

sddec19-09: Finding Hay in the Needle Stack: Big Data, and Doing the Right Thing at the Right Time

Yealim Sung, Mwiza Mireille Iradukunda, Gooi Yin Hong, Austin Konsor

Client: Collins Aerospace

Advisor: Thomas Daniels

Problem Statement

- Collins Aerospace has about 8000 common test files that are spread across multiple directories in an SVN. Among these files, some have the same functions trying to achieve the same thing. However, they produce different results, as some may have had bugs fixed while others did not or some may be outdated.

Solution

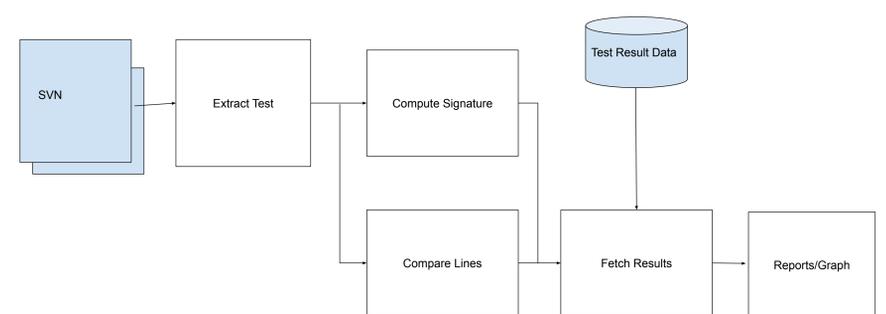
- An application that will recommend and visualize the best file/function to be using given a desired test.

Intended users and uses

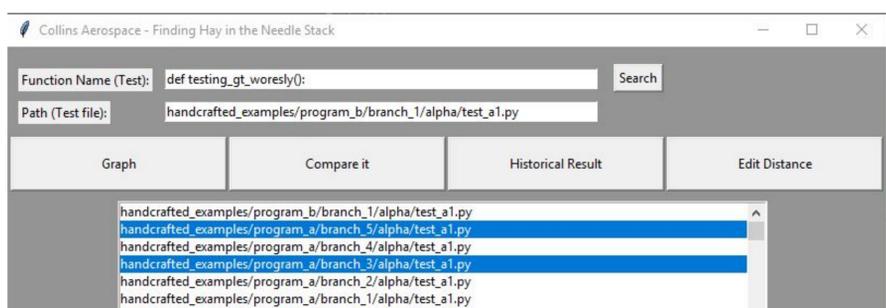
User: Collins Aerospace Engineer.

Use: The product is a visual displaying the performance of scripts, and recommendations given a signature as input.

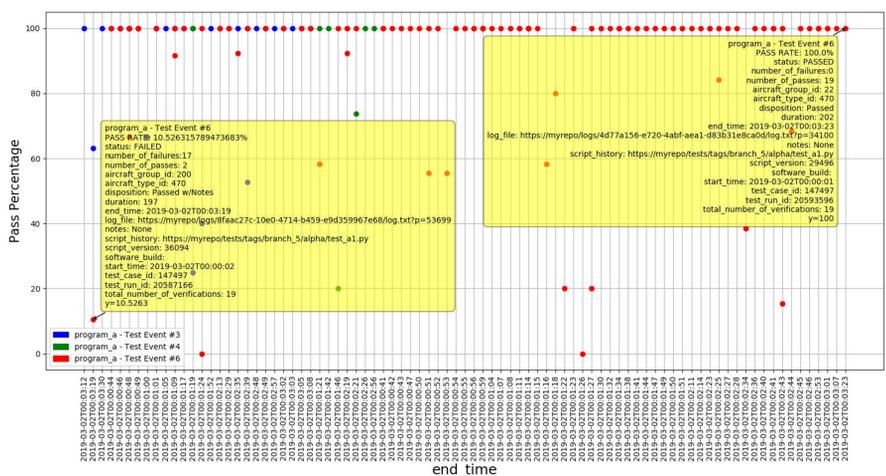
Design Approach



Interface Design



script_history	Program	Last Run Date	Overall Perc	Test runs No	Edit Distance
1 https://myrepo/teststags/branch_4/alpha/test_a1.py	186 program_a_json	2019-03-02 00:01:42	94.82	109	0
2 https://myrepo/teststags/trunk/alpha/test_a1.py	205 program_b_json	2019-03-02 00:02:09	90.65	295	1
3 https://myrepo/teststags/branch_5/alpha/test_a1.py	186 program_a_json	2019-03-02 00:02:44	90.15	103	0
4 https://myrepo/teststags/branch_2/alpha/test_a1.py	205 program_b_json	2019-03-02 00:00:54	88.71	298	3
5 https://myrepo/teststags/branch_3/alpha/test_a1.py	186 program_a_json	2019-03-02 00:03:42	88.32	81	0
6 https://myrepo/teststags/branch_1/alpha/test_a1.py	205 program_b_json	2019-03-02 00:04:17	85.87	283	16



Technical Details

Data Extract Library:

- Extracts test result data from json files.

Directory Crawler:

- Finds all functions and files throughout the SVN directory.

Customizable and Interactive Graph:

- Customizable 2D dynamic graph
- Using Matplotlib

Tables for recommendation:

- Displays data on test cases such as average score, file path, and date of last time used. Great for recommendation of files.

- Tkintertable

File Comparison:

- A page that compares two files side by side showing additions, deletions and changes line by line.
- Subprocess of Beyond Compare

Edit Distance:

- This is a value that, without having to view the files, will display the number of edits required to get from one file to the next.

- Edit_distance

Other Modules used:

- Pandas, Plotly, yapf, pmw, mplcursors

Testing

- Unit
- Regression
- User
- Integration

Standards

- P16085 - ISO/IEC/IEEE International Draft Standard - Systems and Software Engineering - Life Cycle Processes - Risk Management
- P15026-4 - IEEE Draft Standard - Systems and Software Engineering - Systems and Software Assurance - Part 4: Assurance in the Life Cycle