Vijfde college complexiteit

26 februari 2008

Selectie O(n) en Insertion sort

Een adversary argument wordt gebruikt om een ondergrens te vinden voor de worst case complexiteit van een probleem.
De adversary

- "speelt" tegen het algoritme volgens een adversary strategie
- probeert het algoritme zo veel mogelijk vragen te laten stellen
- antwoordt altijd consistent met eerdere antwoorden
- bouwt zo als het ware tegen elk algoritme een bad case invoer op

Het doel is een ondergrens $f(n)$ af te leiden voor het aantal stappen dat elk algoritme tegen de adversary nodig heeft. Dat betekent dat er voor *elk algoritme* een invoer bestaat waarop minstens $f(n)$ stappen nodig zijn. In dat geval is het aantal stappen in de worst case voor elk algoritme dus $\geq f(n)$.

Bepaal of een string bestaande uit n bits twee opeenvolgende nullen bevat.

Basisoperatie: het bekijken van één bitpositie

Laat zien:

1. als $n = 2, 3, 5$ moeten in de worst case alle n bits bekeken worden (*)
2. voor $n = 4$ is er een algoritme dat altijd maar 3 bits hoeft te bekijken

(*) met behulp van een adversary argument

Probleem

Gegeven n verschillende getallen, opgeslagen in een array A: $A[1], A[2], \ldots, A[n]$. Laat verder een geheel getal k met $1 \leq k \leq n$ gegeven zijn. Gevraagd het k-de element in grootte.

Complexiteit

Het selectieprobleem is O(n).
We bewijzen dit door een algoritme te geven dat de k-de in grootte vindt in O(n) vergelijkingen in de worst case.

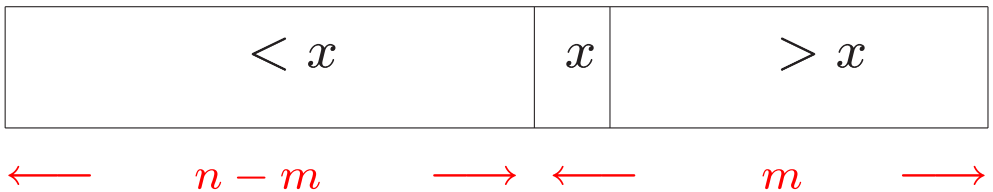
Algoritme:

1. Verdeel de getallen in $\lfloor \frac{n}{5} \rfloor$ groepjes van 5 elementen, en 1 groepje met de resterende n mod 5.

2. Vind de mediaan van elk van de $\lceil \frac{n}{5} \rceil$ groepjes van 5 (of minder), bijvoorbeeld met behulp van Bubblesort of opgave 21.

3. Vind de mediaan x van de in stap 2 gevonden $\lceil \frac{n}{5} \rceil$ medianen: recursie.

4. Partitioneer (ongeveer zoals bij Quicksort) alle elementen rond x. Stel dat m getallen $\geq x$ zijn en $n-m$ getallen $< x$.



5. Vind de k-de in grootte uit m stuks als $k \leq m$, of de $(k-m)$-de uit $n-m$ als $k > m$: recursie

De mediaan van ℓ elementen is de $\lceil \frac{\ell}{2} \rceil$-de in grootte.

Na stap 3. van het algoritme geldt:

- Ten minste $3 * (\lceil \frac{1}{2} \lceil \frac{n}{5} \rceil \rceil - 2) \geq \frac{3n}{10} - 6$ elementen zijn groter (respectievelijk kleiner) dan x.
- Dus er zijn hooguit $\frac{7n}{10} + 6$ elementen $\leq x$ (respectievelijk $\geq x$) (*).

Gevolg: in stap 5. wordt het algoritme derhalve recursief aangeroepen op hooguit $\frac{7n}{10} + 6$ elementen (*).

(*) om preciezer te zijn: $\lceil \frac{7n}{10} \rceil + 6$

Laat $T(n)$ = aantal vergelijkingen in de worst case dat dit (recursieve) algoritme doet.
Onder de aanname dat $T(n)$ stijgend is geldt:

$$T(n) \leq \begin{cases} \texttt{const} & n \leq \ldots \\ T(\lceil \frac{n}{5} \rceil) + T(\underbrace{\lceil \frac{7n}{10} \rceil + 6}_{<\, n \text{ als } n > 20}) + \underbrace{O(n)}_{\leq\, dn} & n > \underset{\substack{\uparrow \\ \text{bijv. } 80}}{\ldots} \end{cases}$$

De term O(n) komt van stap 1, 2 en 4 samen.

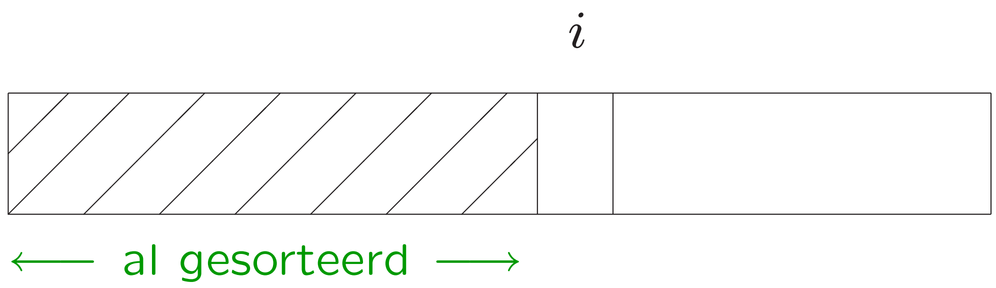
Bewering:

$T(n) \leq cn$ voor geschikte c en $n \geq \cdots$

Probleem

Gegeven een rij (array) A met n elementen $A[1], \ldots, A[n]$. Sorteer A oplopend, dus $A[i] \leq A[i+1]$ voor alle i ($<$ als alle $A[i]$ verschillend zijn).

Insertion sort



$A[i]$ op de juiste plek in het gesorteerde stuk $A[1] \cdots A[i-1]$ invoegen.

Het algoritme is gebaseerd op het doen van arrayvergelijkingen.

```
for i := 2 to n do
    // nu A[i] op de juiste plek in A[1]...A[i − 1] invoegen
        x := A[i];
        j := i − 1;
        while j > 0 and A[j] > x do
            A[j + 1] := A[j];
            j := j − 1;
        od
        A[j + 1] := x;
od
```

- Het aantal arrayvergelijkingen is een goede maat voor de complexiteit.

- Insertion sort doet eigenlijk steeds compare-exchange operaties: vergelijk en verwissel (indien nodig). Deze zijn hier vermomd als verschuivingen, waarna pas in de laatste stap $A[i]$ daadwerkelijk wordt neergezet.

- De verwisselingen zijn steeds buursverwisselingen.

We tellen het aantal vergelijkingen $A[j] > x$.

1. Worst case: $W(n) = \sum_{i=2}^{n}(i-1) = \frac{1}{2}n(n-1)$
2. Best case: $B(n) = \sum_{i=2}^{n} 1 = n-1$
3. Average case (*): $A(n) = \frac{1}{4}n(n-1) + n - \sum_{i=1}^{n}\frac{1}{i} = \Theta(n^2)$

(*) onder de aanname dat alle $A[i]$'s verschillend zijn en dat alle $n!$ permutaties (ordeningen) van $A[1]$ t/m $A[n]$ even waarschijnlijk zijn. We middelen dan over alle mogelijke permutaties en dat zijn in essentie alle mogelijke invoerrijtjes.