MSR 2014

Program of the Eleventh Working Conference on Mining Software Repositories

— Saturday, May 31 —

⚇ Distinguished Paper | 🏛Journal Invitee | **Twitter**: #msr14 |
Facebook: facebook.com/msrconf

MSR 2014 Opening
Sat, May 31, 08:15 – 08:30
Room: MR G.01 & MR G.02 & MR G.03

Prem Devanbu, Sung Kim, Martin Pinzger, Olga Baysal & Emad Shihab

Keynote
Sat, May 31, 08:30 – 09:30
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Prem Devanbu

Is Mining Software Repositories Data Science?

Audris Mockus

Avaya Labs Research

Trick question: what is Data Science? The collection and use of low-veracity data in software repositories and other operational support systems is exploding. It is, therefore, imperative to elucidate basic principles of how such data comes into being and what it means. Are there practices of constructing software data analysis tools that could raise the integrity of their results despite the problematic nature of the underlying data? The talk explores the basic nature of data in operational support systems and considers approaches to develop engineering practices for software mining tools.

Audris Mockus wants to know how and why software development and other complicated systems work. He combines approaches from many disciplines to reconstruct reality from the prolific and varied digital traces these systems leave in the course of operation. Audris Mockus received a B.S. and an M.S. in Applied Mathematics from Moscow Institute of Physics and Technology in 1988. In 1991 he received an M.S. and in 1994 he received a Ph.D. in Statistics from Carnegie Mellon University. He works at Avaya Labs Research. Previously he worked at Software Production Research Department of Bell Labs.

Green mining
Sat, May 31, 09:30 – 10:30
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Mei Nagappan

Mining Energy-Greedy API Usage Patterns in Android Apps: an Empirical Study

Mario Linares-Vásquez, Gabriele Bavota, Carlos Eduardo Bernal Cardenas, Rocco Oliveto, Massimiliano Di Penta and Denys Poshyvanyk (College of William and Mary, USA; University of Sannio, Italy; Universidad Nacional de Colombia, Colombia; University of Molise, Italy)

GreenMiner: A Hardware Based Mining Software Repositories Software Energy Consumption Framework

Abram Hindle, Alex Wilson, Kent Rasmussen, Jed Barlow, Joshua Campbell and Stephen Romansky (University of Alberta, Canada)

Mining Questions About Software Energy Consumption

Gustavo Pinto, Fernando Castor and Yu David Liu (Federal University of Pernambuco, Brazil; SUNY Binghamton, USA)

Coffee Break
Sat, May 31, 10:30 – 11:00

Code Clones and Origin Analysis
Sat, May 31, 11:00 – 12:00
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Massimiliano Di Penta

Prediction and Ranking of Co-change Candidates for Clones

Manishankar Mondal, Chanchal K. Roy and Kevin Schneider (University of Saskatchewan, Canada)

Incremental Origin Analysis of Source Code Files

Daniela Steidl, Benjamin Hummel and Elmar Juergens (CQSE, Germany)

Oops! Where Did That Code Snippet Come From?

Lisong Guo, Julia Lawall and Gilles Muller (INRIA, France; LIP6, France; Sorbonne, France; UPMC, France)

Data showcase teasers
Sat, May 31, 12:00 – 12:30
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Emad Shihab

A dataset for pull-based development research

Georgios Gousios and Andy Zaidman (Delft University of Technology, Netherlands)

The Bug Catalog of the Maven Ecosystem

Dimitris Mitropoulos, Vassilios Karakoidas, Panagiotis Louridas, Georgios Gousios and Diomidis Spinellis (Athens University of Economics and Business, Greece; Delft University of Technology, Netherlands)

A Dataset of Feature Additions and Feature Removals from the Linux Kernel

Leonardo Passos and Krzysztof Czarnecki (University of Waterloo, Canada)

Kataribe:A Hosting Service of Historage Repositories - Fine-Grained Analaysis for all

Kenji Fujiwara, Hideaki Hata, Erina Makihara, Yusuke Fujihara, Naoki Nakayama, Hajimu Iida and Ken-Ichi Matsumoto (NAIST, Japan)

