K8. Median-of-Medians 思想 für Lösung des Auswahlproblems in erwartet linear Zeit

Beobachtung: Wählt man Pivot-Element x in Linearzeit so, dass

$$\max\{|A_{<x}|, |A_{>x}|\} \leq \lfloor d \cdot |A| \rfloor$$

für eine konstante $d < 1$, so hat Auswahl (A, i) Laufzeit $O(n)$

Beweis: Für Laufzeitfunktion $f: \mathbb{N} \to \mathbb{N}$ gilt dann

$$f(n) \leq f(\lfloor d \cdot n \rfloor) + \tilde{c} \cdot n$$

$d \cdot n$: 子问题的大小

$f(\lfloor d \cdot n \rfloor)$: 递归调用的代价，即进入子问题之后的运行问题

$\tilde{c} \cdot n$: partion cost, 划分代价.

为什么 划分代价是 $\tilde{c} \cdot n$?

因为当我们选择一个 Pivot x 之后，需要把集合拆成三部分

$A_{<x}$, $A_{=x}$, $A_{>x}$

我们需要扫描整个数组，对每个元素做一次比较，判断它属于左边还是右边

即每个元素最多和 Pivot 比较一次

所以 代价是 $O(n)$

对于 $f(n)$ 的 Laufzeit，有两种证法：

方法一：

$$\begin{aligned} f(n) &\leq f(dn) + cn \\ &\leq (f(d^2 n) + c\cdot(dn)) + cn \quad \text{* 代入 } n = dn \\ &\leq (f(d^3 n) + c\cdot(d^2 n)) + c\cdot(dn) + cn \\ &\dots \\ &\leq f(d^k n) + cn\cdot(d^0 + d^1 + d^2 + \dots d^{k-1}) \end{aligned}$$

weil $d < 1 \Rightarrow d^0 + d^1 + \dots d^{k-1} = \sum_{j=0}^{\infty} d^j = \frac{1}{1-d}$

$\Rightarrow f(n) \leq cn \cdot \frac{1}{1-d} + f(d^k n)$

weil $d^k \to 0$, $k \to \infty$

$\Rightarrow f(n) \leq cn \cdot \frac{1}{1-d} + t$, $t \in \mathbb{R}$

$\Rightarrow f(n) \in O(n)$

方法2：

use verallgemeinerter Aufteilungs-Beschleunigungs-Satz

Es sei $f: \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ monoton wachsend, $a \in \mathbb{R}_{>1}$, $b, c \in \mathbb{R}_{>0}$, so dass $f(1) \leq \frac{c}{a}$, und die folgende Ungleichung gilt:

$$f(a \cdot n) \leq b \cdot f(n) + c \cdot n, \quad \text{für alle } n \geq 1$$

Dann gilt:

$$f(n) = \begin{cases} O(n) & , a > b \\ O(n \log n) & , a = b \\ O(n^{\log_a b}) & , a < b \end{cases}$$

对于这个问题：

$f: N \to N$, dann gilt $f(n) \leq f(\lfloor d \cdot n \rfloor) + \tilde{c} n$

我们把公式放一起比较

$$f(n) \leq f(\lfloor d \cdot n \rfloor) + \tilde{c} n$$
$$f(a \cdot n) \leq b \cdot f(n) + cn$$

$\Rightarrow b = 1 \quad \lfloor d \cdot n \rfloor \cdot a = an$

$\Rightarrow \quad a = \left\lceil \frac{1}{d} \right\rceil$

denn $d < 1 \Rightarrow a > 1$, $\Rightarrow a > b$

$\Rightarrow f(n) \in O(n)$.

$\boxed{\text{Satz:}}$ Wählt man in Algorithmus Auswahl (A, i) das Pivot-Element $x \in A$ zufällig und gleichverteilt, so ist erwartete Laufzeit $O(n)$

证明见课件