

**2011 27th IEEE  
International Conference  
on Software Maintenance  
(ICSM 2011)**

**Williamsburg, Virginia, USA  
25 – 30 September 2011**



**IEEE Catalog Number: CFP11079-PRT  
ISBN: 978-1-4577-0663-9**

## Table of Contents

### Front Matter

Foreword	viii
ICSM 2011 Organizing Committee	ix
ICSM 2011 Program Committee	x
<i>Additional Reviewers</i>	xii
Industry Track Program Committee	xiii
<i>Additional Reviewers</i>	xiii
Early Research Achievements Program Committee	xiv
<i>Additional Reviewers</i>	xiv
Tool Demonstrations Program Committee	xv
<i>Additional Reviewers</i>	xv

### Technical Papers

#### Keynotes

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

#### Research Track

##### Faults and Regression Testing

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

##### Impact Analysis

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

##### Dynamic Analysis

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

## Research Track

### Natural Language Analysis

Mining Java Class Naming Conventions	93
<i>Simon Butler, Michel Wermelinger, Yijun Yu and Helen Sharp</i>	
Generating Natural Language Summaries for Crosscutting Source Code Concerns	103
<i>Sarah Rastkar, Gail C. Murphy and Alexander W. J. Bradley</i>	
Expanding Identifiers to Normalize Source Code Vocabulary	113
<i>Dawn Lawrie and Dave Binkley</i>	

### Traceability

Using Tactic Traceability Information Models to Reduce the Risk of Architectural Degradation during System Maintenance	123
<i>Mehdi Mirakhorli and Jane Cleland-Huang</i>	
On Integrating Orthogonal Information Retrieval Methods to Improve Traceability Link Recovery	133
<i>Malcom Gethers, Rocco Oliveto, Denys Poshyvanyk and Andrea De Lucia</i>	
Structural Conformance Checking with Design Tests: An Evaluation of Usability and Scalability	143
<i>João Brunet, Dalton Dario Serey Guerrero and Jorge Figueiredo</i>	

### Migration and Refactoring

MoMS: Multi-objective Miniaturization of Software	153
<i>Nasir Ali, Wei Wu, Giuliano Antoniol, Massimiliano Di Penta, Yann-Gaël Guéhéneuc and Jane Huffman Hayes</i>	
A Method Engineering Based Legacy to SOA Migration Method	163
<i>Ravi Khadka, Slinger Jansen, Amir M. Saeidi, Jurriaan Hage and Gijs Reijnders</i>	
Identifying Overly Strong Conditions in Refactoring Implementations	173
<i>Gustavo Soares, Melina Mongiovi and Rohit Gheyi</i>	

### Components

Graph-based Detection of Library API Imitations	183
<i>Chengnian Sun, Siau-Cheng Khoo and Shao Jie Zhang</i>	
Crossing the Boundaries While Analyzing Heterogeneous Component-Based Software Systems	193
<i>Amir Reza Yazdanshenas and Leon Moonen</i>	
On the Maintenance of UI-integrated Mashup Applications	203
<i>Maxim Shevertalov and Spiros Mancoridis</i>	

### Program Comprehension

An Exploratory Study of Feature Location Process: Distinct Phases, Recurring Patterns, and Elementary Actions	213
<i>Jinshui Wang, Xin Peng, Zhenchang Xing and Wenyun Zhao</i>	
Exploiting Text Mining Techniques in the Analysis of Execution Traces	223
<i>Heidar Pirzadeh, Abdelwahab Hamou-Lhadj and Mohak Shah</i>	
An Evaluation of the Strategies of Sorting, Filtering, and Grouping API Methods for Code Completion	233
<i>Daqing Hou and Dave Pletcher</i>	

