CSC, KTH **Betygsgradering:** Grunduppgift lätt, betyg E-A
DD1310/DD1314 våren 2014 (Python)

189 Skräppostmärkare

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

Varudeklaration: Datastrukturer, stränghantering

Idag består en stor del av mängden epost (eng. email) som cirkulerar av s.k. skräppost (eng. spam). Det är mer eller mindre seriös reklam, propaganda eller försök till dataintrång. Ett sätt att som användare slippa en del av detta är att filtrera bort de oönskade breven. Vanligtvis installerar man ett program som märker upp inkommande epost som skräp ifall den innehåller viss typ av information. Ofta är det flera olika typer av kännetecken man tittar efter och poängsätter brevet för varje typ. Om brevet får poäng över en viss gräns märks det som skräp.

Din uppgift är att i en text som läses in från fil, leta efter kännetecken och poängsätta brevet för varje kännetecken. Sedan summeras poängen ihop och om poängen blir 10 eller mer så skrivs ordet ”SPAM” ut, följt av en lista med det som gav poäng.

Du ska skapa en fil med minst 20 ord som kännetecknar skräppost. I filen anger du för varje ord:
hur många poäng ordet ger
om stora/små bokstäver ska spela roll för ordet
om upprepade förekomster ska öka poängen, och i så fall hur mycket den ska ökas

Exempel på brev:

```
Hej!
Har du för hög ränta???
Behöver du skuldsanering???
Har du bra inkomst och vill slippa skatt???

Tveka inte utan kontakta oss för avgiftsfri rådgivning!!!
Endast seriösa kunder...

Ripp-Off Ekonomi AB
```

När programmet körs med detta brev som indata blir utskriften

```
SPAM 14p:
ränta 6p
skuldsanering 3p
skatt 5p
```

Tips: Förslag till svenska ord som kännetecknar skräp: skuldsanering, inkomst, ränta, sex, porr, viagra, tillväxthormon, examensdiplom, rea, virusvarning.

Förslag till engelska ord som kännetecknar skräp: debt, income, mortage, sex, porn, viagra, growth hormone, diploma, sale, virus warning.

Extrauppgift, betyg C: Låt användaren själv ange poänggräns för skräppost, samt namn på tre filer. Första filen är en infil som innehåller alla brev. Skriv ut skräpet på den andra filen och de riktiga breven på den tredje.

Inför även felkontroll för användarens inmatning (poänggränsen samt filernas existens).

Extrauppgift, betyg B: Ge programmet möjlighet att läsa från flera olika brevlådor.

Låt även användaren specifiera vilka mail som ska tolkas som spam och spara informationen till följande körningar.

Extrauppgift, betyg A: Gör ett grafiskt användargränssnitt, där man kan välja fil, och få se ett brev i taget där skräpkännetecknen markerats.