AVERAGE SEARCH AND UPDATE COSTS IN SKIP LISTS*

THOMAS PAPADAKIS, J. IAN MUNRO and PATRICIO V. POBLETE

Abstract.

Skip lists, introduced by Pugh, provide an alternative to search trees, although a precise analysis of their behaviour had been elusive. The exact value of the expected cost for the search of the mth element in a skip list of n elements is derived first in terms of previously studied functions, and secondly as an asymptotic expression. The latter suggests that Pugh's upper bound of the expected search cost is fairly tight for the interesting cases. Assuming a uniform query distribution, the exact and an asymptotic value of the average (over all m) expected search cost in a skip list of n elements is also derived. Finally, all insert and delete costs are obtained.

CR classification: E.1, F.2.2.
AMS classification: 68E05.
Keywords: skip lists, dictionary, probabilistic algorithms, asymptotic analysis.

1. Introduction.

The skip list, recently introduced by W. Pugh [10], is an interesting and practical alternative to search trees. The approach is to store each of its n (n ≥ 1) elements in one or more of a set of sorted linear linked lists. All elements are stored in sorted order in a linked list denoted as level 1, and each element in the linked list at level i (i = 1, 2, . . .) is included with (independent) probability, p (0 < p < 1) in the linked list at level i + 1. A header contains the references to the first element in each linked list. The height of the data structure, that is, the number of linked lists, is also stored.

As will become apparent later, the number of linear linked lists which an element belongs to remains fixed as long as the element exists in the skip list. It therefore

* This research was supported in part by the Natural Science and Engineering Research Council of Canada under grant No. A-8237, the Information Technology Research Centre of Ontario, and FONDECYT (Chile) under grant 91-1252.
makes sense to store each element in the skip list once, together with an array of horizontal pointers, as shown in Figure 1.

A search for an element begins at the header of the highest numbered linked list. This linked list is scanned, until it is observed that its next element is greater than or equal to the one sought (or the reference is null). At that point, the search continues one level below until it terminates at level 1 (see the search for the 6th element in Figure 1). We have adopted the convention that an equality test is done only at level 1 as the last comparison. This is the usual choice in a standard binary search and avoids two tests (or a three-way branch) at each step.

The search cost is defined as the number of pointer inspections excluding the last one for the equality test. For the search in Figure 1, this is 9. Of principal interest is $C_p(m, n)$ ($m = 1, 2, \ldots, n + 1$), which, assuming that the 0th element is $-\infty$ and the $(n + 1)$st element is $+\infty$, is defined as the expected cost when searching successfully for the $m$th element, or unsuccessfully for an element between the $(m - 1)$st and the $m$th in a skip list of $n$ elements. This expected value is taken over all possible skip lists, for fixed $m$, $n$ and $p$.

It is worthy of note at this early stage that the expected search cost will depend on the relation between the element sought and the other elements in the skip list. The expected cost of a search for the element that turns out to be the $m$th smallest, will be an increasing function of $m$. In particular, searching for the smallest element, or one less than the smallest, will have cost equal to the height of the structure, while searching for other elements will have the additional cost of following links horizontally. It is clear that the height of a skip list is a lower bound on the search cost for any element. The spirit of the structure is that behaviour be "reasonable" for any values of $m$ and $n$; this suggests that $p$ be a probability of moderate size, say $\frac{1}{2}$ or $\frac{1}{3}$.

A summary statistic of interest is the average expected cost for a successful search, $C_p(\ast, n)$. This average is taken over all $C_p(m, n)$, for $m = 1, 2, \ldots, n$, assuming a uniform distribution of them. The average expected cost for an unsuccessful search, $C'_p(\ast, n)$, is defined similarly over all $n + 1$ expected costs of the $n + 1$ possible unsuccessful searches. These two costs differ slightly, as the latter includes a search for a value greater than the last element in the skip list.

---

1 Note that our search cost is 1 greater than the search cost analyzed by Pugh. Our references to his results are translated into our terms.