



## **PRÀCTICA DE SCALA**

Jan Puig Martin  
u1973798

Girona, Maig de 2024

## Abast de la solució.

Les diverses funcions estan comentats de forma adient al codi, tot això es farà una breu explicació d'aquestes mateixes:

- **freq(x: String, n: Int = 1): List[(String, Int)]**: Aquesta funció rep un valor x, el qual és una cadena de strings i pot rebre opcionalment un valor n que de forma per defecte esta establert a 1, aquest segon valor és la quantitat de paraules que s'han d'agrupar per tal de formar, o no, n-grames. Aquesta funció també aplanar el text deixant-lo només amb paraules minuscules i traient qualsevol caràcter que no formi part d'una paraula.
- **quitstop(x: String, y: List[String]): String**: Donat un String x s'hi treuen les paraules de la llista y, havent previament aplanat el text igual que a la funció freq.
- **nonstopfreq(x: String, y: List[String]): List[(String, Int)]**: Aquesta funció aplica les dues anteriors juntament
- **paraulafreqfreq(x: String): Unit**: Utilitzant la funció freq es mostren les 10 majors freqüències i les 10 menors.
- **showfreq(x: List[(String, Int)]): Unit**: Una funció simple per tal de mostrar per pantalla la sortida de les funcions **freq** i **nonstopfreq**.
- **cosinesim(x: String, y: String): Double**: Aquesta última funció retorna un valor entre 0 i 1 el qual descriu la similitud entre dos fitxers. Els fitxers que entren han de ser aplanats i sense *stop-words*.

La sortida de les diferents funcions sembla correcte, tot i això no s'ha assolit un resultat idèntic amb les funcions freq, nonstopfreq, i paraulafreqfreq, amb comparació amb les proves que el professorat ens ha donat.