Marianne Månsson
Matematisk statistik D3, vt05

Grupparbete: Skiplistor

Det här projektet handlar om så kallade skiplistor, som är en relativt ny (introducerades 1990) form av datastruktur för lagring av elementen i en ordnad lista. En ordentlig beskrivning av hur skiplistor fungerar hittar ni på sidan 362 i Data Structures and Algorithms in Java av Goodrich, som var kurslitteratur på Datastrukturer. Nedan följer en kortfattad icke-datalogisk beskrivning av hur listan ser ut, mest för att introducera terminologin som används här.

Antag att vi har n stycken tal vi vill lagra i en skiplista. Ordna talen och skapa en lista med n noder, där varja nod innehåller ett av talen. Detta är nivå 1 i skiplistan. Singla en slant för varje nod: blir det krona så skapas en ny nod på nivå 2 som innehåller samma tal, dvs med sannolikhet 1/2 kopieras en nod upp till nivå 2, oberoende av vad som sker med övriga noder. Fortsätt sedan på samma sätt: varje nod på nivå i får följa med till nivå $i + 1$ med sannolikhet 1/2. Varje nivå ska dessutom börja med $-\infty$ och sluta med $+\infty$.

1. Minnesutrymme

   (a) Vad är förväntat antal noder på nivå i, om man inte räknar med ändnoderna $-\infty$ och $+\infty$? Ledtråd: Utnyttja binomialfördelningen.

   (b) Vad är förväntat totalantal noder i skiplistan, även här exklusive $-\infty$ och $+\infty$?

   (c) Hur stor plats kräver en skiplista med n stycken tal? (I termer av stora ordo O.)

2. Värsta fallet

   Det värsta som kan inträffa i en skiplista, med avseende på tidsåtgång vid sökning, är att alla tal når upp till samma nivå. Vad är sannolikheten för att detta ska inträffa i en lista med n stycken tal?

   Ledtråd: Om alla når upp till samma nivå, betyder det att alla når nivå 1 men ej 2, eller att alla når nivå 2 men ej 3, eller att alla når nivå 3 men ej 4, eller ...

3. Skiplistans höjd

   (a) Förklara varför höjden, dvs antal nivåer, i en skiplista som består av n stycken tal, H_n, kan beskrivas som $H_n = \max\{X_1, \ldots, X_n\}$, där $X_1, \ldots, X_n$ är oberoende och geometriskt fördelade stokastiska variabler med parameter 1/2.

   (b) Bestäm fördelningsfunktion och frekvensfunktion för H_n.

(c) Beräkna $E[H_{2^i}]$, $i = 1, 2, ..., 7$. Ser du något mönster? Utnyttja detta för att motivera att $E[H_n] = O(\log_2 n)$. (Låt $n = 2^i$ så att $i = \log_2 n$.)

Ett explicit uttryck för $E[H_n]$ för godtyckligt n finns inte. Se dock gärna Theorem 1 på http://www.wits.ac.za/helmut/postscriptfiles/florenz.ps där ett asymptotiskt uttryck är givet.

4. Tidsåtgång för sökning i listan (frivillig)

(a) Följande är lite trassligt. Rita en bild så blir det nog lättare.

Antag att vi letar efter talet b, och att det vid "framåtscanningen" på nivå $i+1$ visat sig att största tal $\leq b$ är a och minsta tal $\geq b$ är c. Hur många steg framåt behöver vi då ta på nivå i? Jo, vi måste gå igenom alla noder på nivå i med tal mellan a och c. (Om talet b eller något tal mellan b och c finns på nivå i, så blir det någon/några noder färre, men det behöver vi inte bry oss om.) Dessa noder är alltså sådana att de finns på nivå i men ej på nivå $i+1$. Det betyder att för varje sådan nod har det blivit klave i slantsinglingen. Vad är då det förväntade antalet steg man behöver ta på nivå i (med andra ord förväntat antal klave innan det blir krona för första gången)?

(b) Argumentera med hjälp av 3(c) och 4(a) ovan för att den förväntade tidsåtgången vid sökning asymptotiskt är $O(\log_2 n)$.

Tidsåtgången för insättningen och borttagning av tal är av samma storleksordning, men det går vi ej närmre in på här.