2011 27th IEEE
International Conference
on Software Maintenance

(ICSMM 2011)

Williamsburg, Virginia, USA
25 – 30 September 2011
# Table of Contents

## Front Matter
- Foreword
  - viii
- ICSM 2011 Organizing Committee
  - ix
- ICSM 2011 Program Committee
  - x
  - Additional Reviewers
  - xii
- Industry Track Program Committee
  - Additional Reviewers
  - xiii
- Early Research Achievements Program Committee
  - Additional Reviewers
  - xiv
- Tool Demonstrations Program Committee
  - Additional Reviewers
  - xv

## Technical Papers

### Keynotes
- How to Steal a Botnet and What Can Happen When You Do
  - Richard A. Kemmerer
  - 1
- Useful Software Engineering Research - Leading a Double-Agent Life
  - Lionel C. Briand
  - 2

### Research Track
- Faults and Regression Testing
  - Fault Interaction and its Repercussions
    - Nicholas Digiuseppe and James Jones
    - 3
  - A Novel Approach to Regression Test Selection for J2EE Applications
    - Sheng Huang, Zhong Jie Li, Jun Zhu, Yanghua Xiao and Wei Wang
    - 13
  - Localizing Fault-Inducing Program Edits Based on Spectrum Information
    - Lingming Zhang, Miryung Kim and Sarfraz Khurshid
    - 23

### Impact Analysis
- Automated Change Impact Analysis for Agent Systems
  - Hoa Khanh Dam and Aditya Ghose
  - 33
- ImpactScale: Quantifying Change Impact to Predict Faults in Large Software Systems
  - Kenichit Kobayashi, Akihiko Matsuo, Katsuro Inoue, Yasuhiro Hayase, Manabu Kamimura and Toshiaki Yoshino
  - 43
- A Seismology-inspired Approach to Study Change Propagation
  - Salima Hassaine, Ferdaous Boughanmi, Yann-Guil Guéhéneuc, Sylvie Homel and Giuliano Antoniol
  - 53

### Dynamic Analysis
- SCOTCH: Test-to-Code Traceability using Slicing and Conceptual Coupling
  - Abdallah Qusef, Gabriele Bavota, Rocco Oliveto, Andrea De Lucia and Dave Binkley
  - 63
- Identifying Distributed Features in SOA by Mining Dynamic Call Trees
  - Anis Yousefi and Kamran Sartipi
  - 73
- Identifying Performance Deviations in Thread Pools
  - Mark D. Syer, Bram Adams and Ahmed E. Hassan
  - 83
## Research Track

### Natural Language Analysis

Mining Java Class Naming Conventions  
*Simon Butler, Michel Wermelinger, Yijun Yu and Helen Sharp*  
Page 93

Generating Natural Language Summaries for Crosscutting Source Code Concerns  
*Sarah Rasikar, Gail C. Murphy and Alexander W. J. Bradley*  
Page 103

Expanding Identifiers to Normalize Source Code Vocabulary  
*Dawn Lawrie and Dave Binkley*  
Page 113

### Traceability

Using Tactic Traceability Information Models to Reduce the Risk of Architectural Degradation during System Maintenance  
*Mehdi Mirakhorli and Jane Cleland-Huang*  
Page 123

On Integrating Orthogonal Information Retrieval Methods to Improve Traceability Link Recovery  
*Malcom Gethers, Rocco Oliveto, Denys Poshvanyk and Andrea De Lucia*  
Page 133

Structural Conformance Checking with Design Tests: An Evaluation of Usability and Scalability  
*João Brunet, Dalton Dario Serey Guerrero and Jorge Figueiredo*  
Page 143

### Migration and Refactoring

MoMS: Multi-objective Miniaturization of Software  
*Nasir Ali, Wei Wu, Giuliano Antoniol, Massimiliano Di Penta, Yann-Gaël Guéhéneuc and Jane Huffman Hayes*  
Page 153

A Method Engineering Based Legacy to SOA Migration Method  
*Ravi Khadka, Slinger Jansen, Amir M. Saeidi, Jurriaan Hage and Gijs Reijnders*  
Page 163

Identifying Overly Strong Conditions in Refactoring Implementations  
*Gustavo Soares, Melina Mongiovi and Rohit Gheyi*  
Page 173

### Components

Graph-based Detection of Library API Imitations  
*Chengnian Sun, Siau-Cheng Khoo and Shao Jie Zhang*  
Page 183

