Lezione 5: TABELLE HASH

Una funzione hash è una mappa $h: U \to \{0, 1, \dots, m-1\}$

collisione: $\exists\, k_1, k_2 \in K \subseteq U$, $k_1 \neq k_2$ e $h(k_1) = h(k_2)$.

Lavoreremo sempre con funzioni hash:

- uniformi : $\Pr[h(k) = q] = \frac{1}{m}$ $\quad q \in \{0, 1, \dots, m-1\}$
- indipendenti : $\Pr[h(k_1) = q_1 \wedge h(k_2) = q_2] = \Pr[h(k_1) = q_1] \cdot \Pr[h(k_2) = q_2]$

5.1. TABELLE HASH CON CONCATENAZIONE

Il fattore di carico (load factor), $\alpha = \frac{n}{m}$ = # medio di elementi memorizzati in una lista collegata.

n : # di chiavi memorizzate

m : # di slot della tabella

$$\alpha \;\substack{<\\=\\>}\; 1$$

j-esimo slot della tabella

Per $j = 0, 1, \dots, m-1$, $T[j]$ ha lunghezza n_j

$n = n_0 + n_1 + \dots + n_{m-1}$, $\mathbb{E}[n_j] = \alpha = \frac{n}{m}$.

Assumiamo di potere calcolare/accedere ad $h(k)$ e $O(1)$ (costante).

→ il tempo necessario alla ricerca di un elemento con chiave k dipende dalla lunghezza $n_{h(k)}$ della lista $T[h(k)]$.

La ricerca di un elemento in una tabella hash in cui le collisioni sono risolte per concatenazione richiede tempo $\Theta(1 + \alpha)$.

5.1.1 RICERCA SENZA SUCCESSO

Thm: Una ricerca senza successo richiede tempo $\Theta(1+\alpha)$.

Proof: Bisogna calcolare $h(k)=j \rightsquigarrow$ tempo costante

ed esaminare n_j elementi, $\mathbb{E}[n_j]=\alpha$

$\Rightarrow \Theta(1+\alpha)$.

5.1.2 RICERCA CON SUCCESSO

Thm: Una ricerca con successo richiede tempo $\Theta(1+\alpha)$.

Proof: Enumeriamo gli elementi della tabella $x_1, \dots, x_n$.

Sia x_i l'elemento ricercato. $k_i = x_i.key$

Il # di elementi esaminati è 1+ il numero di elementi che precedono x_i nella lista collegata (sono stati inseriti successivamente).

Per ogni slot $q \in \{0, \dots, m-1\}$, $\forall k_i, k_j \in K$, $k_i \neq k_j$ sia

$$X_{ijq} = \begin{cases} 1 & \text{se la ricerca è per } x_i,\ h(x_i)=q,\ h(x_j)=q \\ 0 & \text{altrimenti} \end{cases}$$

$$\Pr[X_{ijq}=1] = \Pr\left[\begin{matrix} x_i \text{ è l'elemento} \\ \text{cercato} \end{matrix}\right] \Pr[h(x_j)=q] \Pr[h(x_i)=q] =$$

$$= \frac{1}{n} \cdot \frac{1}{m} \cdot \frac{1}{m} = \frac{1}{nm^2}$$

$$\mathbb{E}[X_{ijq}] = \underbrace{0 \cdot \Pr[X_{ijq}=0]}_{=0} + 1 \Pr[X_{ijq}=1] = \frac{1}{nm^2}.$$

Definiamo $\forall j=1,\dots,n$

$$Y_j = \begin{cases} 1 & x_j \text{ appare nella stessa lista di } x_i \text{, prima} \\ 0 & \text{altrimenti} \end{cases}$$

$$Y_j = \sum_{q=0}^{m-1} \sum_{i=1}^{j-1} X_{ijq}$$

perché $X_{ijq}=1$ per al più un valore di (i,q), $\Leftrightarrow h(x_i)=h(x_j)$ e $i<j$.

Def. $Z = \sum_{j=1}^{n} Y_j$ (# elementi nella stessa lista di x_i e lo precedono)

$$T(n) = 1 + \mathbb{E}[Z] = 1 + \mathbb{E}\left[\sum_{j=1}^{n} \sum_{q=0}^{m-1} \sum_{i=1}^{j-1} X_{ijq}\right] =$$

$$1 + \sum_{j=1}^{n} \sum_{q=0}^{m-1} \sum_{i=1}^{j-1} \mathbb{E}[X_{ijq}] = 1 + \sum_{j=1}^{n} \sum_{q=0}^{m-1} \sum_{i=1}^{j-1} \frac{1}{n\, m^2} =$$

$$1 + \frac{1}{n\, m^2}\left(\underbrace{\sum_{q=0}^{m-1}}_{m} \sum_{j=1}^{n} \underbrace{\sum_{i=1}^{j-1} 1}_{j-1}\right) = 1 + \frac{1}{n\, m^{\cancel{2}}}\, \cancel{m} \left(\sum_{j=1}^{n} j-1\right) =$$

$$= 1 + \frac{1}{n\, m} \sum_{j=0}^{n-1} j = 1 + \frac{1}{\cancel{n}\, m} \frac{\cancel{n}(n-1)}{2} =$$

$$1 + \frac{n-1}{2m} = 1 + \frac{\overset{=\alpha}{n}}{2m} - \frac{1}{2m}\frac{n}{n} = 1 + \frac{\alpha}{2} - \frac{\alpha}{2n} = \Theta(1+\alpha). \square$$

5.2. Tabelle Hash a indirizzamento aperto

Una funzione hash di permutazione

$$h: K \longrightarrow \text{Permutazioni di } (\underset{\{0,\dots,m-1\}}{m}) \qquad (m!)$$

$$\underset{x}{k} \longmapsto \langle h(k,1), \dots, h(k,m) \rangle$$

Lavoriamo con hash di permutazione che sono uniformi e indipendenti.

