2011 27th IEEE International Conference on Software Maintenance

(ICSM 2011)

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

## Front Matter

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreword</td>
<td>viii</td>
</tr>
<tr>
<td>ICSM 2011 Organizing Committee</td>
<td>ix</td>
</tr>
<tr>
<td>ICSM 2011 Program Committee</td>
<td>x</td>
</tr>
<tr>
<td>Additional Reviewers</td>
<td>xii</td>
</tr>
<tr>
<td>Industry Track Program Committee</td>
<td>xiii</td>
</tr>
<tr>
<td>Additional Reviewers</td>
<td>xiii</td>
</tr>
<tr>
<td>Early Research Achievements Program Committee</td>
<td>xiv</td>
</tr>
<tr>
<td>Additional Reviewers</td>
<td>xiv</td>
</tr>
<tr>
<td>Tool Demonstrations Program Committee</td>
<td>xv</td>
</tr>
<tr>
<td>Additional Reviewers</td>
<td>xv</td>
</tr>
</tbody>
</table>

## 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  
  *Kenichi 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-Gaël Guéhéneuc, Sylvie Hamel and Giuliano Antoniol*  
  53

#### Dynamic Analysis

- SCOTCH: Test-to-Code Traceability using Slicing and Conceptual Coupling  
  *Abdallah Qasef, Gabriele Bavota, Rocco Oliveto, Andrea De Lucia and Dave Binkley*  
  63
- Identifying Distributed Features in SOA by Mining Dynamic Call Trees  
  *Antis 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 93

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

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

Traceability

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

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

Structural Conformance Checking with Design Tests: An Evaluation of Usability and Scalability
João Brunet, Dalton Darío Serey Guerrero and Jorge Figueiredo 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 153

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

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

Components

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

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

On the Maintenance of UI-integrated Mashup Applications
Maxim Shevertalov and Spiros Mancoridis 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 213

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

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

Reliability and Quality

A Probabilistic Software Quality Model
Tibor Bakota, Péter Hegediš, Péter Körtvélyesi, Rudolf Ferenc and Tibor Gyimóthy 243

Predicting Post-release Defects Using Pre-release Field Testing Results
Foutse Khomh, Brian Chan, Ying Zou, Anand Sinha and Dave Dietz 253

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

Clones

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

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

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

Metrics

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

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

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

Clustering and Categorization

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

Categorizing Software Applications for Maintenance
Collin Mcmillan, Mario Linares-Vásquez, Denys Poshvanyk and Mark Grechanik

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

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

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

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

Reverse Engineering

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

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

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

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

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

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

Program Analysis and Verification

Precise Detection of Un-initialized Variables in Large, Real-life COBOL Programs in Presence of Unrealizable Paths 448
Rahul Jiresal, Adnan Contractor and Ravindra Naik

Type-preserving Heap Profiler for C++ 457
József Mihalicza, Zoltán Porkoláb and Ábel Gábor

Analyzing the Effects of Formal Methods on the Development of Industrial Control Software 467
Jan Friso Groote, Ammar Osaiweran 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 473
Kazuki Nishizono, Shuji Morisaki, Rodrigo Vivanco and Kenichi Matsumoto

Estimating Software Maintenance Effort from Use Cases: An Industrial Case Study 482
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 492
Lauren R. Biggers, Brian P. Eddy, Nicholas A. Kraft and Letha H. Etzkorn

A Comparison of Stemmers on Source Code Identifiers for Software Search 496
Andrew Wiese, Valerie Ho and Emily Hill

Clustering and Lexical Information Support for the Recovery of Design Pattern in Source Code 500
Simone Romano, Giuseppe Scanniello, Michele Risi and Carmine Gravino

Code Convention Adherence in Evolving Software 504
Michael Smit, Barry Gergel, H. James Hoover and Eleni Stroulia

Software Changes and Maintainability

SE² Model to Support Software Evolution 512
Huzefa Kagdi, Malcom Gethers and Denys Poshvanyk

Measuring Maintainability of Spreadsheets in the Wild 516
José Pedro Correia and Miguel Alexandre Ferreira

Using Stereotypes to Help Characterize Commits 520
Natalia Dragan, Michael L. Collard, Maen Hammad and Jonathan I. Maletic

