Wybór k-tego elementu

niedziela, 16 czerwca 2024 15:37

Zadanie

Dane: $T[1 \dots n]$
k - liczba z przedziału $\langle 1, n \rangle$

Wynik: k-ty co do wielkości element ciągu T

wszystkie elementy w ciągu różne

Szczególne przypadki:

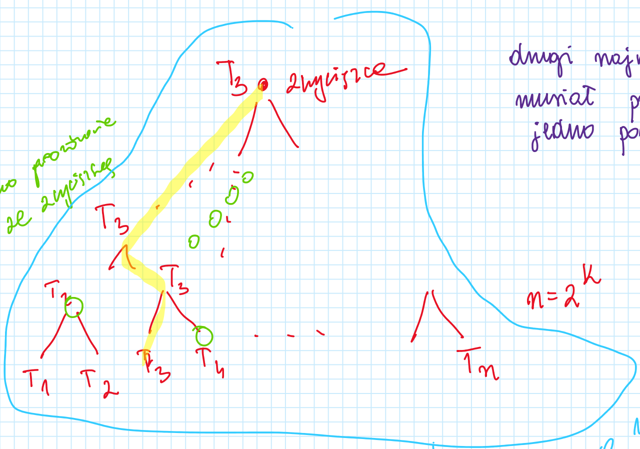
$k=1$

konieczna i wystarczająca liczba porównań to $n-1$

wszystkie elementy oprócz jednego muszą przegrać jakieś porównanie
potrzebujemy $n-1$ pojedynków by w każdym raz pokonało $n-1$ elementów

pierwszy element jest porównywany z drugim, jeśli drugi jest większy, to teraz z nim będziemy porównywać kolejne elementy; czyli będzie $n-1$ porównań

$k=2$ potrzeba i wystarczy $n-2+\lceil \log n \rceil$



log n

o przegrały jedno porównanie ze zwycięzcą

drugi największy element musiał przegrać DOKŁADNIE jedno porównanie

$n=2^k$

tych elementów jest tyle, co wysokość drzewa czyli log n

zatem spośród tych wierzchołków trzeba wybrać drugiego zwycięzcę

a na to potrzeba kolejnych $\lceil \log n \rceil - 1$ porównań

drzewo binarne o n liściach, zatem wierzchołków wewnętrznych jest $n-1$ = liczba porównań

