

Skiplistor

Grupparbete i Matematisk statistik och diskret matematik, ht 05

Det här projektet handlar om så kallade skiplistor, som är en datastruktur för lagring av elementen i en ordnad lista. En ordentlig beskrivning av hur skiplistor fungerar hittar ni i avsnitt 8.6 i *Data Structures and Algorithms in Java* av Goodrich and Tamassia, som är kurslitteratur på Datastrukturer. Chalmersbiblioteket har boken i elektronisk form. (Välj Books24x7 BusinessPro bland chalmersbibliotekets E-boksamlingar eller gå in i Chans.) 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 varje nod innehåller ett av talen. Detta är nivå 0 i skiplistan. Singla en slant för varje nod: blir det krona så skapas en ny nod på nivå 1 som innehåller samma tal, dvs med sannolikhet $1/2$ kopieras en nod upp till nivå 1, 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 sannolikheten att det finns k stycken noder på nivå i , där $k = 0, 1, \dots, n$ och $i = 0, 1, \dots$, om man inte räknar med ändnoderna $-\infty$ och $+\infty$? Ledtråd: Utnyttja binomialfördelningen.
- (b) Vad är förväntat antal noder på nivå i , $i = 0, 1, \dots$, även här exklusive ändnoder? Ledtråd: Utnyttja väntevärdet i binomialfördelningen.
- (c) Vad är förväntat totalantal noder i skiplistan, exklusive ändnoder?
- (d) 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.

- (a) Vad är sannolikheten att alla når nivå i men ej $i + 1$, där $i = 0, 1, 2, \dots$?
- (b) Vad är sannolikheten att alla tal når upp till samma nivå i en lista med n stycken tal?

Ledtråd: Om alla når upp till samma nivå, betyder det att ingen når nivå 1, eller 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 ...

Utnyttja (a) och att "eller = $\cup$ ".

3. Skiplistans höjd

- (a) Förklara varför höjden, dvs antal nivåer, i en skiplista som består av n stycken tal, H_n , kan beskrivas som $H_n = \max\{X_1, \dots, X_n\} - 1$, där $X_1, \dots, X_n$ är oberoende och geometriskt fördelade stokastiska variabler med parameter $1/2$.
- (b) Bestäm fördelningsfunktionen för H_n .
- (c) Bestäm frekvensfunktionen för H_n . Utnyttja att $f(k) = F(k) - F(k-1)$ för heltalsvärda stokastiska variabler.
- (d) Beräkna $E[H_{2^i}]$, $i = 1, 2, \dots, 7$. 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://math.sun.ac.za/prodinger/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.

Sista inlämningsdag 12/9.