



LE TEXT MINING AVEC R

Le text mining (ou fouille de textes) est un sous domaine de l'intelligence artificielle. Cette technique permet d'extraire de l'information dans des données textuelles grâce à des méthodes et algorithmes d'analyse et d'interprétation. R dispose de nombreux packages dédiés au text mining. Cette formation fera de vous un 'text miner' aguerri.



Durée

- 1 jour (possibilité 2 jours)



Tarif

- Inter : 550€ HT / participant (6 participants maximum)
- Intra : nous consulter



Lieu

- Locaux de Dcid Consulting (Orléans)
- Locaux de votre entreprise (Région Centre et Paris)



Public

Tout public ayant déjà des connaissances en programmation et/ou analyse de données

- Analyse de données ★★★★★
- Statistique ★★★★★
- Programmation ★★★★★
- Base de données ★★★★★
- Curiosité ★★★★★



Renseignements

- contact@dcid-consulting.fr
- 02 38 25 33 90



Prérequis

- Ordinateur individuel
- Connexion internet



Ressources utiles

- [Wikipedia](#)
- [R project, NLP](#)
- [Edutech wiki, U. de Genève](#)
- [Bradley Boehmke, U. de Cincinnati](#)



Objectifs pédagogiques

- Maîtriser les fondamentaux théoriques des données textuelles
- Maîtriser l'import et la manipulation de données au format texte
- Savoir prétraiter les données textuelles (nettoyage, tokenisation, stemming, etc.) et les représenter graphiquement
- Modéliser des données textuelles : analyse sémantique et topic modeling



Programme

- Découvrir les données textuelles et ses principes fondamentaux
Définition, description et représentation
- Importer et manipuler les données textuelles
Les sources de données texte
Traitement des chaînes de caractères et expression régulière
- Prétraiter les données textuelles
Le nettoyage, la ponctuation et les 'stopwords'
La tokenisation ou unité basique lexicale
Le stemming ou l'extraction du radical
La transformation en 'term-document matrix'
- Modéliser les données textuelles
Analyse sémantique
Topic modeling et clustering
- Représenter graphiquement les données textuelles
Représentation graphique avancée
Les packages de représentation dynamique
- Et après ? Perspectives et cas d'usage



Méthodes et moyens pédagogiques

- Alternance d'exposés et d'applications pratiques
- Possibilité de travailler sur les jeux de données des participants
- Codes source et supports de formation remis aux participants
- Formation dispensée par un data scientist expérimenté (pratique de R depuis 7 ans minimum)