

FICHE : LANCEMENT "KICKOFF" DU PROJET

IDENTIFIER VOTRE CHOIX DE SUJET

Titre du projet : Cinemates

URL si applicable : <https://github.com/cegepmatane/projet-specialise-2022-Lechiffre4>

DÉCRIVEZ VOTRE PROJET (2 PARAGRAPHES MINIMUM)

Mon projet est un site en python. Il utilisera une bibliothèque de python nommée BeautifulSoup permettant d'effectuer du « Web Scraping ».

J'ai pour objectif de réunir des informations utiles sur des films, séries, animation japonaise ou autre. Dans un premier temps, j'ai pu remarquer que les sites de notation de film sur internet ne sont pas tous en accord avec les autres. Par exemple, un film peut avoir la note de 2/10 sur le site RottenTomatoes et 9/10 sur IMDb. Mon projet a donc pour objectif de récupérer plusieurs informations de plusieurs sites spécialisés dans la notation de film afin d'établir une moyenne de ses notes, et ainsi avoir une note plus fidèle à la réalité. L'utilisateur pourra avoir un avis objectif sur le film. Les avis les plus intéressants trouvés sur internet concernant les films recherchés (en fonction du nombre de likes qu'il obtient sur les différents sites) seront affichés.

Concernant l'interface, celle-ci sera assez facile à comprendre. L'utilisateur disposera d'un champ de texte où il pourra écrire le nom du film. En validant sa recherche, une liste de films pouvant correspondre apparaît. En cliquant sur le film de son choix, l'utilisateur sera renvoyé sur la page du film avec tous les détails le concernant.

Si le temps m'en donne l'occasion, je pourrai ajouter à mon site plusieurs autres modules de WebScraping, comme un sélecteur aléatoire de film en fonction de catégories afin d'aider l'utilisateur à choisir son film.

Les échanges de données sur le site se feront uniquement avec des requêtes HTTP sous le format JSON. Le site ne comportera pas de base de données attitrée puisque toutes les données nécessaires seront récupérées sur d'autres sites.

NOMMEZ 3 COMPÉTITEURS À VOTRE PROJET

qui vous inspireront pour les fonctionnalités (avec les liens)

Nom du compétiteur 1	https://www.rottentomatoes.com
Nom du compétiteur 2	https://www.imdb.com
Nom du compétiteur 3	https://www.allociné.com

LISTER LE MATÉRIEL DE CE PROJET

- Un ordinateur
-

Est-ce tous les éléments matériels nécessaires au projet sont en votre possession ? Sinon, lesquels ?

Oui, je possède tout le matériel nécessaire

Aimeriez-vous que le département vous prête un composant, une pièce ? Si oui laquelle ?

IDENTIFIER DEUX ASPECTS TECHNOLOGIQUES

Aspects : Serveurs, Services, Algorithmes, Web, IoT, etc

Aspect 1 : Python web (Flask)

Aspect 2 : Bibliothèque de scrapping python (BeautifulSoup)

PRÉCISEZ L'APPRENTISSAGE TECHNIQUE QUI SE FERA PENDANT LE PROJET

soyez aussi précis que possible (nom de la technologie, de la librairie, url)

Technologie : BeautifulSoup

URL : <https://beautiful-soup-4.readthedocs.io/en/latest/>

Quelle utilisation en ferez-vous à l'intérieur du projet ?

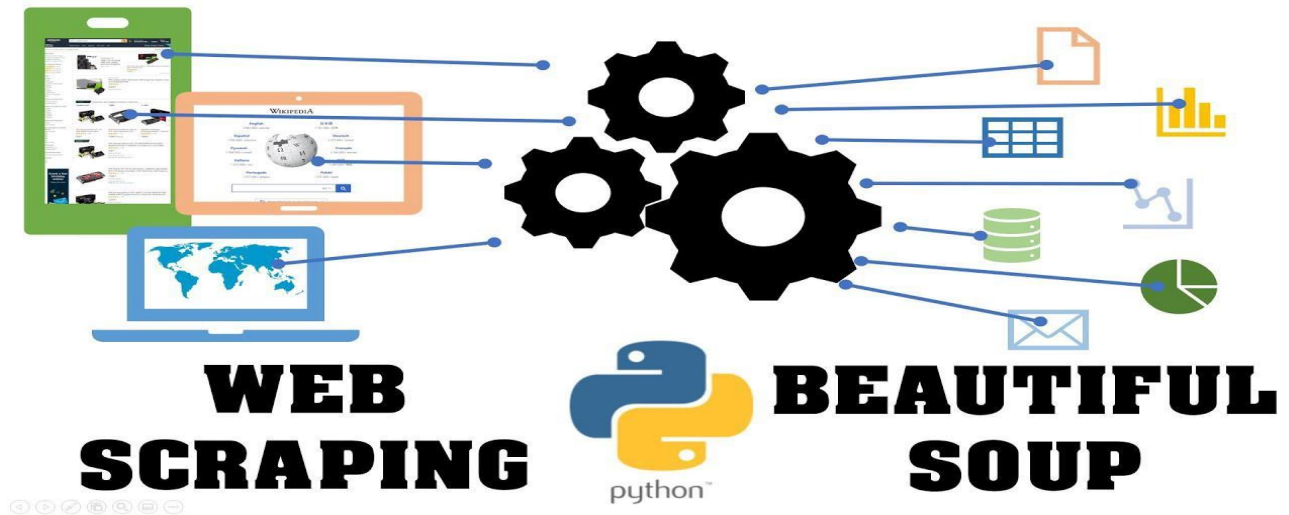
Je possède déjà de petites compétences en Python néanmoins, je compte utiliser pour ce projet Python afin de faire du Web ce qui est une découverte complète pour moi. Je vais utiliser le framework(cadriciel) Flask parce qu'il est considéré comme un framework web simple pour le Python.

Pour récupérer les données présentes sur mon site je compte utiliser la bibliothèque BeautifulSoup4(une découverte pour moi) qui permet de prendre dans le HTML d'une page web des informations.

Pour utiliser BeautifulSoup il faut indiquer au programme python, l'url du site dont on veut récupérer les informations. Par la suite y faut parser le fichier html de la page pour le rendre compréhensible par le programme. Ensuite, il est possible de récupérer n'importe quelle information contenu dans le html.

ILLUSTRATION

Si une illustration peut nous aider à mieux comprendre le projet, merci de l'inclure ici.



What is Web Scraping?



Website Pages,
Unstructured Data



Web Scraping/
Data Extraction



Structured Data