→ zatem otrzymujemy $n-1+\lceil \log n \rceil - 1$

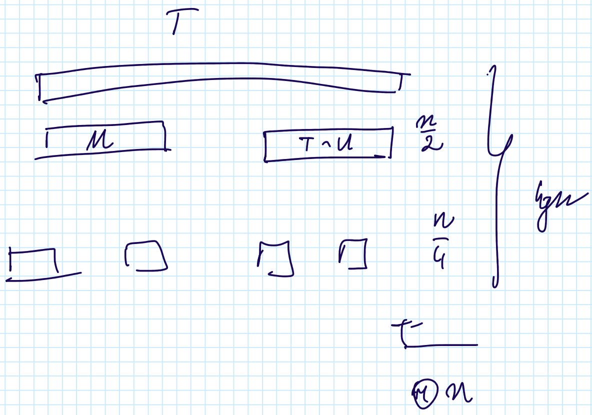
czyli $n-2+\lceil \log n \rceil$

to jest granica górna tyle wystarcza

Przypadek ogólny

Przypadek ogólny

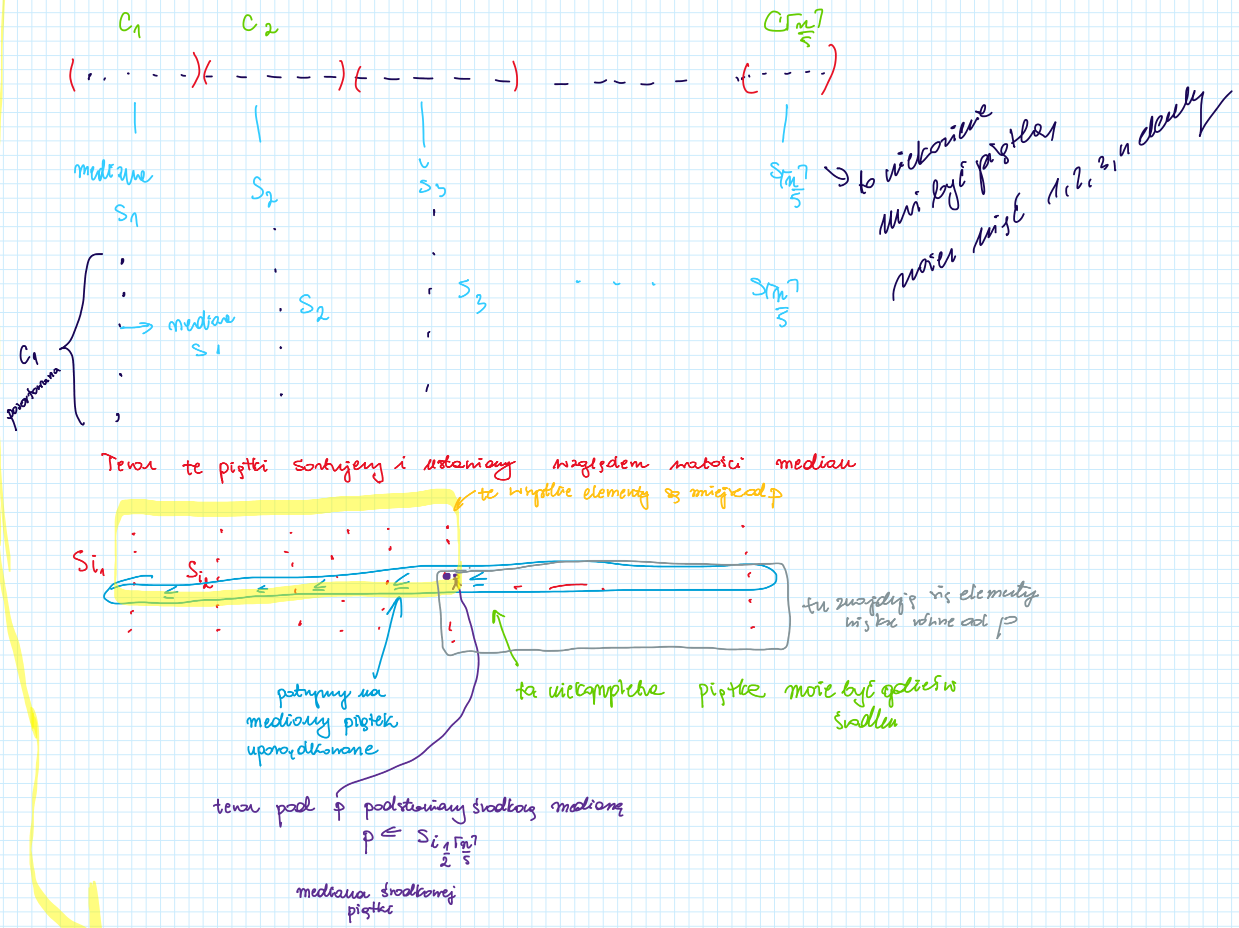
procedure SELECTION(k, T)
if $|T|$ mały to sort(T) & return T[k]
$p \leftarrow$ jakiś element z T
$U \leftarrow$ elementy z T mniejsze od p
if $k \le |U|$ then return (SELECTION(k, U))
else return (SELECTION($k-|U|$, $T \setminus U$))



Szukamy 20 elementu z zbiór
U ma 7 ⇒ szukamy potem 20 − 7 = 13
elementu w zbiorze T.

czas $\Theta(n)$

Algorytm magicznych piątek



function Pseudomed(T)

