

dt006g Java I

## Laboration:

Att räkna antalet förekomster av ord i en text

Daniel Bosk, Martin Kjellqvist

map.tex 2232 2015-04-01 11:45:27Z martin

## Innehåll

1	Introduktion	1
2	Syfte	2
3	Läsanvisningar	2
4	Uppgift	2
5	Resultat	3
6	Examination	3

## 1 Introduktion

År 1986 gav Jon Bentley<sup>1</sup> en utmaning till Donald Knuth<sup>2</sup>. Utmaningen var att med *literate programming* skapa ett program som skulle läsa innehållet i en text och bestämma de  $n$  mest förekommande orden och skriva ut en sorterad lista tillsammans med antalet förekomster av respektive ord. Knuth löste naturligtvis problemet, men likaså gjorde Douglas McIlroy<sup>3</sup> fast med ett mycket kort skalskript. McIlroys lösning följer:

```
1 n=10
2 cat hitchhikersguide.txt | \
3   tr -cs A-Za-z '\n' | \
4   tr A-Z a-z | \
5   sort | \
6   uniq -c | \
7   sort -rn | \
8   head -n \n
```

<sup>1</sup>Skapare av bland annat datastrukturen  $k$ -dimensionella träd.

<sup>2</sup>Bland annat fader till modern datalogi och skapare av T<sub>E</sub>X.

<sup>3</sup>Skapare av bland annat UNIX:s *pipelines*, på svenska närmast kallat rör.

Denna kan köras direkt i terminalen på ett UNIX-lik system och varje kommando har en manualsida där du kan läsa om vad kommandot gör med de givna argumenten.

## 2 Syfte

Syftet med laborationen är att du ska få bekanta dig med ytterligare några Java-konstruktioner och tillämpa dessa för att lösa ett problem. En av de Javaklasser som är nya för denna laboration är `java.util.Map`.

## 3 Läsanvisningar

Du bör ha läst om följande Collections:

- `java.util.ArrayList`,
- `java.util.TreeMap`,
- `java.util.Set` och
- `java.util.Collections`.

Det finns många användbara metoder i `java.util.Collections`

- `Collections.max`,
- `Collections.sort`,
- `Collections.frequency`

Även följande metoder från Stringklassen:

- `String.toLowerCase()`,
- `String.toUpperCase()`.

Inläsning görs med fördel med hjälp av en Scanner. Man kan för ett Scanner-objekt ange egna avgränsare, såsom kommatecken, punkt osv.

- `Scanner.useDelimiter`.

## 4 Uppgift

Du ska lösa Bentleys problem till Knuth. Som indata har du en textfil innehållandes ett kortare utdrag ur *The Hitchhiker's Guide To The Galaxy*, du finner den på lärplattformen. Utdata från ditt program ska vara de 10 mest förekommande orden sorterade i ordningen mest förekommande först.

Studera först Collectionsklasserna de underlättar enormt för att finna en lösning. Därefter tänker du ut vilka delsteg du måste lösa. Då det är gjort kan du inleda programmeringen.

## 5 Resultat

Ditt program ska redovisa de 10 vanligaste orden i ordnat så att det vanligaste ordet står på första radent tillsammans med antalet förekomster.

## 6 Examination

Om du läser på campus kontrollerar du med din handledare att din lösning är rimlig innan du lämnar in uppgiften. Om du läser på distans gör du en extra koll att din lösning uppfyller kraven i uppgiften.

Då du löst uppgiften laddar du upp din källkod till inlämningslådan i lärplattformen.

## Referenser