EAR212 - Web data

Mis à jour le 16-04-2024

Code : EAR212

Unité d'enseignement de type cours

4 crédits

Volume horaire de référence (+/- 10%) : **40 heures**

Responsabilité nationale :
EPN09 - Economie Finance Assurance Banque (EFAB) / Héloïse PETIT

Contact national :

EPN09 - EFAB

292 rue Saint-Martin
Accès 3
75003 Paris
01 40 27 23 66
Virginie Moreau
virginie.moreau@lecnam.net

Présentation

Prérequis

Aucun pré-requis sauf des notions de mathématiques basiques.

Le cours nécessite d'avoir un ordinateur portable sur lequel Python peut être installé (gratuit).

Objectifs pédagogiques

Ce cours est une introduction à la collecte, au traitement et à l'analyse des données en Python. Ce cours ne nécessite aucun pré-requis sauf des notions de mathématiques basiques.

Programme

Contenu

Dans un premier temps, nous présenterons les concepts de base en programmation, à savoir la création de variables, les conditions et les boucles. Dans une seconde partie, nous présenterons les principales structures de données intégrées à Python telles que les listes et les dictionnaires qui permettent de structurer des données. Nous verrons également comment ouvrir, lire, écrire et sauvegarder des objets ou structures de données dans des fichiers. Dans une troisième partie, nous verrons comment traiter, fusionner et analyser des bases de données dans différents formats. Enfin, dans une quatrième partie, nous présenterons des méthodes de collecte de données sur internet via du web scraping et l'utilisation des API web.

Différents modules seront utilisés. En particulier:
- *pandas* qui permet la manipulation et l'analyse des données. Ce module propose en particulier des structures de données et des opérations de manipulation de tableaux numériques
- *matplotlib* qui permet de tracer et visualiser des données sous forme de graphiques
- *numpy* qui permet de manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux
- *requests* qui permet d'effectuer programmatiquement des requêtes HTTP (comme celles que fait un navigateur)
- *re* qui permet de travailler sur les expressions régulières qui sont utilisées pour la lecture, la modification, la manipulation et l'analyse de textes

Modalités de validation

- Contrôle continu
- Projet(s)

Bibliographie

Titre	Auteur(s)
Apprenez à programmer en Python	Vincent Le Goff