

# Python.O8: Filer

*I denna uppgift kommer du att öva på att använda filer i Python.*

## Mål

Du ska efter denna uppgift känna till och kunna använda:

- filer och olika filmoder (file mode: skrivning, läsning, etc.) i Python

## Uppgift

I denna uppgift ska du skriva programmet Ordstatistik. Som vanligt gäller det att utskrifterna ska stämma *exakt* med körexemplet, även på fil.

Den här gången finns det inga krav på rimlighetskontroller alls. Du kan antaga att filnamnen som matas in är korrekta.

## Ordstatistik

En ålderdomlig statistikbyrå har anställt dig som konsult. Ett av deras huvuduppdrag är att räkna förekomster av ord i filer och dokument. Då de som jobbar där börjar bli lite till åren komna tycker de att det börjar bli väldigt svårt att hålla reda på det i huvudet medan de läser, så ditt jobb är att skriva ett Pythonprogram som gör arbetet åt dem. Deras behov är följande:

- Ett underprogram som tar emot ett filnamn och läser in alla ord från den korresponderande filen, lagrar dem i lämplig datastruktur tillsammans med en räknare för varje ord som håller reda på hur många gånger det förekommer i filen. Returnera sedan variabeln som innehåller resultaten.
- Ett underprogram som tar emot ett filnamn samt resultatet från föregående underprogram och skriver ut alla resultaten, sorterade lexikografiskt, till filen som korresponderar mot det givna filnamnet, ett resultat per rad. Om filen har innehåll sedan tidigare ska detta skrivas över.
- Ett underprogram som tar emot ett filnamn vi vill lagra resultatet på, resultatet från den första funktionen, samt filnamnet på filen vi hämtade orden ifrån. Underprogramet tar sedan reda på vilket ord i resultatet som har flest förekomster för att skriva ut filnamnet ordet hämtades ifrån, ordet, samt hur många gånger det förekommer på filen som korresponderar mot resultat-filnamnet. I de fall flera ord delar förstaplatsen ska det ord som är "minst lexikografiskt" (det första man stöter på om orden är sorterade) väljas. Om det sedan tidigare finns resultat lagrade på denna fil ska det nya resultatet läggas till på en ny rad. Se körexempel för format.

## Python.O8: Filer

I huvudprogrammet ska vi fråga användaren vilken fil som ska sökas igenom, vilken fil vi vill skriva ut det fullständiga resultatet till, samt göra alla anrop till underprogrammen. Den sista filen, den som innehåller bara de ord som förekommit flest gånger (ett per körning) ska heta "most\_frequent.txt", men det ska gå att ändra enkelt genom att modifiera *huvudprogrammet*.

### TIPS!

- Du kan skapa tomma filer i terminalen, om behov finns, genom att navigera till din labb-mapp och skriva:  

```
touch filnamn.txt
```
- Vi behöver inte hantera att python gör skillnad på små och stora bokstäver, och vi behöver heller inte ta bort komma och punkt: "Hej" och "hej" är olika ord i detta program, det är även "hej,", "hej." och "hej".
- Du kan sortera *listor* genom en av:  

```
# vi skapar en sorterad version av listan, och sparar den  
seq = sorted(seq)  
  
# vi sorterar listan internt, seq är nu sorterad  
seq.sort()
```
- Vi kan inte sortera en *dictionary*, men vi kan titta på dess nycklar i sorterad ordning:  

```
sorted_keys = sorted(my_dict.keys())
```

### Givna filer:

- Filen testning.txt innehåller en kort text som kan användas för testning av ert program.
- Filen lorem\_ipsum.txt innehåller fem lite längre paragrafer text.

## Körexempel

### Körexempel 1:

Mata in vilken fil du vill läsa: **given\_files/testning.txt**

Mata in vilken fil du vill spara till: **result.txt**

Första femton raderna ur result.txt:

Att - 1

Det - 1

Givetvis - 1

Heja - 1

I - 1

Ju - 1

Om - 1

Testning - 1

accepterad - 1

alltid - 2

analys - 1

approacher, - 1

assistenternas - 1

att - 2

automatiska - 1

Följande rad bör vara sista raden i most\_frequent.txt:

testning.txt: det - 5

*Körexempel 2 på nästa sida*

**Körexempel 2:** Vi kör programmet två gånger i rad:

Mata in vilken fil du vill läsa: **given\_files/testning.txt**  
Mata in vilken fil du vill spara till: **result.txt**

Mata in vilken fil du vill läsa: **given\_files/testning.txt**  
Mata in vilken fil du vill spara till: **result.txt**

**Första femton raderna ur result.txt:**

Att - 1  
Det - 1  
Givetvis - 1  
Heja - 1  
I - 1  
Ju - 1  
Om - 1  
Testning - 1  
accepterad - 1  
alltid - 2  
analys - 1  
approacher, - 1  
assistenternas - 1  
att - 2  
automatiska - 1

**Följande rader bör vara de sista raderna i most\_frequent.txt:**

testning.txt: det - 5  
testning.txt: det - 5

*Körexempel 3 på nästa sida*

**Körexempel 3:**

Mata in vilken fil du vill läsa: **given\_files/lorem\_ipsum.txt**

Mata in vilken fil du vill spara till: **result.txt**

**Första femton raderna ur result.txt:**

Aenean - 2  
Aliquam - 3  
Curabitur - 2  
Donec - 3  
Duis - 2  
Etiam - 4  
In - 3  
Integer - 3  
Lorem - 1  
Maecenas - 1  
Mauris - 1  
Morbi - 4  
Nam - 3  
Nulla - 4  
Nullam - 1

**Följande rad bör vara den sista raden i most\_frequent.txt:**

lorem\_ipsum.txt: et - 10