

1st International Workshop on Software Analytics (SWAN)

co-located with SANER 2015, Montréal, Canada

March 02, 2015

Olga Baysal* and Latifa Guerrouj†

**Université de Montréal, Montréal, QC, Canada*

†*Concordia University, Montréal, QC, Canada*

Email: **olga.baysal@umontreal.ca*, †*latifa.guerrouj@polymtl.ca*

Abstract—Software practitioners make technical and business decisions based on the understanding they have of their software systems. This understanding is grounded in their own experiences, but can be augmented by studying various kinds of development artifacts, including source code, bug reports, version control meta-data, test cases, usage logs, etc. Unfortunately, the information contained in these artifacts is typically not organized in the way that is immediately useful to stakeholders everyday decision making needs. To handle the large volumes of data, many practitioners and researchers have turned to analytics — the use of analysis, data, and systematic reasoning for making decisions. Thus, software analytics is an emerging field of modern data mining and analysis.

The workshop on Software Analytics (SWAN) aims at providing a common venue for researchers and practitioners across software engineering, data mining and mining software repositories research domains to share new approaches and emerging results in developing and validating analytics rich solutions, as well as adopting analytics to software development and maintenance processes to better inform their everyday decisions.

I. MOTIVATION

Software development projects generate impressive amounts of data. For example, source code, check-ins, bug reports, work items and test executions are recorded in software repositories such as version control systems (Git, Subversion, Mercurial, CVS) and issue-tracking systems (Bugzilla, JIRA, Trac), and the information about user experiences of interacting with software is typically stored in log files. While vast quantities of information are generated during development, very little is organized, stored and presented in a way that is immediately useful to developers and managers to support their decision-making. But at the same time, the information contained in various development artifacts including source code, bug report data, commit history, test suits, documentation, etc. could provide valuable insights about software projects.

Many prominent tech companies including IBM, Microsoft, and Google have embraced an analytics-driven culture to help improve their decision making. Analytics include methods of gathering, preprocessing, transforming and modelling raw data with the purpose of highlighting useful information and drawing conclusions from it. Software analytics are used to leverage large volumes of data from multiple sources to help practitioners make informed decisions about their projects.

While analytics solutions demonstrated promising results, there are many challenges left concerned with developing,

integrating, adopting analytics into software development processes.

The goals of the SWAN 2015 workshop are to discuss progress on software analytics, data mining and analysis; to gather empirical evidence on the use and effectiveness of analytics; and to identify priorities for a research agenda.

We aim at meeting these goals by offering the following activities. First, we invite researchers to submit a position/short paper (2–4 page long) describing their experiences with software analytics, sharing the tools and techniques they used, the challenges they faced, and the solutions that they found useful and successful. Second, we facilitate a discussion and dissemination of the presented research by opening up a discussion and involving participants in sharing their opinions. Third, we organize a panel discussion to identify and discuss hot topics that are most relevant to the workshop participants. This panel would include both researchers and practitioners interested in improving, adopting, and applying software analytics to the software development and maintenance activities, and beyond. Thus, the workshop invites both academic researchers and industrial practitioners for an exchange of ideas and collaboration.

II. TOPICS

The main theme of the SWAN 2015 workshop is to exchange ideas from both academia and industry to form a consolidated view regarding how good existing software analytics and tools are and how to benefit from them for different software development and maintenance activities. The topics of discussion include (but not limited to) the following:

- 1) Applications of software and data analytics to support decision making;
- 2) Data-driven approaches for data exploration and analysis;
- 3) Predictive analytics;
- 4) Data mining and visualization techniques;
- 5) Web analytics, development analytics, business intelligence tools, Hadoop tools;
- 6) Quantitative vs. qualitative analytics;
- 7) Large-scale data mining, analysis and analytics;
- 8) Software analytics for various stakeholders (e.g., managers vs. developers);

- 9) Methods of integrating data from multiple sources (applications, interfaces, mobile apps);
- 10) Empirical studies on how software analytics are used in practice and their effectiveness;
- 11) Negative results (“what did not work”) when adopting software analytics, and experience reports;
- 12) Identification of open research challenges and proposed solutions.

