

Le Web-scraping

Niveau : Avancé

Prérequis : Programmation HTML, XML.

Objectifs

Cette formation vise à fournir aux participants des compétences dans l'extraction automatique des données à partir de page web. Cette formation présente plusieurs approches. Ces approches font l'objet d'une présentation théorique et d'une application empirique. A la fin de la formation, les participants seront capables de :

- ✓ Récupérer des données à partir d'une page Web;
- ✓ Récupérer des informations par une interface de programmation (API).



Durée

5 jours (30 h)
Hors examen
13 – 17 Janvier 2020



Lieu

ENSEA,
Abidjan, Côte d'Ivoire

Prix : 650 000 F CFA



Programme

Module	Durée
Droit des données	½ jour
Concept de licence sur les données	
Open data	
Récupération des informations par API	½ jour
Définition d'une API	
Requêtage	
Exemple avec Python et R	
Extraction de données du Web	1 jour
Définition du web-scraping	
Switching de pages web	
Exemple avec Python et R	
Présentation des Apis Web	
Outils pour le web-scraping	2 jours
Outils pour sites statiques	
Outils pour sites dynamiques	
Quelques problèmes d'extraction	½ jour
Ordonnancement	
Proxy	
Authentification	
Erreurs http	
Nouveaux développements	½ jour