Skip Lists

Andreas Klappenecker

Texas A&M University

Goal

Consider a set

$$S = \{x_1 < x_2 < \ldots < x_n\}$$

from a totally ordered universe. This set can dynamically change by adding or removing elements. Our goal is to search S for an element k.

Searching

Goal

Consider a set

$$S = \{x_1 < x_2 < \ldots < x_n\}$$

from a totally ordered universe. This set can dynamically change by adding or removing elements. Our goal is to search S for an element k.

Bottom and Top Elements

We add a bottom element $-\infty$ and top element $+\infty$ to the universe such that

$$-\infty < x_1 < x_2 < \ldots < x_n < +\infty.$$

These elements can simplify the implementation of the search.

Linked List Representation

Implementation

We can represent the set S by an ordered linked list. The problem is that we cannot index into this list, so the search is slow.

Search Trees

A search tree can speed up the search, but can be a bit awkward to maintain under insert and delete operations.

Idea Behind Skip Lists

We want to obtain the speed of a binary search tree but combine it with the ease of maintaining a sorted linked list.

Skip lists were invented by Bill Pugh in 1990.

They offer an expected search time of $O(\log n)$.

They generalize linked lists and are easy to implement.

Structured Search using Filtrations

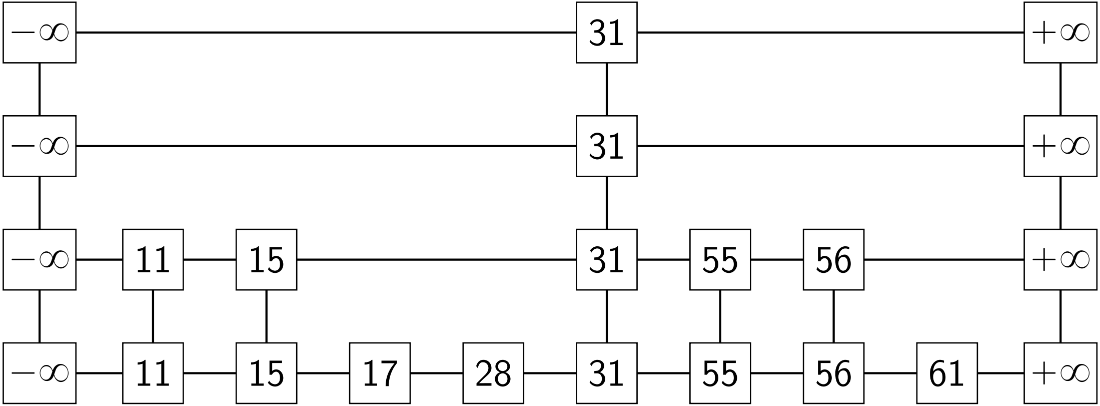
A **descending filtration** is a sequence S_i of subsets of S such that

$$\varnothing = S_r \subseteq S_{r-1} \subseteq \cdots \subseteq S_1 = S.$$

In computer science, the S_i are called levels. The idea is that S_k for a large k is easy to search, since it has fewer elements than S_1.

The idea is that we implement each S_i by a sorted linked list. Each element x in S_i is also linked to the element x in the finer level S_{i-1}.

Example


$-\infty$
31
$+\infty$
$-\infty$
31
$+\infty$
$-\infty$
11
15
31
55
56
$+\infty$
$-\infty$
11
15
17
28
31
55
56
61
$+\infty$

