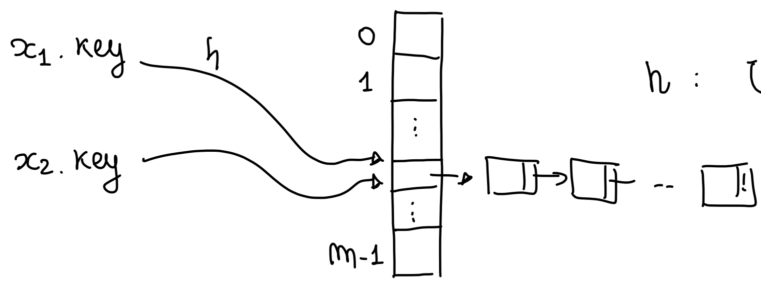
Algoritmi e Strutture Dati (02/11/2021)

* Tabelle hash

Insiemi dinamici di elementi x con chiave $x.key$ in U



$h : U \to \{0, 1, \dots, m-1\}$

$m \ll |U|$

↳ collisioni

* Chaining

Chaining : ogni entry contiene la lista degli elementi con lo stesso hashing della chiave

tempo medio

α = fattore di carico = $\frac{n}{m}$ $\qquad 0 \leq \alpha \begin{cases} < 1 \\ = 1 \\ > 1 \end{cases}$

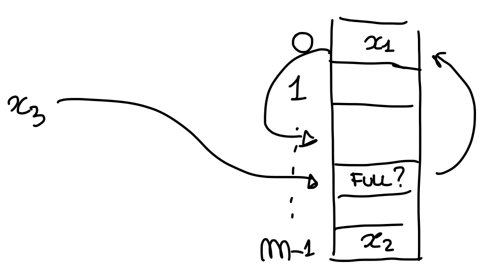
↳ - ricerca elemento assente $\qquad 1 + \alpha$

- ricerca elemento presente $\qquad 1 + \alpha/2$ $\qquad O(1+\alpha)$

se $n \leq c\,m$ ⇝ $O(1)$

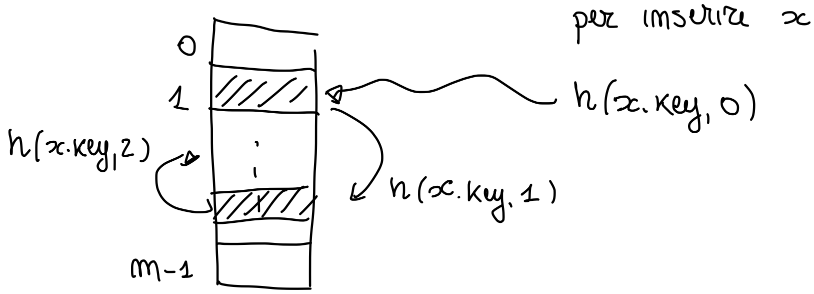
* Open Addressing

Idea : memorizzo gli elementi dell'insieme dinamico solo nello spazio della tabella



funzione di hash $h(k, i)$

chiave → k, # tentativo → i



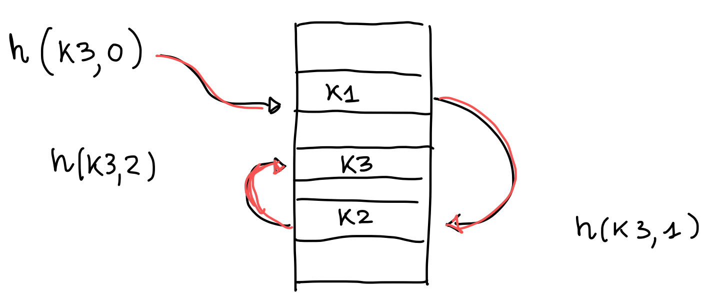
$$\underbrace{h(k,0)\quad h(k,1)\quad h(k,2)\quad \cdots\quad h(k,m-1)}_{\text{sequenza di ispezione}}$$

