

## ВЫБОР АЛГОРИТМА ПОИСКА МЕДИАНЫ ПРИ НЕБОЛЬШОЙ РАЗМЕРНОСТИ ЗАДАЧИ

к.т.н. А.В. Шостак

(представил д.ф.-м.н., проф. С.В. Смеляков)

*Приводится анализ алгоритмов поиска медианы при небольшой размерности задачи.*

**Введение.** Медианой для  $n$  чисел (пусть  $n$  нечетно) называется число, меньшее (или равное) половине из  $n$  чисел и большее (или равное) другой половине из  $n$  чисел, т.е. если  $n$  нечетно, то медиана – это порядковая статистика номер  $i = (n + 1)/2$  [1, 2].

Нахождение медианы является весьма распространенным видом вычислений при статистической обработке данных, а также при фильтрации изображений [1 – 4]. В последнем случае в качестве оценки значения свойства цвета в  $r$ -й точке изображения берется медиана среди  $n$  чисел, описывающих значения свойства цвета точек вокруг точки  $r$ . Особенностью медианной фильтрации является то, что величина  $n$  обычно не превышает 100, а числа, описывающие значения свойства цвета, как правило, принадлежат диапазону (0 ... 255), имеющему мощность  $k = 256$ . При этом естественным требованием к алгоритму поиска медианы является высокое быстродействие.

Очевидно, что ввиду небольшого  $n$  определяющей при выборе наиболее быстродействующего алгоритма поиска медианы становится мультипликативная константа, а не сложность алгоритма поиска.

**Основная часть.** Были рассмотрены алгоритмы поиска медианы с помощью алгоритмов сортировки сложности  $O(n^2)$  (сортировки простыми обменом, вставкой и выбором [2]), с помощью поразрядной сортировки [3] сложности  $O(n)$ , с помощью основанной на методе быстрой сортировки процедуры поиска медианы Select сложности  $O(n)$  [4] и с помощью основанной на идее сортировки подсчетом [1] процедуры поиска медианы Count сложности  $O(n)$ .

При  $n < 100$  и  $k = 256$  для неупорядоченной исходной последовательности лучшее быстродействие показали алгоритмы поиска медианы Select и Count. В качестве оценки быстродействия использовалась сумма

количества сравнений и присваиваний при выполнении алгоритма поиска медианы.

Элемент деления в алгоритме Select выбирался как медиана из первого, последнего и среднего элементов исходного массива. Наилучшее быстродействие алгоритм Select показал при полностью отсортированном (по убыванию или по возрастанию) исходном массиве. Наихудшее быстродействие – при равенстве всех элементов исходного массива. Поэтому данный алгоритм поиска медианы необходимо дополнить процедурой проверки равенства элементов исходного массива.

В табл. 1 приведена зависимость количества операций при поиске медианы алгоритмом Select от размерности массива (элементы массива расположены случайным образом, диапазон чисел (0..255), количество усреднений 1000).

Таблица 1  
Зависимость числа операций алгоритма Select от размерности массива

Размерность массива	9	19	29	39	49	59	69	79	89	99
Число операций	91	199	311	421	535	634	740	857	963	1090

Линия регрессии для зависимости суммы количества операций сравнения и присваивания как функция от размерности исходного массива  $n$  для алгоритма Select, построенная в соответствии с данными из табл. 1, имеет вид

$$f_s(n) = 10,98 \cdot n - 9,08.$$

В алгоритме Count на основании исходного массива  $A = \{a_1, \dots, a_n\}$  размерности  $n$  определяется массив  $C = \{c_1, \dots, c_k\}$  размерности  $k$ , причем  $i$ -й элемент этого массива равен количеству чисел  $i$ , находящихся в массиве  $A$ . Далее на основании массива  $C$  начинает рассчитываться новый массив  $C$  по правилу  $c[i] = c[i] + c[i - 1]$ , то есть теперь  $c[i]$  равно количеству элементов исходного массива  $A$ , не превосходящих значения  $i$ . Поиск медианы состоит в определении индекса первого элемента нового массива  $C$ , который больше или равен  $(n + 1)/2$  (после этого формирование массива  $C$  прекращается).