1. Podziel T na rozłączne podzbiory 5-elementowe C_j ($j = 1, \dots, \lceil |T|/5 \rceil$)
   {jeśli T nie dzieli się przez 5, to $C_{\lceil |T|/5 \rceil}$ zawiera mniej niż 5 elementów}
2. for $i \leftarrow 1$ to $\lceil |T|/5 \rceil$ do $s_i \leftarrow$ adhocmed(C_i) ← wyznacz medianę w 5 el. zbiorze
3. $S \leftarrow \{s_i \mid i = 1 \dots \lceil |T|/5 \rceil\}$
4. return (SELECTION($\lceil \frac{|S|}{2} \rceil$, S)) ← wybierz środkową medianę

function *Pseudomed(T)*

1. Podziel T na rozłączne podzbiory 5-cioelementowe C_j ($j = 1, \dots, \lceil |T|/5 \rceil$)
   {jeśli $|T|$ nie dzieli się przez 5, to $C_{\lceil |T|/5 \rceil}$ zawiera mniej niż 5 elementów }
2. **for** $i \leftarrow 1$ **to** $\lceil |T|/5 \rceil$ **do** $s_i \leftarrow adhocmed(C_i)$
3. $S \leftarrow \{s_i \mid i = 1, \dots, \lceil |T|/5 \rceil\}$
4. **return** $(SELECTION(\lceil \frac{|S|}{2} \rceil, S))$

$\lceil n \rceil$

4. **return** $(SELECTION(\lceil \frac{|S|}{2} \rceil, S))$

Zauważ, że jest $\lceil \frac{n}{5} \rceil$ median i mniejszych od p $\frac{1}{2}$, w każdej

piątce 3 elementy mniejsze : $3 \cdot \left(\frac{1}{2} \cdot \lceil \frac{n}{5} \rceil - 1\right) \geq \frac{3}{10} n - 3$ ← tyle jest elementów niemniejszych od p

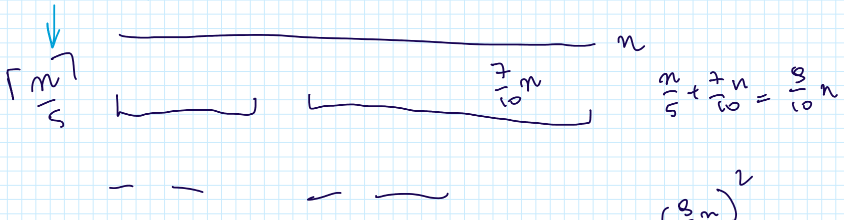
↓ być może poza 1 bo jedna piątka jest niepełna

↓ czyli $\frac{3}{10} n - 4$ jest mniejszych od p

tak samo $\frac{3}{10} n - 4$ elementów będzie większych od p

$$T(n) = T\left(\left\lceil \frac{n}{5} \right\rceil\right) + T\left(\frac{7}{10} n\right) + \Theta(n)$$

w ilu zbiorach się rekurencyjnie wywołamy



$$\frac{n}{5} + \frac{7}{10} n = \frac{9}{10} n$$

$$\left(\frac{9}{10} n\right)^2$$

$$O(\log n)$$

$$\underbrace{O\left(n\left(\frac{9}{10} + \left(\frac{9}{10}\right)^2 + \ldots\right)\right)}_{O(n)}$$

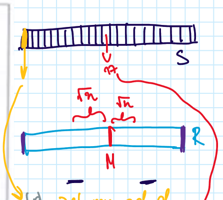
Zrandomizowany algorytm wyznaczania mediany

Dane wejściowe: zbiór S złożony z n elementów z dobrze uporządkowanej przestrzeni.

Wynik: mediana zbioru S oznaczona przez m.

1. Wybierz (multi)zbiór R złożony z $\lceil n^{3/4} \rceil$ elementów z S wybranych niezależnie i w sposób jednostajny z powtórzeniami.
2. Sortuj zbiór R.
3. Niech d będzie $(\lfloor \frac{1}{2} n^{3/4} - \sqrt{n} \rfloor)$-tym najmniejszym elementem w posortowanym zbiorze R.
4. Niech u będzie $(\lceil \frac{1}{2} n^{3/4} + \sqrt{n} \rceil)$-tym najmniejszym elementem w posortowanym zbiorze R.
5. Porównując każdy element w S z u i d, wyznacz zbiór $C = \{x \in S : d \leq x \leq u\}$ oraz liczby $\ell_d = |\{x \in S : x < d\}|$ i $\ell_u = |\{x \in S : x > u\}|$.
6. Jeśli $\ell_d > n/2$ lub $\ell_u > n/2$, to PORAŻKA.
7. Jeśli $|C| \leq 4n^{3/4}$, to sortuj zbiór C, a w przeciwnym wypadku PORAŻKA.
8. Wypisz element na pozycji $\lfloor n/2 \rfloor - \ell_d + 1$ w posortowanym zbiorze C.

