

Third International Workshop on Software Clones (IWSC)

Rainer Koschke
University of Bremen, Germany
koschke@tzi.de

James Cordy
Queen's University, Canada
cordy@cs.queensu.ca

Stanislaw Jarzabek
National University of Singapore
dcssj@nus.edu.sg

Katsuro Inoue
Osaka University, Japan
inoue@ist.osaka-u.ac.jp

Software clones are identical or similar pieces of code. They are often a result of copying and pasting as an act of ad-hoc reuse by programmers. Software clones research is of high relevance for industry. Many researchers have reported high rates of cloning in both industrial as well as open-source systems.

Many techniques exist that try to detect clones. There are also lines of research in clone detection that evaluate these approaches, reason about ways to remove clones, assess the effect of clones on maintainability, track their evolution, and investigate root causes of clones. Today, research in software clones is an established field with more than 100 publications in various conferences and journals.

History

Beginning in 2002 with the first edition of the *International Workshop on Detection of Software Clones (IWDSC)* as part of ICSM, this community has started to form itself and to find ways to meet and collaborate. Several events have followed since. The increase of participation shows how this field has gathered momentum. When Rainer Koschke organized the *First International Workshop on Detection of Software Clones (IWDSC)*, the community was still modest in size (about 15 people). The workshop drew clone detection researchers from Europe, North America, and Japan. Seven months later, Andrew Walenstein and Arun Lakhota organized a working session on clone detection benchmarking at IWPC'2003. In November of 2003, the *Second International Workshop on Detection of Software Clones (IWDSC'2003)* was held in conjunction with

WCRE. This workshop drew about 40 participants and resulted in a summary writeup which was published in ACM Software Engineering Notes [6]. This third workshop saw a broadening of scope and participant background. The workshop included a research team focused on refactoring of redundant code for use within agile development methods, a team exploring code sharing within and between open-source software, several teams investigating duplication within WWW sites, and an expert witness involved in copyright litigation. Clearly, finding and eliminating code copying and redundancy can be expected to be a common and ongoing concern for all of these groups in the foreseeable future. The summit so far was the Dagstuhl seminar "Duplication, Redundancy, and Similarity in Software" in 2006 where more than 40 international experts in the field met for several days [4]. How much this field has matured and how active it has become is demonstrated by the journal and conference publications related to software clones. In recent years, there is no conference in our field that does not have at least one contribution related to software clones.

At the *International Conference in Software Maintenance (ICSM)* in Beijing in 2008, more than 15 researchers have gathered informally to look into the next opportunity to meet. They have chosen CSMR in Kaiserslautern as their primary venue.

Goals

The purpose of the workshop is, essentially, to solidify and give shape to this research area and community. More specifically, the goals are to bring together researchers in order to evaluate the cur-

rent state of research, discuss common problems, discover new opportunities for collaboration, exchange ideas, and envision new areas of research, applications, empirical research, and approaches. The expected outcome is a list of open issues that should be addressed in the near future and concrete plans on how to conduct research to address the issues.

Topics

Several recent publications have summarized many open questions related to software clones. Among these are the summary of the Dagstuhl seminar [1], a book chapter in *Software Evolution* [3], a contribution in *Frontiers of Software Maintenance at ICSM 2008* [2] and a technical report by Roy and Cordy [5]. The short list of the open questions is as follows:

- the definition of software redundancy and clones
- the types, distribution, and nature of clones in software systems
- root causes for clones
- effects of clones
- techniques for software clone management
- cost/economic and trade-off models for clone removal
- clone and pattern detection techniques and algorithms
- clone/redundancy and pattern visualization tools
- relevance ranking for detectors
- evaluation and benchmarking issues
- language design and modeling techniques for redundancy elimination
- evolution of clones

This list is not exhaustive. It is a goal of this workshop to extend this list with new relevant and interesting topics.

Program Committee:

Giulio Antoniol	Jim Cordy
Massimiliano Di Penta	Michael Godfrey
Katsuro Inoue	Stanislaw Jarzabek
Miryung Kim	Rainer Koschke
Jens Krinke	Ettore Merlo
Andrew Walenstein	Michel Wermelinger

Workshop Format

In this whole-day workshop, we will have one invited presentation to set the scene so that all participants get an overview of the field. Then, we will have paper presentations in the morning sessions. These papers will raise important open research issues. These issues will be collected and be discussed in the afternoon in working sessions. Depending upon the number of participants and issues, we expect to split the audience into many groups. We will rotate groups and issues in the second half of the afternoon. The goal of these working groups is to define concrete research hypotheses and plans to address the issues. In a final plenary session, we will collect the results of the working groups.

References

- [1] R. Koschke. Survey of research on software clones. In R. Koschke, E. Merlo, and A. Walenstein, editors, *Duplication, Redundancy, and Similarity in Software*, number 06301 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2007. Dagstuhl.
- [2] R. Koschke. Frontiers in software clone management. In *Frontiers in Software Maintenance, Proceedings of the International Conference on Software Maintenance*. IEEE Computer Society Press, 2008.
- [3] R. Koschke. *Software Evolution*, chapter Identifying and Removing Software Clones, pages 15–39. Springer Verlag, 2008. Editors: Serge Demeyer and Tom Mens.
- [4] R. Koschke, E. Merlo, and A. Walenstein, editors. *Duplication, Redundancy, and Similarity in Software*, number 06301 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2007. Dagstuhl.
- [5] C. Roy and J. Cordy. A survey on software clone detection research. Technical Report 541, Queen’s University, 2007.
- [6] A. Walenstein, A. Lakhotia, and R. Koschke. Second international workshop on detection of software clones: workshop report. *ACM SIGSOFT Software Engineering Notes*, 29(2):1–5, Mar. 2004.