

УДК 004

МЕТОДИКА ИССЛЕДОВАНИЯ РЕКУРСИВНЫХ АЛГОРИТМОВ
СОРТИРОВКИ

Е. А. ВАСИЛЬКОВ

Научный руководитель И. А. БЕККЕР

Белорусско-Российский университет

Могилев, Беларусь

Рассмотрим методику исследования рекурсивных алгоритмов сортировки на примере сортировки слиянием (Merge Sort). Исследование алгоритма состоит из двух частей: аналитической и эмпирической.

Merge Sort – сортировка типа «разделяй и властвуй». Каждый раз массив делится пополам, что дает $\log_2 n$ операций. Поскольку логарифмы с разными основаниями отличаются на константу, основание логарифма в асимптотическом анализе опускают.

Затем разделённые фрагменты последовательности «сливаются», на что тратится порядка n операций. Итоговая сложность алгоритма выходит $O(n \log(n))$. Алгоритм является стабильным, поскольку каждый раз сравнивается элемент из левого и из правого массива, следовательно, при равенстве значений порядок элементов останется без изменения. Сложность алгоритма не зависит от порядка входных данных, поскольку каждый раз массив делится пополам. Есть алгоритмы, сложность которых зависит от порядка входных данных. Сложность таких алгоритмов рассматривается в трёх случаях: при практически отсортированных данных, при случайных и при реверсивно отсортированных данных.

Для слияния элементов требуется дополнительный массив такого же размера, как первоначальный, следовательно, требования по памяти (емкостная сложность) – $O(n)$.

Для получения эмпирических результатов можно использовать два метода. Первый метод заключается в том, чтобы при разных значениях размера массива измерять время выполнения программы. Очень важно брать большие размеры массивов, поскольку если ограничиться малыми, то результат сильно исказится, большая часть ресурсов будет затрачена на запуск программы и выделения памяти. Второй способ: встроить счётчики сравнения, копирования и присваивания. Второй метод хорош тем, что полученные таким образом данные не зависят от фоновых процессов. После получения численных значений используются статистические методы для аппроксимации функции. Например, метод наименьших квадратов (МНК), суть которого заключается в минимизации суммы квадратов отклонений значений от значений искомой функции. После находится коэффициент корреляции. Если значение больше 0,9, то можно сделать вывод о том, что аппроксимирующая кривая хорошо описывает функцию трудоёмкости.

Имея как аналитические, так и эмпирические результаты, можно обоснованно судить о применимости алгоритма в различных условиях.