Order-Statistics

Eingabe:

Eine Menge von n Schlüsseln aus einer geordneten Menge
Eine Zahl k, $1 \leq k \leq n$

Ausgabe:

Der k-te Schlüssel (nach Größe)

Spezialfall:

Median, der Schlüssel in der Mitte.

Einfachste Lösung:

1. Sortieren
2. Den Schlüssel an Postition k zurückgeben

Quickselect

Wie Quicksort, aber nur die richtige Seite rekursiv behandeln:

Algorithmus

procedure $quickselect(k, L, R)$:
  if $R \le L$ **then return** $a[k]$ **fi**;
  $p := a[L]; l := L; r := R + 1;$
  do
    do $l := l + 1$ **while** $a[l] < p$;
    do $r := r - 1$ **while** $p < a[r]$;
    vertausche $a[l]$ *und* $a[r]$
  while $l < r$;
  $a[L] := a[l]; a[l] := a[r]; a[r] := p;$
  if $k = r$ **then return** p
  else if $k < r$ **then return** $quickselect(k, L, r)$
  else return $quickselect(k, r, R)$ **fi**

Quickselect – Analyse

Bei Quicksort hatten wir diese Rekursionsgleichung für die Anzahl der Vergleiche:

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

Für Quickselect gilt:

- Mit W'keit $1/n$ ist das Pivotelement der gesuchte Schlüssel
- Mit W'keit $(k-1)/n$ ist der gesuchte Schlüssel links
- Mit W'keit $(n-k)/n$ ist der gesuchte Schlüssel rechts

Quickselect – Analyse

Für $n > 1$ haben wir:

$$\begin{aligned} C_n &= n + 1 + \frac{1}{n}\sum_{i=1}^{k-1} C_{n-i} + \frac{1}{n}\sum_{i=k+1}^{n} C_{i-1} \\ &\le n + 1 + \frac{2}{n}\sum_{i=\lceil n/2\rceil}^{n-1} C_i \end{aligned}$$

Außerdem ist $C_0 = C_1 = 0$.

Zeige mit Induktion:

$$C_n \le 4n$$

(Einfach $\longrightarrow$ Übungsaufgabe)

Quickselect

Theorem

Quickselect findet den Schlüssel mit Rang k in einer n-elementigen Menge in $O(n)$ Schritten – im Erwartungswert.

Ein deterministischer Algorithmus

1. Falls $n < 30$: Sortiere und finde so das Ergebnis.
2. Bilde $m = \lfloor n/5 \rfloor$ Gruppen mit je fünf Schlüsseln. Bis zu vier Schlüssel bleiben übrig.
3. Berechne den Median jeder Gruppe.
4. Berechne rekursiv den Median p dieser $\lfloor n/5 \rfloor$ Mediane.
5. Verwende p als Pivotelement und führe damit Quickselect aus.

Welchen Rang r hat p?

p ist der Median von $m = \lfloor n/5 \rfloor$ vielen kleinen Medianen
$\rightarrow$ mindestens $m/2 - 1$ kleine Mediane sind kleiner als p
Jeder kleine Median hat zwei Schlüssel die kleiner sind
$\rightarrow$ mindestens $3m/2 - 1$ kleinere Schlüssel als p

Ebenso: Mindestens $3m/2 - 1$ größere Schlüssel als p
Das sind jeweils $3\lfloor n/5 \rfloor/2 - 1 \geq 3n/10 - 3/2 \geq n/4$

Deterministisches Selektieren – Analyse

Die Anzahl der Vergleiche ist jetzt

$$C_n \leq n + 1 + C_{\lfloor n/5 \rfloor} + C_{\lfloor 3n/4 \rfloor}$$

falls $n > 30$ und $O(1)$ falls $n \leq 30$:

- $C_{\lfloor n/5 \rfloor}$ für das rekursive Finden der Mediane
- $C_{\lfloor 3n/4 \rfloor}$ für die nächste Suche

Es folgt $C_n = O(n)$, da $1/5 + 3/4 < 1$.

Wir können also den Schlüssel mit Rang k in linearer Zeit finden.