

FORMATION COURTE
FOUILLE D'INFORMATION TEXTUELLE SUR DE
GRANDES
COLLECTIONS DE DOCUMENTS ET SUR LE WEB –
PARTIE 1

DUREE

14 heures sur 2 jours

DATE

Nous consulter

LIEU

Site Saint Charles, Marseille
Aix-Marseille Université

COÛT

1 200 € HT

Établissement non soumis à la
TVA.

ORGANISATION

Inter et intra entreprise

OBJECTIFS

- Maîtriser les composants logiciels les plus courants dans une architecture séquentielle pour le traitement automatique de documents textuels.
- Exploiter et paramétrer des environnements standards pour la recherche d'information, la classification et la fouille de textes sur des collections propriétaires et le Web.
- Savoir estimer les performances à attendre d'un système automatique et les coûts associés pour les atteindre.
- Savoir évaluer les approches implémentées et d'identifier les verrous scientifiques et techniques posés par des données ou des applications nouvelles.

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).

PROGRAMME

1- Introduction (1/2 jr) :

- Historique et état de l'art en recherche d'information et traitement automatique des langues.
- Bibliothèques numériques et Web.
- Modules logiciels de base (analyse linguistique de surface).

2- Modèles de recherche d'information (1/2 jr) :

- Du modèle booléen aux modèles probabilistes ;
- Modèles numériques adaptés au Web ;
- Application avec l'environnement de recherche européen ouvert Terrier sur des collections de grande taille du Web.

3- Méthodes non supervisées de classification automatique (1/2 jr) :

- Applications avec l'environnement 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