

Søgemaskineprojekt

Thomas Halkier Nicolajsen

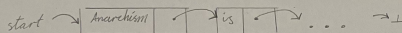
December 14, 2023

Indhold

- Demo
- Basal del
- Index1 og hægtet liste
- Index2
- Index3
- Index4
- Evaluering af Basal del
- Avanceret del
- Index5
- Index6
- Index7
- Index8

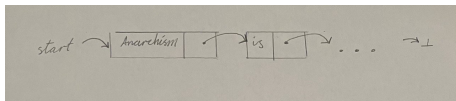
Index1 og Hægtet liste

- Byg: start $O(n)$
- Søg: $O(n)$
- Hægtet liste start



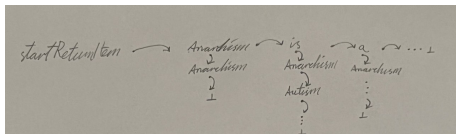
Index2

- Byg: start
- Søg: $O(n)$ sammenligninger
- String returnString: $O(d)$
- Søg: $O(n + d)$
- Hægtet liste start



Index3

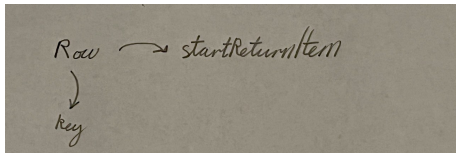
- Byg: start, startDistinct, startReturnItem
- Tid: $O(n + n^2 + u * d) = O(n^2)$
- Plads: $O(n * d)$
- Søg på unikt ord: $O(n + d)$



- Unikke ord på vandret, og dokumentnavne på lodret.

Index4

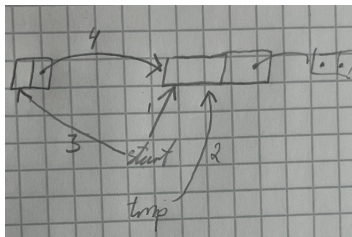
- Byg: start, startDistinct, startReturnItem, ht
- Søg på unikke ord



- ReturnItems med samme hashværdi hægtes på hinanden i samme Row.

Indsæt i hashtable

$O(1)$ Indsæt som stak:



-
- `tmp=start;`
- `start=new ReturnItem("...",null);`
- `start.next=tmp;`
- Alt.: Hold reference til halen.

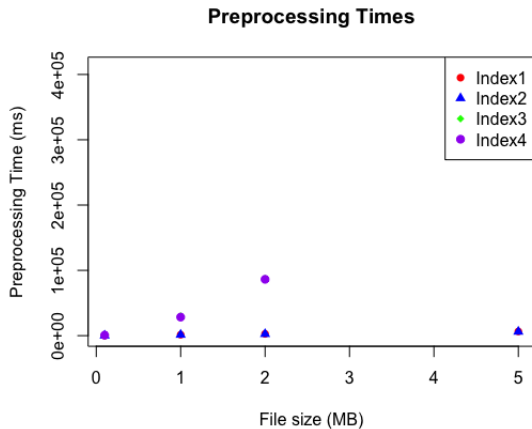
Hashfunktion: Prehash i for-løkken

```
1:  $H := 10$ 
2: procedure HASH(String s, int H)
3:   int hash = 7
4:   for int i in 1 to s.l do
5:      $hash = hash * 31 + s.charAt(i)$ 
6:   end for
7:    $hash = |hash| \bmod H$ 
8:   Return hash
end procedure
```

