Venu Vardhan Reddy Tekula
venuvardhanreddytekula8@gmail.com
Hyderabad, India
Timezone: Indian Standard Time (UTC+5:30)

Creating Quality models using GrimoireLab and CHAOSS metrics

GSoC Project proposal for CHAOSS

Personal Details and Contact Information

Full name: Venu Vardhan Reddy Tekula

GitHub username: vchrombie

Email: venuvardhanreddytekula8@gmail.com

University: Amrita Vishwa Vidyapeetham, Amritapuri.

Timezone: Kolkata, India (GMT +5:30)

IRC nickname: vchrombie

Website: vchrombie.github.io/

Blog: vchrombie.github.io/blog/

Twitter: twitter.com/vchrombie

Microtasks: vchrombie/chaoss-microtasks/

Project Link: chaoss/grimoirelab/issues/287

Synopsis

GrimoireLab is a powerful open-source platform that provides support for monitoring and in-depth analysis of software projects. It produces a rich set of dashboards, which can be easily inspected by decision-makers to help them understand the evolution and health of their projects. Despite the large set of dashboards available in GrimoireLab, comparing projects between each other is not straightforward since it requires navigating and drilling down the data in different dashboards.

Prosoul is a web application that empowers decision-makers with the means to create and manage their own quality models, which are useful means to evaluate and compare software projects. This project idea is about supporting the definition of Quality Models using GrimoireLab data and Prosoul.

The main aim of the project is to design an approach to shape the GrimoireLab data in a format that can easily be consumed by Prosoul and implement it on the data obtained from a few data sources like git, github and mailing list repositories to obtain simple quality models.

I would be closely working with Perceval, ELK, Mordred and Prosoul repositories.

The mentors for this project are David Moreno Lumbreras (@dlumbrer), Aniruddha Karajgi (@Polaris000), Santiago Dueñas (@sduenas) and Valerio Cosentino (@valeriocos).

Benefits to the Community

- Quality Models is a powerful way to evaluate and compare software projects and Prosoul helps to create and manage these quality models and it will be a great addition to the GrimoireLab components.
- CHAOSS metrics can be used to create the quality models for the different GrimoireLab data sources.
- This project can be extended to other CHAOSS metrics and GrimoireLab data sources like gitlab, slack in future.

Current Status of the Project

Most of the work regarding this project idea revolves around Prosoul which is a software quality models manager to create, import/export, view, and edit quality models. It has been implemented in the context of the EU funded research project CROSSMINER.

In the initial version, the metrics supported will be the CROSSMINER ones and the Kibana dashboards will be created using GrimoireLab. Based on the models, visualizations and assessment for the projects are generated.

Goals

Goal 1

Designing an approach to shaping GrimoireLab data in a format that can be easily consumed by Prosoul.

Goal 2

Implementing the approach with GrimoireLab data obtained from git, github, mailing lists repositories to obtain simple quality models using the CHAOSS metrics.

Goal 3

Improving the Prosoul UI to simplify the management of quality models.

Stretch Goals: Add support for multiple formats for importing/exporting quality models.

Deliverables

Deliverable 1

By the end of the first phase, the evaluation of both the approaches, studies & enrichers would be completed and one approach will be selected based on the report. Using the selected approach the enriched indexes for git, github and mailing list repositories will be created.

Deliverable 2

By the end of the second phase, simple quality models are created for the git, github and mailing repositories. The assessments for these quality models will be completed using the generated enriched indexes.

Deliverable 3

By the end of the project, the quality models using the GrimoireLab metrics for git, github and mailing list repositories will be completed. There will be changes in the UI of Prosoul to simplify the management of quality models.

Expected Results

The final goal is to have a tool for "Automatic Project Assessment and Visualization based on Evolved Quality Models" using the GrimoireLab metrics.

Approach

Goal 1: Designing an approach to shaping GrimoireLab data in a format that can be easily consumed by Prosoul.

In the current version of the Prosoul, the metrics supported are of the CROSSMINER ones like scava-metrics. The project dashboard for scava-metrics is available here, Scava Project Dashboard.

The current GrimoireLab data cannot be fed to Prosoul directly. So, we need a way to process the data into enriched indexes in a required format that can be fed to Prosoul directly, without much change to the current application architecture. There are two possible approaches for this task, using the studies or writing a new enricher for processing the data.

The ELK connectors are basically pipelines to connect the Perceval, raw and enrichers. We can define a new enricher for the required data sources and create a new connector from the parent connector replacing only the enricher.

