Exploring Software Library Metrics

Instructor: Sarah Nadi
Student: Monica Bui

Fall 2018

1 Course Objectives

Application Programming Interfaces (APIs) are a reuse mechanism that allows developers to exploit the functionality offered by an existing library or to extend the functionality of an existing framework. Developers often have the choice between the APIs of several libraries or frameworks that provide similar functionality. However, it is not clear how a developer can determine which library is more suitable for their needs?

The goal of the course is to explore and compare different metrics that can be used to measure various characteristics of a software library. Specifically, goals include:

- Exploring literature related to API/library usage and library quality.

- Mining different data sources such as bug repositories, version control repositories, and question-answer websites to extract library metrics

- Conducting a project related to the topic of the course and writing a research paper about it

2 Prerequisites

Previous software engineering course. Experience working on large projects.

3 Workload

In this course, the student will:

- Read/discuss work related to library quality and metrics, in terms of bug reports, evolution, and usage.

- Apply mining software repositories techniques to extract data used to calculate library metrics

- Develop github badges for these metrics

4 Timeline

There will be weekly meetings to discuss papers, reviews, and project progress with the instructor. The student is expected to attend the weekly scrum meeting of the Software Maintenance and Reuse (SMR), which happens Mondays 1pm in CSC 2-59. The individual meeting will the student will then happen after the scrum meeting.

- Weeks 1 - 4: Read related literature and explore the existing library metrics and data sources.

- Week 5: Decide on a at least one new metric to extract in the course

- Weeks 6 - 8: Implement the extraction for the new library metric(s) and integrate in `http://smr.cs.ualberta.ca/comparelibraries/`

- Weeks 9 - 12: Create GitHub badges for all existing metrics, as well as the new one developed in the course

- Weeks 13 - 14: Write final course report describing complete methodology of the new metric and the description of the created badges. Report must include a link to a Github repository with all scripts, code, and collected data during the course.

5 Evaluation

- 15% Literature review (due Oct. 1st)

- 5% 1 page description of the proposed new metric and its extraction method (due Oct. 8th)

- 15% Scripts and demo of the extracted metric (due Oct. 29th)

- 15% Code and demo of created GitHub badges (due Nov. 26th)

- 15% Final presentation (during the week of Dec. 3 - 7)

- 10% Code deliverable: final polished Github repository with all code and data. Must have a clear ReadMe file with all instructions needed for running and/or replicating things (due Dec. 14)

- 25% Final project write-up (due Dec. 14)

6   Literature

[1] C. Chen, S. Gao, and Z. Xing, "Mining analogical libraries in q&a discussions–incorporating relational and categorical knowledge into word embedding," in *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, vol. 1.   IEEE, 2016, pp. 338–348.

[2] Y. M. Mileva, V. Dallmeier, M. Burger, and A. Zeller, "Mining trends of library usage," in *Proceedings of the joint International and Annual ERCIM workshops on Principles of software evolution (IWPSE) and software evolution (Evol) workshops.*   ACM, 2009, pp. 57–62.

[3] Y. M. Mileva, V. Dallmeier, and A. Zeller, *Mining API Popularity.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 173–180.

[4] A. Hora and M. T. Valente, "apiwave: Keeping track of api popularity and migration," in *Proceedings of the IEEE International Conference on Software Maintenance and Evolution (ICSME).*   IEEE, 2015, pp. 321–323.

[5] D. Pletea, B. Vasilescu, and A. Serebrenik, "Security and emotion: sentiment analysis of security discussions on github," in *Proceedings of the 11th Working Conference on Mining Software Repositories.* ACM, 2014, pp. 348–351.

[6] F. L. de la Mora and S. Nadi, "Which library should i use?: a metric-based comparison of software libraries," in *Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results.*   ACM, 2018, pp. 37–40.

[7] A. Trockman, S. Zhou, C. Kästner, and B. Vasilescu, "Adding sparkle to social coding: An empirical study of repository badges in the npm ecosystem," in *Proceedings of the 40th International Conference on Software Engineering.* ACM, 2018, pp. 511–522.

[8] F. L. de la Mora and S. Nadi, "An empirical study of metric-based comparisons of software libraries," in *Proceedings of the 14th International Conference on Models and Data Analytics in Software Engineering*, 2018.

[9] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. C. Gall, "How can i improve my app? classifying user reviews for software maintenance and evolution," in *Proceedings of the IEEE International Conference on Software Maintenance and Evolution (ICSME).* IEEE, 2015, pp. 281–290.

[10] "Stackoverflow APIs." [Online]. Available: https://api.stackexchange.com/

[11] "Stackoverflow data dump." [Online]. Available: https://archive.org/details/stackexchange

[12] T. F. Bissyandé, D. Lo, L. Jiang, L. Réveillere, J. Klein, and Y. Le Traon, "Got issues? who cares about it? a large scale investigation of issue trackers from github," in *Proceedings of the IEEE International Symposium on Software Reliability Engineering (ISSRE).* IEEE, 2013, pp. 188–197.