

# **2009 6th IEEE International Working Conference on Mining Software Repositories**

**(MSR 2009)**

**Vancouver, BC, Canada  
16 – 17 May 2009**



**IEEE Catalog Number: CFP0978C-PRT  
ISBN: 978-1-4244-3493-0**

# TABLE OF CONTENTS

## VERSION CONTROL AND INFRASTRUCTURE

<b>The Promises and Perils of Mining Git</b> .....	1
<i>Christian Bird, Peter C. Rigby, Earl T. Barr, David J. Hamilton, Daniel M. German, Prem Devanbu</i>	
<b>Amassing and Indexing a Large Sample of Version Control Systems: Towards the Census of Public Source Code History</b> .....	11
<i>Audris Mockus</i>	
<b>MapReduce As a General Framework to Support Research in Mining Software Repositories (MSR)</b> .....	21
<i>Weiyi Shang, Zhen Ming Jiang, Bram Adams, Ahmed E. Hassan</i>	
<b>A Platform for Software Engineering Research</b> .....	31
<i>Georgios Gousios, Diomidis Spinellis</i>	

## DEFECT PREDICTION

<b>Evaluating the Relation Between Coding Standard Violations and Faults Within and Across Software Versions</b> .....	41
<i>Cathal Boogerd, Leon Moonen</i>	
<b>Tracking Concept Drift of Software Projects Using Defect Prediction Quality</b> .....	51
<i>Jayalath Ekanayake, Jonas Tappolet, Harald C. Gall, Abraham Bernstein</i>	
<b>Does Calling Structure Information Improve the Accuracy of Fault Prediction?</b> .....	61
<i>Yonghee Shin, Robert Bell, Thomas Ostrand, Elaine Weyuker</i>	

## TEXT ANALYSIS

<b>Mining Source Code to Automatically Split Identifiers for Software Analysis</b> .....	71
<i>Eric Enslin, Emily Hill, Lori Pollock, K. Vijay-Shanker</i>	
<b>Code Siblings: Technical and Legal Implications of Copying Code Between Applications</b> .....	81
<i>Daniel M. German, Massimiliano Di Penta, Yann-Gael Gueheneuc, Giuliano Antoniol</i>	

## MSR CHALLENGE

<b>Author Entropy vs File Size in the GNOME Suite of Applications</b> .....	91
<i>Jason R. Casebolt, Jonathan L. Krein, Alexander C. MacLean, Charles D. Knutson, Daniel P. Delorey</i>	
<b>Evaluating Process Quality in GNOME Based on Change Request Data</b> .....	95
<i>Holger Schackmann, Horst Lichter</i>	
<b>Mining the Coherence of GNOME Bug Reports with Statistical Topic Models</b> .....	99
<i>Erik Linstead, Pierre Baldi</i>	
<b>Visualizing Gnome with the Small Project Observatory</b> .....	103
<i>Mircea Lungu, Jacopo Malnati, Michele Lanza</i>	

<b>On the Use of Internet Relay Chat (IRC) Meetings by Developers of the Gnome Gtk+ Project</b> .....	107
<i>Emad Shihab, Zhen Ming Jiang, Ahmed E. Hassan</i>	

## **TOPIC MINING**

<b>Mining Search Topics from a Code Search Engine Usage Log</b> .....	111
<i>Sushil Bajracharya, Cristina Lopes</i>	

## **DEVELOPERS**

<b>From Work to Word: How Do Software Developers Describe Their Work?</b> .....	121
<i>Walid Maalej, Hans-Jorg Happel</i>	
<b>Assigning Bug Reports Using a Vocabulary-based Expertise Model of Developers</b> .....	131
<i>Dominique Matter, Adrian Kuhn, Oscar Nierstrasz</i>	
<b>Mining the History of Synchronous Changes to Refine Code Ownership</b> .....	141
<i>Lile Hattori, Michele Lanza</i>	

## **POSTER PAPERS**

<b>Using Association Rules to Study the Co-Evolution of Production &amp; Test Code</b> .....	151
<i>Zeeger Lubsen, Andy Zaidman, Martin Pinzger</i>	
<b>On What Basis to Recommend: Changesets Or Interactions?</b> .....	155
<i>Sarah Rastkar, Gail C. Murphy</i>	
<b>Mining the Jazz Repository: Challenges and Opportunities</b> .....	159
<i>Kim Herzig, Andreas Zeller</i>	
<b>Using Latent Dirichlet Allocation for Automatic Categorization of Software</b> .....	163
<i>Kai Tian, Meghan Revella, Denys Poshyvanyk</i>	
<b>Evolution of the Core Team of Developers in Libre Software Projects</b> .....	167
<i>Gregorio Robles, Jesus M. Gonzalez-Barahona, Israel Herraiz</i>	
<b>On Mining Data Across Software Repositories</b> .....	171
<i>Prasanth Anbalagan, Mladen Vouk</i>	
<b>Automatic Labeling of Software Components and Their Evolution Using Log-Likelihood Ratio of Word Frequencies in Source Code</b> .....	175
<i>Adrian Kuhn</i>	
<b>Learning from Defect Removals</b> .....	179
<i>Nathaniel Ayewah, William Pugh</i>	
<b>SourcererDB: An Aggregated Repository of Statically Analyzed and Cross-Linked Open Source Java Projects</b> .....	183
<i>Joel Osher, Sushil Bajracharya, Erik Linstead, Pierre Baldi, Cristina Lopes</i>	
<b>On the Transfer of Evolutionary Couplings to Industry</b> .....	187
<i>Pierre van de Laar</i>	

## **Author Index**