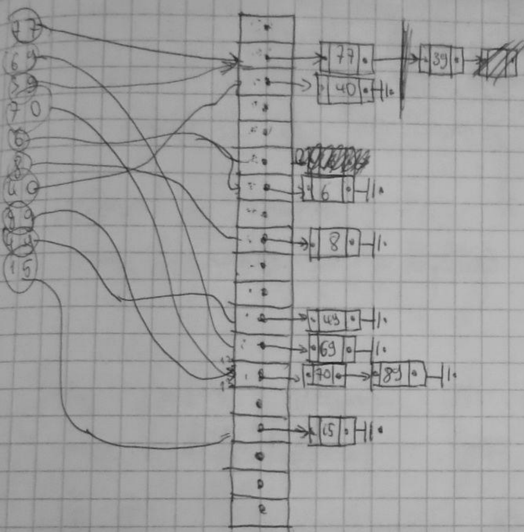
ZADAĆA 1 - HASH TABLICE

Zad. 1 Sljedeći zadatci koji se ilustriraju korištenje funkcije raspršivanja (eng. hash function)

1) Demonstrirajte ubacivanje ključeva 77, 69, 39, 70, 6, 8, 40, 89, 49, 15 u hash tablicu veličine $m = 19$.

a) Ako se kolizije rješavaju ulančavanjem, a $h(k) = k \bmod m$.



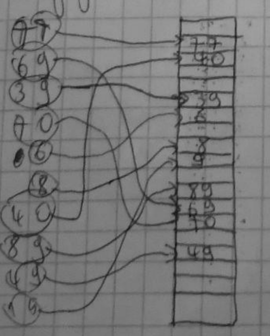
Pr. $77 \bmod 19 = 1$

Pr. $69 \bmod 19 = 12$

b) U kojem se kolizije rješavaju za $i = 0, 1, 2, \ldots, m-1$, koristeći dvostruko pretraživanje:

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

gdje su $h_1(k) = k \bmod m$, $h_2(k) = 1 + (k \bmod (m-1))$ pomoćne hash f-je



Pr. $h(77, 0) = (77 \bmod 19 + \cancel{0 \cdot (1 + [77 \bmod 18])}) \bmod 19$

$= 1$

Pr. $h(69, 0) = (69 \bmod 19 + \cancel{0 \cdot (1 + [69 \bmod 18])}) \bmod 19$

$= 12$

Pr. $h(39, 0) = (39 \ldots + 0 \cdot (1 + [39 \ldots]) \ldots$

$= 1$ ⊗ kolizija

$h(39, 1) = \ldots = 6$

Zad 1.2. $f(x) = \sum_{i=1}^{n} a_i x_i \pmod 8$. $a_i \in \{0, \dots, 7\}$, $x_i \in \{0, \dots, 9\}$.

Je li ovakva f-ja univerzalna hash f-ja? Ako nije, dajte kontraprimjer.

U materijalima danima u predlošku zadaće, pokazano je da je skup ovakvih f-ja univerzalan za prost m. Međutim, 8 nije prost. Još gore, 8 je potencija broja 2.

To znači da ne vrijedi svojstvo ograničenje gdje se $a_0 \dots a_n$ za x i $a_0, \dots a_n$ za y mogu poništavati samo za jedan odabir a_i, tj. vjerojatnost da se hashevi poklapaju za različite $x \neq y$ bit će potencijalno veća od $\frac{|H|}{m}$.

Primjerice:

$x = 0$, $y = 8$ ~~...~~

Za $a_0 = 1$: $f(0) = 0$, $f(8) = 0$

Za $a_0 = 2$: $f(0) = 0$, $f(16) = 0$

[illegible]

Zad. 2 Pretp. da koristimo hash f-ju h za raspršivanje n razl. ključeva u tablici T duljine m. Uz pretp. uniformnog raspršenja, koliki je očekivani broj kolizija?

Neka je h f-ja jednostavnog uniformnog raspršenja koja raspršuje n razl. ključeva u m mjesta u tablici.

Neka je I_{ij} indikator sl. var. t.d. $I_{ij} = \begin{cases} \frac{1}{m}, & h(i)=h(j) \\ 0, & \text{inače} \end{cases}$, jer je vjerojatnost da je $h(i)=h(j) = \frac{1}{m}$ za m mjesta u tablici. Također, $E[I_{ij}] = \frac{1}{m}$ (indik. sl. var.)

Postoji $\binom{n}{2} = \frac{n(n-1)}{2}$ takvih parova ključeva, svaki sa vjerojatnošću kolizije $\frac{1}{m}$, pa je očekivani broj podudaranja:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} E[I_{ij}] = \sum_{1 \le i \le j \le n} E[I_{ij}] = \frac{1}{m} \sum_{i,j \in [1,n]} 1 = \frac{n(n-1)}{2} \cdot \frac{1}{m} = \frac{n(n-1)}{2m}$$

Zad. 3.

