Quality Software Engineering (QSE) Research

Research Overview
Crowdsourced Software Inspection (CSI)

Stefan Biffl    Marta Sabou    Dietmar Winkler
Institute of Software Technology and Interactive Systems
Vienna University of Technology
http://qse.ifs.tuwien.ac.at
Quality Software Engineering (QSE) Research
Research Topics & Major Contributors

Research Lines
- Multi-Disciplinary Systems Engineering
- Software systems that efficiently coordinate humans/disciplines

Research Areas
- Model Quality Assurance and Process Improvement in Multi-Disciplinary Engineering
  *Dietmar Winkler*
- Semantic Web Technologies and Human Computation
  *Marta Sabou*
- Collective Intelligence Systems & Software Architecture
  *Angelika & Jürgen Musil*
- Concepts, Methods, Tools for Empirical Software/System Engineering & Didactics
  *Stefan Biff*
Research Topic
Crowdsourcing in Software Engineering

“The act of undertaking any external software engineering tasks by an undefined, potentially large group of online workers in an open call format.” (Mao et al., 2016)

Use of Crowdsourcing in Software Engineering

Model Quality Inspection (MQI)

Goal

- Identify defects in (large-scale) models early, effectively, and efficiently.
- Foundations: Best-Practice Software Inspection, Crowdsourcing Methods.

Example Use Case

**Restaurant**

- Introduction
- Scenario: Order management
- Scenario: Recipe management
- Scenario: Shopping and Storage Management

System EER Diagram Model

Does the model completely and correctly represent the specification?

Are there defects in the model?
Crowdsourced Software Inspection (CSI)  
(Winkler et al., 2017a)

- **Text Analysis**
  - Identification of *Expected Model Elements*,
    e.g., entities, attributes, relationships.
  - Candidate task for automation with *natural language processing* approaches.

- **Model Analysis**
  - Model analysis based on *Expected Model Elements* (EMEs).
  - Sample defects: Missing EMEs, wrongly modeled EMEs.
  - Challenging: Synonyms, singular/plural, domain knowledge, etc.

Controlled Experiment Setup

Crowdsourcing-Based Inspection (CSI)

- 30 min: Tutorial Group A & B
- 60 min: Text Analysis Group A → Model Analysis Group A
- 60 min: Model Analysis Group B → Text Analysis Group B

Traditional Best-Practice Inspection (Pen&Paper)

- 120 min: Tutorial Group C → Traditional Inspection (Group C)

- **Crowdsourced Software Inspection** vs. Traditional Best-Practice Inspection (control group) with cross-over design.
- 75 participants in academic course in 4 sessions (63 crowd workers; 12 inspectors).
- Study Material:
  - Design Specification: 3 pages, 7 scenarios and 110 EMEs
  - EER Diagram: 9 entities, 13 relationships, 32 attributes; 33 re-seeded defects.
  - Supporting material: Questionnaires (experience and feedback), guidelines for task execution
  - Tool: *Crowdflower* application.
Initial CSI Experiment Results
(Winkler et al., 2017b)

- Effectiveness (number of true defects)

<table>
<thead>
<tr>
<th>Group</th>
<th>No. part.</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSI</td>
<td>63</td>
<td>15</td>
<td>6.5</td>
<td>7</td>
<td>4.9</td>
</tr>
<tr>
<td>P&amp;P</td>
<td>12</td>
<td>21</td>
<td>5.7</td>
<td>10</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Pen & Paper (P&P) inspectors reported more (true) defects (due to more inspection time; 2 hours vs. 1 hour for CSI)

- Defect Detection Efficiency (true defects per hour)

<table>
<thead>
<tr>
<th>Group</th>
<th>No. Part.</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSI</td>
<td>63</td>
<td>3.5</td>
<td>2.46</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>P&amp;P</td>
<td>12</td>
<td>5.7</td>
<td>2.17</td>
<td>2</td>
<td>9</td>
</tr>
</tbody>
</table>

Higher efficiency for P&P (with 1 hour EME penalty for CSI).

- Better control with CSI on inspection coverage, a foundation for the inspection of large models.

QSE Research – Summary and Cooperation Options

- Close fundamental gaps in scientific knowledge on
  - Multi-Disciplinary Systems Engineering
  - Software systems that efficiently coordinate humans/disciplines

- Evaluation in empirical studies in several application domains.

- Application examples
  - Model Quality Assurance + Human Computation
  - Collective Intelligence Systems

- We are interested in research cooperation
  - Systematic literature reviews
  - Method application and evaluation
  - Prototype design and evaluation
  - Empirical studies with practitioners and students

- Visit us online at http://qse.ifs.tuwien.ac.at
  - http://cdl.ifs.tuwien.ac.at/downloads