Search

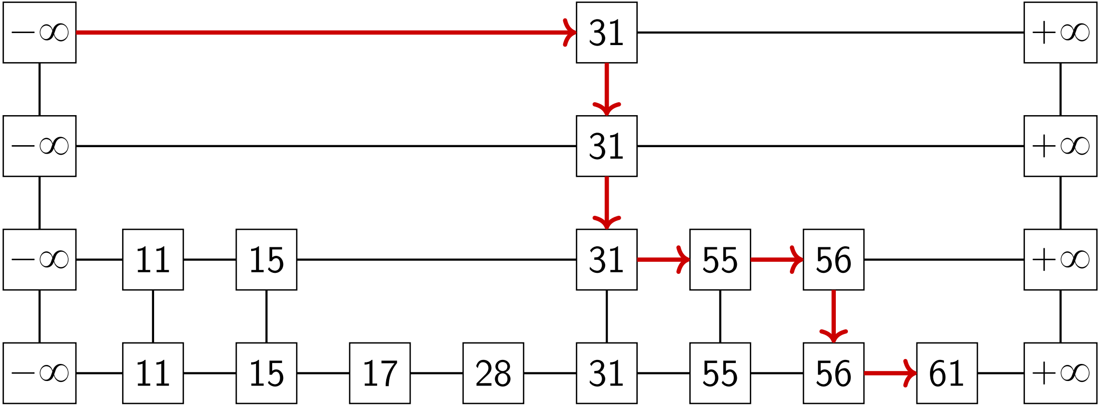
When we search for k:

If $k = \text{key}$, done!

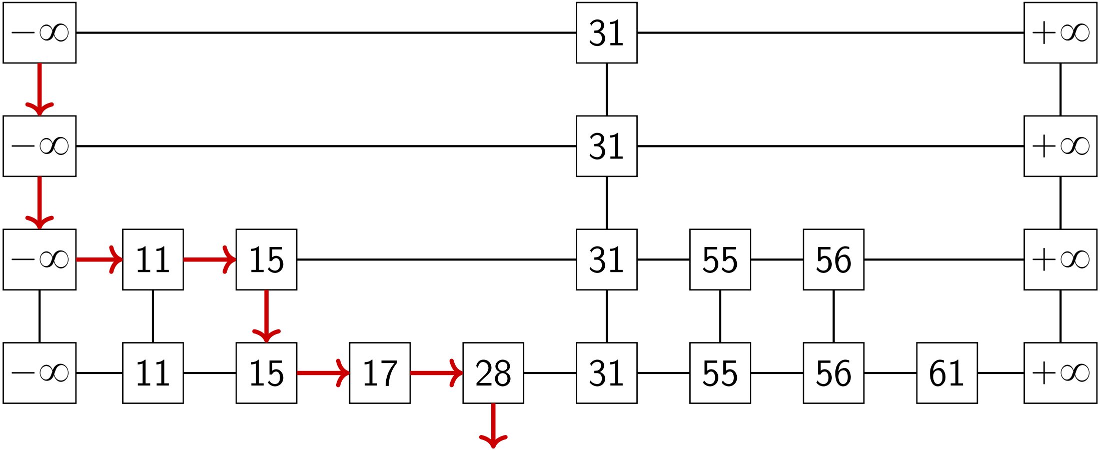
If $k < \text{next key}$, then k is not in this list, so go down a level

If $k \geqslant \text{next key}$, then go right

Example: Search for 61



Example: Search for 29



Deterministic Construction

Construction

If the set $S_1 = S$ is fixed, then we could choose to include every other element into S_2. Next, put every other element of S_2 into S_3, and so forth.

Problem

We want to be able to insert and delete elements. These operations destroy the nice structure!

Randomized Construction

Construction

Let $S_1 = S$. For every element x in S_k, include x in S_{k+1} with probability $1/2$.

Expected Number of Elements

$$\begin{aligned} \mathrm{E}[|S_1|] &= n, \\ \mathrm{E}[|S_2|] &= n/2, \\ \mathrm{E}[|S_3|] &= n/4, \\ &\vdots \end{aligned}$$

Height of an Element

We say that an element x_k has **height** ℓ if and only if

$$x_k \in S_\ell, \quad \text{but} \quad x_k \notin S_{\ell+1}.$$

Let X_k be the random variable that gives the height of the element x_k. We have

$$\Pr[X_k = \ell] = p(1-p)^{\ell-1}.$$

So for $p = 1/2$, we have

$$\Pr[X_k = \ell] = (1/2)^\ell = 2^{-\ell}.$$

Interlude: Jensen's Inequality

Proposition (Jensen's Inequality)

Let X be a random variable with $\mathrm{E}[X] < \infty$. If $f : \mathbf{R} \to \mathbf{R}$ is a convex function, so $\smile$, then

$$f(\mathrm{E}[X]) \leqslant \mathrm{E}[f(X)].$$

Jensen's Inequality for Convex Functions

Proposition (Jensen's Inequality)

Let X be a random variable with $\mathrm{E}[X] < \infty$. If $f : \mathbf{R} \to \mathbf{R}$ is a convex function, so $\smile$, then

$$f(\mathrm{E}[X]) \leqslant \mathrm{E}[f(X)].$$

Proof.