1) Uz pretpostavku uniformnog hashiranja, pokažite da je za i-to ubacivanje vjerojatnost više od k strogih pokušaja ~~[illegible]~~ najviše 2^{-k}.

Neka je X_k sl. varijabla koja modelira broj pokušaja napravljenih u neuspješnoj pretrazi. Također, modeliramo događaj A_k, za $k = 1, 2, 3, \ldots$ da bude događaj gdje je k-ti probe i bio je zauzet mjesto u tablici.

Tada je događaj $\{X_k > k\}$ presjek događaja $A_1 \cap A_2 \cap \ldots \cap A_k$. Znamo:

$$\underbrace{P\{A_1 \cap A_2 \cap \ldots \cap A_k\}}_{\text{UVJETNA VJEROJATNOST}} = P\{A_1\} \cdot P\{A_2 | A_1\} \cdot \ldots \cdot P\{A_k | A_1 \cap A_2 \cap \ldots \cap A_{k-1}\}$$

Vjerojatnost da je, za $j > 1$, j-ti probe završio na zauzetom mjestu, pod pretpostavkom da su prethodnih $j-1$ probova ~~[illegible]~~ bili na zauzetim slotovima iznosi:

$$\frac{(n-j+1)}{(m-j+1)}$$

broj zauzetih slotova koje nismo pretražili

— u — ukupno — u —

$$P\{X_k \geq k\} = \frac{n}{m} \cdot \frac{n-1}{m-1} \cdot \frac{n-2}{m-2} \cdot \ldots \cdot \frac{n-(k-1)}{m-(k-1)} < \frac{n^k}{m^k} = \left(\frac{n}{m}\right)^k \quad \left(n = \frac{m}{2}\right)$$

$$= \left(\frac{\frac{1}{2} \cdot \cancel{m}}{\cancel{m}}\right)^k = \left(\frac{1}{2}\right)^k = 2^{-k}.$$

(2) ~~Korištenjem prethodnog dokaza, imamo:~~

~~$k =$~~

Za $i = 1, 2, 3, \ldots, n$, vjerojatnost da je i-ti insert zahtijeva više od $2 \lg n$ probova je najviše $\frac{1}{n^2}$.

Korištenjem preth. izraza

$X_k \Leftrightarrow X_{2 \lg n}$

$$P\{X_{2\lg n} \geq k\} = \frac{n}{m} \cdot \frac{n-1}{m-1} \cdot \ldots \cdot \frac{n-(2\lg n - 1)}{m-(2\lg n-1)} \cdot \leq \left(\frac{n}{m}\right)^{2\lg n} = 2^{-2\lg n}$$

$$= 2^{\lg \frac{1}{n^2}} = \frac{1}{n^2} = O\left(\frac{1}{n^2}\right)$$

* Neka je sl. var X_i označava broj pokušaja potrebnih za i-to ubacivanje i neka je sl. var $X = \max_{1 \le i \le n} X_i$ najveći broj pokušaja potrebnih za bilo koje od n ubacivanja.

(3) Pokažite da je $P\{X \geq 2\lg n\} = O(1/n)$

Potrebno je samo modificirati prethodni zadatak. Prethodni zadatak smo gledali za specifični insert. Ovdje želimo vidjeti vjerojatnost da je maksimalan broj probanja za bilo koji insert jednak $\frac{1}{n}$.

Kako znamo da je za i-ti insert vjerojatnost $\frac{1}{n^2}$, pustimo da i „šeta" od 1 do n i zbrojimo vjerojatnosti, jer bilo koji od tih inserta može biti onaj koji nam garantira tu gornju granicu ...

$$P\{X \geq 2\lg n\} = P\{X_1 \geq 2\lg n\} + P\{X_2 \geq 2\lg n\} + \cdots + P\{X_n \geq 2\lg n\}$$

$$= \sum_{i=1}^{n} \frac{1}{n^2} = n \cdot \frac{1}{n^2} = \frac{1}{n} \; // = O\left(\frac{1}{n}\right)$$

(4) Pokažite da je očekivana duljina $E[X]$ najdužeg niza pokušaja $O(\lg n)$

Znajući da je najdulja moguća probe sekvenca n, te da je vjerojatnost da je sekvenca duljine veće od $2\lg(n)$ $\frac{1}{n}$ [iz preth. zad.]. Dakle, imamo potpunu vjerojatnost, tj:

$$P\{X \leq 2\lg(n)\} + P\{X > 2\lg n\} = 1$$

$$\Rightarrow E[X] \leq \underbrace{P\{X \leq 2\lg n\} \cdot 2\lg n}_{E[X \leq 2\lg n]} + \underbrace{P\{X > 2\lg n\} \cdot n}_{E[X > 2\lg n]} =$$

$$= \frac{n-1}{n} \cdot 2\lg n + \frac{1}{\cancel{n}} \cdot \cancel{n} = \frac{2\cancel{n}\lg n}{\cancel{n}} - \frac{2\lg n}{n} + 1 =$$

$$= 2\lg n - \frac{2\lg n}{n} + 1 = O\left(\frac{\lg n}{n}\right) = O($$