Lean GHTorrent: GitHub data on demand

Georgios Gousios, Bogdan Vasilescu, Alexander Serebrenik and Andy Zaidman (Delft University of Technology, Netherlands; Eindhoven University of Technology, Netherlands)

A Code Clone Oracle

Daniel Krutz and Wei Le (Rochester Institute of Technology, USA)

Generating Duplicate Bug Datasets

Alina Lazar, Sarah Ritchey and Bonita Sharif (Youngstown State University, USA)

FLOSS 2013: A survey dataset about free software contributors. Challenges for curating, sharing and combining

Gregorio Robles, Laura Arjona-Reina, Bogdan Vasilescu, Alexander Serebrenik and Jesús M. González-Barahona (Universidad Rey Juan Carlos, Spain; Universidad Politécnica de Madrid, Spain; Eindhoven University of Technology, Netherlands)

A Green Miner's Dataset: Mining the Impact of Software Change on Energy Consumption

Chenlei Zhang and Abram Hindle (University of Alberta, Canada)

Gentoo Package Dependencies over Time

Remco Bloemen and Chintan Amrit (University of Twente, Netherlands)

Models of OSS Project Meta-Information: A Dataset of Three Forges

James Williams, Davide Di Ruscio, Nicholas Matragkas, Dimitris Kolovos and Juri Di Rocco (University of York, UK; University of L'Aquila, Italy)

A Dataset of Clone References with Gaps

Hiroaki Murakami, Yoshiki Higo and Shinji Kusumoto (Osaka University, Japan)

Maven Components and Bug Patterns in them

Hitesh Sajnani, Vaibhav Saini, Joel Ossher and Cristina Lopes (University of California at Irvine, USA)

OpenHub: A scalable architecture for the analysis of software quality attributes

Gabriel Farah, Juan Tejada and Dario Correal (Universidad de los Andes, Colombia)

Understanding software evolution: the Maisqual Ant data set

Boris Baldassari and Philippe Preux (SQuORING Technologies, France; LIFL, France; CNRS, France; INRIA, France; University of Lille, France)

The Promises and Perils of Mining GitHub

Eirini Kalliamvakou, Georgios Gousios, Kelly Blincoe, Leif Singer, Daniel German and Daniela Damian (University of Victoria, Canada; Delft University of Technology, Netherlands)

Mining StackOverflow to Turn the IDE into a Self-confident Programming Prompter

Luca Ponzanelli, Gabriele Bavota, Massimiliano Di Penta, Rocco Oliveto and Michele Lanza (University of Lugano, Switzerland; University of Sannio, Italy; University of Molise, Italy)

Mining Questions Asked by Web Developers

Kartik Bajaj, Karthik Pattabiraman and Ali Mesbah (University of British Columbia, Canada)

Process Mining Multiple Repositories for Software Defect Resolution from Control and Organizational Perspective

Monika Gupta, Ashish Sureka and Srinivas Padmanabhuni (IIIT Delhi, India; Infosys, India)

A Study of External Community Contribution to Open-Source Projects on GitHub

Rohan Padhye, Senthil Mani and Vibha Singhal Sinha (IBM Research, India)

Understanding "Watchers" on GitHub

Jyoti Sheoran, Kelly Blincoe, Eirini Kalliamvakou, Daniela Damian and Jordan Ell (University of Victoria, Canada)

Do Developers Discuss Design?

João Brunet, Gail C. Murphy, Ricardo Terra, Jorge Figueiredo and Dalton Serey (Federal University of Campina Grande, Brazil; University of British Columbia, Canada; Federal University of Lavras, Brazil)

Magnet or Sticky?: An OSS Project-by-Project Typology

Kazuhiro Yamashita, Shane McIntosh, Yasutaka Kamei and Naoyasu Ubayashi (Kyushu University, Japan; Queen's University, Canada)

Security and Emotion: Sentiment Analysis of Security Discussions on GitHub

Daniel Pletea, Bogdan Vasilescu and Alexander Serebrenik (Eindhoven University of Technology, Netherlands)

Sentiment Analysis of Commit Messages in GitHub: An Empirical Study

Emitza Guzman, David Azócar and Yang Li. (TU München, Germany)

