

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Université de M'sila  
Faculté des Mathématiques et de l'Informatique  
Département d'informatique



جامعة المسيلة  
كلية الرياضيات والإعلام الآلي  
قسم الإعلام الآلي

# Web Mining

Le fouille du web

**Domaine:** Mathématique et informatique  
**Filière:** informatique  
**Master :** RTIC

**Présenté par:** Dr. Benazi Makhoulf  
**Année d'étude:** deuxième année  
**Année universitaire:** 2023/2024

# Informations Pratiques



- **Intitulé du Matière** : Web Mining
- **Unité d'enseignement**: Fondamentale (UEF 3.2)
- **Semestre**: 03
- **Intitulé du Master** : RTIC
- **Coefficient** : 2
- **Crédits** : 4
- **V.H.** 1.5h cours, 1.5h TD 4 Travail personnel
- **Mode d'évaluation** : TD :50%, Examen Final: 50%.
- **Enseignant responsable de la matière**: Benazi Makhlouf
- **Email**: [makhlouf.benazi@univ-msila.dz](mailto:makhlouf.benazi@univ-msila.dz)
- **Sites utiles** : <https://elearning.univ-msila.dz/moodle/>

# **Objectifs de l'enseignement**

L'objectif de ce module est de donner divers points relatifs à la fouille du web, notamment le contenu du web, sa structure et son usage. Ce module abordera également les différentes techniques d'analyse des réseaux sociaux.

## **Connaissances préalables :**

L'étudiant doit avoir une bonne connaissance des concepts de base de la fouille de données, ses approches et ses algorithmes, notamment les règles d'association, la classification et le clustering.

## **Références:**

Yanchun Zhang, éditeur « Web Mining and Social Networking », 2011, Springer.

Anthony Scime, « Web Mining », 2005, Idea Group Publishing.

# Contenu de la matière :

1. **Introduction:** fouille de données et fouille du web, communauté web et réseaux sociaux.
2. **Rappels des concepts de base:** modèle de données web, fonctions de similarité, recherche d'information et évaluation des performances, concepts de base des réseaux sociaux.
3. **Fouille du contenu web:** modèle d'espace vectoriel, recherche web, indexation sémantique latente (LSI), extraction automatique de thèmes.
4. **Fouille de structure web:** algorithmes Pagerank et HITS, découverte de communauté web, modélisation par les graphes, classification par information des liens.
5. **Fouille d'usage web:** modélisation d'intérêts de l'internaute par clustering, analyse de sémantique latente, découverte des patterns d'accès de l'utilisateur, exploitation des fichiers logs (weblogs).
6. **Extraction et analyse des web de réseaux sociaux:** évolution des communautés web, analyse du comportement social.