Urejanja

1 Quicksort - hitro urejanje

1.1 Sled algoritma

0	5	3	4	8	9	6	2	1	7
				l	lr	l	r	r	
				1	2		9	8	
	2				5				
1	2	3	4	1		6	9	8	7
		lr	l	r		r	l		
		1		3					
	1	2				6			
2			4	3			9	8	7
				lr					lr
			3	4			7		9
3							7	8	
							r	l	
							7		

1.2 Zahtevnost v najboljšem primeru

Najboljši primer se zgodi, kadar se tabela vedno deli na enako veliki polovici. Za zahtevnost tako velja

$$T(n) = 2T(\frac{n}{2}) + O(n).$$

S pomočjo "master izreka" vemo, da $T(n) = O(n \lg n)$. Intuitivno to lahko ugotovimo tudi z naslednjim sklepanjem. Če tabelo vedno delimo na enaki polovici je globina rekurzije kvečjemu $\lg n$. Na vsaki globini je potrebno skupno (vse porazdelitve skupaj) opraviti $O(n)$ dela.

1.3 Zahtevnost v najslabšem primeru

Najslabši primer se zgodi, kadar kot pivot vedno izberemo najmanjši (ali največji) element. Takrat je delitev tabele zelo neenakomerna: en del je prazen, v drugem delu je pivot in tretji deli je skoraj tako velik kot prvotna tabela.

Velikost prvotne tabele je sprva n. Toliko je tudi potrebnega dela za porazdelitev: torej prva delitev iz prvotne tabele velikosti n ustvari tabelo velikosti $n-1$, nato $n-2, n-3$, itd. Kar seštejemo in dobimo $\frac{n(n+1)}{2} = O(n^2)$

1.4 Zahtevnost v povprečju

Pri analizi časovne zahtevnosti štejemo primerjave elementov. Predpostavimo, da so vsi elementi različni oz. da urejamo naključno permutacijo števil od 1 do n od katerih je vsaka enaka verjetna. Označimo z C_n povprečno število primerjav, ki jih naredi Quicksort pri izvajanju na tabeli velikosti n. Velja $C_0 = C_1 = 0$.

Ena porazdelitev (obe notranjih zanki skupaj) naredita $n+1$ primerjav. Porazdelitev lahko za razdeli tabelo na poljubnem mestu. Tabela tako razpade na tri dele: levi del velikosti i, potem sledi pivot in desni del velikosti $n-1-i$. Pri analizi povprečne zahtevnosti moramo upoštevati vsa mesta delitve, verjetno vsake od delitev je $\frac{1}{n}$. V splošnem lahko zapišemo rekurenčno enačbo

$$C_n = (n+1) + \frac{1}{n}\sum_{i=0}^{n-1}(C_i + C_{n-1-i}). \tag{1}$$

Razpišimo vsoto, da se bo lažje videlo, kako jo poenostaviti:

$$\sum_{i=0}^{n-1}(C_i + C_{n-1-i}) = \sum_{i=0}^{n-1} C_i + \sum_{i=0}^{n-1} C_{n-1-i}) = (C_0 + C_1 + C_2 + \cdots + C_{n-1}) + (C_{n-1} + C_{n-2} + C_{n-3} + \cdots + C_1 + C_0).$$

Ugotovimo, da gre za enaki vsoti. Upoštevajmo to in enačbo (1) še množimo z n in dobimo

$$nC_n = n(n+1) + 2\sum_{i=0}^{n-1} C_i.$$

V dobljeno enačbo namesto n vstavimo $n-1$ in dobimo

$$(n-1)C_{n-1} = (n-1)n + 2\sum_{i=0}^{n-2} C_i.$$

Odštejemo dobljeno enačbo od prejšnje in dobimo

$$\begin{aligned} nC_n - (n-1)C_{n-1} &= n(n+1) - (n-1)n + 2\sum_{i=0}^{n-1} C_i - 2\sum_{i=0}^{n-2} C_i \\ &= 2n + 2C_{n-1}. \end{aligned}$$

Sedaj C_n pustimo na levi, C_{n-1} prestavimo na desno stran enačbe, nato še delimo enačbo z $n(n+1)$ in dobimo

$$nC_n = (n+1)C_{n-1} + 2n \quad \text{oz.} \quad \frac{C_n}{n+1} = \frac{C_{n-1}}{n} + \frac{2}{n+1}.$$

Zaporedoma razvijamo $C_n/(n+1)$ in dobimo

$$\begin{aligned} \frac{C_n}{n+1} &= \frac{2}{n+1} + \frac{2}{n} + \frac{2}{n-1} + \cdots + \frac{2}{3} + \frac{C_1}{2} \\ &= 2\left(\frac{1}{n+1} + \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{3}\right) \\ &= 2\left(\frac{1}{n+1} + \frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{3} + \frac{1}{2} + \frac{1}{1} - \frac{1}{2} - \frac{1}{1}\right) \\ &= 2\left(H_{n+1} - \frac{1}{2} - \frac{1}{1}\right) \\ &= 2H_{n+1} - 3. \end{aligned}$$

Povprečno število primerjav je torej enako

$$C_n = 2(n+1)H_{n+1} - 3(n+1).$$

Ker velja $H_n \sim \ln n$ (za dokaz tega glej naslednje razdelke), velja

$$C_n = 2(n+1)\ln n - 3(n+1) \sim 2n\ln n \approx 1.39n\lg n.$$

Quicksort je torej v povprečju 39% počasnejši kot v najboljšem primeru.

1.5 Aproksimacija z integralom

Kadar vsoto lahko izrazimo kot

$$\sum_{i=m}^{n} f(i)$$

kjer je $f(i)$ monotono naraščajoča funkcija, potem lahko vsoto aproksimiramo z integralom

$$\int_{m-1}^{n} f(x)dx \le \sum_{i=m}^{n} f(i) \le \int_{m}^{n+1} f(x)dx.$$

Podobno je, kadar je $f(i)$ monotono padajoča funkcija

$$\int_{m}^{n+1} f(x)dx \le \sum_{i=m}^{n} f(i) \le \int_{m-1}^{n} f(x)dx.$$

1.6 Harmonična števila in naravni logaritem

Za pozitivno celo število n definirajmo *harmonično število* kot

$$H_n = \sum_{i=1}^{n} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n}.$$

V nadaljevanju bomo pokazali, da H_n narašča enako hitro kot $\ln n$, torej $H_n \sim \ln n$. Funkcija $\frac{1}{i}$ je monotono padajoča, zato bi jo lahko aproksimirali po zgornjem integralu, vendar se pojavi težava, ker integral ne konvergira – spodnja meja določenega integrala je 0, kar zahteva izračun $\ln 0$. V izogib težavi se raje lotimo aproksimacije funkcije

$$H'_n = H_n - 1 = \sum_{i=2}^{n} \frac{1}{i}.$$

Za H'_n velja naslednje

$$\begin{aligned}
\int_2^{n+1} \frac{dx}{x} \quad &\leq H'_n \leq \quad \int_1^n \frac{dx}{x} \\
\ln x \Big|_2^{n+1} \quad &\leq H'_n \leq \quad \ln x \Big|_1^n \\
\ln(n+1) - \ln 2 \quad &\leq H'_n \leq \quad \ln n - \ln 1 \\
\ln(n+1) - 1 \quad &\leq H'_n \leq \quad \ln n
\end{aligned}$$

in posledično

$$\ln(n+1) \leq H_n \leq \ln n + 1.$$