Analysing the `Biodiversity' of Open Source Ecosystems: The Github Case
Nicholas Matragkas, James Williams, Dimitris Kolovos and Richard Paige (University of York, UK)

Co-Evolution of Project Documentation and Popularity within Github
Karan Aggarwal, Abram Hindle and Eleni Stroulia (University of Alberta, Canada)

An Insight into the Pull Requests of GitHub
Mohammad Masudur Rahman and Chanchal K. Roy (University of Saskatchewan, Canada)

Bug Characterizing
Sat, May 31, 17:00 – 18:00
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Abram Hindle

Works For Me! Characterizing Non-reproducible Bug Reports
Mona Erfani Joorabchi, Mehdi Mirzaaghaei and Ali Mesbah (University of British Columbia, Canada)

Characterizing and Predicting Blocking Bugs in Open Source Projects
Harold Valdivia Garcia and Emad Shihab (Rochester Institute of Technology, USA)

An Empirical Study of Dormant Bugs
Tse-Hsun Chen, Meiyappan Nagappan, Emad Shihab and Ahmed E. Hassan (Queen's University, Canada; Rochester Institute of Technology, USA)

Conference Dinner
Sat, May 31, 19:00
Hall 2 at Hyderabad International Convention Centre
Welcoming at 7pm
Performance by Mangala Bhatt & troupe *(7.15pm – 7.40pm)*
Wine Service *(7.45 pm onwards)*
Dinner Service *(7.50 pm onwards)*

— *Sunday, June 01* —

Awards/announcement
Sun, June 01, 08:30 – 09:00
Room: MR G.01 & MR G.02 & MR G.03
MSR Most Influential Paper Award
Distinguished paper awards
MSR Challenge winners, and Best data showcase award

Mining Applications
Sun, June 01, 09:00 – 10:30
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Chris Bird

MUX: Algorithm Selection for Software Model Checkers
Varun Tulsian, Aditya Kanade, Rahul Kumar, Akash Lal and Aditya Nori (Indian Institute of Science, India; Microsoft Research, India)

Improving the Effectiveness of Test Suite Through Mining Historical Data
Jeff Anderson, Saeed Salem and Hyunsook Do (Microsoft, USA; North Dakota State University, USA)

Finding Patterns in Static Analysis Alerts
Quinn Hanam, Lin Tan, Reid Holmes and Patrick Lam (University of Waterloo, Canada)

Impact Analysis of Change Requests on Source Code based on Interaction and Commit Histories
Motahareh Bahrami Zanjani, George Swartzendruber and Huzefa Kagdi (Wichita State University, USA)

Coffee Break
Sun, June 01, 10:30 – 11:00

Defect Prediction
Sun, June 01, 11:00 – 11:45
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Romain Robbes

⬛ An Empirical Study of Just-In-Time Defect Prediction Using Cross-Project Models
Takafumi Fukushima, Yasutaka Kamei, Shane McIntosh, Kazuhiro Yamashita and Naoyasu Ubayashi (Kyushu University, Japan; Queen's University, Canada)

⬛ ⚙ Towards Building a Universal Defect Prediction Model
Feng Zhang, Audris Mockus, Iman Keivanloo and Ying Zou (Queen's University, Canada; Avaya Labs Research, USA)

Short papers
Sun, June 01, 11:45 – 12:35
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Girish Maskeri Rama

⬛ Innovation diffusion in Open Source Software
Remco Bloemen, Chintan Amrit, Stefan Kuhlmann and Gonzalo Ordóñez–Matamoros (University of Twente, Netherlands)

Improving the Accuracy of Duplicate Bug Report Detection Using Textual Similarity Measures
Alina Lazar, Sarah Ritchey and Bonita Sharif (Youngstown State University, USA)

Undocumented and Unchecked: Exceptions that Spell Trouble
Maria Kechagia and Diomidis Spinellis (Athens University of Economics and Business, Greece)

New Features for Duplicate Bug Detection
Nathan Klein, Christopher Corley and Nicholas Kraft (Oberlin College, USA; University of Alabama, USA)

Mining Modern Repositories with Elasticsearch
Oleksii Kononenko, Olga Baysal, Reid Holmes and Michael W. Godfrey (University of Waterloo, Canada)