$\Updownarrow$

Ogni sequenza di prove per ogni chiave ha la stessa probabilità, e non dipende da quella di altre chiavi.

Il fattore di carico è $n \le m \Rightarrow \alpha \le 1$.

Noi assumiamo che $\alpha<1$. $\rightarrow$ Almeno uno slot è vuoto.

```
HASH SEARCH(T,k)

i = 0
repeat
      q = h(k,i)
      if T[q]==k
          return q
      i = i+1
until T[q]==NIL or i==m
return NIL
```

No cancellazioni.

Dobbiamo quantificare il # di prove.

Proveremo che esso è limitato superiormente da

$$1+\alpha+\alpha^2+\alpha^3+\cdots = \frac{1}{1-\alpha}$$

5.2.2.

Thm: Se $\alpha<1$, il numero atteso di prove in una ricerca senza successo è al più $\frac{1}{1-\alpha}$.

Proof:

X: numero di prove fatte in una ricerca senza successo

$\forall i=1,2,\dots,m$

A_i: "La i-esima prova avviene e incontra uno slot occupato".

$\{X\geq i\} = A_1\cap A_2\cap\cdots\cap A_{i-1}$

$$Pr[\{X \geq i\}] = Pr[A_1 \cap A_2 \cap \cdots \cap A_{i-1}] = Pr[A_1] \cdot Pr[A_2 | A_1] \cdots$$

$$\cdots Pr[A_{i-1} | A_1 \cap \cdots \cap A_{i-2}] \leq \underbrace{\alpha \cdot \alpha \cdots \alpha}_{i-1} = \alpha^{i-1}$$

$$Pr[A_1] = \frac{n}{m} = \alpha$$

$j>1 \quad Pr[A_j | A_1 \cap \cdots \cap A_{j-1}] = \dfrac{n-(j-1)}{m-(j-1)} \underset{2 \leq j < m}{\leq} \dfrac{n}{m} = \alpha \qquad i \leq n+1$

$Pr[X \geq i] = 0$ se $i > n+1$

$$T(n) = E[X] = \sum_{i=1}^{\infty} Pr[X \geq i] = \sum_{i=1}^{n+1} \underbrace{Pr[X \geq i]}_{\leq \alpha^{i-1}} + \underbrace{\sum_{i > n+1} Pr[X \geq i]}_{= 0} \leq$$

$$\leq \sum_{i=1}^{n+1} \alpha^{i-1} \leq \sum_{i=0}^{\infty} \alpha^i = \frac{1}{1-\alpha} \,. \quad \square$$

$$\sum_{i=1}^{n+1} \alpha^{i-1} = \sum_{i=0}^{n} \alpha^i$$

COROLLARIO: Inserire un elemento se $\alpha < 1$, richiede tempo $\dfrac{1}{1-\alpha}$.

5.2.2 Ricerca con successo

Thm: Il tempo atteso per una ricerca con successo se $\alpha < 1$ è $\leq \dfrac{1}{\alpha} \ln \dfrac{1}{1-\alpha}$.

Proof: Se l'elemento cercato k è l'elemento $(i+1)$-esimo a essere stato inserito, il fattore di carico al tempo dell'inserimento è $\dfrac{i}{m} = \tilde{\alpha}$, quindi il numero atteso di prove in una ricerca per k è $\dfrac{1}{1-\frac{i}{m}} = \dfrac{m}{m-i}$

Facendo una media per tutte le chiavi, otteniamo che il numero atteso è

$$\frac{1}{n}\sum_{i=0}^{n-1}\frac{m}{m-i} = \frac{m}{n}\sum_{i=0}^{n-1}\frac{1}{m-i} \underset{\downarrow}{=} \frac{m}{n}\sum_{k=m-n+1}^{m}\frac{1}{k} =$$

$k = m-i$

$i=0 \rightarrow k=m$

$i=n-1 \quad k=m-n+1$

$$= \frac{1}{\alpha}\sum_{k=m-n+1}^{m}\frac{1}{k} = \frac{1}{\alpha}\int_{m-n}^{m}\frac{1}{x}\,dx = \frac{1}{\alpha}\Big(\ln(m)-\ln(m-n)\Big) =$$

$$= \frac{1}{\alpha}\ln\frac{m}{m-n} =$$

$$= \frac{1}{\alpha}\ln\frac{1}{1-\frac{n}{m}} =$$

$$= \frac{1}{\alpha}\ln\frac{1}{1-\alpha}. \quad \square$$

Se $f(x)$ è monotonicamente decrescente allora

$$\int_{a}^{b+1} f(x)\,dx \le \sum_{j=a}^{b} f(j) \le \int_{a-1}^{b} f(x)\,dx$$