

Алгоритм нахождения медианы медиан является оптимизацией выбора опорного элемента, позволяющей в худшем случае сортировать массив быстрой сортировкой за время $O(N \log N)$.

Алгоритм медианы медиан вычисляет приблизительную медиану, точку, которая гарантированно находится между 30-м и 70-м перцентилями. Таким образом, набор поиска уменьшается как минимум на 30 процентов. Применение этого алгоритма к уменьшенному набору рекурсивно продолжается до тех пор, пока не останется один или два элемента.

Алгоритм нахождения k -ой порядковой статистики через медиану медиан:

1. Все n элементов входного массива разбиваются на $\lfloor n/5 \rfloor$ групп по 5 элементов и одну группу, содержащую оставшиеся $n \bmod 5$ элементов.
2. Каждая из $\lfloor n/5 \rfloor$ групп сортируется методом сортировки вставкой, а затем в каждой отсортированной группе выбирается медиана.
3. Рекурсивно определяется медиана x множества из $\lfloor n/5 \rfloor$ медиан, найденных на шаге 2.
4. Массив разделяется относительно x с использованием операции partition.
5. Рекурсивно вызывается для подмассива, содержащего все элементы меньше x , и для подмассива, содержащего все элементы больше или равные x .

Лемма: Медиана медиан гарантированно делит массив в соотношении не хуже $3 : 7$.

Как минимум половина медиан, найденных на шаге 2, больше или равны медиане медиан x . Таким образом, как минимум $n/5$ групп содержат по 3 элемента, превышающих величину x . Приходим к выводу, что количество элементов, величина которых превышает x , равно как минимум $3N/10$.

Рекуррентное соотношение для времени работы алгоритма медианы медиан в наихудшем случае:

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

При помощи метода подстановки покажем, что время работы линейно зависит от количества входных элементов. Пусть $T(n) \leq nc$, где c — некоторая небольшая константа. Также пусть $O(n)$ ограничено сверху an . Тогда:

$$\begin{aligned}
T(n) &\leq c \cdot \frac{n}{5} + c \cdot \frac{7n}{10} + an \\
&\leq \frac{9cn}{10} + an \\
&\leq cn + (-cn/10 + an)
\end{aligned}$$

Следовательно, $T(n) \leq cn$ верно, если взять $c \geq 10a$. Поскольку такое c можно выбрать, алгоритм работает за линейное время.