Collaboration in Open-Source Projects: Myth or Reality?
Yuriy Tymchuk, Andrea Mocci and Michele Lanza (University of Lugano, Switzerland)

A Dictionary to Translate Change Tasks to Source Code
Katja Kevic and Thomas Fritz (University of Zurich, Switzerland)

Tracing Dynamic Features in Python Programs
Beatrice Åkerblom and Tobias Wrigstad (Stockholm University, Sweden; Uppsala University, Sweden)

It's Not A Bug, It's A Feature: Does Misclassification Affect Bug Localization?
Pavneet Singh Kochhar, Tien-Duy Le and David Lo (Singapore Management University, Singapore)

Classifying Unstructured Data into Natural Language Text and Technical Information
Thorsten Merten, Bastian Mager, Simone Bürsner and Barbara Paech (Bonn-Rhein-Sieg University of Applied Sciences, Germany; University of Heidelberg, Germany)

Working Lunch with short paper posters
Sun, June 01, 12:35 – 14:00

Mining Mix
Sun, June 01, 14:00 – 15:30
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Vibha Sinha

Syntax Errors Just Aren't Natural: Improving Error Reporting with Language Models
Joshua Campbell, Abram Hindle and José Nelson Amaral (University of Alberta, Canada)

Do Developers Feel Emotions? An Exploratory Analysis of Emotions in Software Artifacts
Alessandro Murgia, Parastou Tourani, Bram Adams and Marco Ortu (University of Antwerp, Belgium; Polytechnique Montréal, Canada; University of Cagliari, Italy)

How Does a Typical Tutorial for Mobile Development Look Like?
Rebecca Tiarks and Walid Maalej (University of Hamburg, Germany)

Unsupervised Discovery of Intentional Process Models From Event Logs
Ghazaleh Khodabandelou, Rebecca Deneckère, Camille Salinesi and Charlotte Hug (Sorbonne, France)

Coffee Break
Sun, June 01, 15:30 – 16:00

Code Review and Code Search
Sun, June 01, 16:00 – 17:00
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Chanchal Roy

🖺 ⚥ The Impact of Code Review Coverage and Code Review Participation on Software Quality: A Case Study of the Qt, VTK, and ITK Projects
Shane Mcintosh, Yasutaka Kamei, Bram Adams and Ahmed E. Hassan (Queen's University, Canada; Kyushu University, Japan; Polytechnique Montréal, Canada)

Modern Code Reviews in Open-Source Projects: Which Problems Do They Fix?
Moritz Beller, Alberto Bacchelli, Andy Zaidman and Elmar Jürgens (Delft University of Technology, Netherlands; CQSE, Germany)

Thesaurus-Based Automatic Query Expansion for Interface-Driven Code Search
Otavio Lemos, Adriano de Paula, Felipe Zanichelli and Cristina Lopes (Federal University of São Paulo, Brazil; University of California at Irvine, USA)

Effort Estimation and Reuse
Sun, June 01, 17:00 – 18:00
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Tim Menzies

Estimating Development Effort in Free/Open Source Software Projects by Mining Software Repositories
Gregorio Robles, Jesús M. González-Barahona, Carlos Cervigón-Ávila, Andrea Capiluppi and Daniel Izquierdo-Cortázar (Universidad Rey Juan Carlos, Spain; Brunel University, UK; Bitergia, Spain)

An Industrial Case Study of Automatically Identifying Performance Regression-Causes
Thanh H. D. Nguyen, Meiyappan Nagappan, Ahmed E. Hassan, Mohamed Nasser, and Parminder Flora (Queen's University, Canada; BlackBerry, Canada)

Revisiting Android Reuse Studies in the Context of Code Obfuscation and Library Usages
Mario Linares-Vásquez, Andrew Holtzhauer, Carlos Eduardo Bernal Cardenas and Denys Poshyvanyk (College of William and Mary, USA; Universidad Nacional de Colombia, Colombia)

Wrap-Up
Sun, June 01, 18:00 – 18:10
Room: MR G.01 & MR G.02 & MR G.03
Session Chair: Martin Pinzger

MSR 2015 *(May 16-17 2015)* **announcement**
Co-located with ICSE 2015 in Firenze, Italy
http://msrconf.org