An empirical investigation into the usage of a live debugger

Master Thesis
Roger Stebler
Supervisor: Andrei Chiș
Problem

Applications are complex; debugging is difficult

Using the debugger in an efficient way is often not an easy task

How to improve the debugger?
How do developers use the debugger?
Research Questions

- **RQ1. Debugging Strategies**
- **RQ2. Domain-specific information**
- **RQ3. Usage of extensions**
Experiment Setup

**Group 1**
- 5 participants
- can use any tools

**Group 2**
- 5 participants
- can use any tools
  + domain-specific extensions

Participants randomly selected
∅ programming experience: 10 years
∅ experience with Pharo: 4.5 years
Experiment: Solve a bug

Bug in *Rubric* – default text editor in Pharo
Optimal path

evaluateSelectionAndDo: aBlock

self lineSelectAndEmptyCheck: [^ ''].
^ self
  evaluate: self selectionForDoitAsStream
  andDo: aBlock
Code of the faulty method

```python
def computeSelectionIntervalForCurrentLine(i, left, right, s):
    s := self string.
    i := s encompassParagraph: self selectionInterval.
    left := i first.
    right := i last.
    whileTrue: [ left <= right and: [ (s at: left) =$" or: [(s at: left) isSeparator] ] ]
        left := left + 1.
    whileTrue: [ left <= right and: [ (s at: right) =$" or: [(s at: right) isSeparator] ] ]
        right := right - 1.
    left to: right
```
Optimal path

evaluateSelectionAndDo: aBlock

self announce: (GLMEvaluateSelection new action: aBlock; makeSynchronous)
Domain-specific extensions: Announcements
**Optimal path**

```
evaluateSelectionAndDo: aBlock
self lineSelectAndEmptyCheck: [''].
  ^ self
  evaluate: self selectionForDoitAsStream
  andDo: aBlock
```
Domain-specific extensions: Stream

Group 1

Group 2
Optimal path

**evaluateSelectionAndDo:** aBlock
self lineSelectAndEmptyCheck: [^ ''].
^ self
evaluate: self selectionForDoitAsStream
andDo: aBlock
Navigation Paths

evaluateSelectionAndDo: aBlock
self lineSelectAndEmptyCheck: [^ ''].
^ self
  evaluate: self selectionForDoitAsStream
  andDo: aBlock
<table>
<thead>
<tr>
<th>Group</th>
<th>Realistic?</th>
<th>Difficulty</th>
<th>Duration</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>agree</td>
<td>difficult</td>
<td>40:00</td>
<td>C</td>
</tr>
<tr>
<td>P2</td>
<td>strongly agree</td>
<td>neutral</td>
<td>23:20</td>
<td>A</td>
</tr>
<tr>
<td>P3</td>
<td>agree</td>
<td>difficult</td>
<td>40:54</td>
<td>F</td>
</tr>
<tr>
<td>P4</td>
<td>agree</td>
<td>difficult</td>
<td>45:18</td>
<td>B</td>
</tr>
<tr>
<td>P5</td>
<td>strongly agree</td>
<td>easy</td>
<td>30:36</td>
<td>A</td>
</tr>
<tr>
<td>P6</td>
<td>agree</td>
<td>easy</td>
<td>34:34</td>
<td>B</td>
</tr>
<tr>
<td>P7</td>
<td>agree</td>
<td>difficult</td>
<td>39:41</td>
<td>B</td>
</tr>
<tr>
<td>P8</td>
<td>strongly agree</td>
<td>difficult</td>
<td>48:30</td>
<td>B</td>
</tr>
<tr>
<td>P9</td>
<td>agree</td>
<td>neutral</td>
<td>25:17</td>
<td>B</td>
</tr>
<tr>
<td>P10</td>
<td>strongly agree</td>
<td>difficult</td>
<td>52:44</td>
<td>B</td>
</tr>
</tbody>
</table>

A. Participant solved the bug  
B. Participant discovered that there was a problem in computing the selection  
C. Participant successfully navigated the propagation of the announcement, but made no other progress  
F. Participant made no real progress towards solving the bug and was unable to navigate the propagation of the announcement
RQ1. Debugging Strategies

