

## Guest Editorial: Special section on mining software repositories

Romain Robbes<sup>1</sup> · Emily Hill<sup>2</sup> · Christian Bird<sup>3</sup>

Published online: 24 March 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Developers use a variety of software tools in order to develop software systems. In turn, these tools record this usage in *software repositories*. The Mining Software Repositories (MSR) field analyzes the rich data available in software repositories to uncover interesting and actionable information about software systems and their developers. Some commonly explored areas include software evolution, models of software development processes, characterization of developers and their activities, prediction of future software qualities, use of machine learning techniques on software project data, software defect prediction, analysis of software change patterns, and analysis of code clones. This special issue highlights six recent MSR regular research papers, and one data showcase paper. The goal of data showcase papers is to describe at length a valuable Software Engineering dataset, in the hope that it allows prospective users of these datasets a smooth start with them.

The paper “The Impact of Rapid Release Cycles on the Integration Delay of Fixed Issues” by Alencar da Costa, McIntosh, Treude, Kulesza, and Hassan, presents a mixed-method study: a quantitative study of the Firefox project, followed up by a qualitative survey of 37 open-source contributors. The study finds that, contrary to common knowledge, adopting a rapid release cycle may not result in issues being integrated more quickly than when using a more traditional release process. The study then explores the possible reasons for this observation.

---

✉ Romain Robbes  
rrobbes@unibz.it

Emily Hill  
emhill@drew.edu

Christian Bird  
Christian.Bird@microsoft.com

<sup>1</sup> Software and Systems Engineering Research Group (SwSE), Free University of Bozen-Bolzano, Bolzano, Italy

<sup>2</sup> Drew University, Madison, USA

<sup>3</sup> Microsoft Research, Redmond, WA, USA

The paper “FEVER: An Approach to Analyze Feature-Oriented Changes and Artefact Co-Evolution in Highly Configurable Systems” by Dintzner, van Deursen, and Pinzger, focuses on the evolution of highly configurable systems, such as the Linux Kernel. The paper proposes an approach to capture feature-related changes in software evolution, validating the approach against 810 manually inspected commits, across 15 releases of the Linux Kernel, finding that FEVER had an overall accuracy of more than 85%.

In the paper “How the R Community Creates and Curates Knowledge: An Extended Study of Stack Overflow and Mailing Lists”, the authors Zagalsky, Germán, Storey, Teshima, and Poo-Camaño report on another mixed-method study of two communication channels of the R software development community: Stack Overflow and the R-Help mailing list. The study finds that there are two modes of knowledge creation: participatory (involving collaboration), and crowd sourced (with independent work). The study also investigates community participation patterns, finding for instance that some prolific users act as bridge of knowledge between the two channels.

The paper “Aggregating Association Rules to Improve Change Recommendation” by Rolfsnes, Moonen, Di Alesio, Behjati, and Binkley, improves on the state of the art in change recommendation algorithms by aggregating the results of multiple mined association rules together, instead of considering each rule in isolation. The approach is validated on 15 open-source software systems, and two industrial systems, finding that between 13 and 90% of change recommendations can be improved by rule aggregation.

In the paper “Addressing Problems with Replicability and Validity of Repository Mining Studies Through a Smart Data Platform”, the authors Trautsch, Herbold, Makedonski, and Grabowski present SmartSHARK, a cloud-based platform which is aimed at improving the replicability of MSR studies. The approach is validated through an experience report, in which several MSR approaches are replicated using the platform, and the experience is discussed.

The paper “Domain-Specific Cross-Language Relevant Question Retrieval” by Xu, Xing, Xia, Lo, and Li, introduces a cross-language information retrieval approach to support Chinese software developers when querying Stack Overflow. The approach allows the developers to query Stack Overflow in Chinese, formulating an English query from the initial query through a domain-specific translation. The approach is evaluated on 120 query Chinese questions, and compared with 4 baseline approaches.

Finally, the data showcase paper “Data Sets Describing The Circle of Life in Ruby Hosting, 2003-2016” by Megan Squire is a dataset covering more than a decade of evolution of two software ecosystems, RubyForge, and its successor, RubyGems. The dataset is extensively described and is part of the FLOSSmole repository.

**Acknowledgments** We are grateful for the continuous support and encouragement offered by the editorial board for the Journal of Empirical Software Engineering and by the Editors-in-Chief: Robert Feldt, Thomas Zimmermann, and Lionel Briand (Editor in Chief Emeritus). We also thank the authors for keeping up with the review schedule and the reviewers for their detailed and constructive comments which helped to shape the papers.