

# Introduction to MSR 2006

Stephan Diehl  
FB IV – Informatik  
Universität Trier  
Trier, Germany  
diehl@acm.org

Harald Gall  
Martin Pinzger  
Department of Informatics  
University of Zürich  
Zürich, Switzerland  
{gall,pinzger}@ifi.unizh.ch

Ahmed E. Hassan  
Performance Engineering  
Research In Motion (RIM)  
Waterloo, Canada  
ahmed@alumni.uwaterloo.ca

## Categories and Subject Descriptors

D.2 [Software Engineering]: Miscellaneous

## General Terms

Algorithms, Management, Measurement

## ABSTRACT

Software repositories such as source control systems, defect tracking systems, or archived communications between project personnel are used to help manage the progress of software projects. Software practitioners and researchers are beginning to recognize the potential benefit of mining this information to support the maintenance of software systems, improve software design/reuse, and empirically validate novel ideas and techniques. Research is now proceeding to uncover the ways in which mining these repositories can help to understand software development, to support predictions about software development, and to plan various aspects of software projects.

Following the success of the first two iterations of the MSR workshop in 2004 and 2005, MSR 2006 attracted even more submissions: We received 45 papers from 15 different countries. The international program committee accepted 16 full and 12 short papers for presentation at the workshop and inclusion in the proceedings. We are grateful for the excellent and professional review job done by the reviewers on such a tight schedule.

## 1. GOAL AND TOPICS

The goal of this two-day workshop is to establish a community of researchers and practitioners who are working to recover and use the data stored in software repositories for further understanding of software development practices. We expect the presentations and discussions in this workshop to continue on a number of general themes and chal-

lenges, from the previous MSR workshops held at ICSE 2004 and 2005, a recent TSE special issue on the MSR topic, and the Dagstuhl-Seminar on Multi-Version Program Analysis held in Summer 2005. The workshop covers themes and topics such as:

- Engineering tasks related to the infrastructure and tools needed to recover useful data from repositories
- Methods of integrating mined data from various data sources
- Development and validation of approaches to visualize and present such data
- Use of recovered history for system understanding and analysis of change patterns
- Models of defects and software reliability using data from such repositories
- Uncovering of the social processes and interaction between the development community
- Discovery of techniques to facilitate software reuse

As the field of mining software repositories has become more mature, we expect MSR 2006 to be a forum for exploratory work as well as continuing work. In 2006 we want to foster systematic comparisons of different approaches in our field. To this end, MSR 2006 includes a challenge session in addition to the demo session.

## 2. MSR 2006 CHALLENGE

The MSR Mining Challenge brings together researchers and practitioners who are interested in applying, comparing, and challenging their mining tools and approaches on software repositories. The this year's challenge covers the two well known open source software projects PostgreSQL and ArgoUML. 12 mining reports address the development process, team structure, change coupling, bug resolution, and cross-cutting concerns. 4 reports concentrated on analyzing ArgoUML, 5 on PostgreSQL, and 3 on analyzing both projects. The results of all 12 reports present valuable insights into both open source projects; for instance, did you know that the main contributors to PostgreSQL comprise only two people?

The reports and the program of the MSR Mining Challenge are the results of hard work. First of all, we would like to thank the authors of submitted reports. Many thanks goes to the open source community and in particular to the

ArgoUML and PostgreSQL project teams for sharing their project data. They enable us to develop, compare, and challenge our mining approaches and tools.

We further would like to thank Beat Fluri, Patrick Knab, and Sandro Bocuzzo for helping with the reviews, and Martin Pinzger, Harald Gall, Michele Lanza, and Marco D'Ambros for organizing the MSR challenge. We are looking forward to the participation in the final round of the MSR Mining Challenge to see the best mining tools.

For more information on MSR and the MSR Challenge we refer to <http://msr.uwaterloo.ca/>.