deRSE25 and SE25 Timetables



Contribution ID: 152

Type: Poster

Who Do You Think You Are? Identifying Research Software Engineering Personas From Developer / Repository Interaction Data

Wednesday 26 February 2025 19:40 (20 minutes)

'Personas'are widely used within traditional software contexts during design to represent groups of users or developer types by generating descriptive concepts from user data.

'Social coding'practices and version control 'code forges'including GitHub allow fine-grained exploration of developers'coding behaviours though analysis of commit data and usage of repository and development management features such as issue tickets.

By combining software repository mining techniques with persona concepts, we have generated a novel taxonomy of Research Software Engineering Personas (RSE Personas).

This gives us insight into collaborative development best practices and helps represent developer-repository interactions. This work has initially been done on a dataset gathered from 10, 000 research software GitHub repositories which have been deposited with Zenodo.

RSE Personas identify distinct groupings of development behaviours by applying clustering analysis to data from larger collaborative research software project repositories. Correlations between groups of 'best practice' behaviours and common development and repository management activities are examined.

This poster explains the RSE Personas methodology, describes important Personas and their properties, shows key emerging findings at developer / repository levels, and explores future applications and potential caveats for this novel method.

Classification methods from Open Source Software research such as commit message keyword classification [1] and commit file type classification [2] are combined with factors such as review of pull requests, issue tickets, commit size and frequency data, commit 'activity types'(for example: coding new features, versus documenting or managing the repository) and other contributions information from GitHub to build a picture of developers'engagement with their repositories.

An example RSE Persona identified is the "Active Leader": responsible for a high proportion of commits, frequently assigned to issue tickets and pull request reviews; they modify files across the entire codebase, with strong contributions to activities such as developer documentation.

Personas demystify subtle interactions between researchers and their code, unlocking insights into the day-today behaviours of RSEs and the different contributions they make to their projects, and how those managing such projects could identify ways to better support their teams towards effective research software development.

Personas could allow RSEs to interact more effectively by understanding their current practices in relation to their teams and communities, helping them identify 'next steps best practices', boosting their professional - as well as software - development.

We agree that while RSEs are certainly far more than their code and the digital footprints they leave on their repositories, the Research Software Engineering Personas methodology now allows us to describe, explore and investigate the current real-world practices of contributors to research software, and we invite you to engage with our work.

References:

[1] L. P. Hattori and M. Lanza, 'On the nature of commits', in 2008 23rd IEEE/ACM International Conference on Automated Software Engineering - Workshops, Sep. 2008, pp. 63–71. DOI: 10.1109/ASEW.2008.4686322.

[2] B. Vasilescu, A. Serebrenik, M. Goeminne, and T. Mens, 'On the variation and specialisation of workload —A case study of the Gnome ecosystem community', Empir Software Eng, vol. 19, no. 4, pp. 955–1008, Aug. 2014, DOI: 10.1007/s10664-013-9244-1.

I want to participate in the youngRSE prize

no

Primary author: ANDERSON, Felicity (EPCC, University of Edinburgh)

Co-authors: Dr SINDT, Julien (EPCC, University of Edinburgh); CHUE HONG, Neil (University of Edinburgh)

Presenter: ANDERSON, Felicity (EPCC, University of Edinburgh)

Session Classification: Poster and Demo Session together with Reception

Track Classification: Research Software: RSE research