

FORMATION COURTE  
**FOUILLE D'INFORMATION TEXTUELLE SUR DE  
GRANDES  
COLLECTIONS DE DOCUMENTS ET SUR LE WEB –  
PARTIE 2**

## DUREE

14 heures sur 2 jours

## DATE

Nous consulter

## LIEU

Site Saint Charles, Marseille  
Aix-Marseille Université

## COÛT

1 250 € HT

Établissement non soumis à la  
TVA.

## ORGANISATION

Inter et intra entreprise

## OBJECTIFS

- Implémenter une méthode probabiliste applicable à grande échelle pour l'analyse d'opinion.
- Implémenter et configurer une méthode probabiliste séquentielle pour la fouille de texte et l'annotation automatique.

## PUBLIC

Ingénieurs R&D, Chercheurs, Ingénieurs veille autour des SI et du Web.

## PRE-REQUIS

Modèles statistiques de base, notions de probabilité.  
Connaissance de langages informatiques (ex. Perl, Python, Java).

La partie 2 nécessite la partie 1.

## PROGRAMME

### 1- Fouille de textes et extraction d'information (1 jr) :

- Méthodes symboliques à base d'automates.
- Méthodes probabilistes.
- Application à la fouille de documents issus d'une bibliothèque numérique.

### 2- Application des méthodes pré-citées à l'analyse d'opinion et à la recommandation automatique (1 jr) :

- Analyse de champs libres de formulaires d'enquête ou de commentaires Web.
- Analyse de logs de serveurs Web...

Tous les cours sont accompagnés d'ateliers pour applications à des cas d'utilisation réels (bibliothèques numériques, web, etc.). Les ateliers se déroulent avec les logiciels Weka, Terrier et R.

## Les + de la formation

- Formation dispensée par le laboratoire d'excellence Archimède.

## Responsable de la formation

M. Patrice BELLOT  
Professeur AMU Informatique

## Renseignements et inscription

fpc-formationcourtes@univ-amu.fr  
04. 42. 60. 42. 80