PIC 16, Winter 2018 – Assignment 9W

Assigned 3/7/2018. Code (a `.zip`ped Scrapy project folder file) due by the end of class 3/12/2018 on CCLE. Hand in a printout of this document with the self-assessment portion completed by the end of class on 3/12/2018

In this assignment, you will use Scrapy to help you download the names, numbers, and descriptions of Pokemon.

Task

Here some steps to help you divide the work into smaller chunks.

1. First, you might want to review video of Lecture 8W 2016 and/or Discussion 8R 2016. As in Lecture 8W this year, I reviewed how to inspect elements of an HTML document in the Chrome browser, generate ideas for potential CSS expressions, how to test CSS expressions in a Scrapy shell, and how to set up a Scrapy project. However, I also showed how you can use Scrapy to parse links to other pages, generate new `Request`s, and parse these additional pages. That's why they're called Spiders, after all: they crawl the web.
2. Create your own Scrapy project to produce a `.csv` file with just the names and numbers of the Pokemon listed at https://pokemondb.net/pokedex/national. Note that the command for producing a `.csv` is identical to that of producing a `.json` (except for the file extension `.csv` rather than `.json`). This is about as simple as it gets, so getting the bugs worked out here will help a lot before moving on.
3. *Recommended: find* `settings.py` *within your Scrapy project and uncomment all lines that begin with* `AUTOTHROTTLE`. *This should slow down your spider to avoid upsetting the server when it starts visiting links.*
   Modify your spider to visit the link corresponding with each Pokemon. That is, your `parse` method should `yield` a `Request` rather than a dictionary with the name and number.
   Then, instead of producing the list of names and numbers, just collect the first "Pokedex Entry" of each Pokemon. That is, in your method for parsing each Pokemon's individual page, `yield` a dictionary with just the Pokemon's first Pokedex entry.
   (Don't worry if the word "Pokemon" doesn't look right when you open the CSV file in Excel; this is Excel's fault.)
   Recommended: debug visiting the links to only the first 10 Pokemon to save time. Before turning this in, make sure it works will all 802 Pokemon.
4. One way of collecting the name, number, and description is to collect all of that information from the Pokemon's individual page. However, the way I'd like you to do it (that is, the way it needs to be done for credit on this assignment) is to combine the Pokedex entry from the individual Pokemon's page with the name and number from the main page using the trick shown here. The idea is that you can pass "meta" information (also a dictionary) within a `Request` object that will be included in the corresponding `Response`.

Self-Assessment

Print the assignment document and check off the steps 2-4 that you completed successfully. Give yourself 1/3 credit for each and write your total score below: