

MSR1 Writeup

Part 1

Methods

After installing required dependencies and importing the libraries, I authenticated the Github API in order to have a higher-rate limit for submitting requests to the API. Due to the size of the data I was extracting from the discourse/discourse repository for this assignment, I wrote helper functions to read the fetched JSON data into a local file. This helped with run-time efficiency and validity of the data.

Q1

Methods

I extracted all the pull requests and commits made for discourse/discourse by using a GET command for “/repos/discourse/discourse/pulls” and “/repos/discourse/discourse/commits”. In order to identify the unique contributors for this project (where a contribution includes commits and pull requests), I stored all the usernames for each pull request in an array. I repeated the same for the commits, but also mined the author’s name for the commits that didn’t include the username metadata. Then, I used the `most_common()` function in the Counter class from the “collections” library to return a unique list with frequencies of each username, and took the length of that list to find the number of unique contributors.

Results

There are a total of 1251 unique contributors in the discourse/discourse project.

Q2

Methods

By passing in a parameter in order to find the largest count in the `most_common()` function I found the tuple pair (username, count) for the user with the most contributions in discourse/discourse.

Results

In the discourse/discourse project, the user “eviltrout” made 6556 contributions, which is the most for this repository.

Q3

Methods

I extracted all the commits made to the badge.rb file by using a GET command for “/repos/discourse/discourse/commits?path=app/models/badge.rb” - in this url, I passed the path parameter in order to navigate to the badge file. I repeated the methods for Q1 and Q2 in order to find the top contributor to this file. As I was mining for commits, I ran into the same issue where some author metadata was missing, so I mitigated that by mining the names as before.

Results

The user “SamSaffron” made 55 contributions to the badge.rb file in the discourse/discourse project, which is the most.

Part 2

Methods

I first mined all repositories owned by Dan Abramov by calling a GET to “users/gaearon/repos” and stored a list of the repo names in an array.

Q1

Methods

In order to find how many of his own projects Dan committed to the most often in the past 3 years, I first got a list of all the commits he made to his repositories in the prescribed time range (June 1, 2017 to June 1, 2020) by iterating through all the repos and storing the number of commits made to each repository in a dictionary. From there, I identified the repository with the most number of commits.

Results

Dan made the most number of commits (250) to the overreacted.io repository.

Q2

Methods

In order to identify when Dan first made a release in the facebook/react project, I pulled all the releases made in the facebook/react repository (by calling a GET on “repos/facebook/react/releases”) and stored in a JSON file. From there, I mined for releases made by “gaearon” (Dan’s username) and stored the tag name and corresponding timestamp in a dictionary. With this information, I took the earliest timestamp from the dictionary.

Results

Dan made his first release (v0.14.8) on March 29, 2016 at 5:35 PM in the facebook/react repository.