Group 11 Final Project Design Document

Samantha Sample, Matthew Connealy, Michelle Huang, Sweta Pragyan Praharaj

Overall Design:

For our project, we'd like to create a website that uses existing endpoints from the Augur API, as well as new ones, to give insight to the health of a certain repo or repo group. Since this is a complicated and arbitrary determination, we will be focusing on a select few metrics that can help in determining overall code and project health.

Metric Designs:

1) **Contributor Occupation/Contributor Location**
    a) Integrate Google Maps on website
        i) Many users are already familiar with Google Maps
        ii) Allows for zooming and panning without affecting the pinned locations
    b) Users will pick a repository group through a selection box or module which will then pin all contributors within that repository group on the map
        i) This will reduce the number of pins on the map and allows for the user to distinguish between repository groups
    c) Clicking on pins will return the contributor's profile
        i) The profile will reveal additional information such as email, github ID, occupation, company, start date, and activity

    Data Used:
    augur_data.contributors: contrb_long, contrib_lat, cntrb_company
    augur_data.contributor_affiliations: ca_affiliation, ca_start_date

2) **Testing Coverage**
    a) User will select a particular repo they wish to view
    b) Matching the repo id with the repo and its respective test coverage will enable the subroutine and statement data to be extracted
    c) We can then calculate the subroutine test coverage using (subroutines tested/total subroutines)*100 and the same formula for statement coverage from the extracted data.
    d) Once the coverages are calculated, we can display the data to the user through some type of chart.

    Data Used:
    Augur_data.repo_test_coverage: file_subroutine_count
    Augur_data.repo_test_coverage: file_subroutines_tested

Augur_data.repo_test_coverage: file_statement_count
Augur_data.repo_test_coverage: file_statements_tested
Augur_data.repo_test_coverage: testing_tool

3) **Gender and Ethic Diversity Among Committers**
    a) Overall Procedure:
        i) The user will select a particular repo or repo group from which they would like to see the data.
        ii) Using the Augur_data.commits table, author name data will be collected for that repo (group) for each unique committer id. This will be conducted using an SQL api call to the Augur database
        iii) Using the NamSor API, each name will be passed to the API, which returns a predicted gender and value from -1 to 1 that correlates to the strength of the prediction, with -1 being male and 1 being female.
        iv) Using a threshold of abs(.4), the percentage of male vs female contributors will be returned as a json response
    b) Necessary Steps:
        i) Create a new metric function which will access and evaluate the data, returning a convenient data form that can be easily used to create visualizations.
        ii) Create a new endpoint with which the metric can be accessed for a particular repo (group).
    c) Necessary Tools:
        i) NamSor API
        ii) Python, SQL

Data Used:
Augur_data.commits: cmt_author_name
Augur_data.commits: cmt_ght_committer_id
Augur_data.commits: repo_id