



ORGANISME DE FORMATION AUX TECHNOLOGIES ET METIERS DE L'INFORMATIQUE

Formation Web Scraping, récolter des données sur le web avec Python

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

Vous devez extraire des données du web, les manipuler, les vérifier ou les archiver ? Pour être plus performant, automatisez vos récoltes, élargissez le champ de ces opérations. Optez pour le web scraping avec Python, ses bibliothèques de scraping et sa simplicité permettent rapidement d'industrialiser les processus.

Référence	WPY
Durée	4 jours (28h)
Tarif	2 470 €HT

Objectifs

- | Maîtriser les bases du langage Python
- | Connaître des éléments de programmation avancée en Python
- | Posséder une vue d'ensemble des principales librairies Python disponibles pour gérer tous types de données de sites
- | Sélectionner la bonne librairie Python pour votre projet de web scraping et être capable de la mettre en oeuvre
- | automatiser des récoltes d'envergure (large-scale web scraping) avec des scripts

Public

| Développeurs, consultants, analystes, chefs de projet et toute personne souhaitant automatiser la récolte de données sur le web.

Prérequis

- | Maîtriser les bases de l'algorithmique ou savoir programmer.
- | Avoir des connaissances en HTML et CSS est recommandé

Programme de la formation

Les bases du langage Python

- | Les principaux types de variables.
- | Effectuer des opérations et travailler sur les chaînes de caractères.
- | Les structures de données en Python.
- | Comprendre les types mutable et immutable.
- | Les structures de contrôle.
- | Créer et utiliser des fonctions
- | Lire et écrire des fichiers textes ou binaires.

Éléments de programmation avancée en Python

- | Les fonctions anonymes lambda.
- | Comprendre l'utilité des générateurs et savoir en créer.
- | Traiter les erreurs avec la gestion des exceptions.
- | Créer de nouveaux types de données avec la programmation objet.
- | Notions d'héritage en programmation objet.
- | Utiliser une librairie.
- | Sélectionner et évaluer les librairies développées en open source

Extraire des données via des API Restful

- | Se connecter à une API web.
- | Effectuer différents types de requêtes HTTP.
- | Le format JSON, analyser les données extraites.
- | Rechercher des données spécifiques dans du texte avec des expressions régulières.
- | Gérer les erreurs de connexion.

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

le 8 septembre 2026

PARIS

le 8 septembre 2026

[VOIR TOUTES LES DATES](#)

Capter des tableaux de données sur internet et les traiter avec Pandas

- | Rappel des bases HTML et CSS.
- | Les fondamentaux de Pandas.
- | Importer et exporter des données dans différents formats.
- | Manipuler des données avec Pandas.
- | Scraper des tableaux de données sur le web.

Scraper des sites web avec Beautiful Soup

- | Scraping facile : Beautiful Soup.
- | Mettre en oeuvre le parser.
- | Rechercher dans l'arborescence du parser.

Automatiser des récoltes d'envergure avec Scrapy

- | Le fonctionnement de base du framework Scrapy.
- | Identifier du contenu à scraper.
- | Structurer une spider.
- | Automatiser une récolte Scrapy et enregistrer les résultats.
- | Évaluer la performance d'une campagne.

Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

- | Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
- | Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
- | Privilégier une connexion filaire plutôt que le Wifi.
- | Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
- | Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
- | Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
- | Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
- | Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
- | Horaires identiques au présentiel.

Accessibilité



Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.
Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.