

SYLLABUS

Domaine : Mathématique-informatique **Filière :** Informatique

Spécialité : Master SID (SYSTEMES INFORMATIQUES ET DECISIONS)

Semestre : 3

Année universitaire : 2025/2026

Identification de la matière d'enseignement

Intitulé : WEB MINING

Unité d'enseignement : UEF2

Nombre de Crédits : 4

Coefficient : 2

Volume horaire hebdomadaire total : 3h

- Cours (nombre d'heures par semaine) : 1h30
- Travaux dirigés (nombre d'heures par semaine) : /
- Travaux pratiques (nombre d'heures par semaine) : 1h30

Responsable de la matière d'enseignement

Nom, Prénom, Grade : Dr. Samira LAGRINI MCB

Localisation du bureau (Bloc, Bureau) : Laboratoire Labged

Email : samira.lagrni@univ-annaba.dz

Horaire du cours : Mercredi 11h30 - 13h00 J23

Horaire du TP : Mercredi 13h00 - 14h30 J23

Description de la matière d'enseignement

Prérequis : savoir programmer avec Python 3 ; Compréhension des bases de données relationnelles (SQL) et non relationnelles (NoSQL), **Familiarité avec les librairies Python pour la data science (Pandas, NumPy)**, Familiarité avec les outils de web scraping (BeautifulSoup, Scrapy)

Objectif général du la matière d'enseignement

L'objectif principal du cours de Web Mining est de former les étudiants à l'exploitation des données disponibles sur le web à des fins d'analyse et de prise de décision. Le cours vise à enseigner les techniques et outils nécessaires pour collecter, structurer et analyser des données web de manière automatisée, ainsi qu'à comprendre les dynamiques des réseaux sociaux à travers des méthodologies d'exploration des données. À la fin du cours, les étudiants seront capables de :

- Utiliser des outils pour extraire et manipuler des données web à grande échelle (web scraping, API).
- Structurer et nettoyer des bases de données complexes provenant du web pour les préparer à l'analyse.
- Appliquer des méthodes d'analyse statistique et de réseau social pour extraire des informations pertinentes à partir des données web.
- Modéliser les interactions et les processus sociaux dans les réseaux sociaux en ligne, et comprendre leur dynamique.
- Maîtriser l'écosystème Python pour la visualisation et l'analyse des réseaux sociaux (NetworkX, Matplotlib, Seaborn, Plotly)

Le cours se concentre également sur les enjeux éthiques et légaux liés à la collecte et à l'utilisation des données web.

Contenu de la matière

1. Introduction au Web Mining :

- Définition et importance du Web Mining
- Concepts clés du Web Mining et ses applications
- Différences entre les types de Web Mining

2. Collecte de données web :

- Techniques de collecte automatisée de données web (Web scraping)
- Administration en ligne d'une collecte de données
- Outils pour la collecte de données (ex. BeautifulSoup, Scrapy)
- Aspects éthiques et légaux de la collecte de données web

3. Structuration et Manipulation des Bases de données complexes

- Nettoyage des données après extraction
- Structuration de bases de données complexes
- Programmation avec Python
- Préparation des données pour l'analyse

4. Analyse des réseaux sociaux :

- Introduction aux concepts de réseaux sociaux
- Métriques en analyse des réseaux (centralité, densité, etc.)
- Utilisation de la librairie NetworkX pour l'analyse des réseaux
- Représentation des réseaux sociaux sous forme de matrices (matrices d'adjacence)
- Manipulation des graphes et calcul des métriques avec Python
- Visualisation interactive des réseaux avec Plotly et Matplotlib

5. Structure et dynamique des réseaux sociaux :

- Description des réseaux sociaux en ligne (ex :Twitter, Facebook, LinkedIn).
- Analyse des dynamiques et interactions dans les réseaux
- Calcul matriciel appliqué aux réseaux sociaux avec NumPy
- Modélisation de l'action collective en ligne
- Étude des interactions entre les réseaux sociaux en ligne et hors ligne
- Détection de communautés avec des algorithmes Python (Louvain, Girvan-Newman)

6. Modélisation des processus sociaux d'action collective :

- Introduction à l'action collective
- Modélisation des processus sociaux
- Exemples pratiques d'action collective

Modalité d'évaluation

Nature du contrôle	Pondération en %
Examen	60%
Micro – interrogation	
Travaux dirigés	
Travaux pratiques	40%
Projet personnel	
Travaux en groupe	
Sorties sur terrains	
Assiduité(Présence /Absence)	
Autres (à préciser)	
Total	100%

Références & Bibliographie

Textbook (Référence principale) :		
Titre de l'ouvrage	Auteur	Éditeur et année d'édition
Web Data Mining : Exploring Hyperlinks, Contents, and Usage Data	Bing Liu	Springer Science & Business Media; 2011
Mining the Social Web	Matthew A. Russell	O'Reilly Media, Inc. 2019
Les références de soutien si elles existent :		
Titre de l'ouvrage(1)	Auteur	Éditeur et année d'édition
Practical Web Scraping for Data Science: Best Practices and Examples with Python	Seppe vanden Broucke, Bart Baesens	Apress 2018

Planning du déroulement du cours

Semaine	Titre du Cours	Date
1	Séance 1 : Introduction au Web Mining	
2	Séance 2 : Collecte de données web	
3	Séance 3 : Collecte de données web	
4	Séance 4 : Structuration et Manipulation des Bases de données complexes	
5	séance 5 : Structuration et Manipulation des Bases de données complexes	
6	Séance 6 : Analyse des réseaux sociaux	
7	Séance 7 : Analyse des réseaux sociaux	
8	Séance 8 : Structure et dynamique des réseaux sociaux	
9	Séance 9 : Structure et dynamique des réseaux sociaux	
10	Séance 10 : Modélisation des processus sociaux d'action collective	
	Examen de fin de semestre	

Liste des étudiants

	Nom	prénom	Emargement
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			