Source Code Survival with the Kaplan Meier Estimator 524
Giuseppe Scanniello

Managing and Supporting Software Maintenance Activities

Tracking Technical Debt: An Exploratory Case Study 528
Yuepu Guo, Carolyn Seaman, Rebeka Gomes, Antonio Cavalcanti, Graziela Tonin, Fabio Q. B. Da Silva, André L. M. Santos and Clairtorn Siebra

Measuring Disruption from Software Evolution Activities Using Graph-Based Metrics 532
Prashant Paymal, Rajvardhan Patil, Sanjukta Bhosmick and Harvey Stiy

Clustering and Recommending Collections of Code Relevant to Tasks 536
Seonah Lee and Sungwon Kang

Dependency Profiles for Software Architecture Evaluations 540
Eric Bouwers, Arie Van Deursen and Joost Visser
### Post Doctoral Symposium Track

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical Assessment of UML Class Diagram Layouts Based on Architectural Importance</td>
<td>544</td>
</tr>
<tr>
<td>Bonita Sharif</td>
<td></td>
</tr>
<tr>
<td>Emergent Laws of Method and Class Stereotypes in Object Oriented Software</td>
<td>550</td>
</tr>
<tr>
<td>Natalia Dragan</td>
<td></td>
</tr>
<tr>
<td>A Logic Meta-Programming Foundation for Example-Driven Pattern Detection in Object-Oriented Programs</td>
<td>556</td>
</tr>
<tr>
<td>Coen De Roover</td>
<td></td>
</tr>
<tr>
<td>Evidence-based Software Process Recovery: A Post-Doctoral View</td>
<td>562</td>
</tr>
<tr>
<td>Abram Hindle</td>
<td></td>
</tr>
<tr>
<td>Sociotechnical Coordination and Collaboration in Open Source Software</td>
<td>568</td>
</tr>
<tr>
<td>Christian Bird</td>
<td></td>
</tr>
</tbody>
</table>

### Tool Demonstrations Track

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Tool for Combinatorial-based Prioritization and Reduction of User-Session-Based Test Suites</td>
<td>574</td>
</tr>
<tr>
<td>Sreedevi Sampath, Renee Bryce, Sachin Jain and Schuyler Manchester</td>
<td></td>
</tr>
<tr>
<td>MARBLE. A Business Process Archeology Tool</td>
<td>578</td>
</tr>
<tr>
<td>Ricardo Pérez-Castillo, Maria Fernández-Ropero, Ignacio García-Rodríguez de Guzmán and Mario Piattini</td>
<td></td>
</tr>
<tr>
<td>Program Querying with a SOUL: The Barista Tool Suite</td>
<td>582</td>
</tr>
<tr>
<td>Carlos Noguera, Coen De Roover, Andy Kellens and Viviane Jonckers</td>
<td></td>
</tr>
<tr>
<td>WebDiff: A Generic Differencing Service for Software Artifacts</td>
<td>586</td>
</tr>
<tr>
<td>Nikolaos Tsantalis, Natalia Negara and Eleni Stroudia</td>
<td></td>
</tr>
<tr>
<td>EQ: Checking the Implementation of Equality in Java</td>
<td>590</td>
</tr>
<tr>
<td>Chandan R. Rupakheti and Daqing Hou</td>
<td></td>
</tr>
<tr>
<td>Maleku: An Evolutionary Visual Software Analytics Tool for Providing Insights into Software Evolution</td>
<td>594</td>
</tr>
<tr>
<td>Antonio González Torres, Roberto Therón, Francisco José García Peñalvo, Michel Wermelinger and Yijun Yu</td>
<td></td>
</tr>
</tbody>
</table>

### Tutorials

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective Mining of Software Repositories</td>
<td>598</td>
</tr>
<tr>
<td>Marco D’Ambros and Romain Robbes</td>
<td></td>
</tr>
<tr>
<td>Practical Combinatorial (t-way) Methods for Detecting Complex Faults in Regression Testing</td>
<td>599</td>
</tr>
<tr>
<td>Rick Kuhn and Raghu Kacker</td>
<td></td>
</tr>
<tr>
<td>Research Methods in Computer Science</td>
<td>600</td>
</tr>
<tr>
<td>Serge Demeyer</td>
<td></td>
</tr>
</tbody>
</table>