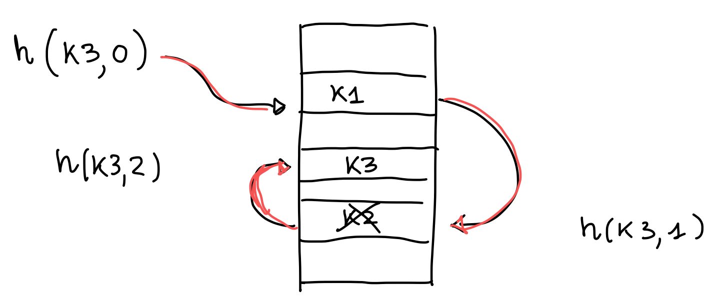
permutazione di $0, 1, \ldots, m-1$

* <u>operazioni</u>

```
Insert (T, x)
    i = 0
    repeat
        j = h(x.key, i)
        if (T[j] == nil) or (T[j] == deleted)
            T[j] = x
            return j
        i = i + 1
    until i == m
    error "tabella piena"
```

```
Search (T, k)
    i = 0
    repeat
        j = h(k, i)
        if T[j].key == k
            return j
        i = i + 1
    until (i == m) or (T[j] == nil)
    return nil
```

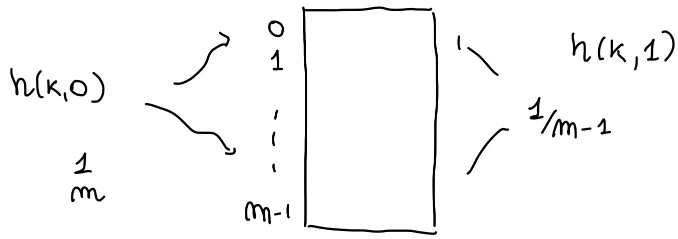





```
Delete (T, j)
    T[j] = deleted
```

* Hashing Uniforme

ogni elemento determina con la stessa probabilità una qualunque delle $m!$ sequenze di ispezione

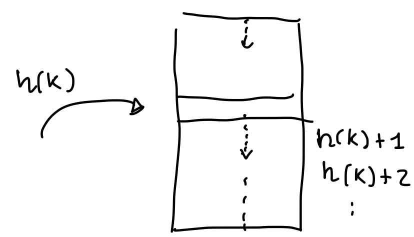


* Come definire funzioni di hash?

① Ispezione lineare

fisso $h(k)$ funzione di hash

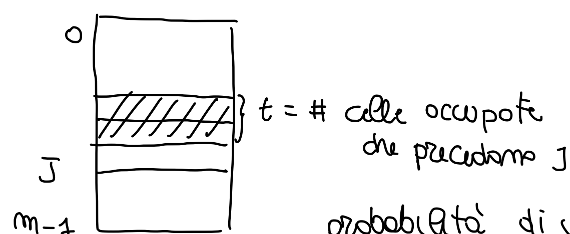
$h'(k,i) = (h(k)+i) \bmod m$



vantaggio : semplice

svantaggio : addensamento primario

prob. di inserire in J è "proporzionale" al numero di celle occupate che la precedono



probabilità di usare J è

$$\frac{t+1}{m}$$

\# di sequenze di ispezione = m (dipende solo da $h(k) \bmod m$)

② Ispezione quadratica

data $h(k)$ funzione di hash

$$h'(k,i) = (h(k) + c_1 i + c_2 i^2) \bmod m \qquad c_1, c_2 \text{ opportune}$$

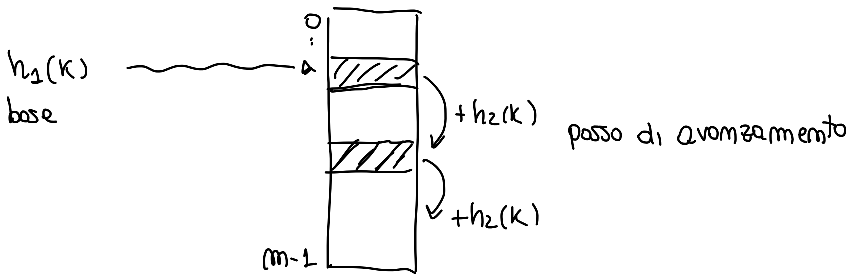
→ addensamento secondario

→ # sequenze di ispezione: dipende solo da $h(k) \bmod m$ ⇒ m possibilità

③ Doppio hashing

Siano $h_1(k)$, $h_2(k)$ funzioni di hash di un solo argomento

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



* come mi assicuro che vengono ispezionate tutte le celle?