Algorytm 3.1. Zrandomizowany algorytm wyznaczania mediany



$\ell_d \Rightarrow$ el. mn. od d

$\ell_u \Rightarrow$ el. większ od u

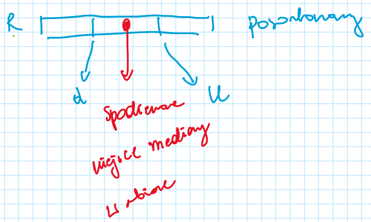
$C \Rightarrow$ el. między d i u

Lazy Select : Cel ⇒ działanie w $\Theta(n)$

n $\quad 4 \cdot n^{\frac{3}{4}} \cdot \log n \Rightarrow \Theta(n)$

1) Wybieramy losowo próbkę R z

2) sortujemy próbkę R (bo ma bardzo mało elementów, więc możemy to zrobić) czyli próbka może zawierać $\frac{n}{\log n}$ elem. bo sortowanie $\frac{n}{\log n} \cdot \log n = n$



Cel próbkowania: wyznaczenie d i u t.że prawd. że mediana jest znikome

$C = \{T[i] : d \leq T[i] \leq u\}$ jest mało

Cel próbkowania: wyznaczenie d i u t.że ppb braku mediany jest znikome

$C = \{T[i] \; d \le T[i] \le u\}$ jest mało

by C dało się posortować w czasie liniowym

by w C było wystarczająco dużo elementów

by do C wpadła mediana z T

miejscami $\frac{n}{\log n}$

Update: mam wybrany by w R było $n^{\frac{3}{4}}$ elementów z powtórzeniami

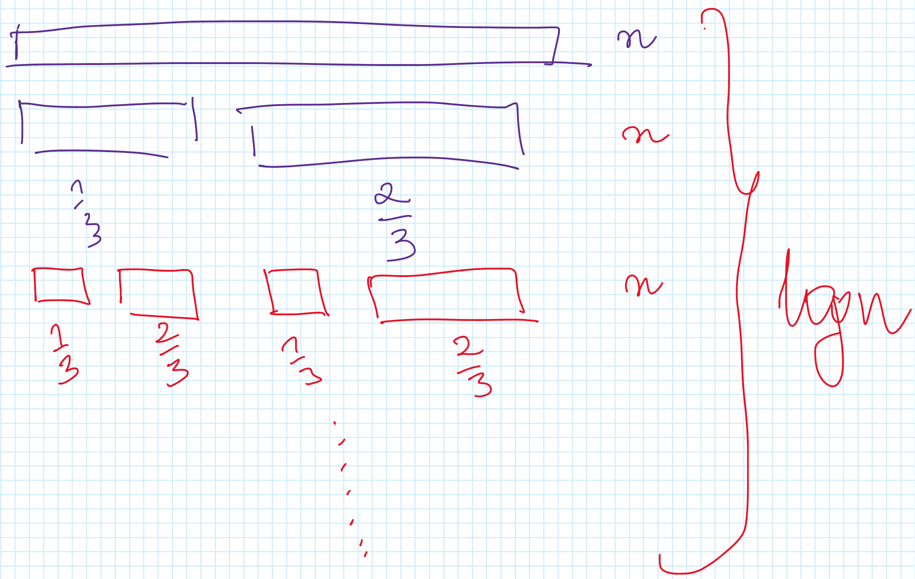
po posortowaniu oczekiwane miejsce mediany to $\frac{1}{2}$

pozycje małych liczb d i u wybieramy na $\sqrt{n}$ od spodziewanego miejsca mediany

(związek z odchyleniem standardowym)

Oczekiwany czas algorytmu: liniowy

Oczekiwana liczba prób: 2



$O(n \log n)$

$C = 1 \; 2 \; 3 \; 4 \; 5 \; 6 \; 7 \; 8 \; 9$

$H \Rightarrow \{5, \underset{d}{5}, 5, \underset{u}{5}, 5\}$

ld ⇒ 1 2 3 4

w ⇒ 6 7 8 9

c' ⇒ 5