



Text Mining avec Python

Référence et durée

PYTHM

2 jours, 14 heures

Objectif

- Acquérir la méthodologie d'étude à mettre en œuvre pour étudier les données textuelles
- Savoir utiliser les bibliothèques Python de traitement d'analyse de données textuelles

Public

Data Scientists.

Prérequis

Connaissances de base de Python.

Programme

- Introduction et domaines d'application
- Représentation des données
 - Loi de Zipf
 - Attributs
 - Valeurs et pondération : sacs de mots, vecteurs
 - Mesures de distance et similarité
- Recherche d'information (RI)Re
 - Indexation
 - RI booléenne
 - RI vectorielle
 - Page Rank
- Classification
 - Classification supervisée
 - Classification non supervisée
- Annotation
 - HMM
 - CRF
- Exemples d'applications
 - Moteur de recherche
 - Analyse de sentiments
- Exercices avec Python
 - Librairie NLTK
 - Librairie TextBlob
 - Librairie Stanford's CoreNLP