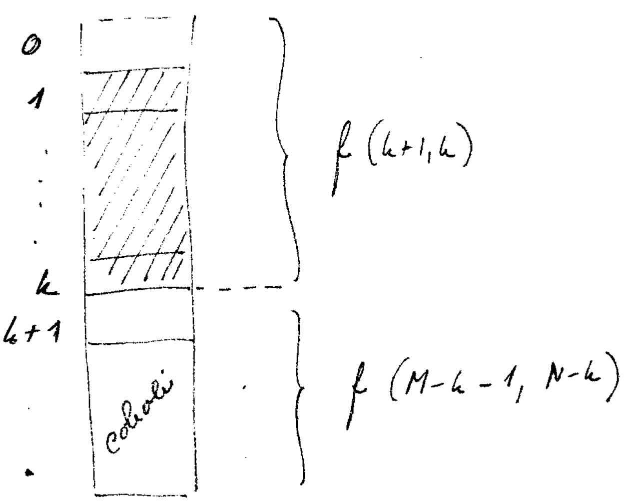
kde $\alpha = \frac{N}{M}$ faktor naplnění označíme: $f(M,N)$ = počet haš. posloupností takových, že adresa 0 v tabulce je prázdná

$$P(\text{adresa 0 je prázdná}) = 1 - \frac{N}{M} = \frac{f(M,N)}{M^N}$$

$$\Rightarrow f(M,N) = \left(1 - \frac{N}{M}\right) M^N$$

$$f(0,0) = 1$$

ozn. $g(M,N,k)$ = počet haš. posl. takových, že adresa 0 je prázdná, adresy $1, \dots, k$ obsazené, adresa $k+1$ prázdná



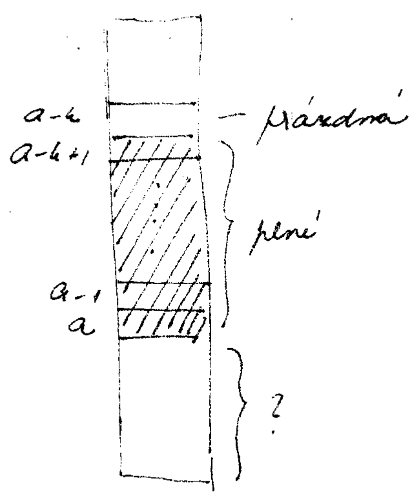
$\binom{N}{k}$ možností, jak obsadit adresy $1, \dots, k$

z toho: $g(M,N,k) =$

$$= \binom{N}{k} f(k+1,k)\, f(M-k-1, N-k)$$

ozn. $P_k = P$(při Insertu $(N+1)$-ního prvku potřebuju $k+1$ pokusů)

vkládám prvek na adresu a



co může být na adresách $a+1$, $a+2$ atd.?

buď $a+1$ prázdná, za ní cokoli

nebo $a+1$ plná, $a+2$ prázdná, cokoli

nebo $a+1, a+2$ plné, $a+3$ prázdná, cokoli

atd.

počet všech případů: M^N

$$P_k = \frac{1}{M^N}\left(g(M,N,k) + g(M,N,k+1) + \dots + g(M,N,N)\right)$$

spočteme $c_N' = \sum_{k=0}^{N} (k+1) P_k$ a z toho potom c_N

$$M^N c_N' = M^N \sum_{k=0}^{N} (k+1) P_k = \sum_{k=0}^{N-1} \sum_{j=0}^{N-k} g(M,N,k+j) =$$

$$= g(M,N,0) + g(M,N,1) + \dots + g(M,N,N) +$$

$$+ 2g(M,N,1) + \dots + 2g(M,N,N) +$$

$$+ 3g(M,N,2) + \dots + 3g(M,N,N) +$$

$$+ \dots =$$

$$= \sum_{k=0}^{N} \sum_{j=0}^{k} (j+1)\, g(M,N,k) = \sum_{k=0}^{N} \frac{(k+2)(k+1)}{2}\, g(M,N,k) =$$

$$= \underbrace{\frac{1}{2} \sum_{k=0}^{N} (k+1)\, g(M,N,k)}_{1.} + \underbrace{\frac{1}{2} \sum_{k=0}^{N} (k+1)^2 g(M,N,k)}_{2.}$$