"git": [Git, GitOcean, GitEnrich, GitCommand],
"git_qm": [Git, GitOcean, GitEnrichQM, GitCommand],

"github": [GitHub, GitHubOcean, GitHubEnrich, GitHubCommand],
"github_qm": [GitHub, GitHubOcean, GitHubEnrichQM, GitHubCommand],

Using these new connectors, we can generate the new enriched indexes (qm_git, qm_github) which can be further used for the assessment of the Quality Models. The indexes can be tested using the dev tools in Kibiter.

After evaluating both approaches, I plan to submit a report based on the results of the two approaches and decide to choose the best approach possible along with the mentors. Using the best approach, I plan to create the enriched indexes for the git, github and mailing list repositories.

Goal 2: Implementing the approach with GrimoireLab data obtained from git, github, mailing lists repositories to obtain simple quality models.

A QM is identified by a title and composed of several Goals, which aim at measuring specific aspects of your projects, such as Activity, Community and Processes. Each goal is defined by a set of Attributes, they aim at characterizing a given goal. For instance, the Activity goal could include Code and Bugs, while Community could have Coders and Reporters as attributes. Each attribute is then mapped to one or more Metrics. A Metric has a name, a 5-level threshold (used to rate the project wrt a metric) and the data implementing the metric.

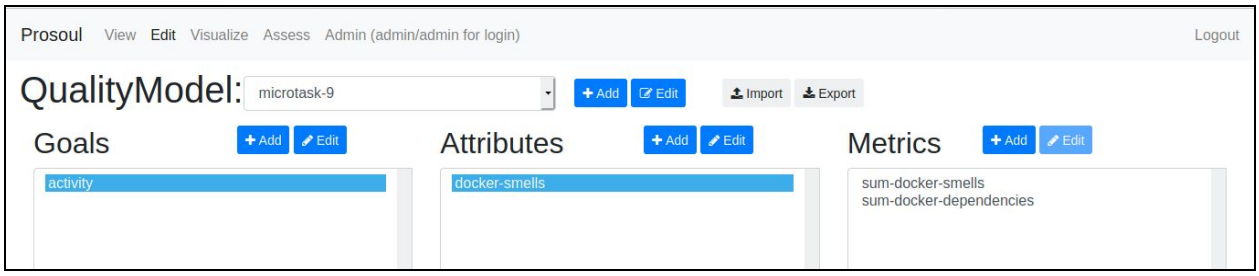
In the case of CROSSMINER, the metric implementation is the metric_name field of the metrics extracted from SCAVA and stored in the index scava-metrics. For example, the attribute Code previously created could be mapped on the SCAVA metric Patches, while the attribute Bugs could be represented by the SCAVA metrics Bugs, Won't-Fix Bugs and Fixed Bugs.

The process of creating a quality model, visualization and assessment using the enriched index is clearly explained in the microtask-9. Here, I used the scava-metrics index for the assessment.

In this phase, I would work on creating simple quality models using the CHAOSS metrics for GrimoireLab data (git, github and mailing list repositories). The assessments can be completed using the enriched indexes created in the first coding phase.

Goal 3: Improving the Prosoul UI to simplify the management of quality models.

There is a huge scope of improving the UI of Prosoul. One plan is to simplify the management of quality models.



First, you need to select the quality model and there will be three boxes for managing the Goals, Attributes and Metrics. One approach to simplify this is by having a tree structure.

→ 1 quality model
- →1.1 goal-1
  - → 1.1.1 attribute-1
    - → 1.1.1.1 metric-1
    - → 1.1.1.2 metric-2
  - → 1.1.2 attribute-2
    - → 1.1.2.1 metric-1
- → 1.2 goal-2
  - → 1.2.1 attribute-1
    - → 1.2.1.1 metric-1
    - → 1.2.1.2 metric-2
    - → 1.2.1.3 metric-3

The approach to this is having an Add or Edit button to add a configuration or edit an existing configuration of a particular field. So, you need to select a field in the above tree and use the Add/Edit buttons.

As of now, the quality models can be exported to JSON files. This can be further extended to multiple formats for importing/exporting quality models. As stated in the Stretch Goals, the support can be extended to YML and plain text files with a definite format (Bitergia/prosoul#4).

Timeline

Until May 4: During the application review period, I plan to discuss and work on improvements that can be made in Prosoul like helping with the unit testing Bitergia/prosoul#53. This would help me in the final weeks of the coding period. I would also work on a few existing issues like chaoss/grimoirelab-perceval#630 and chaoss/grimoirelab-elk#803 and understand the GrimoireLab toolchain better. I will be watching out for the CFP opening for the CHAOSScon NA 2020 conference and submit a talk proposing the work I am planning to do for this summer. Also, I would be watching out for the Python India 2020 conference too.

May 4 - June 1 (Community Bonding): This phase would involve understanding how the community works and how I can participate in the discussion that happens, setting-up a clear way of communication with the mentors, determining the tasks that are of highest urgency to complete.