### Reliability and Quality

A Probabilistic Software Quality Model	243
<i>Tibor Bakota, Péter Hegedüs, Péter Körtvélyesi, Rudolf Ferenc and Tibor Gyimóthy</i>	
Predicting Post-release Defects Using Pre-release Field Testing Results	253
<i>Fouse Khomh, Brian Chan, Ying Zou, Anand Sinha and Dave Dietz</i>	
Sahara: Guiding the Debugging of Failed Software Upgrades	263
<i>Rekha Bachwani, Olivier Crameri, Ricardo Bianchini, Dejan Kostic and Willy Zwaenepoel</i>	

## Research Track

### Clones

- Late Propagation in Software Clones 273  
*Liliane Barbour, Foutse Khomh and Ying Zou*
- File Cloning in Open Source Java Projects: The Good, The Bad, and The Ugly 283  
*Joel Ossher, Hitesh Sajani and Cristina Lopes*
- An Automatic Framework for Extracting and Classifying Near-Miss Clone Genealogies 293  
*Ripon K. Saha, Chanchal K. Roy and Kevin A. Schneider*

### Metrics

- Using Source Code Metrics to Predict Change-Prone Java Interfaces 303  
*Daniele Romano and Martin Pinzger*
- You Can't Control the Unfamiliar: A Study on the Relations Between Aggregation Techniques for Software Metrics 313  
*Bogdan Vasilescu, Alexander Serebrenik and Mark Van Den Brand*
- Understanding a Developer Social Network and its Evolution 323  
*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 333  
*Tejinder Dhaliwal, Foutse Khomh and Ying Zou*
- Categorizing Software Applications for Maintenance 343  
*Collin Mcmillan, Mario Linares-Vásquez, Denys Poshyvanyk and Mark Grechanik*
- Evaluating Software Clustering using Multiple Simulated Authoritative Decompositions 353  
*Mark Shtern and Vassilios Tzerpos*

## Industry Track

### Testing

- Industrial Experiences with Automated Regression Testing of a Legacy Database Application 362  
*Erik Rogstad, Lionel Briand, Ronny Dalberg, Marianne Rynning and Erik Arisholm*
- Regression Testing in Software as a Service: An Industrial Case Study 372  
*Hema Srikanth and Myra Cohen*
- A Clustering Approach to Improving Test Case Prioritization: An Industrial Case Study 382  
*Ryan Carlson, Hyunsook Do and Anne Denton*

### Reverse Engineering

- Code Hot Spot: A Tool for Extraction and Analysis of Code Change History 392  
*Will Snipes, Brian Robinson and Emerson Murphy-Hill*
- An Integration Resolution Algorithm for Mining Multiple Branches in Version Control Systems 402  
*Alexander Tarvo, Thomas Zimmermann and Jacek Czerwonka*
- Relating Developers' Concepts and Artefact Vocabulary in a Financial Software Module 412  
*Tezcan Dilshener and Michel Wermelinger*

### Evolution and Migration

- Incremental and Iterative Reengineering towards Software Product Line: An Industrial Case Study 418  
*Gang Zhang, Liwei Shen, Xin Peng, Zhenchang Xing and Wenyun Zhao*
- The Evolution of Information Systems. A Case Study on Document Management 428  
*Paolo Salvaneschi*
- Testing & Quality Assurance in Data Migration Projects 438  
*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

- SE<sup>2</sup> Model to Support Software Evolution 512  
*Huzefa Kagdi, Malcom Gethers and Denys Poshyvanyk*
- 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 Claurton Siebra*
- Measuring Disruption from Software Evolution Activities Using Graph-Based Metrics 532  
*Prashant Paymal, Rajvardhan Patil, Sanjukta Bhowmick and Harvey Siy*
- 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

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

## Tool Demonstrations Track

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

## Tutorials

Effective Mining of Software Repositories <i>Marco D'Ambros and Romain Robbes</i>	598
Practical Combinatorial (t-way) Methods for Detecting Complex Faults in Regression Testing <i>Rick Kuhn and Raghu Kacker</i>	599
Research Methods in Computer Science <i>Serge Demeyer</i>	600