Since f is convex, we can find a linear function $g(x) = ax + b$ which lies entirely below the graph of f, but touches f at $\mathrm{E}[X]$. In other words, we can choose real numbers a and b such that

$$f(\mathrm{E}[X]) = g(\mathrm{E}[X])$$

and $g(x) \leqslant f(x)$ for all $x \in \mathbf{R}$.

Proof. (Continued).

Since $f(x) \geqslant g(x)$ for all $x \in \mathbf{R}$, it follows that

$$
\begin{aligned}
\mathrm{E}[f(X)] &\geqslant \mathrm{E}[g(X)] \\
&= \mathrm{E}[aX + b] = a\mathrm{E}[X] + b \\
&= g(\mathrm{E}[X]) = f(\mathrm{E}[X]).
\end{aligned}
$$

$\square$

Jensen's Inequality for Concave Functions

Proposition (Jensen's Inequality)

Let X be a random variable with $\mathrm{E}[X] < \infty$. If $f : \mathbf{R} \to \mathbf{R}$ is a concave function, so $\frown$, then

$$\mathrm{E}[f(X)] \leqslant f(\mathrm{E}[X]).$$

Proof.

If f is concave, then $-f$ is convex. So

$$-f(\mathrm{E}[X]) \leqslant \mathrm{E}[-f(X)] = -\mathrm{E}[f(X)]$$

by Jensen's inequality for convex functions. Thus,

$$f(\mathrm{E}[X]) \geqslant \mathrm{E}[f(X)]. \qquad \square$$

Back to Skip Lists

Expected Height

Proposition

The expected maximum height of a skip list with n elements is given by

$$\mathrm{E}\left[\max_{1 \leqslant k \leqslant n} X_k\right] \in O(\log n).$$

Proof.

Let α be a real number in the range $1 < \alpha < 2$. Then

$$\begin{aligned} \mathrm{E}\left[\max_{1\leqslant k\leqslant n} X_k\right] &\leqslant \log_\alpha \mathrm{E}\left[\alpha^{\max_{1\leqslant k\leqslant n} X_k}\right] \\ &= \log_\alpha \mathrm{E}\left[\max_{1\leqslant k\leqslant n} \alpha^{X_k}\right]. \end{aligned}$$

Since $\alpha^{X_k} \geqslant 1$, we can estimate the right-hand side by the sum

$$\mathrm{E}\left[\max_{1\leqslant k\leqslant n} X_k\right] \leqslant \log_\alpha \mathrm{E}\left[\sum_{k=1}^{n} \alpha^{X_k}\right].$$

Continued.

$$\mathrm{E}\left[\max_{1\leqslant k\leqslant n} X_k\right] \leqslant \log_\alpha \mathrm{E}\left[\sum_{k=1}^{n} \alpha^{X_k}\right] = \log_\alpha \left(\sum_{k=1}^{n}\sum_{k\geqslant 1} \alpha^k 2^{-k}\right)$$
$$= \log_\alpha \left(\sum_{k=1}^{n} \frac{1}{1-\alpha/2}\right)$$
$$= \log_\alpha n + \log_\alpha \frac{1}{1-\alpha/2} = O(\log n),$$

which is what we wanted to show. □

Proposition

The number of levels of a skip list of a set with n elements satisfies $O(\log n)$ with high probability.

Proof.

Let X_k denote the random variable giving the number of levels of the k-the element of S. Then

$$\Pr[X_k > t] \leqslant (1-p)^t.$$

So

$$\Pr[\max_k X_k > t] \leqslant n(1-p)^t = \frac{n}{2^t}$$

for $p = 1/2$. Choosing $t = a \log n$ and $r = \max_k X_k$, we can conclude that

$$\Pr[r > a \log n] \leqslant \frac{1}{n^{a-1}}$$

for any $a > 1$. $\square$