



CAHIER DES CHARGES

Aymeric Guillochet

(Apprentissage par Projet / © Active Learning by Gaming Campus)

NOM DU PROJET : Algo 2 + Programmation Python

Date : du 30/01/23 au 17/02/2023

Classe ou Groupe concerné/e : Guardia 1

- **Discipline(s) / Enseignement(s) couvert(s) par le projet :**
 - Approfondir les différentes notions vues durant le premier projet en Algorithmie et en Python
 - Scraper du contenu Web avec Python et comprendre comment l'utiliser
 - Gérer la logique de programmation dans Python
 - Créer et exploiter de la data avec Python
 - Comprendre la logique d'une base de données (CSV/XML) dans l'optique de la convertir dans une base de données SQL
- **Nom du professionnel / intervenant :**

Aymeric Guillochet

- **Profil LinkedIn :**

.....

- **Modalités d'apprentissage :**

Pour ce genre de projet, le développement sera individuel il est impératif de comprendre chaque strat du problème.

- **Intitulé, contexte et descriptif du projet (intitulé, contexte, périmètre) :**

Pour ce projet, vous allez être dans une situation particulière. Vous cherchez un extract de toutes les boutiques bio de votre région à des fins commerciales, Cependant, le seul souci apparent est le suivant : Vous ne pouvez pas bénéficier de l'extract gratuitement car la plateforme vous demande une somme onéreuse.

En revanche, vous allez utiliser vos capacités de développeur pour pouvoir vous-même scraper les informations, les rentrer dans une base de données et faire vous-même votre propre annuaire en Python.

- **Problématique :**

Comment approfondir ses notions en Algorithmie et approfondir ses compétences en Python, tout en utilisant la POO. De plus, comment avoir une approche Web avec le Python.

A l'issue du projet, l'étudiant sera capable de :

- Concevoir une application Python
- Créer des algorithmes orientés objet
- Gérer des algorithmes simples et complexes
- Scraper de la data sur le web et la manier
- Comprendre l'approche du scraper

ÉCHÉANCES ET LIVRABLES

Dates échéances	Livrables attendus	Moyens / formats (Comment on transmet Le livrable)	Remarques
Chaque jour	Code source de résolution des exercice	Soumission sur la plateforme de compétition
30/01	Théorie Algo + Python	Création d'un Repo Git et Push à chaque Coaching	
31/01	Scrapping sur le site web	Récupération des informations sur un drive	
1/02	Identification et tri des données		
3/02	Scrapping sur du CSV et sur du XML à partir de Python		
13/02	Intégrer les données python scrappées dans le CSV		
14/02	Utiliser la POO pour mettre des classes (Théorie + Pratique)		

15/02	Finalisation du code final + Préparation de la soutenance.		
17/02	Soutenance	Dépôt du projet sur les ressources du Google Drive.	

Etapes clefs : PRÉSENCE OBLIGATOIRE

30/01 – KickOff Projet + Théorie

- Revoir l’algorithmie rapidement
- Revue des bases en python
 - Variables
 - Constantes
 - Opérateurs
 - Structures conditionnelles
 - Exercices applicatifs pour vérifier le niveau global de la classe et offrir une homogénéisation des niveaux avant le lancement du projet
- Présentation du projet
- Explication de la roadmap

31/01 : Scrapping sur le Site web

- Présentation de BeautifulSoup
- Exploitation de BeautifulSoup sur un site classique pour récupérer des informations rapidement
- Comprendre les fondements et les utilisations du scrapping
- Commencer à scraper les informations voulues
- Préparer les informations scrappées pour traitement

1/02 : Identification et tri des données

- Observer la structure du site
 - Trier avec GET

- Trier avec l'emplacement des balises HTML
- Comprendre où est l'information, l'identifier
- Entraînement puis application par un formulaire (POST)
- Entraînement depuis un GET (URL)
- Entraînement par la sémantique
- Entraînement par l'identificateur de la balise
- Comprendre comment lancer les différents entraînements et savoir utiliser la bonne méthode.

3/02 : Théorie sur les bases de données :

- XML
- CSV avec l'utilisation de Pandas

Application sur ces différents formats de base de données

13/02 : Intégration des données python

- Créer sémantiquement des tableaux de données.
- Comment trier les données avec la gestion Objet

14/02 : Théorie sur la programmation objet

- Notion d'Objet
- Notion de classe
- Syntaxe Python
- Application

15/02 : Développement et finalisation du projet

- Développement des différentes classes
- Intégration des différents objets
- Mises en forme
- Validations finales
- Préparation de la soutenance avec un document spécifique reprenant tout ce qui a été scrappé, vu et traité dans le projet.

17/02 – Soutenance

AVANT LE DÉBUT DU PROJET, TRAVAIL PRÉPARATOIRE À RÉALISER :

Sur les ordinateurs des deux membres du binôme :

- Installer Python 3

- *Installer Visual Studio Code et les extensions Visual Studio Live Share et Python*
- *Installer Git et configurer Git pour GitHub*
- *Installer Scrappy*

RESSOURCES

- **Ressources pour le travail préparatoire :**

Notions de Scrapping : https://fr.wikipedia.org/wiki/Web_scraping

Projet de Scrapping : <https://github.com/pmasiphelps/ScrapingProject>

Notions d'Objet et de classe : <https://www.jedha.co/formation-python/programmation-orientee-objet>

POO en python spécifiquement : <https://courspython.com/classes-et-objets.html>

ACQUIS D'APPRENTISSAGE DÉTAILLÉS

CONNAISSANCES À ACQUÉRIR :

- *Récupérer des informations en masse sur le web*
- *Traiter de l'information*
- *Gérer des algorithmes spécifiques et complexes*
- *Utiliser le Python de manière plus approfondie*

- *Comprendre l'utilisation de python dans une approche orientée Web*

COMPÉTENCES (savoir - faire) À ACQUÉRIR :

A l'issue du projet l'étudiant sait :

- *Créer un algo complexe*
- *Utiliser la POO en Python*
- *Créer et manier des fichiers en*
 - a. *XML*
 - b. *CSV*
 - c. *Markdown*
- *Scraper des données*
- *Comprendre l'approche cyber à partir du scrapping et du traitement d'information*
- *Créer un annuaire manuellement à partir de méthodes de scrapping pré-définies.*

GRILLE D'ÉVALUATION DU PROJET

Date : 24/02/2023

Noms des Intervenants / Professeurs ou Membres du Jury :

Alexandre Quilan Delaistre

Connaissances / Compétences	Barème
Qualité du Scrapping	/20
Utilisation des Objets et Classes en Python	/20
Rapidité du programme	/20
Optimisation de l'algorithme et du code (en termes de requêtes http)	/20
Agencement des données	/20
Qualité de la prestation	/20
TOTAL	/120

Remarques :

.....

.....

.....