

DT028G Introduktion till programmering

## Laboration: Att räkna antalet förekomster av ord i en text

Daniel Bosk\*

map.tex 778 2013-03-19 16:05:27Z danbos

### Innehåll

1	Introduktion	1
2	Syfte	2
3	Läsanvisningar	2
4	Uppgift	2
5	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. 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 | \
```

---

\*Detta verk är tillgängligt under licensen Creative Commons Erkännande-DelaLika 2.5 Sverige (CC BY-SA 2.5 SE). För att se en sammanfattning och kopia av licenstexten besök URL <http://creativecommons.org/licenses/by-sa/2.5/se/>.

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

<sup>2</sup>Bland annat fader till modern datalogi och skapare av  $\text{\TeX}$ .

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

```
8 | head -n $n
```

Denna kan köras direkt i terminalen på ett UNIX-lik system och varje kommando har en manuelsida 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 C++-konstruktioner och tillämpa dessa för att lösa ett problem. En av de C++-konstruktioner som är nya för denna laboration är `std::map`.

## 3 Läsanvisningar

Du bör ha läst om följande datastrukturer i standardbiblioteket för C++:

- `std::string`,
- `std::vector`, och
- `std::map`.

Därefter bör du ha läst om följande funktioner från standardbibliotekets algoritm-del:

- `std::transform`,
- `std::sort`,
- `std::find`,
- `std::count`,
- `std::max_element`,
- `std::unique`.

Följande funktioner från C-biblioteket kan också vara behjälpliga:

- `tolower(3)`,
- `toupper(3)`.

Du kan läsa om standardbiblioteket i kurslitteraturen [2] eller någon referens, exempelvis *C++ Reference* [1], eller manuelsidorna.

## 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  $n$  mest förekommande orden sorterade i ordningen mest förekommande först. Det variabla antalet  $n$  ska enkelt kunna ändras i programmet.

## 5 Examination

Din lösning ska redovisas muntligen för en lärare vid något av kursens redovisningstillfällen. När du redovisat och fått godkänt laddar du upp din källkod (och byggsript) till inlämningslådan i lärplattformen.

## Referenser

- [1] *C++ Reference*. URL <http://www.cplusplus.com/reference/>.
- [2] Deitel, Paul J. och Deitel, Harvey M. *C++ : How To Program*. Pearson/Prentice Hall, Harlow, 8:e internationella utgåvan, 2012.