Studio sulla ricerca tabellare

Benchè si possa risolvere, quando possibile, con la tecnica della dicotomia (e dimostrerò che questa tecnica è la più veloce), ritengo sia utile dare alla ricerca tabellare una base matematica, sia per additare soluzioni di realizzazione più semplici anche se più lente, sia per rendere più veloci quelle già usate, sfruttando le caratteristiche delle informazioni da accoppiare con la tabella.

1 Col termine tabella si intende un insieme di informazioni atte ad integrare o controllare altre informazioni; ogni elemento della tabella è formato, in linea generale, da un codice e da altri dati. Il codice serve da collegamento con le informazioni da elaborare. Tabella può essere un archivio su di una memoria ad accesso casuale come disco o tamburo, una lista di descrizioni, i dati retributivi contrattuali relativi ad ogni categoria.
La tabella può essere ordinata come i dati da elaborare, avere un altro ordine, od essere disordinata.
Il primo caso corrisponde all'accoppiamento di flussi con lo stesso ordine.
Vi sono infine tabelle particolari, il cui elemento generico è individuabile tramite un'informazione contenuta sui dati da elaborare, informazione che, eventualmente moltiplicata per una costante, ci dà la posizione dell'elemento della tabella.

2 La gestione di una tabella disordinata non può che avvenire ricercando elemento per elemento, quello da accoppiare con l'informazione da elaborare. Indichiamo con « n » il numero di elementi della tabella e « p » la probabilità che l'elemento non esista in tabella. Allora il numero di tentativi per accoppiare l'informazione da elaborare sarà:

$$y = (1-p)\frac{n}{2} + pn \qquad 2.1$$

3 Il caso in cui la tabella è ordinata, ma non come le informazioni da elaborare, conduce al problema pratico di ricercare quale elemento di essa corrisponda al dato da elaborare, nel modo più semplice e più veloce possibile.
Il sistema di avanzare di un elemento per volta è inaccettabile poichè si avrebbero $\frac{n}{2}$ tentativi.
Un sistema semplice e più veloce consiste nell'avanzare di un certo numero di elementi per volta e, quando all'ultimo confronto l'elemento in tabella risulti maggiore dell'informazione da elaborare, avanzare di un elemento alla volta a partire dal penultimo confronto.
L'analisi ci fornisce il modo di ottimizzare il passo di avanzamento « x ».

Infatti il numero di tentativi è:

$$y = \frac{1}{2}\frac{n}{x} + \frac{1}{2}x \qquad 3.1$$

$$\frac{dy}{dx} = -\frac{1}{2}\frac{n}{x^2} + \frac{1}{2} \qquad 3.2$$

$$-\frac{1}{2}\frac{n}{x^2} + \frac{1}{2} = 0;\ x = \sqrt{n} \qquad 3.3$$

sostituendo questo valore nella 3.1

$$y = \frac{1}{2}\frac{n}{\sqrt{n}} + \frac{1}{2}.\sqrt{n} = \frac{n}{\sqrt{n}} = \sqrt{n} \quad 3.4$$

Si ottiene un numero di passaggi inferiore estendendo la tecnica precedente, cioè adottando ogni qualvolta l'elemento della tabella risulta maggiore dell'informazione da elaborare un passo d'avanzamento inferiore, sino a giungere in ultimo ad avanzare di un elemento per volta.
Questi passi sono potenze positive di un certo numero che indicheremo con x:

$$x^m, x^{m-1}, \ldots\ldots x, 1 \qquad \text{con } x^{m-i} < n \leq x^m \quad 3.5$$

esempio se $n = 9000$, $x = 10$
i passi di avanzamento saranno: 1000, 100, 10 e 1.
Il numero di questi cicli lo ricaviamo dalla disequazione 3.5 considerando il caso più restrittivo $n = x^m$ cioè

$$\log n = m \log x$$

$$m = \frac{\log n}{\log x} \text{ (logaritmi naturali)} \qquad 3.6$$

allora il numero di tentativi sarà:

$$y = \frac{1}{2}x . \frac{\log n}{\log x} \qquad 3.7$$

ricorrendo all'analisi cerchiamo il valore di x che minimizza la 3.7

$$\frac{dy}{dx} = \frac{\log n}{2} . \frac{\log x - 1}{(\log x)^2} \qquad 3.8$$

$$\log n (\log x - 1) = 0$$
$$\log x = 1 \qquad x = e = 2{,}718 \ldots\ldots \qquad 3.9$$

i valori di x che soddisfano alla condizione di minimo sono 3 e 2.

4 Tabelle probabilistiche.
Con questo nome si designano le tabelle in cui l'elemento n-esimo da ricercare ha probabilità « p » di trovarsi in una posizione determinata dall'elemento (n-1) -esimo, ad esempio subito dopo.
Determiniamo per quali valori di « p », in funzione del numero, di passaggi occorrenti per trovare l'elemento cercato, è conveniente tenerne conto.
Sia z il numero di passaggi avremo:

$y = z$ se non si tien conto della probabilità
$y = (1-p)(Z+1) + p$ se si tien conto della probabilità $\qquad$ 4.1

cioè

$$z \geq (1-p)(z+1) + p \qquad 4.2$$

per = otteniamo il valore limite di p

$$z = (1-p)(z+1) + p \qquad 4.3$$
$$z = z + 1 - p(z+1) + p$$

$$pz = A \qquad p = \frac{1}{z} \qquad 4.4$$

Giovanni Rossati