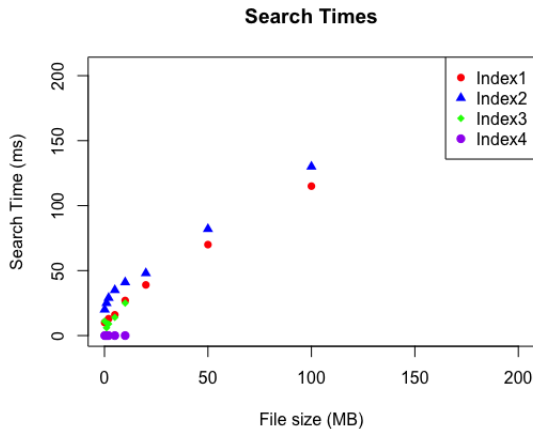
▷ Tabellængde H

▷ $hash \in \{1 \dots H\}$

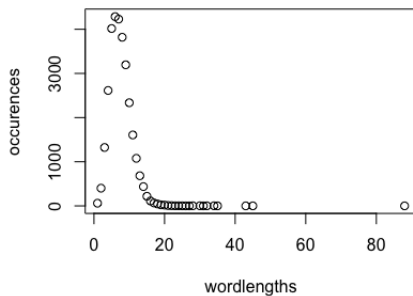
Eksperimentel evaluering: Byggetid som funktion af filstørrelsen n



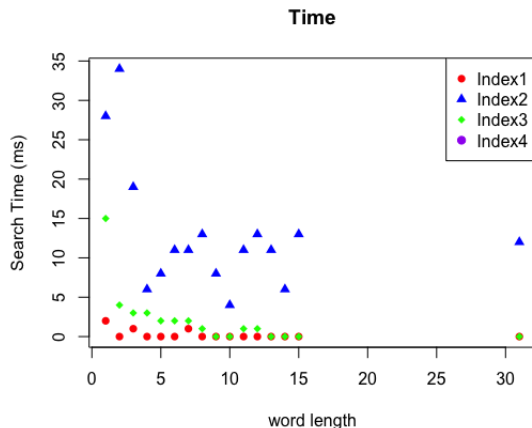
Eksperimentel evaluering: Søgetid som funktion af filstørrelsen n



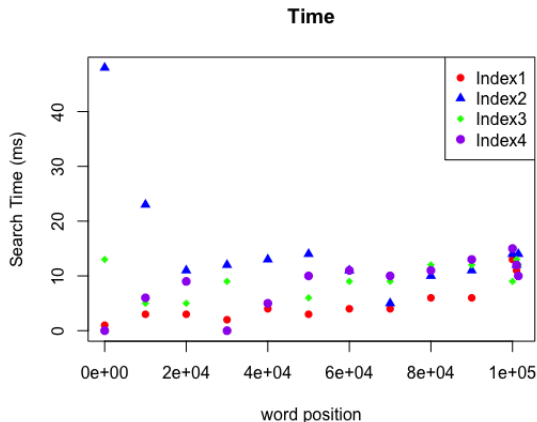
Eksperimentel evaluering: Fordeling af ordlængder I



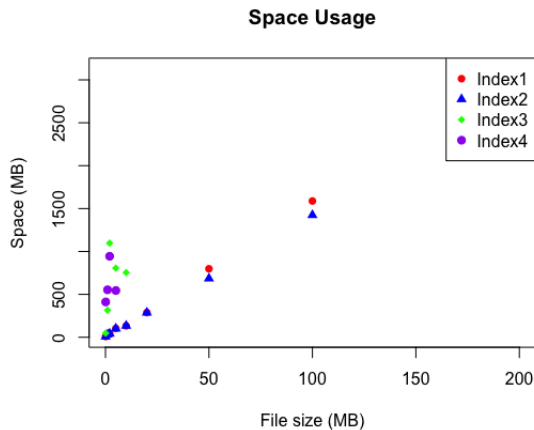
Eksperimentel evaluering: Søgetid som funktion af ordlængde I



Eksperimentel evaluering: Søgetid som funktion af søgeordets placering x



Eksperimentel evaluering: Pladsforbrug som funktion af n

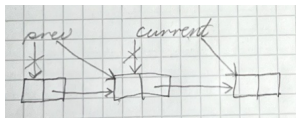
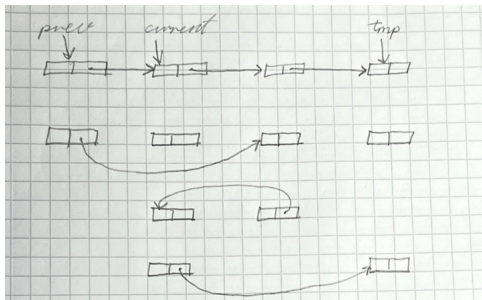