June 1 - June 15 (Week 1 and 2): During this phase, I would work on learning how the studies & enrichers work and how to implement them to convert the existing GrimoireLab data to some format which can be fed to Prosoul. At the end of the two weeks, I would be making a detailed report on both the approaches and select the best one along with the mentors. I will also publish these reports as a blog post explaining the work.

June 15 - June 29 (Week 3 and 4): After selecting the suitable approach after having a discussion with the mentors, I will work on creating the enriched indexes from the raw data during these two weeks. I would complete this work by the end of this period. I would also work on writing tests for the same.

(The above-defined task is expected to be completed before First Evaluation)

June 29 - July 13 (Week 5 and 6): I would be working on creating quality models for the GrimoireLab data, git and mailing lists repositories. I would be doing the assessments for these quality models using the generated enriched indexes. During this period, I will be making a video explaining about the various aspects of a Quality Model.

July 13 - July 27 (Week 7 and 8): I would be working on creating quality models for the GrimoireLab data, github repositories. GitHub needs some extra time as there will be a

lot of things that can be extracted which can be used for making good quality models. I would be doing the assessments for these quality models using the generated enriched indexes.

(The above-defined task is expected to be completed before Second Evaluation)

July 27 - August 10 (Week 9 and 10): I would be concentrating on cleaning up the UI to simplify the management of quality models. I would be discussing the implementation of the idea which I proposed and work on it. If I get some extra time, I plan to work on fixing some UI issues and work on some existing UI enhancements.

August 10 - August 24 (Week 11 and 12): During this week, I would specifically focus on testing the application and the enhancements to Prosoul, including unit testing and integration testing. I would also work on adding support to a few other formats for importing/exporting the quality models.

August 24 - August 31 (Final Evaluation): This would work as a *buffer period* for suggestions and improvements that are needed in the implementation and in case there's a delay in any point in the described timeline. I plan to work on the end evaluation report. I will also be making a demo video of the work I did for the GSoC.

After the GSoC period: I am happy to maintain the project once the GSoC period ends and also helps with adding quality models for a few more other backends like gitlab, slack, etc. I would be contributing to other GrimoireLab tools too. After completing the project, I will be submitting a talk proposal based on the work done for FOSDEM 2021.

About Me

- My name is Venu Vardhan Reddy Tekula. I am a final year undergraduate student pursuing my Bachelor's in Computer Science and Engineering (CSE) from Amrita Vishwa Vidyapeetham, Kollam, India. I was born and brought up in Hyderabad, India.
- I am a member of FOSS@Amrita, (Open Source Club of Amrita) which trains students to encourage contributions to free and open-source software and skill development. I actively mentor juniors and participate in various club activities.
- My areas of interest are Data Analytics, Cyber Security and Machine Learning.
- I have good experience with Python. I worked with Python for almost three years now. A few of my projects include amfoss/GitLit and snitch3s/hackbunch.

- I have been contributing to CHAOSS for almost a year. I tried for GSoC 2019 too and this repo has the micro-tasks for last year, vchrombie/chaoss-microtasks-19.
- Along with the contributions, I have been actively engaged with the CHAOSS community too by attending the weekly calls and working group meetings.
- I was a part of the team who worked extensively for CHAOSS participation in the Google Code-In program (GCI 2019) under the Linux Foundation organization. I am looking forward to driving the CHAOSS participation in GCI again this year.
- I have completed all the 10 microtasks listed on the ideas page. The link to the microtasks repository is vchrombie/chaoss-microtasks. I have been fairly comfortable with the Prosoul's codebase and I have worked on some UI improvements, fixing a build error and some documentation help, contributions to Prosoul.

Some more information

- I intend to continue working with the GrimoireLab project after the Summer of Code program.
- My last semester will be completed by May 2020. I can spend 40+ hours a week for the GSoC project if I get selected. Also, I don't have any commitments other than GSoC.
- I intend to keep in touch with the mentors and update them weekly once through the IRC meeting with my progress and clarifying my doubts through emails.
- I will continue to attend the GrimoireLab Working Group meetings as well as the CHAOSS weekly call and actively be a part of the discussions.
- I plan to write a blog post at the end of every week and put that up on my personal blog. This would be helpful for the mentors to review my work on a weekly basis.

Note of Thanks

I would like to thank Valerio and Aniruddha for their valuable suggestions and continuous help with the microtasks, proposal and also the other members of the CHAOSS community for their suggestions and help.

References:

1. https://github.com/chaoss/grimoirelab/issues/287
2. https://github.com/Bitergia/prosoul/
3. https://github.com/crossminer/crossminer