Crossing the Boundaries While Analyzing Heterogeneous Component-Based Software Systems  
*Amir Reza Yazdanshenas and Leon Moonen*  
Page 193

On the Maintenance of UI-integrated Mashup Applications  
*Maxim Shevertalov and Spiros Mancoridis*  
Page 203

### Program Comprehension

An Exploratory Study of Feature Location Process: Distinct Phases, Recurring Patterns, and Elementary Actions  
*Jinshui Wang, Xin Peng, Zhenchang Xing and Wenyun Zhao*  
Page 213

Exploiting Text Mining Techniques in the Analysis of Execution Traces  
*Heidar Pirzadeh, Abdelwahab Hamou-Lhadj and Mohak Shah*  
Page 223

An Evaluation of the Strategies of Sorting, Filtering, and Grouping API Methods for Code Completion  
*Daqing Hou and Dave Fletcher*  
Page 233

### Reliability and Quality

A Probabilistic Software Quality Model  
*Tibor Bakota, Péter Hegedűs, Péter Körtvélyesi, Rudolf Ferenc and Tibor Gyimóthy*  
Page 243

Predicting Post-release Defects Using Pre-release Field Testing Results  
*Fouuse Khomih, Brian Chan, Ying Zou, Anand Sinha and Dave Dietz*  
Page 253

Sahara: Guiding the Debugging of Failed Software Upgrades  
*Rekha Bachwani, Olivier Crameri, Ricardo Bianchini, Defan Kostic and Willy Zwaenepoel*  
Page 263
Research Track

Clones

Late Propagation in Software Clones
Liliane Barbour, Foutse Khomh and Ying Zou 273

File Cloning in Open Source Java Projects: The Good, The Bad, and The Ugly
Joel Ossher, Hitesh Sajnani and Cristina Lopes 283

An Automatic Framework for Extracting and Classifying Near-Miss Clone Genealogies
Ripon K. Saha, Chanchal K. Roy and Kevin A. Schneider 293

Metrics

Using Source Code Metrics to Predict Change-Prone Java Interfaces
Daniele Romano and Martin Pinzger 303

You Can't Control the Unfamiliar: A Study on the Relations Between Aggregation Techniques for Software Metrics
Bogdan Vasilescu, Alexander Serebrenik and Mark Van Den Brand 313

Understanding a Developer Social Network and its Evolution
Qiaona Hong, Sunghun Kim, Shing-Chi Cheung and Christian Bird 323

Clustering and Categorization

Classifying Field Crash Reports for Fixing Bugs : A Case Study of Mozilla Firefox
Tejinder Dhaliwal, Foutse Khomh and Ying Zou 333

Categorizing Software Applications for Maintenance
Collin Mcmillan, Mario Linares-Vasquez, Denys Poshypynik and Mark Grechanik 343

Evaluating Software Clustering using Multiple Simulated Authoritative Decompositions
Mark Shtern and Vassilios Tzerpos 353

Industry Track

Testing

Industrial Experiences with Automated Regression Testing of a Legacy Database Application
Erik Rogstad, Lionel Briand, Ronny Dalberg, Marianne Rynning and Erik Arisholm 362

Regression Testing in Software as a Service: An Industrial Case Study
Hema Srikanth and Myra Cohen 372

A Clustering Approach to Improving Test Case Prioritization: An Industrial Case Study
Ryan Carlson, Hyunsook Do and Anne Denton 382

Reverse Engineering

Code Hot Spot: A Tool for Extraction and Analysis of Code Change History
Will Snipes, Brian Robinson and Emerson Murphy-Hill 392

An Integration Resolution Algorithm for Mining Multiple Branches in Version Control Systems
Alexander Tarvo, Thomas Zimmermann and Jacek Czerwona 402

Relating Developers' Concepts and Artefact Vocabulary in a Financial Software Module
Tezcan Dilshener and Michel Wermeleigner 412

Evolution and Migration

Incremental and Iterative Reengineering towards Software Product Line: An Industrial Case Study
Gang Zhang, Liwei Shen, Xin Peng, Zhenchang Xing and Wenyun Zhao 418

The Evolution of Information Systems. A Case Study on Document Management
Paolo Salvaneschi 428

Testing & Quality Assurance in Data Migration Projects
Florian Matthes, Christopher Schulz and Klaus Haller 438
Industry Track

