Skip Lists

William Pugh (1989)

- Easy to implement (as compared to balanced trees)
- Maintains a dynamic set of n elements in $O(\log n)$ time per operation in expectation and with high probability (w.h.p.)

One Linked List

One (Sorted) linked list

14 → 23 → 34 → 42 → 50 → 59 → 66 → 72 → 79

Searches take $\Theta(n)$ time in worst case

Suppose we had two sorted linked lists

- each element can appear in one or both lists
Two Linked Lists

Express and local subway lines
(Ca la New York City 7th Avenue Line)
- Express line connects a few of the stations
- Local line connects all stations
- Links between lines at common stations

Searching in Two Linked Lists

Search (x):
- Walk right in top linked list (L1) until going right would go too far
- Walk down to bottom linked list (L2)
- Walk right in L2 until element found (or not)

Search (59)
Analysis

Search cost $\approx |L_1| + \frac{|L_2|}{|L_1|}$

Minimized when terms are equal

$|L_1|^2 = |L_2| = n$

$|L_1| = \sqrt{n}$ Search is $O(\sqrt{n})$

More Linked Lists

2 sorted lists $\Rightarrow a \sqrt{n}$

3 sorted lists $\Rightarrow 3.3\sqrt{n}$

k sorted lists $\Rightarrow k \cdot k\sqrt{n}$

$\lg n$ sorted lists $\Rightarrow \lg n \cdot \sqrt{n}$

$= 2\lg n$

like a binary tree!
Searching in Ign Linked Lists

Try search(72)

Insert (x)

To insert an element x into a skip list
- Search(x) to see where x fits into bottom list
- Always insert into bottom list
- Insert into some of the lists above which ones?
- Flip fair coin
  - if HEADS: promote x to next level up
  - else stop
WHY ARE SKIP LISTS GOOD?

Warmup Lemma: # levels in n-element skip list is $O(\lg n)$ w.h.p.

$c \cdot \lg n \rightarrow \text{prob } 1 - \frac{1}{n^\alpha}$

related

Proof: Failure probability (not $\leq c \lg n$ levels)

\[= \Pr \{ > c \lg n \text{ levels}\} \]
\[= \Pr \{ \text{some element got promoted } > c \lg n \text{ times}\} \]
\[< n \cdot \Pr \{ \text{element x got promoted } > c \lg n \text{ times}\} \]
\[= n \cdot (\frac{1}{2})^{c \lg n} \text{ by union bound} \]
\[= \frac{n^c}{n} = \frac{1}{n^\alpha} \]
\[\alpha = c - 1 \]
Theorem: Any search in an $n$-element skip list costs $O(\log n)$ w.h.p.

Cool idea: Analyze search backwards

- **Search starts** [ends] at node in bottom list.

- At each node visited:
  - If node wasn't promoted higher (tails here)
    - then we go [came from] left
  - If node was promoted higher (heads here)
    - then we go [came from] up

- Stop [start] when we reach top level or $-\infty$

Look at $\leftarrow \uparrow$ arrows on page 4
Proof of Theorem (backwards)

- Search makes "up" and "left" moves each with probability $1/2$
- Number of moves going "up" $< \# \text{levels} \leq c \cdot \log n \text{ w.h.p.} \quad \text{(by Warmup Lemma)}$

- Total number of moves = number of coin flips until you get $c \log n$ heads ("up" moves)

Claim: Number of coin flips until $c \log n$ heads $= \mathcal{O}(\log n) \text{ w.h.p.}$
**CHERNOFF BOUNDS**

Theorem: Let $Y$ be a random variable representing the total number of heads in a series of $m$ independent coin flips, where each flip has a probability $p$ of coming up heads.

Then for all $r > 0$, we have

$$\Pr [Y \geq E[Y] + r] \leq e^{-\frac{2r^2}{m}}$$

Lemma: For any $c$, there is a constant $d$ such that with high probability (w.h.p.) the number of heads in flipping $d \lg n$ fair coins is at least $c \lg n$.

Proof: Let $Y$ be the number of tails when flipping a fair coin $d \lg n$ times. $p = 1/2$

$m = d \lg n$, so $E[Y] = \frac{1}{2} m = \frac{d \lg n}{2}$

We want to bound the probability of fewer than $c \lg n$ heads = the probability of getting at least $d \lg n - c \lg n$ tails.
Proof of Lemma (contd.)

\[ \Pr \left[ Y > (d-c) \lg n \right] = \Pr \left[ E[Y] + (\frac{d}{2} - c) \lg n \right] \]

Choose \( d = 8c \) \( \Rightarrow r = 3c \lg n \)

By Chernoff, prob of \( \leq c \cdot \lg n \) heads

\[
\leq e^{-\frac{2r^2}{m}} \\
= e^{-\frac{2(3c \cdot \lg n)^2}{8c \cdot \lg n}} \\
= e^{-\frac{9c \cdot \lg n}{4}} \\
= e^{-c \cdot \lg n} \\
\leq 2^{-c \cdot \lg n} \quad (e > 2) \\
= \frac{1}{2} \\
= \frac{1}{n^c} \\
\]

😊 for Lemma
Proof of Theorem (finally!)

Event $A$: number of levels $\leq c \lg n$ w.h.p.

Event $B$: number of moves until $c \lg n$ "up" moves $\leq d \lg n$ w.h.p.

Event $A$ and event $B$ are not independent.

Want to show $\Pr(\text{event } A \& \text{event } B)$ high w.h.p.

\[
\Pr(\text{event } A \& \text{event } B) \geq \Pr(\text{event } A + \text{event } B)
\]

\[
\leq \Pr(\text{event } A) + \Pr(\text{event } B)
\]

\[
\leq \frac{1}{h^{c-1}} + \frac{1}{h^c}
\]

\[
= O\left(\frac{1}{h^{c-1}}\right)
\]

\[\Pr(\text{event } A \& \text{event } B)\] w.h.p.

Search in $O(\lg n)$ w.h.p.

\[\text{for Theorem.}\]