

Web Scraping Challenge Homework

Terrence Cummings

DOCUMENTATION

Key Github documents:

1. Mission_to_mars.ipynb
 - a. Jupyter notebook with Mars data scraping code
2. Scrape_mars.py
 - a. Script containing the Python code from the mission_to_mars.ipynb.
 - b. Called by “/scrape” route of app.py.
3. App.py
 - a. Flask app with “/” route that renders index.html.
 - b. “/scrape” route calls scrape_mars.py script.
4. Index.html (in templates directory)
 - a. Main web page
 - b. Rendered by app.py
5. Mars_webpage_screenshot.pdf
 - a. A screenshot of the final webpage.

Key features of webpage (index.html)

1. Jumbotron from Bootstrap
 - a. Refresh Data
 - i. Initiates a new scrape of all the websites.
 - ii. Results are put in a Python dictionary and stored in a MongoDB document which is read into the webpage by app.py “/” route redirect.
 - b. View Featured Image
 - i. Jumbotron background is the featured image from the scrape.
 - ii. Button opens separate window to view the full size feature image.
2. News Articles
 - a. Scrape reads in date, headline, summary, and URL of 40 news articles from the first page of the NASA Mars news site.
 - b. Use Bootstrap carousel to provide access to the articles.
 - c. Used {% for...%} to populate the carousel.
 - d. “NASA” word is a hyperlink to the NASA website articles page.
 - e. Article title is a hyperlink to the full article on the NASA website.
3. Mars hemispheres
 - a. Uses a Bootstrap carousel to access the 4 hemispheres
 - b. Title of each carousel slide is a hyperlink to the full size image of the hemisphere
 - c. Used {% for...%} to populate the carousel.
4. Current Mars weather
 - a. As required, and additionally includes a hyperlink text to the Twitter feed.
 - b. Scrape looks for first weather related post since there are non-weather related posts on this feed.

5. Mars Facts

- a. As required, and additionally includes a hyperlink text to the source website.
- b. Used {% for...%} to populate the table.