Program Analysis and Verification

Precise Detection of Un-initialized Variables in Large, Real-life COBOL Programs in Presence of Unrealizable Paths  
Rahul Jiresal, Adnan Contractor and Ravindra Naik  
Type-preserving Heap Profiler for C++  
József Mihalicza, Zoltán Porkoláb and Ábel Gábor  
Analyzing the Effects of Formal Methods on the Development of Industrial Control Software  
Jan Friso Groote, Ammar Osaíweran and Jacco H. Wesselius

Metrics and Estimation

Source Code Comprehension Strategies and Metrics to Predict Comprehension Effort in Software Maintenance and Evolution Tasks - An Empirical Study with Industry Practitioners  
Kazuki Nishizono, Shuji Morisaki, Rodrigo Vivanco and Kenichi Matsumoto  
Estimating Software Maintenance Effort from Use Cases: An Industrial Case Study  
Yan Ku, Jing Du, Ye Yang and Qing Wang

Early Research Achievements Track

Linguistic Analysis of Software Artifacts

Toward a Metrics Suite for Source Code Lexicons  
Lauren R. Biggers, Brian P. Eddy, Nicholas A. Kraft and Letha H. Etzkorn  
A Comparison of Stemmers on Source Code Identifiers for Software Search  
Andrew Wiese, Valerie Ho and Emily Hill  
Clustering and Lexical Information Support for the Recovery of Design Pattern in Source Code  
Simone Romano, Giuseppe Scanniello, Michele Risi and Carmine Gravino  
Code Convention Adherence in Evolving Software  
Michael Smit, Barry Gergel, H. James Hoover and Eleni Stroulia

Software Changes and Maintainability

SE² Model to Support Software Evolution  
Huzefa Kagdi, Malcolm Gethers and Denys Poshvanyk  
Measuring Maintainability of Spreadsheets in the Wild  
José Pedro Correia and Miguel Alexandre Ferreira  
Using Stereotypes to Help Characterize Commits  
Natalia Dragon, Michael L. Collard, Maen Hammad and Jonathan I. Maletic  
Source Code Survival with the Kaplan Meier Estimator  
Giuseppe Scanniello

Managing and Supporting Software Maintenance Activities

Tracking Technical Debt: An Exploratory Case Study  
Yuepu Guo, Carolyn Seaman, Rebeka Gomes, Antonio Cavalcanti, Graziela Tonin, Fabio Q. B. Da Silva, André L. M. Santos and Clauritton Siebra  
Measuring Disruption from Software Evolution Activities Using Graph-Based Metrics  
Prashant Paymal, Rajvardhan Patil, Sanjukta Bhowmick and Harvey Sty  
Clustering and Recommending Collections of Code Relevant to Tasks  
Seonah Lee and Sungwon Kang  
Dependency Profiles for Software Architecture Evaluations  
Eric Bouwers, Arie Van Deursen and Joost Visser
Post Doctoral Symposium Track

Empirical Assessment of UML Class Diagram Layouts Based on Architectural Importance
Bonita Sharif

Emergent Laws of Method and Class Stereotypes in Object Oriented Software
Natalia Dragan

A Logic Meta-Programming Foundation for Example-Driven Pattern Detection in Object-Oriented Programs
Coen De Roover

Evidence-based Software Process Recovery: A Post-Doctoral View
Abram Hindle

Sociotechnical Coordination and Collaboration in Open Source Software
Christian Bird

Tool Demonstrations Track

A Tool for Combinatorial-based Prioritization and Reduction of User-Session-Based Test Suites
Sreedevi Sampath, Renee Bryce, Sachin Jain and Schuyler Manchester

MARBLE. A Business Process Archeology Tool
Ricardo Pérez-Castillo, María Fernández-Ropero, Ignacio García-Rodríguez de Guzmán and Mario Piattini

Program Querying with a SOUL: The Barista Tool Suite
Carlos Noguera, Coen De Roover, Andy Kellens and Viviane Jonckers

WebDiff: A Generic Differencing Service for Software Artifacts
Nikolaos Tsantalis, Natalia Negara and Eleni Stroulia

EQ: Checking the Implementation of Equality in Java
Chandan R. Rupakheti and Daqing Hou

Maleku: An Evolutionary Visual Software Analytics Tool for Providing Insights into Software Evolution
Antonio González Torres, Roberto Theron, Francisco José García Peñalvo, Michel Wermelinger and Yijun Yu

Tutorials

Effective Mining of Software Repositories
Marco D'Ambros and Romain Robbes

Practical Combinatorial (t-way) Methods for Detecting Complex Faults in Regression Testing
Rick Kuhn and Raghu Kacker

Research Methods in Computer Science
Serge Demeyer