Condizione: $h_2(k)$ e m sono relativamente primi

$MCD(h_2(k), m)$

dati $i, i' < m$ $\quad i \neq i' \Rightarrow h(k,i) \neq h(k,i')$

$(\Rightarrow \{h(k,0), \dots, h(k,m-1)\} = \{0, \dots, m-1\})$

dim

siano $i, i' \in \{0, 1, \dots, m-1\}$ e sia $h(k,i) = h(k,i')$ $\quad$ mostro $i = i'$

assumo $i' \leq i$

$$(h_1(k) + i\, h_2(k)) \bmod m = (h_1(k) + i'\, h_2(k)) \bmod m$$

$$(h_1(k) + i\, h_2(k)) \bmod m - (h_1(k) + i'\, h_2(k)) \bmod m = 0$$

$$(\cancel{h_1(k)} + i\, h_2(k) - \cancel{h_1(k)} - i'\, h_2(k)) \bmod m = 0$$

$$(i - i')\, h_2(k) \bmod m = 0$$

↳ m divisore di $i - i'$ $\quad$ (dato che $MCD(h_2(k), m) = 1$)

ma $0 \leq i - i' \leq m-1$

$\Rightarrow i - i' = 0 \Rightarrow i = i'$

Esempio

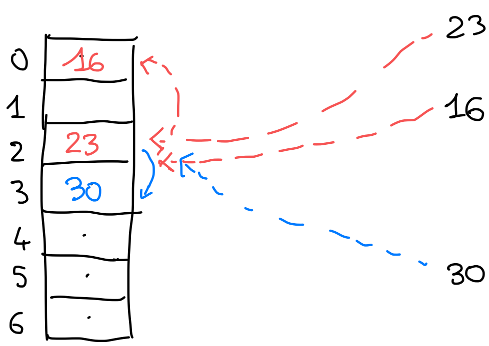
$$\begin{cases} m = 2^p \\ h_2(k) = 2\,h'(k) + 1 \end{cases}$$ con h' funzione di hash qualunque

$$\begin{cases} m \text{ primo} \\ h_2(k) < m \end{cases}$$ $h_2(k) = 1 + k \bmod m'$ con $m' < m$ (es. $m' = m - 1$)

esempio

$m = 7$ $\quad h_1(k) = k \bmod 7$

$m' = 6$ $\quad h_2(k) = 1 + k \bmod 6$



$h_1(k) = 2$

$h_1(k) = 2$

$h_2(k) = 1 + 16 \bmod 6 = 1 + 4 = 5$

$h_1(30) = 2$

$h_2(30) = 1 + 0 = 1$

\# sequenze di ispezione ? dipende da $h_1(k)$ e $h_2(k)$

$\Theta(m^2)$

* ANALISI

caso medio (ip. hashing uniforme)

rispetto $\alpha = \dfrac{n}{m}$ $\quad 0 \leq \alpha \leq 1$

(a) Ricerca di chiave assente

\# medio di celle ispezionate

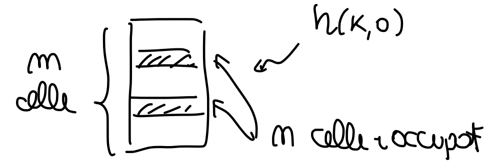
(a) $\dfrac{1}{1-\alpha}$ se $\alpha < 1$

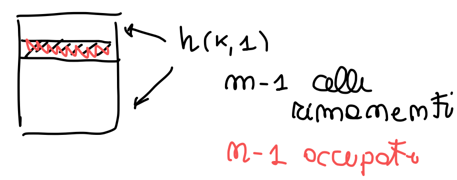
(b) m $\quad \alpha = 1$

dim

seguo la sequenza di ispezione fino ad una cella vuota

		prob
prob	$i = 0$	1
prob	$i = 1$ = prob $i=0$ occupata	$\frac{n}{m}$
prob	$i = 2$ = prob $(i=1)$ · prob. cella occupata	$\frac{n}{m} \cdot \frac{n-1}{m-1}$





prob i

$$\underset{\alpha}{\underset{\shortparallel}{\frac{n}{m}}}\ \underset{\alpha}{\underset{\wedge|}{\frac{n-1}{m-1}}} \cdots \underset{\alpha}{\underset{\wedge|}{\frac{n-i+1}{m-i+1}}} \leq \alpha^i$$

