



8th IEEE Intl. Working Conference on Mining Software Repositories (MSR-11)

Honolulu, Hawaii, USA, May 21-22, 2011

Twitter: @msrconf, <http://msrconf.org>

MSR-11 Program: Saturday, 21 MAY 2011

8:30 – 8:45 **Welcome to MSR-11 from the chairs**

8:45 – 10:00 **Keynote**

Session chair: Thomas Zimmermann (PC Co-Chair)



Fantasy, Farms, and Freemium: What Game Data Mining Teaches Us About Retention, Conversion, and Virality

Prof. Dr. Jim Whitehead (University of California, Santa Cruz)

Abstract: In December of 2010, the new game CityVille achieved 6 million daily active users in just 8 days. Clearly the success of CityVille owes something to the fun gameplay experience it provides. That said, it was far from the best game released in 2010. Why did it grow so fast? In this talk the key factors behind the dramatic success of social network games are explained. Social network games build word-of-mouth player acquisition directly into the gameplay experience via friend invitations and game mechanics that require contributions by friends to succeed. Software analytics (mined data about player sessions) yield detailed models of factors that affect player retention and engagement. Player engagement is directly related to conversion, shifting a free player into a paying player, the critical move in a freemium business model. Analytics also permit tracking of player virality, the degree to which one player invites other players into the game. Social network games offer multiple lessons for software engineers in general, and software mining researchers in particular. Since software is in competition for people's attention along with a wide range of other media and software, it is important to design software for high engagement and retention. Retention engineering requires constant attention to mined user experience data, and this data is easiest to acquire with web-based software. Building user acquisition directly into software provides powerful benefits, especially when it is integrated deeply into the experience delivered by the software. Since retention engineering and viral user acquisition are much easier with web-based software, the trend of software applications migrating to the web will accelerate.

Bio: Jim Whitehead is an Associate Professor and Chair of Computer Science at the University of California, Santa Cruz, where he helped create the Computer Game Design program. He is also the founder and board chair of the Society for the Advancement of the Science of Digital Games, which hosts the yearly Foundations of Digital Games conference. Jim's research interests in the area of games include level design and procedural content generation. In the field of software engineering, Jim performs research on software bug prediction, software repository mining, and software evolution. He runs both the Augmented Design Lab and the Software Introspection Laboratory at UC Santa Cruz. Jim was the Program Co-Chair for the 2009 and 2010 Mining Software Repositories conference. Jim received his PhD in Information and Computer Science in 2000 from the University of California, Irvine, and his BS in Electrical Engineering from Rensselaer Polytechnic Institute in 1989.

10:00 – 10:30 **Coffee break**

10:30 – 12:00 **Paper session #1: LANGUAGE EVOLUTION**

Session chair: Michael Godfrey

Java Generics Adoption: How New Features are Introduced, Championed, or Ignored

Chris Parnin, Christian Bird, and Emerson Murphy-Hill

A Study of Language Usage Evolution in Open Source Software

Siim Karus and Harald Gall

How Developers Use the Dynamic Features of Programming Languages: The Case of Smalltalk

Oscar Callaú, Romain Robbes, Éric Tanter, and David Röthlisberger

An Exploratory Study of Identifier Renamings

Laleh M. Eshkevari, Venera Arnaoudova, Massimiliano Di Penta, Rocco Oliveto, Yann-Gaël Guéhéneuc, and Giuliano Antoniol

12:00 – 13:30 **Lunch break**

13:30 – 15:00 **Paper session #2: RETRIEVAL, REFACTORING, CLONES, READABILITY**

Session chair: Ahmed E. Hassan

Retrieval from Software Libraries for Bug Localization: A Comparative Study of Generic and Composite Text Models

Shivani Rao and Avinash Kak

Comparison of Similarity Metrics for Refactoring Detection

Benjamin Biegel, Quinten David Soetens, Willi Hornig, Stephan Diehl, and Serge Demeyer

Finding Software License Violations Through Binary Code Clone Detection

Armijn Hemel, Karl Trygve Kalleberg, Rob Vermaas, and Eelco Dolstra

A Simpler Model of Software Readability

Daryl Posnett, Abram Hindle, and Premkumar Devanbu

15:00 – 15:30 Coffee break

15:30 – 16:10 **Mining Challenge: Eclipse, Netbeans, Firefox, and Chrome**

Session chair: Adrian Schröter (Challenge Chair)

Operating System Compatibility Analysis of Eclipse and Netbeans Based on Bug Data

Xinlei (Oscar) Wang, Eilwoo Baik, and Premkumar Devanbu

What Topics do Firefox and Chrome Contributors Discuss?

