Algorytmy i Struktury Danych

Lista 4 - Zadanie 7

Kwiecień 2020

Problem

Czy algorytm SELECT powinien mieć polską nazwę "magiczne piątki"? Odpowiedź uzasadnij sprawdzając jaką złożoność miała by zmodyfikowana wersja SELECT'a, która w kroku wyszukiwania mediany median:

- dzieli tablice na $\lceil \frac{n}{3} \rceil$ trójek.
- dzieli tablice na $\lceil \frac{n}{7} \rceil$ siódemek.

Rozwiązanie

Z wykładu wiemy, że SELECT z podziałem na piątki, ma złożoność liniową. Pokażemy, że podział na trójki, nie ma takiej złożoności, a podział na siódemki, ma złożoność liniową. Zaczniemy od podziału na trójki.

Żeby wyznaczyć złożoność, najpierw wyznaczymy minimalną ilość elementów większych od mediany median x. Dla podziału na trójki, każda mediana ma swoją grupę, złożoną z trzech elementów. Wszystkie mediany grup będące na prawo od x są większe. Takich grup jest dokładnie $\lceil \frac{n}{3} \rceil$, z wyłączeniem ostatniej grupy, która nie jest podzielna przez 3 (jeżeli taka jest). Elementów w każdej takiej grupie, większej od x jest 2, co daje nam wzór na minimalną ilość elementów większych od x (z wyłączeniem grupy x):

$$2 \cdot \left(\left\lceil \frac{1}{2} \left\lceil \frac{n}{3} \right\rceil \right\rceil - 2 \right) \geqslant \frac{n}{3} - 4$$

Minimalną ilość elementów, można wyprowadzić w dokładnie taki sam sposób. Teraz spojrzymy na wariant pesymistyczny, w którym, szukana statystyka, znajdzie się w przedziale, mającym dokładnie $n - \left(\frac{n}{3} - 4\right)$ elementów. Stąd otrzymujemy wzór rekurencyjny:

$$T(n) = T\left(\left\lceil \frac{n}{3} \right\rceil\right) + T\left(\frac{2n}{3} + 4\right) + a \cdot n$$

Spróbujemy rozwinąć prawą stronę równania, używając podstawienia. Załóżmy, że $T(n) \leqslant cn$, dla $n \leqslant k$ i dostatecznie dużego c, otrzymamy:

$$T(n) \leqslant \frac{cn}{3} + c + \frac{2cn}{3} + 4c + an = cn + 5c + a \cdot n$$

Jest ograniczone z góry przez cn, jeśli $5c + a \cdot n \leqslant 0$, ale składnik sumy $a \cdot n$ dominuje (nie ma takiego c z założenia), przez co $5c + a \cdot n \geqslant 0$ od jakiegoś n, czyli $T(n)$ jest ograniczony z góry przez złożoność wyższą niż $O(n)$.

Spójrzmy teraz, na drugi przypadek, z podziałem na siódemki. Dla tak samych wykonanych kroków otrzymamy:

$$T(n) = T\left(\left\lceil \frac{n}{7} \right\rceil\right) + T\left(\frac{10n}{14} + 8\right) + a \cdot n$$

Gdzie przy tych samych założeniach:

$$T(n) \leqslant \frac{cn}{7} + c + \frac{10cn}{14} + 8c + an = \frac{12cn}{14} + 9c + a \cdot n = cn + (-\frac{2cn}{14} + 9c + a \cdot n)$$

Jest to ograniczone z góry przez cn, jeśli $(-\frac{2cn}{14} + 9c + a \cdot n) \leqslant 0$, co łatwo sprawdzić, dla tak zadanych założeń jest prawdziwe. Stąd otrzymujemy, że złożoność SELECT'a z podziałem na siódemki, jest o złożoności $O(n)$, tak jak i podział na piątki.

Rzeczywiście, warunek ten jest równoważny

c n - 63 c >= 7 a n

c (n - 63) >= 7 a n

c >= 7 a n/(n - 63)

Np. dla n >= 64 mamy n/(n-63) <= 64, więc wystarczy wziąć c >= 64*7 a