Algoritmer og datastrukturer - Avsnitt 5.2.15 - Algoritmeanalyse

Punkt 5.2.15.1 - Gjennomsnittlig nodedybde i binære søketrær med *n* verdier der alle er forskjellige

En node i et binærtre har en bestemt avstand fra rotnoden. Denne avstanden kalles nodens *dybde*. Summen av alle avstandene kalles treets *indre veilengde*. Treets *gjennomsnittlige nodedybde* er lik forholdet mellom indre veilengde og det antallet noder treet har.

La V_n være gjennomsnittlig indre veilengde for binære søketrær med n verdier. Gjennomsnittet tas over de $n!$ binære søketrærne vi får ved å lage ett tre for hver permutasjon av tallene fra 1 til n. Hvis $n = 0$ har vi et tomt tre og dermed er $V_0 = 0$.

La $n > 0$. Hver av verdiene fra 1 til n vil ha samme sannsynlighet for å ligge i rotnoden. La $V_n(k)$ være gjennomsnittlig indre veilengde for de binære søketrærne som har k, $1 \le k \le n$, som rotverdi. I et slikt tre vil de $k-1$ verdiene som er mindre enn k ligge i rotnodens venstre subtre, k i rotnoden og resten ($n-k$ verdier) i høyre subtre. Gjennomsnittlig indre veilengde for de to subtrærne blir da V_{k-1} og V_{n-k}. Men alle nodene i de to subtrærne får en avstand som er en mer hvis en ser det fra rotnoden. Dermed får vi følgende utrykk:

(1) $$V_n(k) = V_{k-1} + k-1 + V_{n-k} + n-k = V_{k-1} + V_{n-k} + n-1 .$$

Siden alle verdiene har samme sannsynlighet for å ligge i roten, får vi

(2) $$V_n = \frac{1}{n}(V_n(1) + V_n(2) + V_n(3) + \ . \ . \ . \ . + V_n(n)) .$$

Ved å bruke (1) i (2) får vi så :

(3) $$V_n = \frac{2}{n}(V_0 + V_1 + \ . \ . \ . \ . + V_{n-1}) + n-1 .$$

Vi ganger med n på begge sider av likhetstegnet i (3) og får

(4) $$n\,V_n = 2\,(V_0 + V_1 + \ . \ . \ . \ . + V_{n-1}) + n(n-1) .$$

Erstatt så n med $n-1$ i (4)

(5) $$(n-1)V_{n-1} = 2\,(V_0 + V_1 + \ . \ . \ . \ . + V_{n-2}) + (n-1)(n-2) .$$

Ved å ta differansen mellom (4) og (5) får vi

(6) $$n\,V_n - (n-1)V_{n-1} = 2\,V_{n-1} + 2(n-1) ,$$

og når ledd med V_{n-1} samles blir det

(7) $n\,V_n \;=\; 2(n-1) \;+\; (n+1)V_{n-1}$.

Nå deler vi med $n(n+1)$ på begge sider av likhetstegnet i (7)

(8) $$\frac{V_n}{n+1} \;=\; \frac{2(n-1)}{n(n+1)} + \frac{V_{n-1}}{n} \;=\; \frac{4}{n+1} - \frac{2}{n} + \frac{V_{n-1}}{n} \;.$$

Vi bruker (8) gjentatte ganger inntil $n = 1$ og deretter at $V_0 = 0$

(9) $$\frac{V_n}{n+1} \;=\; \frac{4}{n+1} - \frac{2}{n} + \frac{4}{n} - \frac{2}{n-1} + \;.\;.\;.\; + \frac{4}{2} - \frac{2}{1} \;.$$

Positive og negative ledd samles hver for seg

(10) $$\frac{V_n}{n+1} \;=\; 4\left(\frac{1}{n+1} + \frac{1}{n} + \;.\;.\;.\; + \frac{1}{2}\right) - 2\left(\frac{1}{n} + \frac{1}{n-1} + \;.\;.\;.\; + 1\right) \;.$$

Nå bruker vi at $H_n \;=\; \frac{1}{n} + \;.\;.\;.\; + \frac{1}{2} + 1$. Dermed

(11) $$\frac{V_n}{n+1} \;=\; \frac{4}{n+1} + 4(H_n - 1) - 2H_n) \;=\; \frac{4}{n+1} + 2H_n - 4 \,.$$

Ved å gange med $n+1$ på begge sider i (11) ender vi opp med

(12) $V_n \;=\; 2(n+1)H_n - 4n, \quad n > 0$.

La D_n være gjennomsnittlig nodedybde for binære søketrær med n verdier der alle er forskjellige. Gjennomsnittet blir det samme enten vi har n forskjellige verdier eller har verdiene fra 1 til n . Vi får derfor

(13) $$D_n \;=\; 2\left(1+\frac{1}{n}\right)H_n - 4\,, \quad n > 0 \,.$$