



Median og rekkefølge statistikk

29.09.2005

Generelt

- Median: det elementet i en liste som har like mange tall som er større enn seg som er mindre enn seg
- Rekkefølge statistikk: finne det element som ville hatt plass nummer x i en sortert liste
- Er kun aktuelt for usorterte lister. I en sortert liste er det å finne median eller element nummer x trivielt
- Median er et spesialtilfelle av rekkefølgestatistikk der vi ønsker det midterste elementet
- Mest brukt til å finne median

Minste og største

- Trivielt...
- Gå gjennom listen mens du tar vare på minste og største verdi du finner på veien
- Kjøretid: (n) , da en kun trenger å gå gjennom listen en gang

Rekkefølge statistikk forventet i linear tid (SELECT)

- Bruker partition kjent fra quick-sort
- Dersom pivotelementet vi bruker viste seg å ha den plasser vi søker er vi ferdig
- Ellers: Vi vet nå at pivotelementet har riktig plass og vi tar partition i den delen vi vet må inneholde det riktige elementet.
- Forventet kjøretid: (n) , omfattende bevis i boka
- Worst-case: $O(n^2)$, dersom vi velger det minste elementet hver gang som pivot-element vil partition fungere som en svært ineffektiv sorteringsmetode

Rekkefølge statistikk i worst-case lineær tid

- Ideen er å garantere et bra pivot-element
- Del listen i sub-lister av 5 og 5 elementer.
- Sorter de listene og finn deres median
- Finn medianenes median vha SELECT
- Partition hele listen vha medianenes median
- Dersom pivotelementet vi brukte viste seg å ha den plasser vi søker er vi ferdig
- Ellers: Vi vet nå at pivotelementet har riktig plass og bruker denne teknikken til på den delen av listen vi vet at det ønskede element finnes i