

Skapa ett Java-program som gör det möjligt för användaren att mata in en text och så analyserar (gissar) programmet på vilket språk användaren skrivit.

DEL 1 - Analys av procentuell teckenfördelning (12p)

Ladda ner bifogade filer innehållande texter skrivna på 8 olika språk (engelska, tyska, (svenska, norska), (finska, estniska), (franska och italienska).

Gör sedan en algoritm som analyserar den procentuella fördelningen av varje tecken i texterna för varje språk skilt. T.ex. om du har en textfil *svenska.txt* som innehåller texten "Hej Nisse" ska algoritmens output vara följande:

Tecken	Procentuell fördelning
h	$1/8 = 12,5\%$
e	$2/8 = 25\%$
j	$1/8 = 12,5\%$
n	$1/8 = 12,5\%$
i	$1/8 = 12,5\%$
s	$2/8 = 25\%$

Gör det sedan möjligt för användaren att mata in en valfri text. När användaren matat in färdigt ska programmet "analysera" (gissa) vilket språk användaren har skrivit sin text på. Detta görs genom att tillämpa samma algoritm på användarens inmatade text och så jämförs den procentuella teckenfördelningen för den inmatade texten med den procentuella teckenfördelningen för alla språk baserade på textfilerna. Programmets "gissning" blir det språk som har den minsta skillnaden i teckenfördelningen i förhållande till den inmatade texten.

DEL 2 - Analys av procentuell teckenfördelning i kombination av tre tecken (6p)

Samma som DEL 1 men gör en algoritm som analyserar den procentuella fördelningen av tecken i kombination av tre för varje språk. T.ex. om du har en textfil *svenska.txt* som innehåller texten "Kalle skaller" ska algoritmens output vara följande:

3-tecken kombination	Procentuell fördelning
kal	$1/10 = 10\%$
all	$1/10 = 10\%$
lle	$2/10 = 20\%$
les	$1/10 = 10\%$

esk	1/10 = 10%
skä	1/10 = 10%
käl	1/10 = 10%
äll	1/10 = 10%
ler	1/10 = 10%

DEL 3 - Analys av den procentuella fördelningen av första tecknet i varje ord (6p)

Samma som DEL 1 och DEL 2 men gör en algoritm som beaktar första tecknet för varje ord i texten och beräknar den procentuella fördelningen. Om exempelvis texten är följande "Hej Kalle! Hur gammal är du?" blir algoritmens output följande:

1:a tecken	Procentuell fördelning
h	2/6=33%
k	1/6=17%
g	1/6=17%
ä	1/6=17%
d	1/6=17%

DEL 4 - Kombinera alla analyssätt (8p)

Kombinera ovannämnda analyssätt (teckenfördelning, 3-teckenfördelning, fördelning av 1:a tecken) där de olika analyssätten är viktade mot varandra. Viktningen kan ske t.ex. genom att beräkna skillnaden i teckenfördelningen mellan inmatad text och textfilerna för varje analyssätt skilt och sedan kombinerar man dessa skillnader med varandra t.ex. genom att beräkna medeltalet. Det språk med lägsta skillnaden i snitt i förhållande till den inmatade texten blir programmets "språkgissning".

Programmet ska även visa i "tabellformat" resultaten av alla analyser, samt den viktade analysen och ranka alla språk enligt det viktade resultatet, dvs. det språk som har den lägsta snittskillnaden i förhållande till den inmatade texten kommer högst upp i listan. Exempelvis så här:

Språk	Analys 1	Analys 2	Analys 3	Kombinerat	Rangordning
Swe	0.01	0.03	0.08	0.04	1
Eng	0.02	0.03	0.09	0.05	2
..

Siffrorna ovan indikerar alltså att ju närmare 0 desto närmare ligger teckenfördelningen i en inmatad text teckenfördelningen som beräknats för ett specifikt språk (på basen av textfilerna).

DEL 5 - "God objektorienterad design" (8p)

Strukturerat ditt program så att det innehåller flera klasser, t.ex. en klass som tar hand om statistikberäkningarna på texterna, en klass för hantering av användargränssnittet (UI), osv. I din programkod bör även kodupprepningar undvikas.

OBS! Det är strängt förbjudet att kopiera någon annans kod (både från Internet och från en klasskompis). Plagiat leder automatiskt till 0 poäng!

För att få poäng för detta projekt:

- Ladda upp på Itslearning inom utsatt tid din/ditt arbetspars lösning som ett fullständigt Netbeans-projekt
- Beskriv kort vilka av ovannämnda punkter du/ni lyckats lösa och hur.
- Presentera ditt/ditt arbetspars projekt på 1 av 3 alternativa projekt-feedbacktillfällen. Glöm inte att anmäla dej/ditt arbetspar till feedbacktillfället på Itslearning (Anmälningsskylt laddas småningom upp under mappen "PROJEKT")