CAHIER DE CHARGE :

Projet Scrapy python3



Membre du groupe 9 :

BOUILI Espoir Prof :

ADEDOKOU Abdou-Salam M. ANAKPA Daniel

Sommaire

[I – Définition du problème 3](#_Toc136406974)

[II – Objectif du projet 3](#_Toc136406975)

[III – Périmètre 3](#_Toc136406976)

[IV – Description fonctionnelle des besoins 4](#_Toc136406977)

[Ajout d’un site Web : 4](#_Toc136406978)

[Supprimer un site Web : 4](#_Toc136406979)

[Afficher un site Web : 4](#_Toc136406980)

[Afficher tous les sites Web : 5](#_Toc136406981)

[Modifier un site Web : 5](#_Toc136406982)

[Collecte des données : 5](#_Toc136406983)

[V – Les outils de développement : 6](#_Toc136406984)

# I – Définition du problème

Une entreprise de Data Science cherche à collecter des données sur internet pour ses analyses. Le problème qui se pose dans ce cas est la diversité et la complexité des architectures de différents sites Web.

# II – Objectif du projet

Notre objectif est de créer une application qui peut s’adapter pour collecter les données que nous voulons sur les sites Wb de nos choix.

# III – Périmètre

Ce projet qui consiste la mise en place d’une application est conçu pour répondre à un problème particulier a un secteur de l’activité économique de la société. Cette application est seulement adaptée aux administrateurs, techniciens ou consultation d’une organisation qui cherche une solution au problème mentionné ci-dessus.

# IV – Description fonctionnelle des besoins

## Ajout d’un site Web :

Dans la gestion dynamique des sites, notre application permet l’ajout facile de nouveaux sites même avec des architectures différentes. Donc si une entreprise souhaite collecter des informations sur un site récemment publié, elle peut utiliser notre application pour ajouter sans modifier le comportement du site.

## Supprimer un site Web :

Toujours dans la gestion dynamique, notre application permet la suppression facile de sites même avec des architectures différentes. Donc si une entreprise souhaite n’a plus besoin des informations sur un site, elle peut utiliser notre application pour supprimer sans modifier le comportement du site.

## Afficher un site Web :

Notre application permet également d’afficher différents sites même que nous avons afficher. Donc si une entreprise souhaite afficher les sites Web précédemment ajouter, elle peut utiliser notre application pour afficher sans modifier le comportement du site.

## Afficher tous les sites Web :

Notre application vous permet d’afficher tous les sites Web ajouter a la fois. Donc si une entreprise désire afficher plusieurs sites Web au même moment, notre application peut le faire sans toutefois modifier le comportement de ces sites.

## Modifier un site Web :

Notre application permet la modification facile de sites même avec des architectures différentes. Donc si une entreprise souhaite modifier des informations sur un site, elle peut utiliser notre application pour faire cette modification sans changer le comportement du site.

## Collecte des données :

Avec la mise en place d’une API, notre application permet d’avoir accès aux informations d’un site web sous-forme d’articles. Une entreprise souhaite collecter des données pour une analyse par exemple, elle peut utiliser notre application sans toutefois modifier le comportement du site.

## Afficher les articles :

Vous pouvez récupérer des articles à l'aide de Scrapy, les stocker dans une base de données, puis les récupérer et les afficher dans une application Web en utilisant une combinaison d'un Framework backend et d'un Framework frontend.

## Supprimer les articles :

Supprimer des articles dans une application Web construite avec Scrapy, vous devrez généralement implémenter des fonctionnalités sur le backend pour gérer les demandes de suppression d'articles et mettre à jour la source de données sous-jacente (telle qu'une base de données).

|  |  |  |  |
| --- | --- | --- | --- |
| Fonction | Description | Contrainte | Niveau de priorité |
| Ajout d’un Site | La possibilité d’ajouter un site a l’application pour pouvoir faire la collecte de données | Disponible pour les administrateurs seulement | Haut |
| Modifier un Site | La possibilité de changer le libelle du site | Disponible pour les administrateurs seulement | Haut |
| Supprimer un Site | La possibilité de supprimer un site scrape | Disponible pour les administrateurs seulement | Haut |
| Afficher un Site | La possibilité d’afficher un site qu’on a scrape | Disponible pour les administrateurs seulement | Haut |
| Afficher tous les Sites | Afficher tous les sites scrapes | Disponible pour les administrateurs seulement | Haut |
| Collecte des données | La possibilité de collecter les données pour les utiliser pour d’autre fin (scrapy) | Disponible pour les administrateurs et aux clients | Bas |

# V – Les outils de développement :

* Flask:

Flask est un micro Framework open-source de développement web en Python. Il est classé comme micro Framework car il est très léger. Flask a pour objectif de garder un noyau simple mais extensible.

* Flask restful:

Flask-RESTful fournit une extension à Flask pour créer des API REST. Flask-RESTful a été initialement développé en tant que projet interne chez Twilio, conçu pour alimenter leurs API publiques et internes.

* SQL alchemy:

SQLAlchemy est un toolkit open source SQL et un mapping objet-relationnel écrit en Python et publié sous licence MIT.

* sqlalchemy\_utils:

SQL Alchemy-Utils provides custom data types and various utility functions for SQL Alchemy.

* MySQL-connector-python:

Ce manuel décrit comment installer et configurer MySQL Connector/Python, un pilote Python autonome pour communiquer avec les serveurs MySQL, et comment l'utiliser pour développer des applications de base de données.

* MySQL. Connector:

MySQL fournit des pilotes basés sur des normes pour JDBC, ODBC et .Net permettant aux développeurs de créer des applications de base de données dans le langage de leur choix. De plus, une bibliothèque C native permet aux développeurs d'embarquer MySQL directement dans leurs applications.

* Requests:

Requests est une bibliothèque HTTP pour le langage de programmation Python. L'objectif du projet est de rendre les requêtes HTTP plus simples et plus conviviales

* bs4:

Beautiful Soup est une bibliothèque Python permettant d'extraire des données de fichiers HTML et XML. Il fonctionne avec votre analyseur préféré pour fournir des moyens idiomatiques de navigation, de recherche et de modification de l'arbre d'analyse. Cela permet généralement aux programmeurs d'économiser des heures ou des jours de travail.

* html5lib:

Html5lib est une bibliothèque pure-python pour l'analyse HTML. Il est conçu pour se conformer à la spécification WHATWG HTML, telle qu'elle est implémentée par tous les principaux navigateurs Web.

* PyMySQL :

PyMySQL est une bibliothèque client MySQL purement Python, ce qui signifie qu'il s'agit d'un package Python qui crée une interface API pour nous permettre d'accéder aux bases de données relationnelles MySQL.

* Flask-JSON:

Flask-JSON est une extension simple qui ajoute un meilleur support JSON à l'application Flask. Il aide à gérer les requêtes basées sur JSON

* scrapy:

Scrapy est un framework open-source permettant la création de robots d'indexation. Développé en Python, il dispose d'une forte communauté, offrant de nombreux modules supplémentaires.

* Attrs :

Attrs est le package Python qui ramènera la joie d'écrire des classes en vous soulageant de la corvée d'implémentation de protocoles d'objet (alias méthodes dunder). Reconnu par la NASA pour les missions sur Mars depuis 2020 !