Towards an Experiment Line on Software Inspection with Human Computation

Stefan Biffl\textsuperscript{1}, Marcos Kalinowski\textsuperscript{2}, Dietmar Winkler\textsuperscript{1,3}

\textsuperscript{1} TU Wien, Institute of Information Systems Engineering, Vienna, Austria
\textsuperscript{2} Pontifical Catholic University of Rio de Janeiro, Brazil
\textsuperscript{3} Christian Doppler Laboratory for Security and Quality Improvement in the Production System Life Cycle, Vienna, Austria.

qse.ifis.tuwien.ac.at
www.sqi.at
dietmar.winkler@tuwien.ac.at
Motivation & Key Questions

Motivation and Application Context

§ Traditional **Software Inspection** to identify defects in design documents and models.
§ **Limited resources** for inspection processes.
§ **Large-scale software engineering models.**

Key questions

§ How to handle large-scale engineering models with limited resources in a human computation context?
§ How could an experiment line address variability for planning experiments?
§ What are the expected benefits of an experiment line?

Goal of this presentation

§ **Software Inspection Process with Human Computation (Crowdsourcing Supported Inspection (CSI)).**
§ Managing variability of a family of experiments in academic / industry environment.
Illustrative Example .. Starting point

Inspection Task

- Input: Reference document, e.g., requirements specifications.
- Task: Identify defects in (large-scale) models early, effective, and efficient.
- Output: True defects in the model.

Does the model completely and correctly represent the specification?

Are there defects in the model?

System EER Diagram Model
Traditional (Best-Practice) Inspection

Benefits:

- **Formal and structured** process approach (five inspection phases) to identify defect early and efficient in engineering artifacts.
- **Well established** and investigated process approach.
- **Guidelines and reading techniques** support defect detection, e.g., perspectives or scenarios.

Limitations:

- Typically (expensive) **experts** are part of the inspection team.
- **Limited resources** (e.g., 2h of inspection) for inspecting large-scale documents need for several inspection cycles and coordination.
- **Limited tool support.**
“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)

Crowdsourcing (CS) mechanism has been applied in software engineering planning and analysis, implementation, maintenance, and testing...

.. but very limited in the area of Software Quality Assurance and Software Inspection.

Research Questions

Objectives and Approach

- Support of software inspection tasks with crowdsourcing techniques in context of a family of experiment.
- Key Elements:
  - Splitting up inspection tasks into small pieces of work,
  - Distributing inspection work load to a crowd of workers and/or experts within an organization,
  - Providing tool support.
  - Variability model for (inspection) experiment planning.

Research Questions

- RQ.1: How to handle large-scale engineering models with limited resources in a human computation context?
- RQ.2: How could an experiment line address variability for planning experiments?
- RQ.3: What are the expected benefits for such an experiment line?
Software Inspection with Human Computation

RQ.1: How to handle large-scale engineering models with limited resources in a human computation context?

1. **Text Analysis:**
   - Identification of Model Elements (i.e., Entities, Relationships, Attributes) that represent the building blocks of a model ‡ foundation for defect detection.

2. **Model Analysis (Defect Detection):**
   - Based on ME, defect detection in the model under inspection ‡ candidate defect reports by individual crowd workers.

3. **Defect Validation (“Team Meeting”):**
   - Validation of reported defects derived from previous model analysis tasks. ‡ justified and validated defects by crowd workers.
Feature Model for Software Inspection with Human Computation

RQ.2: How could an experiment line address variability for planning experiments?
Study Setup and Design

Basic Study Design

- **Study Type:** Controlled Experiment
- **CSI process vs. Traditional Best-Practice Inspection** (control group) with cross-over design.
- **>75 participants** in academic course in 4 different experiment runs
- **Study Material:**
  - Design Specification: 3 pages, 7 scenarios and 110 MEs.
  - EER Diagram: 9 entities, 13 relationships, 32 attributes; 33 seeded defects.
  - Questionnaires (experience and feedback), guidelines for task execution.
  - Tool: *Crowdflower/Figure Eight* ¹ application and configuration.

¹ www.crowdflower.com / Figure Eight: www.figure-eight.com/
Study History

- Four different experiment runs until today.
  - Fall 2016, spring 2017, fall 2017, spring 2018 (currently running).

- Similar / slightly improved material: Reference Document (Scenarios), Inspection Artifact (EER Model), Experience Questionnaire, Feedback Questionnaires, Artefact Scale, Seeded Defects, Process for traditional (pen & paper inspection)

- Implemented Variations:
  - Focus on CSI process improvement, scope, and tooling improvements

<table>
<thead>
<tr>
<th>Experiment Runs:</th>
<th>Fall 2016</th>
<th>Spring 2017</th>
<th>Fall 2017</th>
<th>Spring 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Steps:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Text Analysis</td>
<td>TASK Ax</td>
<td>TASK Ax</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Model Analysis</td>
<td>TASK Bx</td>
<td>TASK Bx</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Defect Validation</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tooling:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Text Analysis</td>
<td>free-text data collection (Crowdflower)</td>
<td>free-text data collection (Crowdflower)</td>
<td>guided data collection (Crowdflower)</td>
<td>guided data collection (Figure Eight)</td>
</tr>
<tr>
<td>+ Model Analysis</td>
<td>free-text data collection (Crowdflower)</td>
<td>free-text data collection (Crowdflower)</td>
<td>guided data collection (Crowdflower)</td>
<td>guided data collection (Figure Eight)</td>
</tr>
<tr>
<td>+ Defect Validation</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Expected Benefits

RQ.3: What are the expected benefits for such an experiment line?

- Support of **strategically planning** of a family of studies to facilitate the cooperation of research groups.
- Supporting **systematic reuse** of experiment design and material.
- Foundation for an **aggregation strategy** of experiment results (towards a Body of Knowledge).
- Integration of **industry studies** as integral part of the family of experiments based on a proven experiment setup.

**Study Context: Adapted Inspection Process**

1. **Reference Document** → **Text Analysis: Expected Model Elements (EME)** → **Defect Detection (Model Analysis)** → **Defect Validation** → **Validated & Aggregated Defect List**

---

Feature Model for Strategic Planning of Study Variants

Institute of Information Systems Engineering
Summary and Future Work

Summary

- Improving traditional Software Inspection with Human Computation, i.e., Crowdsourcing Based Inspection (CSI).
- Feature Models support planning empirical studies in academia and industry based on a stable study architecture.

Future Work

- Along the planning of a family of empirical studies we will focus on:
  - Needs from candidate industry partners to improve defect detection within a defined context (Case Studies?).
  - Various model types, e.g., structural and behavior models in different domains.
  - Various model sizes towards large-scale software engineering models.
Thank you ...

Towards an Experiment Line on Software Inspection with Human Computation

Stefan Biffl\textsuperscript{1}, Marcos Kalinowski\textsuperscript{2}, Dietmar Winkler\textsuperscript{1,3}

\textsuperscript{1} TU Wien, Institute of Information Systems Engineering, Vienna, Austria
\textsuperscript{2} Pontifical Catholic University of Rio de Janeiro, Brazil
\textsuperscript{3} Christian Doppler Laboratory for Security and Quality Improvement in the Production System Life Cycle, Vienna, Austria.

qse.ifis.tuwien.ac.at/~winkler
www.sqi.at
Dietmar.Winkler@tuwien.ac.at