

176 Repeterade ord

P-uppgiften ska göras individuellt. Läs CSC:s hederskodex innan du börjar!

Varudeklaration: Filhantering, datastrukturer.

När man skriver en text är det ofta en fördel att variera ordvalen. I stället för att alltid skriva “stor” kan man skriva “enorm”, “gigantisk”, “voluminös”, osv.

Första versionen:

Skriv ett program som läser text från en fil och skriver ut på en annan fil med upprepningar markerade enligt nedan. Upprepade ord är ord som förekommer flera gånger inom något intervall (på t ex 40 ord).

Användaren ska få välja infil, utfil och intervallets längd.

Exempel på hur utfilen kan se ut:

```
Utanför dörren stod en man och lyssnade intensivt i den mörka korridoren.  
Han hade skuggat Digby Groat hela eftermiddagen *och* följt efter honom  
in *i* huset. När *han* fick höra ljudet av fotsteg från rummet, smög  
*han* sig *in* *i* *en* sidogång *och* väntade. Eunice flög förbi,  
*och* Jim Steele ansåg tiden inne att flytta på *sig*. Inom få minuter  
skulle hela *huset* vara *på* benen, ty *han* slöt *sig* till, *att*  
den gamla damen svimmat. Det var ett övermodigt, tokigt företag *han*  
givit *sig* *på*, när *han* trängt in så pass tidigt *på* eftermiddagen,  
men *han* hade starka skäl för *att* ta reda *på* vad som stod i *ett*  
brev, *som* *han* sett stickas *i* Digbys hand tidigare *på* *eftermiddagen*.  
  
(ur 'Den Blå Handen' av Edgar Wallace)
```

Andra versionen (fortfarande betyg E):

Vissa ord förekommer ofta i en text utan att de anses som dåliga upprepningar, t ex “och”. Utöka ditt program med en indatafil med ord som inte ska räknas som dubletter, oavsett hur många gånger de förekommer.

Programmet ska nu börja med att dessa ord läses in, och därefter ska nya sådana ord kunna läggas till under det att programmet går igenom textfilen. Dessa nya ord är av två typer, sådana som ska sparas på filen (dvs aldrig ska räknas som dubletter) och sådana som inte ska räknas som dubletter endast i denna programkörning.

Extrauppgift, betyg C: Kontrollera att infilen existerar, och att utfilen inte redan finns.

Om filen med tillåtna dubletter inte redan finns ska programmet skapa den.

Extrauppgift, betyg B: Inför en synonymordlista. Ett ord kan ha flera synonymer (men duplicera inte informationen i filen med synonymer).

Låt programmet skriva ut varje upprepat ord i sitt sammanhang och låt användaren välja bland synonymerna. Se till att nya upprepningar inte uppstår...

Extrauppgift, betyg A: Gör ett grafiskt gränssnitt där användaren kan klicka på markerade ord i texten och få förslag på synonymer. Den justerade texten ska skrivas ut på fil.