- Behavior-driven exploration
- In-depth path exploration
- Code-driven exploration
Navigation P5
Navigation P7
Navigation P3
RQ2. Usage of domain-specific information: Announcements
### RQ2. Usage of domain-specific information: Announcements

<table>
<thead>
<tr>
<th>Group</th>
<th>Time</th>
<th>Method</th>
<th>Experience with announcements</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>16:58</td>
<td>Checking action of announcement</td>
<td>Not at all</td>
</tr>
<tr>
<td>P2</td>
<td>03:03</td>
<td>Stepping through the code using debugger</td>
<td>Extremely knowledgeable</td>
</tr>
<tr>
<td>P3</td>
<td>-</td>
<td>Did not find the method</td>
<td>Slightly knowledgeable</td>
</tr>
<tr>
<td>P4</td>
<td>04:35</td>
<td>Checking action of announcement &amp; extension</td>
<td>Not at all</td>
</tr>
<tr>
<td>P5</td>
<td>05:09</td>
<td>Checking event registration</td>
<td>Extremely knowledgeable</td>
</tr>
<tr>
<td>P6</td>
<td>06:04</td>
<td>Skipped by searching for error message</td>
<td>Slightly knowledgeable</td>
</tr>
<tr>
<td>P7</td>
<td>01:58</td>
<td>Checking event registration</td>
<td>Extremely knowledgeable</td>
</tr>
<tr>
<td>P8</td>
<td>06:50</td>
<td>Checking action of announcement</td>
<td>Moderately knowledgeable</td>
</tr>
<tr>
<td>P9</td>
<td>03:54</td>
<td>Checking event registration</td>
<td>Somewhat knowledgeable</td>
</tr>
<tr>
<td>P10</td>
<td>16:12</td>
<td>Checking action of announcement</td>
<td>Somewhat knowledgeable</td>
</tr>
</tbody>
</table>
RQ2. Usage of domain-specific information: Stream
## RQ2. Usage of domain-specific information: Stream

<table>
<thead>
<tr>
<th>Group</th>
<th>Time</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>-</td>
<td>not checked at all</td>
</tr>
<tr>
<td>P2</td>
<td>01:29</td>
<td>checked contents manually using method #contents</td>
</tr>
<tr>
<td>P3</td>
<td>2</td>
<td>not checked at all</td>
</tr>
<tr>
<td>P4</td>
<td>00:00</td>
<td>extension</td>
</tr>
<tr>
<td>P5</td>
<td>00:00</td>
<td>extension</td>
</tr>
<tr>
<td>P6</td>
<td>01:10</td>
<td>after conversion to String</td>
</tr>
<tr>
<td>P7</td>
<td>00:13</td>
<td>checked contents manually using method #contents</td>
</tr>
<tr>
<td>P8</td>
<td>00:51</td>
<td>after conversion to String</td>
</tr>
<tr>
<td>P9</td>
<td>00:00</td>
<td>extension</td>
</tr>
<tr>
<td>P10</td>
<td>00:00</td>
<td>extension</td>
</tr>
</tbody>
</table>
RQ2. Usage of domain-specific information

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
<th>Used extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Announcements</td>
<td>06:59</td>
<td>07:28</td>
<td>1/5</td>
</tr>
<tr>
<td>Stream</td>
<td>00:56</td>
<td>00:00</td>
<td>5/5</td>
</tr>
</tbody>
</table>

Domain-specific information can improve efficiency in debugging
RQ3. Usage of debugging extensions

Developers use only familiar extensions; ignore new/unknown extensions

→ Encourage to use new extensions
Improvement Opportunities: Highlight Extensions
Improvement Opportunities:
Domain-specific actions & Graphical Components
Improvement Opportunities: Repetitive actions

```
#define factorial

self = 0 ifTrue: [1].
self > 0 ifTrue: [self * (self - 1) factorial].
sel error: 'Not valid for negative integers'
```
Debugging is difficult

Experiment

Results

- Domain-specific information can be useful

- New extensions often ignored

Improvement Opportunities
Tool Usage Timeline
Navigation P1
Navigation P3

![Diagram of navigation process P3]
Navigation P5
Navigation P6
Navigation P7
Navigation P8
Navigation P9
Navigation P10
Improvement Opportunities: Highlight Extensions
Improvement Opportunities:
Graphical Components
Improvement Opportunities: Domain-specific actions
Demo