\# atteso di celle da ispezionare

$$1 + \alpha + \alpha^2 + \cdots + \alpha^{m-1} \quad (*)$$

2 casi

$(\alpha < 1)$ $\quad (*) \leq \sum_{i=0}^{\infty} \alpha^i = \frac{1}{1-\alpha}$

$(\alpha = 1)$ $\quad (*) = \underbrace{1 + \cdots + 1}_{m \text{ volte}} = m$

(b) ricerca di un elemento presente

(a) $\frac{1}{\alpha} \log\left(\frac{1}{1-\alpha}\right)$ $\quad \alpha < 1$

(b) $1 + \log m$ $\quad \alpha = 1$

ho inserito $x_0 \ x_1 \ \_ \ | \ x_i \ .. \ x_{m-1}$

costo ricerca x_i

= costo ricerca di x_i assente dopo che ho inserito $x_0, x_1, \ldots, x_{i-1}$

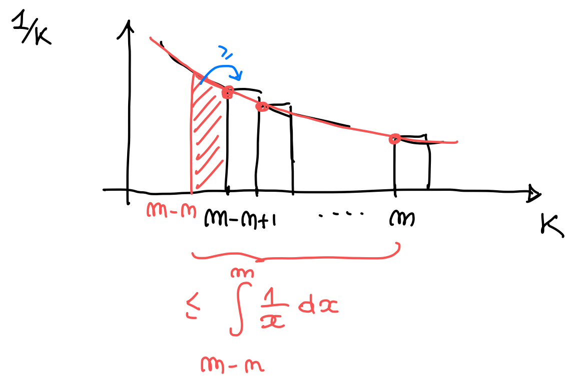
$$= \frac{1}{1-\underbrace{\alpha_i}_{\substack{\text{valore di } \alpha \\ \text{dopo elims. di } x_0, x_1, \dots x_{i-1}}}} = \frac{1}{1-\frac{i}{m}} = \frac{m}{m-i}$$

costo medio per ricerca di un qualunque elemento

$$\frac{1}{n} \sum_{i=0}^{n-1} (\text{costo medio per } x_i)$$

$$= \frac{1}{n} \sum_{i=0}^{n-1} \frac{m}{m-i} \qquad = \underbrace{\frac{m}{n}}_{\frac{1}{\alpha}} \sum_{i=0}^{n-1} \underbrace{\frac{1}{m-i}}_{m-n+1 \; \dots \; m}$$

$$= \frac{1}{\alpha} \sum_{k=m-n+1}^{m} \frac{1}{k}$$



2 casi

$(\alpha < 1)$

$$= \frac{1}{\alpha} \sum_{k=m-n+1}^{m} \frac{1}{k}$$

$$\leq \frac{1}{\alpha} \int_{m-n}^{m} \frac{1}{x}\, dx$$

$$= \frac{1}{\alpha} \left(\log m - \log (m-n) \right)$$

$$= \frac{1}{\alpha} \log \left(\underbrace{\frac{m}{m-n}}_{\frac{1}{1-\frac{n}{m}} = \frac{1}{1-\alpha}} \right) = \frac{1}{\alpha} \log \left(\frac{1}{1-\alpha} \right)$$

$(\alpha = 1)$

$$\underbrace{\frac{1}{\alpha}}_{= 1} \sum_{k=\underbrace{m-n}_{0}+1}^{m} 1/k = 1 + \sum_{k=2}^{m} 1/k \leq 1 + \int_{1}^{m} \frac{1}{x}\, dx$$

separo il primo termine

$$= 1 + \log m - \cancel{\log 1} = 1 + \log m$$

α	ricerca chiave assente $\frac{1}{1-\alpha}$	ricerca chiave presente $\frac{1}{\alpha}\log\left(\frac{1}{1-\alpha}\right)$
0,3	1,43	1,19
0,7	3,33	1,72
0,9	10	2,56
0,99	100	4,65

OPEN ADDRESSING - ASSENTE

OPEN ADDRESSING - PRESENTE

CHAINING - ASSENTE

CHAINING - PRESENTE