

Collecter des données sur le Web avec Python

Antoine Mazières (en visio), Julie Pierson

Atelier, jeudi 15 octobre 9h30-12h15

Présentation générale

Cet atelier vise à initier ses participants à la construction d'une base de données à partir de source hétérogènes et non-structurées. Un intérêt particulier sera porté à l'acquisition de données depuis le Web et des APIs diverses, et aux bibliothèques lxml, Requests et au langage XPATH qui y sont dédiées. Les débutants pourront apprendre quelques bases et découvrir ce à quoi ils peuvent aspirer, tandis que les participants plus expérimentés pourront expérimenter des fonctions plus avancées et faire un tour d'horizon pratique des possibilités offertes par cet ensemble d'outils.

Environnement informatique

Pour participer à cet atelier vous avez besoin d'installer au préalable Anaconda Python : <https://www.anaconda.com/products/individual>

Type de données traitées - droit d'accès

Utilisation de sources de données publiques.

Niveau requis

Tous niveaux.

Objectifs

Avoir une idée de ce qu'il est possible de collecter ou pas sur le web, et pouvoir évaluer l'effort que cela implique. Présentation d'une suite d'outils pour ce faire (lxml, xpath, etc.). Réalisation de plusieurs exemples avec des degrés de difficulté graduels.

Liens

Pour en savoir plus sur l'intervenant : <https://www.antonomase.fr/>