**Module 10 Challenge Report**

**(Francis Odo)**

**Background**

This exercise gives an overview of web scrapping, along with tools and techniques of gathering accessible web information.

**Objective**

Our objective is to utilize tools such as BeautifulSoup, Splinter to automate a web browser and scrape high-resolution images. The information is stored in MongoDB database. We can then update the web application and Flask to display the scraped data using MongaDB as the data storage entity. Bootstrap is applied to style the web app.

**Tools/Environment**

Download and install the following in Python/Panda environment:

Bs4 BeautifulSoup

Splinter/Browser

MongoDB database

PyMongo

Ensure that Flask is operational

**Code Plan**

1. Use the Visual Studio Code environment to create app.py app with the appropriate route.
2. Import all the necessary and related dependencies for the development environment
3. Set up variable/config for a URI to MongoDB for saving scrapped data
4. Define and create functions to scrape with designated app route to target web site. Bind URL to functions.
5. Create Button to scrape
6. Create Button to redirect
7. Scrape data, hold it in MongoDB, update and create new document. Store in JSON Key:Value formart, which is “img\_url : “title” using Pandas dictionary method.

**Approach**

* The app.py application is used to reach a scraping target with the route in conjunction with the index.html file.
* Python program is used for automated programing instruction sets to perform scraping and storing scraped data in MongoDB
* MongDB provides a NoSQL storage using document format.

Target URL and Titles:

1. <https://astrogeology.usgs.gov/search/map/Mars/Viking/cerberus_enhanced>

[Cerberus Hemisphere Enhanced](https://astrogeology.usgs.gov/search/map/Mars/Viking/cerberus_enhanced)

1. <https://astrogeology.usgs.gov/search/map/Mars/Viking/schiaparelli_enhanced>

[Schiaparelli Hemisphere Enhanced](https://astrogeology.usgs.gov/search/map/Mars/Viking/schiaparelli_enhanced)

1. <https://astrogeology.usgs.gov/search/map/Mars/Viking/syrtis_major_enhanced>

[Syrtis Major Hemisphere Enhanced](https://astrogeology.usgs.gov/search/map/Mars/Viking/syrtis_major_enhanced)

1. <https://astrogeology.usgs.gov/search/map/Mars/Viking/valles_marineris_enhanced>

[Valles Marineris Hemisphere Enhanced](https://astrogeology.usgs.gov/search/map/Mars/Viking/valles_marineris_enhanced)

**Risk**

There are a number of reasons why the application may fail to execute successfully. Among them are:

1. Target website is constantly or regularly being updated
2. Local web server (Flask) may be down
3. System may run into memory issue and slow to process programing instructions
4. Database may fail to start for some reason

**Conclusion**

The process is proven to work. However, more time and fine-tuning is needed in order to fully take advantage of the capability of all available resources and get the most out of the application. This is just a scratch in the surface.