

TP2 - Problèmes

IFT3335

Stemmer

- <http://weka.wikispaces.com/file/view/snowball-20051019.jar/82917267/snowball-20051019.jar>
- `java -cp /Users/as/Documents/Work/weka-3-6-10/snowball-20051019.jar:weka-3-6-10/weka.jar weka.gui.Main`

Ohsumed

- Renommez l'attribut "class" en "class-attr" dans le fichier csv sinon votre tokenizer ne va pas marcher

Reuters

- Avant d'appliquer le filtre StringToWordVector il faut appliquer le filtre NominalToString à l'attribut "text"