

Text Mining

1

Erick STATNER

Maître de Conférences en Informatique

Université des Antilles

erick.stattner@univ-antilles.fr

www.erickstattner.com

Démarche à mettre en place

L'analyse de documents suit les étapes

1. Segmentation (*Tokenisation*)
2. Suppression de la ponctuation et des mots vides (stop words)
3. Racinisation / lemmatisation (Stemming/lemmatization)
4. Calculer les fréquences
5. Appliquer des algorithmes de Data Mining
6. Evaluer et visualiser