1.
$$
\begin{array}{llllll}
g(M,N,0) & + g(M,N,1) & + g(M,N,2) & + \dots & = & M^N P_0 \\
 & + g(M,N,1) & + g(M,N,2) & + \dots & = & M^N P_1 \\
 & & + g(M,N,2) & + \dots & = & M^N P_2 \\
 & & + & & & \\
 & & \vdots & & & \vdots
\end{array}
$$

$$= M^N \sum_k P_k = M^N$$

2. dosadíme za $g(M,N,k)$

$$g(M,N,k) = \binom{N}{k} f(k+1,k)\, f(M-k-1,\, N-k) =$$

$$= \binom{N}{k}\left(1-\frac{k}{k+1}\right)(k+1)^{k}\left(1-\frac{N-k}{M-k-1}\right)(M-k-1)^{N-k} =$$

$$= \binom{N}{k}(k+1)^{k-1}\,(M-k-1)^{N-k-1}\,(M-N-1)$$

Pomocné tvrzení:

definujme fci $\displaystyle s(n,x,y) = \sum_{k\geq 0}\binom{n}{k}(x+k)^{k+1}(y-k)^{n-k-1}(y-n)$

platí pro ni $s(n,x,y) = x(x+y)^{n} + n\,s(n-1,\, x+1,\, y-1)$

~~Pro nás: $g(M,N,k) = s(N,1,M-1)\cdot(k+1)^{-2}$~~ $n=N,\ x=1,\ y=M-1$

Z toho: $\displaystyle M^N C_N' = \frac{1}{2}M^N + \frac{1}{2}\underbrace{\sum_k \binom{N}{k}(k+1)^{k+1}(M-k-1)^{N-k-1}(M-N-1)}_{s(N,1,M-1)} =$

$$= \frac{1}{2}\left(M^N + M^N + N\,s(N-1,2,M-2)\right) =$$

$$= \frac{1}{2}\left(M^N + M^N + 2NM^{N-1} + \overset{N(N-1)}{\cancel{2N}}\,s(N-2,3,M-3)\right) = \dots$$

$$\dots = \frac{1}{2}M^N\Big(1 + \underbrace{1 + 2\frac{N}{M} + 3\frac{N(N-1)}{M^2} + \dots}_{Q_1(M,N)}\Big)$$

kde $Q_r(M,N) = \binom{r}{0} + \binom{r+1}{1}\frac{N}{M} + \binom{r+2}{2}\frac{N(N-1)}{M^2} + \dots =$

$$= \sum_{k \geq 0} \binom{r+k}{k} \frac{N}{M} \cdot \frac{N-1}{M} \cdots \frac{N-k+1}{M}$$

(tzv. Ramanujanova Q-funkce – používá se často v analýze algoritmů, např. UNION-FIND)

takže (2) $C'_N = \frac{1}{2}\left(1 + Q_1(M,N)\right)$

dosadíme do (1) $C_N = \frac{1}{N}\sum_{k=0}^{N-1} C'_k$ a vyjde nám

$$C_N = \frac{1}{2}\left(1 + Q_0(M,N-1)\right)$$

$$C_N = \frac{1}{N}\sum_{k=0}^{N-1}\frac{1}{2}\left(1 + Q_1(M,k)\right) = \frac{1}{2}\left(1 + \frac{1}{N}\sum_{k=0}^{N-1} Q_1(M,k)\right)$$

stačí ověřit, že $\sum_{k=0}^{N-1} Q_1(M,k) = N Q_0(M,N-1)$ – indukcí

s použitím: $Q_0(M,N-1) = \frac{M}{N}\left(Q_0(M,N) - 1\right)$

$Q_1(M,N) = (N+1)Q_0(M,N) - NQ_0(M,N-1)$

$Q_1(M,N) = M - (M-N-1)Q_0(M,N)$

(něco z toho se vypočte přímo, něco indukcí)

$$Q_0 = 1 + \frac{N}{M} + \frac{N(N-1)}{M^2} + \dots = \sum_{k \geq 0} \frac{N}{M} \cdot \frac{N-1}{M} \cdot \ldots \cdot \frac{N-k+1}{M}$$

$$Q_1 = 1 + 2\frac{N}{M} + 3\frac{N(N-1)}{M} + \dots = \sum_{k \geq 0} (k+1)\frac{N}{M} \cdot \frac{N-1}{M} \cdot \ldots \cdot \frac{N-k+1}{M}$$

Nyní dosadíme do rovnice (1) výsledek (2):

$$C_N = \frac{1}{N} \sum_{0 \le k < N} \frac{1}{2} (1 + Q_1(M,k)) \ ,$$

do které dále dosadíme za $Q_1(M,k)$ z rovnice (3):

$$\begin{aligned}
C_N &= \frac{1}{N} \sum_{0 \le k < N} \frac{1}{2} \left(1 + \sum_{l \ge 0} (l+1) \frac{k}{M} \frac{k-1}{M} \cdots \frac{k-l+1}{M} \right) \\
&= \frac{1}{2} \frac{1}{N} \sum_{0 \le k < N} \left(1 + \sum_{l \ge 0} (l+1) \frac{k}{M} \frac{k-1}{M} \cdots \frac{k-l+1}{M} \right) \\
&= \frac{1}{2} \left(1 + \frac{1}{N} \sum_{0 \le k < N} \sum_{l \ge 0} (l+1) \frac{k}{M} \frac{k-1}{M} \cdots \frac{k-l+1}{M} \right) \qquad (5)
\end{aligned}$$

