

Uso del Programma

January 5, 2018

Nota : è necessario avere installato il modulo stopwords, la cartella per l'installazione si trova assieme ai file del progetto

- a. eseguire da terminale il file unknown.py e lasciare aperta sia la finestra del terminale (dove compariranno i risultati) sia la GUI che apparirà al lancio del file .
- b. Inserire nel path il percorso alla cartella contenente il DataSet (è importante che tale DataSet sia strutturato in modo da avere i file divisi in categorie, dove ogni categoria è distinta da una specifica cartella).
- c. Inserire nel campo Mode una b o una m a seconda se si vuole testare la versione Bernoulli o quella Multinomiale.
- d. Inserire nel campo Train/Test la percentuale del DataSet che si vuole utilizzare come TrainSet (tutta la restante parte del DataSet verrà di conseguenza impiegata come TestSet).
- e. Inserire il nome del test tenendo conto che il programma non prevede sistemi di controllo quindi se l'utente inserisce lo stesso nome del test per combinazioni DataSet-percentuale diverse, i dati saranno inconsistenti e si avranno quasi certamente errori. Il nome del test può essere lo stesso solo se fra i test diversi con lo stesso nome, i seguenti parametri rimangono invariati : DataSet, Train/Test(%).
- f. Cliccare quindi il pulsante Start e verificare l'esecuzione del programma nel terminale che in caso di corretto funzionamento seguirà le seguenti fasi:
 - (a) se non già presente, durante la creazione del vocabolario appariranno su schermo il numero di documenti analizzati
 - (b) se la tavola non è già presente, verrà quindi notificato l'avvio del train e durante questo sarà possibile visualizzare a schermo la percentuale di completamento del train secondo il seguente schema : categoria-Corrente / totaleCategoria : percentualeFileAnalizzati
 - (c) verrà quindi notificato l'avvio del test vero e proprio e i risultati, mano a mano che vengono prodotti sono da leggersi nel seguente modo:
 - i. categoria : numero effettivo di file per questa categoria
 - ii. per ogni categoria (del DataSet) viene riportato accanto il numero di file (appartenenti alla categoria in analisi) che sono stati predetti essere di tale categoria