

Advanced Web Development and Web Scraping

Dr. Garrett Dancik

Why learn “advanced web development”

- You can develop web pages (obviously)
- You can add features to web pages (i.e., develop browser extensions)
- You can answer questions by extracting and parsing (scraping) web page data
- You can retrieve information and answer questions by using web page application program interfaces (APIs)
- You can automate web-tasks such as logging onto a web site

What did (will) you guys do?

- Entertainment (video, music, and games)
 - Wordle rankings by state
 - Spotify Artist Comparison
 - Artist / Song Lookup (Genius API)
 - Shopping
 - Game Slicer (comparison of video game prices)
 - Sports
 - Top 50 Nationally Ranked Countries in Cricket
 - NBA 2021-2022 Season Standings
 - Planet Football
 - NBA Teams Data Visualizer
- Color legend
 - Browser extension
 - Shiny application
 - Web scraper / API

What did (will) you guys do?

- Information / Knowledge

- Webpage for Stock Information
- Nutritionr – nutritional facts and recipes
- Weather information
- Cancer Data Visualizer

- Color legend
 - Browser extension
 - Shiny application
 - Web scraper / API

- Productivity

- Dictionary / Thesaurus Web Extension
- Steam Store Scraper

Other examples

- Identification of bots on Twitter:
 - <http://www.pewresearch.org/fact-tank/2018/04/19/qa-how-pew-research-center-identified-bots-on-twitter/>
- HiPub Chrome extension for knowledge discovery:
 - <https://www.ncbi.nlm.nih.gov/pubmed/27485446>
- The Cambridge Carcinoma of the Prostate App
 - <https://bioinformatics.cruk.cam.ac.uk/apps/camcAPP/>
- *shinyGEO*: analysis of gene expression data
 - <https://gdancik.github.io/shinyGEO/>
- CongressEdits: Tracks changes to Wikipedia made from IP addresses from the U.S. Congress
 - <https://source.opennews.org/articles/automating-transparency/>
- COVID-19 in CT
 - <https://eida.easternct.edu>

More examples

- Tweepy –
 - https://docs.tweepy.org/en/stable/getting_started.html
 - <https://github.com/tweepy/tweepy>
 - <https://github.com/tweepy/tweepy/blob/master/tweepy/streaming.py>
 - <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/sample-realtime/api-reference/get-statuses-sample>
- Can you develop a Python module (let's call it PokéPy) for accessing the Pokemon API (<https://pokeapi.co/>)?