В алгоритме Count количество присваиваний и сравнений практически не зависит от упорядоченности исходного массива (рассматривались упорядоченные по убыванию и по возрастанию массивы, массивы из равных элементов и массивы со случайным расположением элементов). Лучший по быстродействию случай алгоритма Count состоит в том, что

медиана является первым числом диапазона чисел, составляющих исходный массив, то есть исходный массив состоит из более чем  $(n + 1)/2$  таких чисел. В лучшем случае требуется  $(n + 2)$  присваивания и одно сравнение (сумма присваиваний и сравнений равна  $(n + 3)$ ). Худший случай состоит в том, что медиана является последним числом диапазона чисел и исходный массив состоит только из таких чисел. В этом случае при поиске медианы требуется  $(n + k + 1)$  присваиваний и  $k$  сравнений, т.е. всего  $(n + 2k + 1)$  операций. В среднем случае для поиска медианы с помощью алгоритма Count требуется  $(n + 0,5k + 1)$  присваиваний и  $0,5k$  сравнений, т.е. всего  $(n + k + 1)$  операций.

В табл. 2 приведена зависимость количества операций при поиске медианы алгоритмом Count от размерности массива (элементы массива неупорядочены, диапазон чисел (0..255), количество усреднений 100).

Таблица 2

Зависимость числа операций алгоритма Count от размерности массива

Размерность массива	9	19	29	39	49	59	69	79	89	99
Число операций	236	266	276	282	298	310	324	330	346	352

Линия регрессии для зависимости суммы количества операций сравнения и присваивания как функция от размерности исходного массива  $n$  для алгоритма Count, построенная в соответствии с данными из табл. 2, имеет вид  $f_c(n) = 1,22 \cdot n + 236,15$ .

На рис. 1 приведены графики функций  $f_s(n)$  и  $f_c(n)$ .

При небольшой размерности массива  $A$  и большой размерности массива  $C$  мощность диапазона чисел  $k$  начинает играть основную роль в количестве операций для поиска медианы алгоритмом Count, из этого следует необходимость поиска методов снижения размерности массива  $C$ , т.е. определения реального диапазона чисел массива  $A$ , а не возможного.

В табл. 3 приведена зависимость количества операций при поиске медианы алгоритмом Count от мощности диапазона чисел  $k$  (элементы массива неупорядочены, размерность массива  $A$  равна 49, количество усреднений 100).

Таблица 3

Зависимость числа операций алгоритма Count от  $k$

Мощность диапазона чисел	11	41	71	101	201	256
Число операций	58	88	118	148	248	298

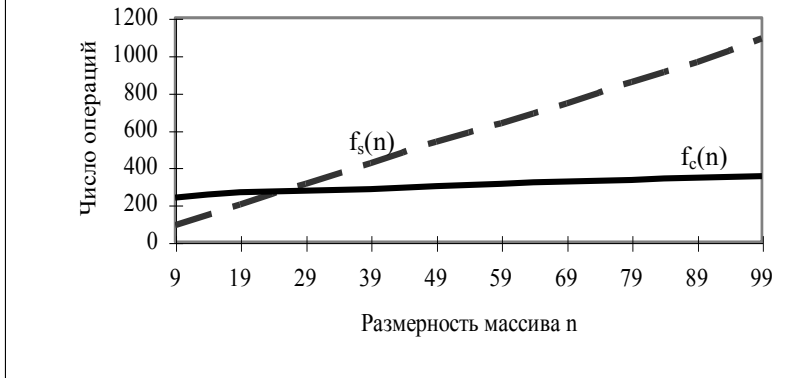


Рис. 1. Зависимость числа операций при поиске медианы с помощью алгоритмов Select и Count от размерности массива

случайного расположения элементов исходного массива при  $n > 25$  алгоритм поиска медианы Count выполняется за меньшее количество операций по сравнению с алгоритмом Select. А при  $n < 25$  более быстрым является алгоритм Select.

#### ЛИТЕРАТУРА

1. Кормен Т., Лейзерсон Ч., Ривест Р. Алгоритмы: построение и анализ. — М.: МЦНМО, 2000. — 960 с.
2. Вирт Н. Алгоритмы и структуры данных. — М.: Мир, 1989. — 360 с.
3. Кнут Д. Искусство программирования для ЭВМ. Т. 3. Сортировка и поиск. — М.: Мир, 1979. — 844 с.
4. Ахо А., Хопкрофт Дж., Ульман Дж. Структуры данных и алгоритмы. — М.: Вильямс, 2000. — 382 с.

Поступила 28.08.2003

**ШОСТАК Анатолий Васильевич**, канд. техн. наук, доцент кафедры ХВУ. Окончил в 1981 году ХАИ. Область научных интересов — синтез топологических структур сетей.