Definice kombinačního čísla praví, že

$$\binom{k}{l} = \frac{k!}{l!(k-l)!} = \frac{k(k-1)\dots(k-l+1)}{l!} \ ,$$

odkud vidíme, že

$$k(k-1)\dots(k-l+1) = \binom{k}{l} l! \qquad (6)$$

Dosadíme-li tento vztah (6) do rovnice (5), upravíme, provedeme prohození sum a z vnitřní sumy vytkneme členy, které nezávisí na sčítacím indexu vnitřní sumy, dostáváme rovnici (7):

$$\begin{aligned}
C_N &= \frac{1}{2} \left(1 + \frac{1}{N} \sum_{0 \le k < N} \sum_{l \ge 0} (l+1) \frac{\binom{k}{l} l!}{M^l} \right) \\
&= \frac{1}{2} \left(1 + \frac{1}{N} \sum_{l \ge 0} \sum_{0 \le k < N} \frac{(l+1)! \binom{k}{l}}{M^l} \right) \\
&= \frac{1}{2} \left(1 + \frac{1}{N} \sum_{l \ge 0} \frac{(l+1)!}{M^l} \sum_{0 \le k < N} \binom{k}{l} \right) \qquad (7)
\end{aligned}$$

Nyní upravíme vnitřní sumu:

$$\begin{aligned}
\sum_{0 \le k < N} \binom{k}{l} &= \binom{0}{l} + \binom{1}{l} + \dots + \binom{N-1}{l} \\
&= \binom{l}{l} + \binom{l+1}{l} + \dots + \binom{N-1}{l} \\
&= \binom{l+1}{l+1} + \binom{l+1}{l} + \binom{l+2}{l} + \dots + \binom{N-1}{l} \\
&= \binom{l+2}{l+1} + \binom{l+2}{l} + \dots + \binom{N-1}{l} \\
&= \dots \\
&= \binom{N}{l+1} \qquad (8)
\end{aligned}$$

Dosadíme výsledek (8) do rovnice (7), dostáváme, že

$$C_N = \frac{1}{2}\left(1 + \frac{1}{N}\sum_{l\geq 0}\frac{(l+1)!}{M^l}\binom{N}{l+1}\right) \tag{9}$$

Opět se podíváme na součin kombinačního čísla a faktoriálu ze vztahu (9):

$$\begin{aligned}(l+1)!\binom{N}{l+1} &= (l+1)!\frac{N!}{(l+1)!(N-l-1)!} = \frac{N!}{(N-l-1)!}\\ &= N(N-1)\cdots(N-l)\end{aligned} \tag{10}$$

Výsledek (10) použijeme na (9), abychom dostali, že

$$\begin{aligned}C_N &= \frac{1}{2}\left(1 + \frac{1}{N}\sum_{l\geq 0} N\frac{(N-1)}{M}\frac{(N-2)}{M}\cdots\frac{(N-l)}{M}\right)\\ &= \frac{1}{2}\left(1 + \sum_{l\geq 0}\frac{(N-1)}{M}\frac{(N-2)}{M}\cdots\frac{(N-l)}{M}\right) = \frac{1}{2}\left(1 + Q_0(M,N-1)\right)\\ &\leq \frac{1}{2}\left(1 + \sum_{l\geq 0}\cancel{\frac{N}{M}\frac{(N-1)}{M}}\cdots\frac{(N-l+1)}{M}\right)\end{aligned} \tag{11}$$

Suma z výrazu (11) je rovna výrazu (4), z čehož dostáváme konečný vztah, že

$$C_N \leq \frac{1}{2}(\cancel{1+Q_0}(M,N)) \tag{12}$$

Nahradíme-li v sumě (4) všechny zlomky jejich horními odhady $\frac{N}{M}$ ($= \alpha$, load factor), dostaneme, že

$$Q_0(M,N) \leq \sum_{k\geq 0}\alpha^k = \frac{1}{1-\alpha} \tag{13}$$

a použijeme-li vzorec (13) v nerovnici (12), dostáváme

$$C_N \leq \frac{1}{2}\left(1 + \frac{1}{1-\alpha}\right),$$

což je přesně vztah (b), který jsme dokazovali.

$$Q_0(M,N) = \sum_{k\geq 0}(k+1)\frac{N}{M}\frac{N-1}{M}\cdots\frac{N-k+1}{N} \leq \sum_{k\geq 0}^{\infty}(k+1)\alpha^k = \sum_{k=0}^{\infty}(\alpha^{k+1})' =$$

$$= \left(\frac{\alpha}{1-\alpha}\right)' = \frac{1-\alpha-\alpha(-1)}{(1-\alpha)^2} = \frac{1}{(1-\alpha)^2}$$

$$Q_0(M,N-1) = \sum_{k\geq 0}\frac{(N-1)}{M}\frac{(N-2)}{M}\cdots\frac{(N-k)}{M} \leq \sum_{k=0}^{\infty}\alpha^k = \frac{1}{1-\alpha}$$