

The cost of trying too hard: Splay trees

 devblogs.microsoft.com/oldnewthing/20060118-16

January 18, 2006



Raymond Chen

Often, it doesn't pay off to be too clever. Back in the 1980's, I'm told the file system group was working out what in-memory data structures to use to represent the contents of a directory so that looking up a file by name was fast. One of the experiments they tried was the splay tree. Splay trees were developed in 1985 by Sleator and Tarjan as a form of self-rebalancing tree that provides $O(\log n)$ amortized cost for locating an item in the tree, where n is the number of items in the tree. (Amortized costing means roughly that the cost of M operations is $O(M \log n)$. The cost of an individual operation is $O(\log n)$ on average, but an individual operation can be very expensive as long as it's made-up for by previous operations that came in "under budget".) If you're familiar with splay trees you may already see what's about to happen. A very common operation in a directory is enumerating and opening every file in it, say, because you're performing a content search through all the files in the directory or because you're building a preview window. Unfortunately, when you sequentially access all the elements in a splay tree in order, this leaves the tree totally unbalanced. If you enumerate all the files in the directory and open each one, the result is a linear linked list sorted in reverse order. Locating the first file in the directory becomes an $O(n)$ operation. From a purely algorithmic analysis point of view, the $O(n)$ behavior of that file open operation is not a point of concern. After all, in order to get to this point, you had to perform n operations to begin with, so that very expensive operation was already "paid for" by the large number of earlier operations. However, in practice, people don't like it when the cost of an operation varies so widely from use to use. If you arrive at a client's office five minutes early for a month and then show up 90 minutes late one day, your explanation of "Well, I was early for so much, I'm actually still ahead of schedule according to amortized costing," your client will probably not be very impressed. The moral of the story: Sometimes trying too hard doesn't work.

(Postscript: Yes, there have been recent research results that soften the worst-case single-operation whammy of splay trees, but these results weren't available in the 1980's. Also, remember that consistency in access time is important.)

[Raymond Chen](#)

Follow