Avanceret del

- DocItem: String, int, next
- Sorterer brugerens svar

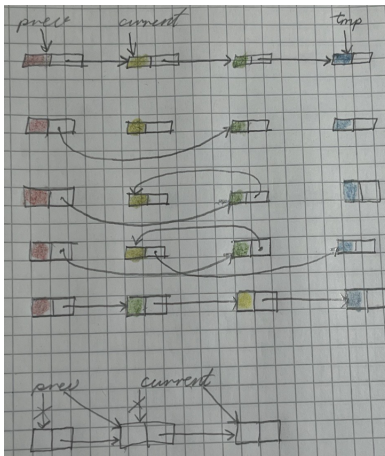
Index5

- Søg sorter brugerens DocItem svar med swap() i sort()
- $O(n + d^2)$ (uden antagelser)
- $O(1 + d^2)$ (SUHA)

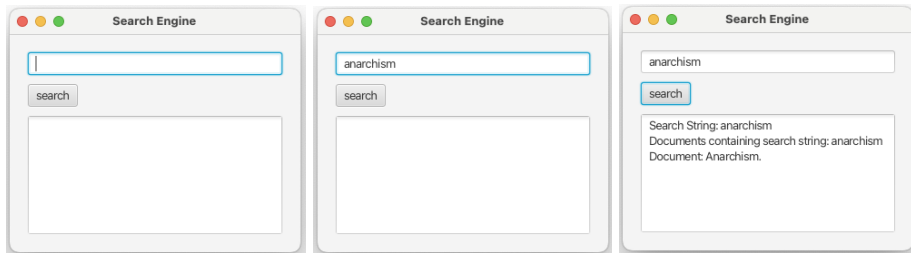


Index5

- Søg sorter brugerens DocItem svar med swap() i sort()
- $O(n + d^2)$ (uden antagelser)
- $O(1 + d^2)$ (SUHA)

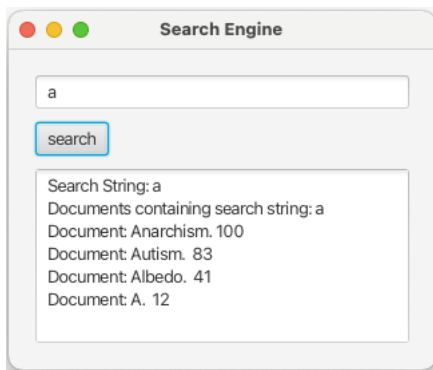


GUI: Java FX



- Søgning på *anarchism*
- Hvert dokument tilføjes som en linie

GUI med rangering (rettet på git)



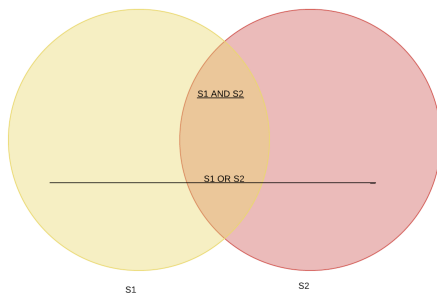
- Søgning på *a*

Index6: Præfiks af søgeordet

- *Anarhism: An, A*

Index7: Boolsk søgning

- To søgninger s_1 og s_2
- "Merge" af dokumentlister under søgning
- alt. n^2 unikke ordpar -*ikke skalerbart*



Index8: Præfikssøgning

- Søg: *startsWith(searchstr)*
- *A**: *Anarchism, An, A*

- Byg: hash med det samme
- Byg: indsæt som startDoc, eller reference til current: $O(1)$ i stedet for $O(d)$