Theme and Goals of Workshop

Software repositories such as

- source control systems,
- archived communications between project personnel, and
- defect tracking systems

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.

Expected Workshop Output

The goal of this one-day workshop is to establish a community of researchers and practitioners who are working to recover and use the data stored in software repositories to further understanding of software development practices. We expect the presentations and discussions in this workshop to continue on a number of general themes and challenges, from the previous workshop held at ICSE 2004, such as:

- Engineering challenges related to the infrastructure and tools needed to recover useful data from these repositories
- Methods of integrating mined data from various historical sources
- Development and validation of approaches to visualize and present such data
- Use of recovered history for system understanding and analysis of change patterns
- Modeling 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

Since MSR 2004 was the first workshop to focus on the field of mining software repositories, it was exploratory in nature with a variety of results in different areas of the field. We expect MSR 2005 to retain an exploratory feel while at the same time promoting collaboration and systematic comparison of approaches and techniques. This mixture of exploratory work with systematic comparisons and collaborations should assist in maturing the field of mining software repository and in directing it to take a central role in supporting software development practices.
**Topics**

We solicit position papers that are no more than 5 pages long. The position papers may address issues along the general themes, including but not limited to the following:

- Approaches to study the quality of the mined data along with guidelines to ensure the quality of the recovered data
- Proposals for exchange formats, meta-models, and infrastructure tools to facilitate the sharing of extracted data and to encourage reuse and repeatability
- Models for social and development processes that occur in large software development projects
- Search techniques to assist developers in finding suitable components for reuse
- Techniques to model reliability and defect occurrences
- Analysis of change patterns to assist in future development
- Case studies on extracting data from repositories of large long lived projects
- Suggestions for benchmarks, consisting of large software repositories, to be shared among the community
- Applications of mined data

**Participants, Selection Process and Activities**

The workshop will be one-day long, preferably before the main ICSE conference. We prefer May 17 since some of the PC members and expected participants will be attending IWPC 2005 which runs on May 15 and 16. Last year the MSR workshop had over 40 participants. We would like to limit the participation this year to around 30 participants to maximize discussion and interaction.

For MSR 2004, we received 38 submissions from 14 countries and we selected 26 papers for publication. Four papers from MSR 2004 were invited for a special issue of the IEEE Transaction on Software Engineering on MSR. We expect several of the attendees of MSR 2004 to attend MSR 2005 along with a number of new attendees. With the rapid growth of open source projects, software repositories are becoming increasingly accessible to researchers. New researchers are continuously joining the MSR topic area. Interesting research results in the area of MSR are appearing in a number of international conferences such as WCRE, ICSM, and ICSE.

Presentations will be grouped into general themes, similar to the ones that emerged from last year’s workshop. Each paper presentation will be assigned a strict time limit, typically 15 minutes including clarification questions. Longer discussion slots, typically 30 minutes, will be scheduled at the end of each group of papers covering a particular theme. Following the format of MSR 2004, we plan to have an hour-long demo session following the lunch hour to encourage researchers to demonstrate their tools and work. The demo session will provide interested parties with an opportunity to learn more details about other work in the field.

**Program Committee**

The following are members of the Program Committee:

1. Harald Gall (U. Of Zurich, Switzerland)
2. Daniel German (U. of Victoria, Canada)
3. Philip Johnson (U. of Hawaii, USA)
4. Dewayne Perry (U. of Texas, USA)
5. Audris Mockus (Avaya Labs, USA)
6. Les Gasser (U. of Illinois, USA)
7. Andreas Zeller (Saarland U., Germany)
Workshop Organizers

Ahmed E. Hassan is a Ph.D. candidate at the University of Waterloo. He received his MMath and BMath from the school of Computer Science at UW in 1999 and 2001. For the last three years, Hassan has been working at Research In Motion (RIM) where he is responsible, along with his team, for the development of protocols, simulation tools, and software to ensure the scalability and reliability of RIM's global infrastructure. His research interests include mining source control data, high redundancy and availability systems, and visualization and migration of web applications. Previously he worked at IBM's Almaden Research Lab in San Jose and at Nortel Networks in Ottawa. He holds numerous patents in the area of wireless communications and distributed systems. (See http://plg.uwaterloo.ca/~aeehassa).

Hassan, along with Ric Holt and Audris Mockus organized MSR 2004, held last year at ICSE. He (along with Ric Holt, Audris Mockus, and Philip Johson) are guest editors for a special IEEE TSE on Mining Software Repositories (expected publication in Oct 2005). Hassan is the principal organizer he can be contacted at aeehassa@plg.uwaterloo.ca

Richard C. Holt is a Professor at the University of Waterloo, where his research interests include visualizing software architecture. This work includes reverse engineering of legacy systems and repairing software architecture. His architectural visualizations have included Linux, Mozilla (Netscape), IBM's TOBEY code generator, and Apache. His previous work includes foundational work on deadlock, development of a number of compilers and compilation techniques, development of the first Unix clone, and authoring a dozen books on programming and operating systems. He is one of the designers of the Turing programming language. (See http://plg.uwaterloo.ca/~holt)

Holt co-organized WoSEF (Workshop on Standard Exchange Format held at ICSE 2000) - http://www.cs.toronto.edu/~simsuz/wosef/. In addition, he was the general chair for IWPC 2001 - http://www.swen.uwaterloo.ca/~iwpc2001/

Stephan Diehl is a Professor of Computer Science at the Catholic University Eichstätt. He received his MSc as a Fulbright scholar at Worcester Polytechnic Institute, MA and his PhD as a DFG scholar from Saarland University. He is author of several books and more than 45 research papers on compilers and abstract machines, internet technology, visualization and software engineering.

He has been general chair of the ACM Web3D Symposium in 2001 and the ACM Symposium on Software Visualization in 2003. He has been a co-organizer of several workshops and Dagstuhl seminars. Currently he is chair of the steering committee of the ACM Symposium series on Software Visualization and heavily involved in the organization of the Symposium on Software Visualization which is collocated with ICSE 2005. Diehl is co-author of an ICSE 2004 paper on mining software archives and in 2004 he received an IBM Eclipse Innovation Award for visual exploration of software histories. (See http://rw4.cs.uni-sb.de/~diehl/index.php)