Mario Luca Bernardi, Carmine Sementa, Quirino Zagarese, Damiano Distanto, and Massimiliano Di Penta

A Tale of Two Browsers

Olga Baysal, Ian Davis, and Michael W. Godfrey

Do Comments Explain Codes Adequately? Investigation by Text Filtering

Yukinao Hirata and Osamu Mizuno

Apples vs. Oranges? An Exploration of the Challenges of Comparing the Source Code of Two Software Systems

Daniel M. German and Julius Davies

16:10 – 17:00 **Short papers**

Session chair: Romain Robbes

Improving Identifier Informativeness Using Part of Speech Information

David Binkley, Matthew Hearn, and Dawn Lawrie

Bug-fix Time Prediction Models: Can We Do Better?

Pamela Bhattacharya and Iulian Neamtii

Integrating Software Engineering Data Using Semantic Web Technologies

Yuan-Fang Li and Hongyu Zhang

Improving Efficiency in Software Maintenance

Sergey Zeltyn, Peri Tarr, Murray Cantor, Robert Delmonico, Sateesh Kannegala, Mila Keren, Ashok Pon Kumar, and Segev Wasserkrug

An Empirical Analysis of the FixCache Algorithm

Caitlin Sadowski, Chris Lewis, Zhongpeng Lin, Xiaoyan Zhu, and E. James Whitehead, Jr.

Visualizing Collaboration and Influence in the Open-Source Software Community

Brandon Heller, Eli Marschner, Evan Rosenfeld, and Jeffrey Heer

17:45 – 20:00 **Conference dinner**

Benihana restaurant

2005 Kalia Rd, Honolulu, HI 96815

<http://www.benihana.com>

8th IEEE Intl. Working Conference on Mining Software Repositories (MSR-11)

Honolulu, Hawaii, USA, May 21-22, 2011, Twitter: @msrconf, <http://msrconf.org>

MSR-11 Program: Sunday, 22 MAY 2011

8:30 – 8:45 **Awards**

MSR-11 Best Paper Award • Mining Challenge Winner

8:45 – 10:00 **Keynote**

Session chair: Tao Xie (MSR PC Co-Chair)



Connecting Technology with Real-world Problems – From Copy-paste Detection to Detecting Known Bugs

Prof. Dr. Yuanyuan Zhou (University of California, San Diego; PatternInsight)

Abstract: In my talk, I will share with you our experience in applying and deploying our source code mining technology in industry. In particular, the most valuable lesson we have learned is that sometimes there is a bigger problem in the real world that can really benefit from our technology but unfortunately we do not know about it until we closely work with industry. In 2004, motivated from some previous research work that pointed out copy-pasting as a major reason for the majority of the bugs in device driver code, my students and I applied data mining technology (specifically frequent subsequence mining algorithms) in identifying copy-pasted code and also detecting forget-to-change bugs introduced during copy-pasting. The benefit of using data mining is that it is highly scalable (20 minutes for 4-5 millions lines of code), and can tolerate changes such as statement insertion/deletion/modification as well as variable name changes. When we released our tool called CP-Miner to the open source community, it attracted some inquiries from industry. These inquiries have motivated us to start a company to commercialize our tools.

During the commercialization process, our customers have taught us that our technology can be applied to solve a major headache faced by many embedded system vendors such as storage companies, network devices, telecommunications, handhelds, electronics, etc. These companies typically have to maintain tens and even hundreds of active branches of similar software, one for slightly different devices or with different customization. These branches are 60–70% similar and are being developed in parallel after split (and are usually never merged back together). So when developers decide to fix a bug or security hole in one branch, it is usually a big challenge for them to check what other branches a similar fix should be also applied. So many costly incidents have happened in the field due to some known bugs (a bug that has already been diagnosed and fixed in some other branches). Since these branches have diverged over the years, many code are similar but are not exactly the same. Also, file names can change, etc. Therefore, it is hard to rely on source control systems such as ClearCase, Subversion to keep track their differences. Being pushed by many customers who suffer this pain, we applied our technology to this problem and build a tool called PatchMiner. Currently PatchMiner has been deployed and widely used in several large companies (more information can be found at <http://www.patterninsight.com/>). It is a very interesting journey. To me personally, I learned that sometimes what I (as an academic) feel as a solution to a major problem may be only a nice-to-have in the real world; and it really requires close interaction with industry to understand their painpoints.