III. GOALS AND EXPECTED RESULTS

The SWAN 2015 workshop will provide an informal interactive forum for researchers and participants to exchange ideas and experiences, streamline research on software analytics, identify some common ground of their work, share lessons and challenges, thereby articulating a vision for the future of software analytics.

The intended outcomes of the SWAN 2015 workshop are:

- 1) Stimulate discussions, interest, and understanding in software analytics and its practical applications through presentations of position (2-page) or short (4-page) papers;
- 2) Bridging the gap between the theory and practice of software analytics by bringing together researchers and practitioners interested in software analytics;
- 3) Discuss challenges, experiences, and lessons learned when conducting research in software analytics, as well as when applying them in industrial settings;
- 4) Debate and argue the different possible strategies to overcome the challenges faced in software analytics and towards promising solutions to essential problems;
- 5) Explore different contemporary software analytics techniques and methods and put them in a common framework to make easy for both the research community and software practitioners finding appropriate methods to address their problems;
- 6) Advance the state of the art in the field of software analytics.

IV. FORMAT

We propose SWAN 2015 as a full-day workshop, consisting of an interesting mix including (1) an introductory presentation by organizers, (2) a keynote, (3) paper presentations, and (4) a lively industry panel and discussion. We invite researchers and practitioners in the field to present and discuss software analytics techniques and methods they applied in their works such as pattern recognition, machine learning, data mining, information visualization and large-scale data computing & processing.

We encourage workshop participants to share their experiences and lessons with the community, as well as fundamental challenges. Results, both positive and negative, will be sought from both industry and academic participants so as to develop an in-depth understanding of the subject in question. Invited experts will attend the discussion panel to provide additional feedback to workshop presenters and moderate a discussion

on the state-of-the-art and state-of-the-practice in software analytics.

The introductory presentation puts SWAN 2015 in context, whereas the keynote speaker transfers his/her knowledge and lessons learned from his/her own experience with software analytics. Invited workshop participants will provide presenters and attendees with feedback, novel ideas, and comments.

Each SWAN paper is allocated 15 minutes to be presented at the workshop. We strongly believe that bringing together people from academia and industry, attracting experts to participate in SWAN 2015, and balancing a keynote and invited experts with group discussions will help investigate better the possibilities of this exciting research area and meeting the ultimate SWAN 2015 goal of establishing a community of software analytics researchers and practitioners.

V. ORGANIZERS

Olga Baysal is an Assistant Professor at the Department of Computer Science and Operations Research (DIRO), Université de Montréal, Canada. She received a Ph.D in Computer Science from the University of Waterloo, Canada. Her research interests span a wide range of software engineering areas, including empirical software engineering, mining software repositories, software analytics, software maintenance and evolution, and human aspects of software engineering.

Latifa Guerrouj is a Postdoctoral Research Fellow at Concordia University, Canada. She received her Ph.D. from the Department of Computing and Software Engineering (DGIGL) of École Polytechnique de Montréal, Canada. Her research work and interests involve empirical software engineering, software analytics, mining software repositories, and software quality.

VI. KEYNOTE SPEAKER

Michele Lanza, University of Lugano, Switzerland

VII. STEERING COMMITTEE

Ayse Bener, Ryerson University, Canada
 Michael W. Godfrey, Waterloo University, Canada
 Ahmed E. Hassan, Queen’s University, Canada
 Tim Menzies, North Carolina State University, USA
 Thomas Zimmermann, Microsoft Research, USA

VIII. PROGRAM COMMITTEE

Bram Adams, Ecole Polytechnique of Montreal, Canada
 Ghizlane El Boussaidi, Ecole de Technologie Supérieure, Canada
 Julius Davies, University of Victoria, Canada
 Reid Holmes, Waterloo University, Canada
 Collin McMillan, University of Notre Dame, USA
 Rocco Oliveto, University of Molise, Italy
 Peter C. Rigby, Concordia University, Canada
 Romain Robbes, University of Chile, Chile
 Emad Shihab, Concordia University, Canada
 Leif Singer, University of Victoria, Canada
 Ashish Sureka, IIIT-Delhi, India