TABELLE HASH : INDIRIZZAMENTO APERTO

$x \in S$

$T[h(x.key)]$

$n \le m$

$|S|$ — dim. tabella hash

$\Downarrow$

$\alpha = \frac{n}{m} \le 1$

T

1	AZALEA	A
2	ZIBIBBO	B
3		C
15	ORTENSIA	O
16	PALMA	P
17	PEPERONE	Q
18	OLMO	R
19	PINO	S
26	ZUCCA	Z

"probing"

ispezione, scansione

$m = 26$

Titolo nota 26/02/2016

$$h : \underbrace{U}_{\text{chiave}} \times \underbrace{[0, m-1]}_{\text{ordine di ispezione}} \longrightarrow \underbrace{[0, m-1]}_{\text{posizione della tabella}}$$

$\forall k \in U$ "sequenza di scansione"

$$\underbrace{\langle h(k,0),\ h(k,1),\ \ldots\ h(k, m-1)}_{\text{permutazione delle } m \text{ posizioni della tabella}} \rangle$$

INSERIMENTO

HASH-INSERT (T, k)

```
i=0
repeat {
    j = h(k, i);
    if (T[j] == NIL) {
        T[j] = k
        return j;
    }
    else i++;
}
until i == m;
error "overflow della tabella hash"
```

$T(n) = O(n)$ caso peggiore

$T(n) = \Theta(1)$ caso ottimo

$$\boxed{T(n) = \Theta\left(\frac{1}{1-\alpha}\right)}$$ caso medio

Titolo nota 26/02/2016

RICERCA (Solo inserimenti e ricerca)

HASH-SEARCH (T, k)

```
i=0;
repeat {
    j = h(k, i);
    if (T[j] == k) return j;
    i++;
}
until (T[j] == NIL OR i == m)
return NIL
```

$T(n) \Rightarrow$ come x l'inserimento

Operazioni di dizionario: analisi al caso medio

IPOTESI di HASHING UNIFORME

ogni chiave ha la stessa probabilità di avere come sequenza di scansione una delle $m!$ permutazioni di $\{0, 1, 2, \ldots, m-1\}$

TEOREMA

Nell'ipotesi di hashing uniforme, data una tabella hash a indirizzamento aperto con un fattore di carico $\alpha = \frac{n}{m} < 1$, il numero atteso di accessi nelle operazioni di dizionario è al massimo $\frac{1}{1-\alpha}$

Dimostrazione

\# accessi per inserimento, caso medio

IPOTESI

- analisi in funzione di α
- hashing uniforme
- tabella mai piena $\quad 0 \le n < m \quad \Rightarrow \alpha < 1$
- nessuna cancellazione

X = numero di accessi

$$E(X) = \sum_{i=1}^{+\infty} i \Pr[X = i] = \underbrace{\sum_{i=1}^{+\infty} \Pr[X \geq i]}$$

Titolo nota 26/02/2016

$i = 1$ $\quad Pr[X \geq 1] = 1$

$i = 2$ $\quad Pr[X \geq 2] =$ prob. di avere la prima cella occupata

$$= \frac{n}{m} = \alpha$$

$Pr[X \geq 3] =$ prob. di avere le prime due celle (della sequenza di scansione) occupate

$$= \frac{n}{m} \cdot \frac{n-1}{m-1} \leq \alpha \cdot \alpha = \alpha^2 \qquad \alpha = \frac{n}{m}$$

$n < m$

$$\boxed{\frac{n-j}{m-j} \leq \frac{n}{m} \qquad \forall \quad 0 \leq j < m} \quad (*)$$

$Pr(X \geq i) =$ probabilità di trovare le prime $i-1$ celle occupate

$$= \underbrace{\frac{n}{m} \cdot \frac{n-1}{m-1} \cdot \frac{n-2}{m-2} \cdots \cdots \frac{n-i+2}{m-i+2}}_{i-1 \text{ termini}} \underset{(*)}{\leq} \left(\frac{n}{m}\right)^{i-1}$$

$$= \alpha^{i-1}$$

$$E(X) \overset{\text{teo}}{=} \sum_{i=1} Pr[X \geq i] \leq \sum_{i=1}^{\text{teo}} \alpha^{i-1} \overset{\text{teo}}{=} \underbrace{\sum_{i=0} \alpha^{i}}_{\alpha < 1} = \frac{1}{1-\alpha}$$

$\square$

$\alpha = \frac{1}{2}$ 50%

$$\#\,\text{accessi in media} \le \frac{1}{1-\alpha} = 2$$

90%

$\alpha = \frac{9}{10}$

$$\#\,\text{accessi in media} \le \frac{1}{1-\alpha} = 10$$

$m = 100$ ⟨∼∼⟩ $m = 10.000$
$n = 90$ $n = 9000$

stesse prestazioni!

