

Introduzione agli Algoritmi e alle Strutture Dati

Tabelle Hash

Dr. Emanuela Merelli

Argomenti della lezione

- Tipi di Dato Astratto
 - Tabella Hash
- *Concetto di*
 - *Struttura dati*

Tabella

La tabella è una successione di elementi; ciascun elemento E_i comprende due parti:

1. la chiave K_i , costituita da un gruppo di caratteri (o più generale di bit) che distinguono l'elemento dagli altri;
 2. l'informazione I_i , associata alla chiave.
- scopo della tabella è la memorizzazione delle I_i ,
 - scopo della chiave è l'individuazione delle I_i dall'esterno.

Operazioni su tabelle

- Ricerca di un elemento E_i con chiave K_i
- Eliminazione di un elemento trovato
- Inserzione di un elemento nuovo nel caso la chiave in esame non risulti associata ad alcun elemento esistente in tabella

Operazioni di ricerca Tabelle

L'operazione fondamentale che si esegue su una tabella è la ricerca di un elemento nota la chiave.

Per ogni elemento E_i si può definire la *lunghezza di ricerca* S_i come un numero di prove necessarie per raggiungere E_i (cioè numero di chiavi lette fino a trovare K_i)

La *lunghezza media* di ricerca $\bar{S}$ è la media delle S_i per tutti gli elementi della tabella

Esempi di Tabelle

Schedari, Elenchi, Dizionari

Schedario della flora italiana

Per ogni vegetale vi sarà una scheda (elemento E_i della tabella)

Si accede alla scheda cercando il nome del vegetale (chiave K_i), che contiene tutte le informazioni I_i d'interesse del vegetale.

Esempio

1	Peperone	Inf su peperone
2	Azalea	Inf su azalea
3	Ortensia	Inf su ortensia
4	Zucca	Inf su zucca
5	Palma	Inf su palma
6	Melanzana	Inf su melanzana
n	...	

Metodo 1: utilizzo di un vettore

senza ordinamento

- **Ricerca Completa**
- Se l'elemento E_i si trova nella posizione $V[j]$ la sua lunghezza di ricerca è $S_i = j$
- Lunghezza media di ricerca sarà

$$\bar{S} = \frac{1}{n} \sum_{j=1}^n j = \frac{n+1}{2}$$

Metodo 1: osservazioni

- Inserimento è semplice
- L'eliminazione crea problemi ...
 - Soluzione, utilizzo dell'elemento $V[1]$ per memorizzare l'indice della prima libera
 - Soluzione è inserire un nuovo campo, marca F che indica se l'elemento è stato cancellato o è attivo → cancellazione logica

Esempio

1	Peperone	0	Inf su peperone
2	Azalea	1	Inf azalea
3	Ortensia	0	Inf su ortensia
4	Zucca	0	Inf su zucca
5	Palma	0	Inf su palma
6	Melanzana	1	Inf melanzana
n	...		

Conseguenza

- Riduzione di efficienza nella ricerca
- Inutilizzazione di alcuni elementi del vettore
- La soluzione proposta è l'utilizzo di una lista concatenata di tutti gli elementi cancellati e quindi liberi di V , utilizzando un puntatore.

Esempio

1	Peperone	0	Inf su peperone
2	6	1	Inf azalea
3	Ortensia	0	Inf su ortensia
4	0	0	Inf su zucca
5	Palma	0	Inf su palma
6	4	1	Inf melanzana
n	...		

Ricerca Binaria

- Memorizzazione in ordine alfabetico
- Algoritmo di ricerca
 - Si confronta la voce centrale con quella cercata
 - Se maggiore si cerca ricorsivamente nella seconda metà in basso
 - Se minore si cerca ricorsivamente nella metà in alto
 - Se uguale trovata.

Codifica hash

- Si basa sull'idea di ricavare direttamente dalla chiave il massimo dell'informazione sulla posizione dell'elemento nella tabella.
- Definire una Funzione d'accesso che permetta di ottenere tale posizione in funzione della chiave, e consenta di ridurre il numero di prove necessarie a individuare la voce cercata

Indirizzi Hash

- $h(K_i)$ è un gruppo di bit ricavati dalla chiave K_i
- $h(k_i)$ è utilizzato per indirizzare direttamente la tabella

Tabella ad accesso diretto

- $h(K_j) = j$ j intero $1 \leq j \leq m$
- $h(K_j) \neq h(K_g)$ $j \neq g$
- La prima relazione vale per ogni chiave K_j
- La seconda per ogni possibile coppia di chiavi distinte K_j e K_g

Indirizzo hash coincide con la posizione nel vettore

Esempio accesso diretto

- $h(\text{azalea}) = 1$
- $h(\text{melanzana}) = 6$
- ...

1	Peperone	Inf su peperone
2	Azalea	Inf su azalea
3	Ortensia	Inf su ortensia
4	Zucca	Inf su zucca
5	Palma	Inf su palma
6	Melanzana	Inf su melanzana
n	...	

Problemi

- Difficoltà di definire una funzione che colleghi la chiave all'indice
- Necessità di conoscere a priori tutte le chiavi
- Possibilità di elementi non utilizzati
- Collisione di elementi $h(K_j) = h(K_g)$

Collisioni di elementi

- Se $h(K_i) = h(K_j)$ come si sceglie la posizione nel vettore?

