

EXAMEN FINAL–PROBLEMA 2

PHASE 2: Lliureu la solució (el fitxer font del programa i fitxers resultats calculat pel programa) del problema2 a la tasca corresponent a Atenea i dins el termini establert.

Problema 2 [3 punts][Temps 50 min.]

El llenguatge utilitzat en les xarxes social inclou cada vegada més abreviacions de paraules o d'expressions de paraules. Per exemple, en anglès s'utilitzen, *PLZ* per indicar *Please* o *YOYO* per dir *You are on your own*. Un problema amb aquest “nou” llenguatge és que per fer processament i anàlisi de missatges, tweets, etc., totes aquestes abreviacions introdueixen un biaix (*soroll*) en el text.

Es demana que implementeu un mètode que:

- donat un fitxer de text *scriptum.txt*, que conté les abreviacions i les seves definicions *<abrv,definició>*, com ara *<YOYO,You are on your own>*
- donat un fitxer *twith.txt*, que conté missatges de text originals de la plataforma TWITH en la qual els usuaris utilitzen abreviacions a gust (podeu suposar que els usuaris de TWITH no usen signes de puntuació)

Calculeu:

- un nou fitxer del text original però corregit, *corrigendum.txt*, on cada abreviació s'hagi substituït per la seva definició segons el *scriptum.txt*.
- un fitxer *freqs.txt* de format *<abrv,freq>* on *freq* és la freqüència de l'abreviació (quantes vegades s'ha substituït en el text l'abreviació).

Cal tenir present que:

- En el text original pot haver abreviacions encara no recollides en el *scriptum.txt*; aquestes abreviacions es quedarien tal com venen en el text original.
- La llista de les abreviacions és ben llarga i el text original a corregir pot ser molt llarg.

Es valorarà l'eficiència de la implementació.

Example:

Fitxer d'entrada *twith.txt*

2DAY class is postponed to 2MORO
For the text using C&P is not recommended
The best criteria for programming is DIY
Once understood implementing a class is E123
EOM for 2DAY
CU F2F

Fitxer d'entrada *scriptum.txt*

2DAY Today
2F4U Too Fast For You
2MORO Tomorrow
2NTE Tonight
4YEO/FYEO For Your Eyes Only
AAMOF As A Matter of Fact
...

Fitxer de sortida *corrigendum.txt*

Today class is postponed to Tomorrow
For the text using Copy and Paste is not recommended
The best criteria for programming is Do it Yourself
Once understood implementing a class is Easy as 1, 2, 3
End of Message for Today See you Face to Face

Fitxer sortida *freqs.txt*

Abrev: 2DAY Freq: 2
Abrev: 2MORO Freq: 1
Abrev: C&P Freq: 1
Abrev: CU Freq: 1
Abrev: DIY Freq: 1
Abrev: E123 Freq: 1
...