Билет 19. Медиана медиан

Основная идея: Способ выбора pivot, который гарантирует $O(n)$ для порядковой статистики и $O(n \log n)$ для быстрой сортировки в худшем случае.

Зачем это нужно?

Проблема: Обычный QuickSelect и QuickSort в худшем случае работают за $O(n^2)$

Решение: Алгоритм медианы медиан (Blum-Floyd-Pratt-Rivest-Tarjan) гарантирует:

- **Поиск k-ой статистики:** $O(n)$ в худшем случае
- **Быстрая сортировка:** $O(n \log n)$ в худшем случае

Алгоритм медианы медиан

Шаг 1: Разбить массив на группы по 5 элементов

Шаг 2: Найти медиану в каждой группе (сортировкой внутри группы)

Шаг 3: Создать массив из найденных медиан

Шаг 4: Рекурсивно найти медиану этого массива медиан

Шаг 5: Использовать найденную медиану медиан как pivot

Пример работы

Дан массив: [3, 1, 4, 1, 5, 9, 2, 6, 5, 3, 5, 8, 9, 7, 9]

```
Шаг 1: Разбиваем на пятерки:
[3, 1, 4, 1, 5]  → медиана: 3
[9, 2, 6, 5, 3]  → медиана: 5
[5, 8, 9, 7, 9]  → медиана: 8

Шаг 2: Массив медиан: [3, 5, 8]
Шаг 3: Медиана медиан: 5
Шаг 4: Используем 5 как pivot
```

Формальное доказательство

- Количество групп: $g = \lceil n/5 \rceil$
- Как минимум $\lceil g/2 \rceil$ групп имеют медианы $\geq x$
- В каждой такой группе 3 элемента $\geq$ её медианы $\geq x$
- Минимум элементов $\geq x$: $3 \cdot \lceil g/2 \rceil \geq 3 \cdot \lceil n/10 \rceil \geq 3n/10$
- Аналогично для элементов $\leq x$

Анализ времени работы

- Время на разделение массивка на пятерки и их сортировка: aN
- Время на поиск медианы медиан T(N/5)
- Время на поиск k-й порядковой статистики не превзойдет времени ее поиска в большей доле, т.е. T(7N/10)

Таким образом,

$$\mathrm{T(N) <= T(N/5) + T(7N/10) + aN}$$

Рекуррентное соотношение:

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

Доказательство линейности времени работы

Рекуррентное соотношение для алгоритма медианы медиан:

$$T(N) \le T\left(\frac{N}{5}\right) + T\left(\frac{7N}{10}\right) + aN$$

Базис индукции

Пусть $T(N) \le cN$ для некоторой константы c и любого $N \le 140$.

Индукционный шаг

$$T(N) \le T\left(\frac{N}{5}\right) + T\left(\frac{7N}{10}\right) + aN$$

$$\le c \cdot \frac{N}{5} + c \cdot \frac{7N}{10} + aN = \frac{9cN}{10} + aN$$

Условие выполнения неравенства

Нам нужно, чтобы:

$$\frac{9cN}{10} + aN \le cN$$

$$aN \le cN - \frac{9cN}{10} = \frac{cN}{10}$$

$$a \le \frac{c}{10} \quad \Rightarrow \quad c \ge 10a$$

Завершение доказательства

Таким образом, если выбрать $c = 10a$, то:

$$T(N) \le \frac{9cN}{10} + aN = \frac{9 \cdot 10a \cdot N}{10} + aN = 9aN + aN = 10aN = cN$$

Что и доказывает, что $T(N) = O(N)$.

Замечание о базисе

Константа 140 возникает из условия, что $\frac{N}{5}$ и $\frac{7N}{10}$ должны быть меньше N и достаточно малы для базиса. Для $N \leq 140$ время работы ограничено константой, так как алгоритм завершается за конечное число шагов.