2009 6th IEEE International Working Conference on Mining Software Repositories

(MSR 2009)

Vancouver, BC, Canada
16 – 17 May 2009
# TABLE OF CONTENTS

## VERSION CONTROL AND INFRASTRUCTURE

- **The Promises and Perils of Mining Git** .........................................................
  Christian Bird, Peter C. Rigby, Earl T. Barr, David J. Hamilton, Daniel M. German, Prem Devanbu .................................................................

- **Amassing and Indexing a Large Sample of Version Control Systems: Towards the Census of Public Source Code History** ..............................................
  Audris Mockus ........................................................................................................

- **MapReduce As a General Framework to Support Research in Mining Software Repositories (MSR)** .................................................................
  Weiyi Shang, Zhen Ming-Jiang, Bram Adams, Ahmed E. Hassan .............................

- **A Platform for Software Engineering Research** ..................................................
  Georgios Gousios, Diomidis Spinellis ....................................................................

## DEFECT PREDICTION

- **Evaluating the Relation Between Coding Standard Violations and Faults Within and Across Software Versions** .........................................................
  Cathal Boogerd, Leon Moonen ............................................................................

- **Tracking Concept Drift of Software Projects Using Defect Prediction Quality** .................................................................
  Jayalath Ekanayake, Jonas Tappolet, Harald C. Gall, Abraham Bernstein ..........

- **Does Calling Structure Information Improve the Accuracy of Fault Prediction?** .................................................................
  Yonghee Shin, Robert Bell, Thomas Ostrand, Elaine Weyuker .............................

## TEXT ANALYSIS

- **Mining Source Code to Automatically Split Identifiers for Software Analysis** .........................................................
  Eric Enslen, Emily Hill, Lori Pollock, K. Vijay-Shanker ........................................

- **Code Siblings: Technical and Legal Implications of Copying Code Between Applications** .................................................................
  Daniel M. German, Massimiliano Di Penta, Yann-Gael Gueheneuc, Giuliano Antoniol

## MSR CHALLENGE

- **Author Entropy vs File Size in the GNOME Suite of Applications** ..................
  Jason R. Casebolt, Jonathan L. Krein, Alexander C. MacLean, Charles D. Knutson, Daniel P. Delorey .................................................................

- **Evaluating Process Quality in GNOME Based on Change Request Data** ...........
  Holger Schackmann, Horst Lichter ....................................................................

- **Mining the Coherence of GNOME Bug Reports with Statistical Topic Models** ........
  Erik Linstead, Pierre Baldt .................................................................................

- **Visualizing Gnome with the Small Project Observatory** .................................
  Mircea Lungu, Jacopo Malnati, Michele Lanza ..................................................
On the Use of Internet Relay Chat (IRC) Meetings by Developers of the Gnome Gtk+ Project
Emad Shihab, Zhen Ming Jiang, Ahmed E. Hassan

TOPIC MINING

Mining Search Topics from a Code Search Engine Usage Log
Sushil Bajracharya, Cristina Lopes

DEVELOPERS

From Work to Word: How Do Software Developers Describe Their Work?
Walid Maalej, Hans-Jorg Happel

Assigning Bug Reports Using a Vocabulary-based Expertise Model of Developers
Dominique Matter, Adrian Kuhn, Oscar Nierstrasz

Mining the History of Synchronous Changes to Refine Code Ownership
Lile Hattori, Michele Lanza

POSTER PAPERS

Using Association Rules to Study the Co-Evolution of Production & Test Code
Zeeger Lubsen, Andy Zaidman, Martin Pinzger

On What Basis to Recommend: Changesets Or Interactions?
Sarah Rastkar, Gail C. Murphy

Mining the Jazz Repository: Challenges and Opportunities
Kim Herzig, Andreas Zeller

Using Latent Dirichlet Allocation for Automatic Categorization of Software
Kai Tian, Meghan Revelle, Denys Poshyvanyk

Evolution of the Core Team of Developers in Libre Software Projects
Gregorio Robles, Jesus M. Gonzalez-Barahona, Israel Herraiz

On Mining Data Across Software Repositories
Prasanth Anbalagan, Mladen Vouk

Automatic Labeling of Software Components and Their Evolution Using Log-Likelihood
Ratio of Word Frequencies in Source Code
Adrian Kuhn

Learning from Defect Removals
Nathaniel Ayewah, William Pugh

SourcererDB: An Aggregated Repository of Statically Analyzed and Cross-Linked Open Source Java Projects
Joel Ossher, Sushil Bajracharya, Erik Linstead, Pierre Baldi, Cristina Lopes

On the Transfer of Evolutionary Couplings to Industry
Pierre van de Laar

Author Index