Web Scraping Éthique & Veille Concurrentielle Digitale Automatisée

Lien : https://innov-systems.com/formation/web-scraping-ethique-veille-concurrentielle-digitale-automatisee

DURÉE
5 jours (35h)

RÉFÉRENCE
BDI219

CATÉGORIE
Big Data - Techniques D'analyse

Innov Systems

OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- Comprendre les principes et les enjeux du Web Scraping éthique
- Identifier les limites juridiques et réglementaires de l'extraction de données
- Développer des outils de collecte automatisée de données web
- Exploiter les principales bibliothèques et frameworks de Web Scraping
- Automatiser la veille concurrentielle et la surveillance de marchés
- Nettoyer, structurer et stocker efficacement les données collectées
- Produire des analyses et tableaux de bord à partir des données extraites
- Industrialiser les processus de collecte et de surveillance
- Mettre en œuvre des solutions robustes, sécurisées et maintenables
- Optimiser les performances des outils de Web Scraping
- Développer une stratégie de veille digitale orientée intelligence économique
- Réaliser des projets concrets de collecte et d'analyse automatisées de données

POUR QUI ?

- Développeurs Python
- Développeurs Full Stack
- Data Engineers
- Data Analysts
- Data Scientists
- Business Intelligence Analysts
- Responsables veille stratégique
- Responsables marketing digital
- Consultants en intelligence économique
- Chefs de projets Data
- Entrepreneurs et responsables innovation
- Toute personne souhaitant automatiser la collecte de données publiques et développer des solutions de veille concurrentielle dans un cadre éthique et conforme

Innov Systems

Programme détaillé

1 / Introduction au Web Scraping et à la veille digitale

- Principes du Web Scraping et de l'extraction de données
- Cas d'usage en veille concurrentielle et intelligence économique
- Limites techniques, juridiques et éthiques

2 / Cadre légal et bonnes pratiques

- Respect des conditions d'utilisation des sites web
- Protection des données personnelles et conformité (RGPD)
- Robots.txt, copyright et pratiques responsables

3 / Technologies et outils de Web Scraping

- Bibliothèques Python (Beautiful Soup, Scrapy, Requests)
- Automatisation avec Playwright et Selenium
- Extraction de données depuis des sites statiques et dynamiques

4 / Collecte et extraction des données

- Analyse de la structure HTML et du DOM
- Sélection des éléments avec XPath et sélecteurs CSS
- Gestion des API publiques et des formats JSON/XML

5 / Gestion des défis techniques

- Pagination, chargement asynchrone (AJAX) et JavaScript
- Gestion des sessions, cookies et authentification
- Optimisation des performances et gestion des erreurs

Innov Systems

6 / Stockage, nettoyage et qualité des données

- Structuration et normalisation des données collectées
- Nettoyage, validation et déduplication
- Stockage dans des bases de données relationnelles et NoSQL

7 / Automatisation de la veille concurrentielle

- Planification des collectes et surveillance continue
- Détection des évolutions de prix, produits et contenus
- Alertes automatiques et workflows de veille

8 / Analyse et visualisation des données

- Création de tableaux de bord et d'indicateurs
- Analyse comparative des concurrents
- Reporting décisionnel et visualisation des tendances

9 / Industrialisation et gouvernance des collectes

- Déploiement sur le Cloud et orchestration des tâches
- Monitoring des collectes et maintenance des scrapers
- Sécurité, traçabilité et gouvernance des données

10 / Cas pratiques et projet de veille automatisée

- Développement d'une solution complète de Web Scraping
- Mise en place d'une plateforme de veille concurrentielle
- Bonnes pratiques d'exploitation, d'évolution et de conformité

Innov Systems

Approche pédagogique

- Support Ecrit et Projection
- Exposés Interactifs, Podcasts et Vidéos
- Brainstorming et Jeux de Rôle
- Cas Pratiques et Labs inclus pour leur impact opérationnel
- Test de Validation des Acquis des Connaissances

Prochaines dates programmées

01 au 05 Fév. 2027	Présentiel - Casablanca
29 Mars au 02 Avr. 2027	Présentiel - Casablanca
24 au 28 Mai 2027	Présentiel - Casablanca
19 au 23 Juil. 2027	Présentiel - Casablanca
13 au 17 Sep. 2027	Présentiel - Casablanca

Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Réservation & Renseignements

Téléphone : +212 522 247 210
Email : contact@innov-systems.com
Web : https//www.innov-systems.com

Innov Systems — Tous droits réservés