Bio: Yuanyuan Zhou is currently a Qualcomm Chair Professor at UC-San Diego. Prior to UCSD, she was a tenured associate professor at University of Illinois at Urbana Champaign. She has also worked at NEC Research Institute as a scientist after completing her Ph.D at Princeton in 2000. Her research interests span the areas of operating systems, architecture, system reliability and maintainability. She was the recipient for the Alfred Sloan Fellowship 2007, UIUC Gear Faculty Award 2006, NSF Career-2004 award, the CRA-W Anita-Borg Early Career Award 2005, the DOE Early Career Principle Investigator Award 2005, the IBM Faculty Award 2004 2005, the IBM SUR-2003 award and NetApp Faculty Fellowship 2010. She has 3 papers selected into the IEEE Micro Special Issue on Top Picks from Architecture Conferences and one best paper in SOSP 2005. She and her students have released several software quality assurance tools that are currently been used many developers in many commercial companies as well as open source projects.

10:00 – 10:30 **Coffee break**

10:30 – 12:00 **Paper session #3: SOFTWARE QUALITY**

Session chair: Christian Bird

Comparing Fine-Grained Source Code Changes And Code Churn For Bug Prediction

Emanuel Giger, Martin Pinzger, and Harald Gall

Security Versus Performance Bugs: A Case Study on Firefox

Shahed Zaman, Bram Adams, and Ahmed E. Hassan

Empirical Evaluation of Reliability Improvement in an Evolving Software Product Line

Sandeep Krishnan, Robyn R. Lutz, and Katerina Goševa-Popstojanova

Implementing Quality Metrics and Goals at the Corporate Level

Pete Rotella and Sunita Chulani

12:00 – 13:30 **Lunch break**

13:30 – 15:00 **Paper session #4: DEVELOPERS**

Session chair: Sung Kim

How Do Developers Blog? An Exploratory Study

Dennis Pagano and Walid Maalej

Entering the Circle of Trust: Developer Initiation as Committers in Open-Source Projects

Vibha Singhal Sinha, Senthil Mani, and Saurabh Sinha

Social Interactions around Cross-System Bug Fixings: The Case of FreeBSD and OpenBSD

Gerardo Canfora, Luigi Cerulo, Marta Cimitile, and Massimiliano Di Penta

Do Time of Day and Developer Experience Affect Commit Bugginess?

Jon Eyolfson, Lin Tan, and Patrick Lam

15:00 – 15:30 Coffee break

15:30 – 16:45 **Paper session #5: DEVELOPMENT SUPPORT**

Session chair: Martin Pinzger

Automated Topic Naming to Support Cross-project Analysis of Software Maintenance Activities

Abram Hindle, Neil A. Ernst, Michael W. Godfrey, and John Mylopoulos

Modeling the Evolution of Topics in Source Code Histories

Stephen W. Thomas, Bram Adams, Ahmed E. Hassan, and Dorothea Blostein

Software Bertillonage: Finding the Provenance of an Entity

Julius Davies, Daniel M. German, Michael W. Godfrey, and Abram Hindle

Supporting Software History Exploration

Alexander W. J. Bradley and Gail C. Murphy

16:45 – 17:00 **Closing remarks and Announcement of MSR-12**

MSR 2012: 9th Working Conference on Mining Software Repositories

June 2–3, 2012. Zurich, Switzerland.

<http://msrconf.org>

Twitter: @msrconf

MSR 2012 General Chair

Michele Lanza

University of Lugano, Switzerland

MSR 2012 Program Co-Chairs

Massimiliano Di Penta

University of Sannio, Italy

Tao Xie

North Caroline State University, USA

MSR 2012 Challenge Chair

Emad Shihab

Queen's University, Canada