CALCOLO DELLA SEQUENZA DI SCANSIONE

1 SCANSIONE LINEARE

m = numero primo

$$h(k,i) = \left(h'(k) + i \right) \bmod m$$

h' ≡ funzione hash ausiliare $\qquad h' : U \to [0, m-1]$

$$\langle \underbrace{h'(k)}_{\bmod m},\ \underbrace{h'(k)+1}_{\bmod m},\ \underbrace{h'(k)+2}_{\bmod m},\ \dots,\ m-1,\ 0,\ 1,\ \dots \rangle$$

$m = 7 \qquad h'(k) = 2$

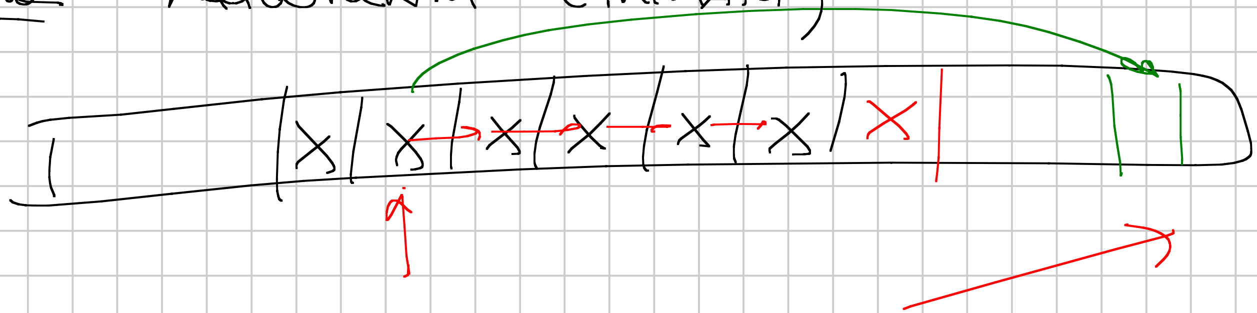
$$\langle 2, 3, 4, 5, 6, 0, 1 \rangle$$

Punto di partenza: "casuale" $h'(k)$, dipende da k

Passo: $+1$, non dipende dalla chiave k

Si generano m sequenze $\neq$ $\quad m << m!$

Problema: AGGLOMERATI (PRIMARI)



Titolo nota 26/02/2016

(2) Scansione Quadratica

m = numero primo

$c_1, c_2 \neq 0$ costanti

$h' : U \rightarrow [0, m-1]$

$\boxed{c_1 = c_2 = \frac{1}{2}}$

$\boxed{m = 2^s}$

$$h(k,i) = (h'(k) + c_1 \cdot i + c_2 i^2) \bmod m$$

PUNTO DI PARTENZA: "casuale": $h'(k)$

PASSO: non è più costante, dipende in modo quadratico da i

c_1 e c_2 devono essere scelti in modo che la sequenza di scansione sia una permutazione di tutte le m posizioni della tabella.

- se due chiavi hanno lo stesso valore hash $\underline{\underline{h'}}$, le loro sequenze di scansione sono identiche

  $\hookrightarrow$ agglomerati secondari

m sequenze $\neq$ $\ll$ $m!$

3 DOPPIO HASH

h_1, h_2 funzioni hash ausiliarie

$$\boxed{h(k,i) = (h_1(k) + i\, h_2(k)) \bmod m}$$

PUNTO DI PARTENZA : "casuale" via $h_1(k)$

PASSO : dipende in modo "casuale" da k, via $h_2(k)$

$$\langle h_1(k) \bmod m,\ (h_1(k) + h_2(k)) \bmod m,\ (h_1(k) + 2\, h_2(k)) \bmod m,\ \dots \rangle$$

$MCD(h_2(k), m) = 1$ per avere una permutazione di tutte le posizioni

Titolo nota 26/02/2016

1) $m =$ potenza di 2

h_2: assume solo valori dispari

2) m numero primo

$h_1(k) = k \bmod m$

$h_2(k) = \underline{1} + \underbrace{k \bmod (m-1)}_{[0,\ m-2]}$

$\forall k$

$\underbrace{1 \leq h_2(k) \leq m-1}$

gli interi in $[1, m-1]$ sono tutti co-primi con m.

$$\boxed{h(k,i) = \left(k \bmod m + i\left(1 + k \bmod (m-1)\right)\right) \bmod m}$$

\# sequenze $\neq$ generate ? $\Theta(m^2)$

$h_1(k), h_2(k)$ determina la sequenza.

$$m^2 < m!$$

Numero medio di passi in una ricerca con successo

α	10%	50%	75%	90%
scansione lineare	1.06	1.50	2.50	5.50
scansione quadratica	1.05	1.44	1.99	2.79
doppio hash	1.05	1.38	1.83	2.55

ESEMPIO

$$h(k,i) = (k \bmod 11 + i) \bmod 11 \qquad 0 \leq i \leq 10$$

$$m = 11$$

1) Scansione lineare

chiavi	sequenza di scansioni
10	10
22	0
31	9
4	4
15	4, 5
28	6
17	6, 7
88	0, 1
59	4, 5, 6, 7, 8
26	4, 5, 6, 7, 8, 9, 10, 0, 1, 2

T

0	22
1	88
2	26
3	
4	4
5	15
6	28
7	17
8	59
9	31
10	10