Тураев Тимур, 504 (SE)
24.09.2013

3.4* По двум отсортированным массивам длины n и m найти k-ый по величине элемент в их объединении.

Пусть наши массивы называются a и b и пусть они нумеруются с единицы. Тогда номер элемента равен его индексу в массиве.

Рассмотрим два элемента a[i] и b[j]. Пусть a[i] > b[j].

Тогда несложно отметить, что a[i] может находиться не ближе, чем на месте (i+j) в объединении массивов, потому что он не меньше всех своих предшествующих элементов a[1..i-1] (таких уже i-1 штук) и не меньше всех элементов b[1..j] (таких j штук). Значит, слева от a[i] (i+j-1) элемент и поэтому он занимает (i+j)-е место.

Рассмотрим элемент b[j]. Рассуждая аналогично, можно прийти к выводу, что этот элемент находится не дальше чем на месте (i+j-1) в объединении массивов, потому что все элементы из a[i+1..n] и из b[j+1..m] – не меньше чем b[j].

То есть мы для каждой пары элементов можем определить «крайнее левое и крайнее правое» возможное положение в массиве для соответствующих элементов.

Далее, если вдруг k < i+j при тех же условиях (a[i] < b[j]), то очевидно, что k-ый элемент это не a[i] и никакой после a[i]. Значит, можем отсечь весь хвост a после i-го элемента.

А если k >= i+j, то искомый k-ый элемент – не b[j] и никакой до b[j]. Значит, мы можем отсечь все начало второго массива. Эти два утверждения прямо следуют из предыдущих замечаний.

Алгоритм становится прост – смотрим на два средних элемента массива, смотрим какой больше и отсекаем соответствующую половину массива (причем, если отсекаем нижнюю половину второго, то не забудем уменьшить k на длину этой отсеченной части).

Действуем до тех пор, пока один из массивов не станет нулевой длины – и тогда просто выводим k-й элемент второго массива.

Алгоритм своим поведением жутко похож на обычный бинарный поиск, только тут поиск «одновременно» в двух массивах, откуда «угадывается» сложность.

Она – очевидно логарифм суммарной длины, т.к. на каждом этапе мы уменьшаем длину одного из массивов в 2 раза. В случае одинаковых длин, получается рекуррентное соотношение T(n) = T(3/4 * n) + 1, откуда T(n) = O(log n), причем асимптотика не меняется при неодинаковых длинах.

А O(log(n+m)) = O(log(n*m)) = O(log n + log m)