

1º) Tras haber usado los distintos prototipos del programa que calculaba las frecuencias de aparición de palabras en un texto y tras haber analizado sus resultados, se detecta que aparecen muchas palabras que podríamos llamar comodines o demasiado comunes o poco interesantes para el estudio que se quiera realizar y que además puede ser un incordio obtener listados de palabras demasiado largos. Para dar solución a estas limitaciones se pide un programa que calcule las frecuencias (absolutas y relativas -en tanto por 100-) de las palabras que aparezcan en un determinado fichero texto y que no aparezcan en otro fichero glosario y muestre las de las N palabras más (o menos) frecuentes. El glosario es un fichero en el que hay una y solo una palabra por línea y no es obligatorio. El número de palabras N a mostrar tampoco es obligatorio. El proyecto se desarrollará en **python** con una interfaz de consola con el siguiente uso:

wfreq.py [-h] [-t] [-e] [-n palabras] [-g glosario] texto

- | | |
|--------------------|---|
| -h | se muestra la ayuda y sale (la ayuda también se mostrará siempre que haya algún error con las opciones y/o argumentos). |
| -q | no se muestra el total de palabras distintas del texto y el tanto por ciento de ellas que no estaba en el glosario (por defecto, se muestra). |
| -r | se ordena las frecuencias de menor a mayor (por defecto, de mayor a menor). |
| -n N | se muestra las N palabras más (o menos, si -r) frecuentes (por defecto, todas). |
| -g glosario | fichero conteniendo las palabras a ignorar en el análisis de frecuencias (por defecto, ninguna palabra). |

Y se desarrollará siguiendo los siguientes pasos:

- Diseño de un prototipo, en el que se usará como único texto el fichero *greatworks.txt* y como glosario el fichero *glossary.txt*, mostrándose sólo las 30 palabras más frecuentes, el total de palabras distintas del texto y el tanto por ciento de ellas que no estaba en el glosario.
- Escritura del programa de manera que procese correctamente todas las opciones y argumentos¹
- Escritura de una mejora del programa para que si texto no se indica se lea de la entrada estándar y si se indica más de un texto se procesen todos ellos. Su uso, por tanto, sería:

wfreq.py [-h] [-t | -e] [-n palabras] [-g glosario] [texto ...]

Para la entrega se creará el repositorio **wfreq** dentro de \$HOME/repositorio que contendrá, al menos, dos ficheros: un LEEME, que explicará el proyecto, enfoque, uso, etc.; y un **wfreq.py** conteniendo el script. Las versiones que se vayan desarrollando de estos ficheros se controlarán con la herramienta **Mercurial-Hg**

¹ Se debe usar el **módulo getopt**