Comparing Trace Visualizations for Program Comprehension through Controlled Experiments

Florian Fittkau, Santje Finke, Wilhelm Hasselbring, and Jan Waller

2015-05-19
Introduction

- Appropriate tools for efficient and effective program comprehension
- Empirical studies required
- Trace visualizations
- Cornelissen et al. [CZvDvR09] compared IDE, and IDE + Extravis versus
We compared Extravis to our web-based tool EplorViz in two experiments.

- Reusable experimental design (experimental package [FFHW15])
- Employed strategies by the subjects
- Common challenges at conducting visualization experiments
ExplorViz in a Nutshell

Tools

Fittkau, Finke, Hasselbring, Waller
Comparing Trace Visualizations
2015-05-19
Alternative Hypotheses

Controlled Experiments

- **H1**: Extravis and ExplorViz require different times for completing typical program comprehension tasks.

- **H2**: The correctness of solutions to typical program comprehension tasks differs between Extravis and ExplorViz.
First experiment

Controlled Experiments

- **PMD**\(^1\) with 30 subjects
- Source code analyzer for, e.g., Java
- Used trace: 279 classes and about 420,000 method calls
- Fairness concern: Tasks similar to Cornelissen et al. (Checkstyle)

\(^1\)http://pmd.sf.net
### Descriptive Statistics for First Experiment

<table>
<thead>
<tr>
<th>PMD</th>
<th><strong>Time Spent</strong></th>
<th><strong>Correctness</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>EXTRAVIS</strong></td>
<td><strong>ExplorViz</strong></td>
</tr>
<tr>
<td>mean</td>
<td>47.65</td>
<td>34.27</td>
</tr>
<tr>
<td>difference</td>
<td><strong>-28.06%</strong></td>
<td></td>
</tr>
<tr>
<td>sd</td>
<td>9.96</td>
<td>3.14</td>
</tr>
<tr>
<td>min</td>
<td>23.04</td>
<td>29.43</td>
</tr>
<tr>
<td>median</td>
<td>48.89</td>
<td>33.84</td>
</tr>
<tr>
<td>max</td>
<td>65.07</td>
<td>38.99</td>
</tr>
</tbody>
</table>

- **Analyzed users**: 12
- **Shapiro-Wilk W**: EXTRAVIS = 0.8807, ExplorViz = 0.9459
- **Levene F**: EXTRAVIS = 2.4447, ExplorViz = 2.0629
- **Student’s t-test**:
  - df = 22
  - t (EXTRAVIS) = 4.4377, t (ExplorViz) = -3.6170
  - p-value (EXTRAVIS) = 0.0002, p-value (ExplorViz) = 0.0015
Replication: **Babsi**\(^2\) with 50 subjects
- Small Android app for Antibiotic Stewardship
- Used trace: 42 classes and 388 method calls (smaller than PMD)
- Used task definition framework by Pacione et al. [PRW04]

\(^2\)http://babsi.sf.net
# Descriptive Statistics for Replication

## Controlled Experiments

<table>
<thead>
<tr>
<th>Babsi</th>
<th>Time Spent</th>
<th>Correctness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EXTRAVIS</td>
<td>ExplorViz</td>
</tr>
<tr>
<td>difference</td>
<td>-7.64%</td>
<td>+38.72%</td>
</tr>
<tr>
<td>sd</td>
<td>7.25</td>
<td>6.48</td>
</tr>
<tr>
<td>min</td>
<td>18.94</td>
<td>19.38</td>
</tr>
<tr>
<td>median</td>
<td>31.27</td>
<td>27.19</td>
</tr>
<tr>
<td>max</td>
<td>43.20</td>
<td>41.56</td>
</tr>
</tbody>
</table>

- Shapiro-Wilk W: EXTRAVIS = 0.9618, ExplorViz = 0.9297
- Levene F: EXTRAVIS = 0.4642, ExplorViz = 0.0527

**Student’s t-test**

- df: EXTRAVIS = 45, ExplorViz = 45
- t: EXTRAVIS = 1.2006, ExplorViz = -3.6531
- p-value: EXTRAVIS = 0.2362, ExplorViz = 0.0007

Analyzed users: EXTRAVIS = 24, ExplorViz = 23
Recommendations and Challenges

Controlled Experiments

Recommendations

▶ Automated tutorial
▶ Electronic questionnaire
▶ Recordings of the PC screen

Challenges

▶ We had to implement an input file generator for Extravis
▶ Tutorial material and source code for Extravis unavailable
▶ Research prototypes should be available as open source
Summary and Outlook

Conclusions

- Significant decrease in time spent and increase in correctness utilizing ExplorViz (PMD)
- Significant increase in correctness utilizing ExplorViz (Babsi)
- Open source web-based tool³

![ExplorViz](http://www.explorviz.net)

Future Work:

- Professionals as subjects
- Comparison to other techniques, e.g., Trümper et al. [TBD10]
- Development of validated programming experience questionnaire

³[http://www.explorviz.net](http://www.explorviz.net)

