The Second International Workshop on Data Analysis Patterns in Software Engineering

June 2, 2014 - Hyderabad, India

Data scientists in software engineering seek insight in data collected from software projects to improve software development. The demand for these data scientists is growing rapidly and there is already a shortage of them. Data science is a skilled art with a steep learning curve. To reduce the learning curve, this workshop will collect best practices in form of data analysis patterns that lead to meaningful conclusions and can be reused in the context of similar data.

In the workshop, we will compile a catalog of such patterns that will help novice and experienced data scientists to better communicate about data analysis. The workshop is intended for anyone interested in how to analyze software engineering data correctly and efficiently in a community accepted way.

The major themes of this workshop are: big data, business intelligence, replication of experiments, theory building, automated data analysis, and their application in software engineering.

SUBMISSIONS

We are interested in patterns used to analyze data related to software development and maintenance (e.g. project plans, code, bugs, reviews, social networks) as well as generated with the use of software, (e.g. performance data, runtime data, usage data, user profiles).

We solicit papers of 3 pages maximum + one-page index card that summarizes the proposed data analysis pattern. In their papers, we encourage authors to describe patterns including the following information:

- Pattern name: title of the pattern
- Problem: why and when to apply the pattern
- Solution: how to apply the pattern
- Consequence: results and trade-offs of applying the pattern, common mistakes to be avoided while applying the pattern, etc.
- Examples: brief summary and/or cite example applications of the pattern in literature; if possible, R snippets or Weka code to apply the pattern, etc.

To develop the one-page index card, authors should use the DAPSE template (available in both Word and LaTex formats).

Authors can submit two types of papers: archival papers and non-archival papers. The program committee will review both types of papers. Accepted archival papers will be published in the workshop proceedings and the ACM and IEEE Digital Libraries. Accepted non-archival papers will be published in the DAPSE web site only.

Submissions as archival papers undergo a two-stage process: review and feedback stages. During the review stage, papers will be assessed based on the 1) clarity of description, 2) singular purpose of the pattern, 3) relevance of the pattern to address real SE problems, and 4) reusability of the pattern. During the feedback stage, authors will be contacted and guided to a more mature understanding of their patterns. Accepted papers will get a positive evaluation at the end of the whole two-stage process.

Prior application of the pattern by the authors is preferred but not mandatory. This workshop is more interested in the mechanics and choice of the data analysis than in the impact of published results.

All submissions must be in English. Papers must follow the ICSE 2014 formatting and submission instructions. Each paper must be accompanied by a pattern index card. Each contribution must be submitted electronically as one single PDF file including both paper and index card, using the submission site hosted by EasyChair.

It is the desire of the organizers that discussion of research at the workshop does not preclude publication of closely related material at conferences or journals. Authors of accepted papers will be able to choose whether to include their papers in the workshop proceedings.

WORKSHOP FORMAT

Before the workshop, there will be a blog to promote and discuss accepted patterns.

The workshop will consist of the following sessions:

- *Lightning session.* Authors will present their proposed pattern in lightning talks (5-10 minutes).
- *Discussion session.* Groups of participants will discuss the purpose, relevance, and reusability of the proposed patterns. This group work will eventually identify pattern types and classify/group patterns.
- *Breakout session.* Groups of participants will use the data analysis patterns from the previous session to solve data science tasks provided by the workshop organizers. The tasks will come from academic research but also from industry.

In both discussion and breakout sessions, each group will present their findings – applicability and usefulness of patterns – in a 5 minutes blitz presentation. The discussion will be run according to the Delphi method to help participants to reach a common agreement on the proposed patterns and their structure.

After the workshop, the organizers will propose a follow-up workshop at ISERN'14 meeting. The organizers also plan to submit a paper titled "Analysis Patterns: Elements of Reusable Data Analysis in SE" to ESEM'15. Selected authors from the workshop will be invited to contribute to the article.

IMPORTANT DATES

Archival Papers:
 Paper Submission: January 31, 2014
 PC review: February 17, 2014
 PC feedback: February 18-23, 2014
 Author Notification: February 24, 2014
 Camera-ready Version: March 14, 2014
Non-Archival Papers:
 Paper Submission: April 18, 2014
 Author Notification: May 5, 2014
Workshop: June 2, 2014

PROGRAM COMMITTEE

Alberto Bacchelli, Delft University of Technology, The Netherlands

Anita Sarma, University of Nebraska, USA

Burak Turhan, University of Oulu, Finland

Carolyn Seaman, University of Maryland, USA

Chris Bird, Microsoft Research, USA

Daryl Posnett, University of California, USA

David Lo, Singapore Management University, Singapore

Earl Barr, University College London, UK

Forrest Shull, Fraunhofer Center, USA

Harald Gall, University of Zurich, Switzerland

Lionel Briand, University of Luxembourg, Luxembourg

Martin Shepperd, Brunel University, UK

Massimiliano Di Penta, University of Sannio, Italy

Nachiappan Nagappan, Microsoft Research, USA

Patrick Wagstrom, IBM TJ Watson Research Center, USA

Pete Rotella, Cisco Systems, USA

Rob DeLine, Microsoft Research, USA

Roberto Di Cosmo, INRIA, France

Tim Menzies, West Virginia University, USA

Tom Zimmermann, Microsoft Research, USA

Tracy Hall, Brunel University, UK

Sandro Morasca, Università dell'Insubria, Italy

Shi Han, Microsoft Research, China

Stefan Wagner, University of Stuttgart, Germany

Vibha Sinha, IBM Research, India

Ye Yang, Chinese Academy of Sciences, China

Yuanfang Cai, Drexel University, USA

ORGANIZERS

General chair: Barbara Russo, Free University of Bozen-Bolzano, Italy

Program co-chairs: Davide Falessi, Fraunhofer CESE, USA; Venkatesh-Prasad Ranganath, Kansas State University, USA

Publicity and social media co-chairs: Laura Inozemtseva, University of Waterloo, Canada; Stefan Wagner, University of Stuttgart, Germany

Webmaster: Rodrigo Rocha Gomes e Souza, Federal University of Bahia, Brazil