Preface

As a continuation of the last two years' special section on software systems, this special section is another effort to encourage and promote research to address challenges from the perspective of software systems. The goal of this special section is to present state-of-the-art and high-quality original research in the area of software systems.

Similar to the last two years' special section, this special section includes two major themes: data-driven software engineering and software testing and analysis.

Data-Driven Software Engineering. With the evolution of software systems and how software is used, data has started to play an increasingly important role in software development. Data, i.e., large-scale software artifacts of various types, is pervasive at different software development stages, and hidden in the data is information about the software quality, user experience, and the dynamics of software development.

Software Testing and Analysis. The theme of software testing and analysis encompasses all the research challenges that concern software quality achieved through software testing and analysis techniques that contribute in detecting/fixing defects in the end-product software. The ever increasing diversity, ubiquity, and dynamism of modern software systems are making such software testing and analysis more challenging.

The special section received 33 submissions. Many of them are of highly competitive quality. After the thorough review and revision process, we had to make very difficult decisions to accept only five articles to assemble this special section.

The accepted five peer-reviewed articles in this special section cover the major hot topics in the two major themes: data-driven software engineering and software testing and analysis.

The first article “What Are They Talking About? Analyzing Code Reviews in Pull-Based Development Model” by Li *et al.* discusses their qualitative study on three popular open-source software projects hosted on GitHub along with their fine-grained taxonomy and their preliminary quantitative analysis on a large set of review comments automatically labeled by a hybrid classification algorithm.

The second article “CSLabel: An Approach for Labelling Mobile App Reviews” by Zhang *et al.* analyzes 3 902 user reviews from 11 mobile apps in a Chinese app store, and finds out 17 issue types, along with presenting their proposed CSLabel approach for labeling user reviews based on the raised issue types.

The third article “A Cluster Based Feature Selection Method for Cross-Project Software Defect Prediction” by Ni *et al.* presents a cluster-based approach called FeSCH (Feature Selection Using Clusters of Hybrid-Data) to alleviate the distribution differences by feature selection in cross-project defect prediction.

The fourth article “On Locating Malicious Code in Piggybacked Android Apps” by Li *et al.* presents a tool-based approach called HookRanker, which provides ranked lists of potentially malicious packages based on the way malware behavior code is triggered.

The fifth article “Automated String Constraints Solving for Programs Containing String Manipulation Functions” by Zhang *et al.* presents an approach to automatically solve program constraints that involve strings and string manipulation functions.

We thank all the authors who submitted to this special section. We appreciate great help from the guest editors for the theme of data-driven software engineering: Yuanfang Cai (Drexel University, Philadelphia), Xuanzhe Liu (Peking University, Beijing), and Xiaoyin Wang (University of Texas at San Antonio, San Antonio), and the guest editors for the theme of software testing and analysis: Mithun P. Acharya (ABB Corporate Research, North Carolina), Marcelo d'Amorim (Federal University of Pernambuco, Recife), and Xiaoxing Ma (Nanjing University,

Nanjing). We are also deeply indebted to the reviewers who provided constructive reviews on the submissions in a tight schedule. Their contributions make this special section possible.

Leading Editor:

Tao Xie, Professor and Willett Faculty Scholar, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana taoxie@illinois.edu

Guest Editors:

Yuanfang Cai, Associate Professor, Drexel University, Philadelphia yfcai@cs.drexel.edu

Xuanzhe Liu, Associate Professor, Peking University, Beijing liuxuanzhe@pku.edu.cn

Xiaoyin Wang, Assistant Professor, Department of Computer Science, University of Texas at San Antonio
San Antonio xiaoyin.wang@utsa.edu

Mithun P. Acharya, Lead Principal Scientist, ABB Corporate Research, North Carolina
mithun.acharya@us.abb.com

Marcelo d'Amorim, Associate Professor, Federal University of Pernambuco, Recife damorim@cin.ufpe.br

Xiaoxing Ma, Professor and Deputy Director, Institute of Computer Software, Nanjing University, Nanjing
xxm@nju.edu.cn

Tao Xie is a professor and Willett Faculty scholar in the Department of Computer Science at the University of Illinois at Urbana-Champaign, Urbana. He received his Ph.D. degree in computer science from the University of Washington at Seattle in 2005. He received his M.S. degree in computer science from Peking University, Beijing, in 2000, and his B.S. degree in computer science from Fudan University, Shanghai, in 1997. He worked as a visiting researcher at Microsoft Research. His research interests are in software engineering, focusing on software testing, program analysis, software analytics, software security, and educational software engineering. He is a distinguished scientist of ACM and a senior member of IEEE.

Yuanfang Cai is an associate professor at Drexel University, Philadelphia. In 2006, she received her Ph.D. degree in computer science from the University of Virginia, Virginia, and started her career at Drexel upon graduation. She was granted tenure and promoted to associate professor in 2013. Her research focuses on software design, software architecture, software evolution, and software economics. Her recent work investigates architecture issues that are the root cause of software defects. Her research results have been licensed and adopted by multiple multinational corporations.

Xuanzhe Liu is an associate professor in Peking University, Beijing. His research interests are mainly focused on software analytics for the mobile/Web/services computing systems. He has published more than 60 referred papers, including papers in premier conferences such as WWW, ICSE, ASE, OOPSLA, UbiComp, and IMC, and high-impact journals such as ACM TOIS/TOIT and IEEE TSE/TMC/TSC.

Xiaoyin Wang is an assistant professor at the Department of Computer Science, University of Texas at San Antonio, San Antonio, since August 2013. Before then, he worked as a postdoctoral researcher in UC Berkeley starting from January 2012. From September 2006 to January 2012, he was a Ph.D. candidate in the Software Engineering Institute (SEI) of Peking University, Beijing. From October 2008 to September 2009, he visited Singapore Management University as a research fellow. He has more than 30 publications, most of which are in top software engineering venues.

Mithun P. Acharya is a lead principal scientist at ABB Corporate Research, North Carolina. His expertise and research interests are in the areas of data analytics and software engineering. Specific topics include: 1) artificial intelligence, machine learning, and cloud/edge computing impacting Industrial Internet of Things (IIoT), 2) software analytics, program analysis, and software testing impacting software quality and software/DevOps engineering productivity. His most recent work in Analytics and AI within ABB is related to digital services, platforms, and applications spanning diverse ABB businesses with a strong customer focus. He is the lead author of the Code Drones framework, which recently received the Best Visions of 2025 and Beyond Award, Runner-Up, at the International Conference on Software Engineering. He has his Ph.D. degree in computer science from North Carolina State University, North Carolina.

Marcelo d'Amorim is an associate professor at the Federal University of Pernambuco (UFPE), Recife. The overarching goal of his research is to improve software quality through program analysis and testing. His team has developed systems to find bugs, prevent them from happening, and diagnose errors. Marcelo is recipient of multiple awards, including an ACM Distinguished Reviewer Award, for his work at ASE'16, and a Microsoft Software Engineering Innovation Award (Microsoft SEIF'13), for his work on human-supported software engineering. Marcelo holds his Ph.D. degree from the University of Illinois at Urbana-Champaign (Urbana-Champaign) and his B.S. and M.S. degrees in computer science from UFPE (Recife). His research is funded by CNPq, FACEPE, Microsoft, and RNP.

Xiaoxing Ma is a professor and the deputy director of the Institute of Computer Software at Nanjing University, Nanjing. He received his Ph.D. degree from the same university in 2003. Xiaoxing's research interests include self-adaptive software systems, software architectures, and middleware systems. On these topics he co-authored more than 60 peer-reviewed papers and won two national Sci-Tech awards of China.