Meta-Heuristic Generation of Robust XPath Locators for Web Testing

Maurizio Leotta  Andrea Stocco  Filippo Ricca  Paolo Tonella
Outline

- The robust locator problem in web testing
- Locator generation as graph reachability problem
- Algorithms for locator generation
  - Greedy optimal
  - Meta heuristic (GA) suboptimal
Web testing

Name: John
Surname: Doe
Mobile: 123456789

<html>
<body>
<table id="userInfo">
  <tr><td>Name: </td><td title="name">John</td></tr>
  <tr><td>Surname: </td><td title="surname">Doe</td></tr>
  <tr><td>Mobile: </td><td title="mobile">123456789</td></tr>
</table>
</body>
</html>

private final WebDriver driver;
void testMobileNumber(String name, String surname) {
  // insert name, surname; submit
  // get result page
  assertEquals(driver.findElement(By.xpath("/html/body/table/tr[3]/td[2]")).getText, "123456789");
}
Web element locators

<table>
<thead>
<tr>
<th>Tool</th>
<th>Kind</th>
<th>Generated XPath Locators for the Target Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>FirePath</td>
<td>rel</td>
<td>//*[@id=&quot;userInfo&quot;]/tr[3]/td[2]</td>
</tr>
<tr>
<td>Chrome</td>
<td>rel</td>
<td>//*[@id=&quot;userInfo&quot;]/tr[3]/td[2]</td>
</tr>
<tr>
<td>XPath Helper</td>
<td>abs</td>
<td>/html/body/table[@id=&quot;userInfo&quot;]/tr[3]/td[@title=&quot;mobile&quot;]</td>
</tr>
<tr>
<td>XPath Checker</td>
<td>rel</td>
<td>id('userInfo')/tr[3]/td[2]</td>
</tr>
<tr>
<td>ROBULA</td>
<td>rel</td>
<td>//td[@title=&quot;mobile&quot;]</td>
</tr>
</tbody>
</table>

Target Element

Name: John
Surname: Doe
Mobile: 123456789
# Robust locators

<table>
<thead>
<tr>
<th>Tool</th>
<th>XPath Locators Robustness</th>
<th>✓ robust</th>
<th>× broken</th>
</tr>
</thead>
<tbody>
<tr>
<td>FirePath</td>
<td>✗ //*[@id=&quot;userInfo&quot;]/tr[3→4]/td[2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrome</td>
<td>✗ //*[@id=&quot;userInfo&quot;]/tr[3→4]/td[2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XPath Helper</td>
<td>✗ /html/body/table[@id=&quot;userInfo&quot;]/tr[3→4]/td[@title=&quot;mobile&quot;]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XPath Checker</td>
<td>✗ id('userInfo')/tr[3→4]/td[2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROBULA</td>
<td>✓ //td[@title=&quot;mobile&quot;]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Target Element: Name: John  
Surname: Doe  
Gender: Male  
Phone: 123456789
Locator generation

transfAddName
  ///*/td -> //tr/td

transfAddPredicate
  //tr/td -> //tr[@name='data']/td
  //tr/td -> //tr[2]/td

transfAddLevel
  //tr/td -> ///*/tr/td

**Completeness**: repeated application of these three transformations to "///*") generate all unique locators for each web page element e. XPaths that do not include e in their result set are discarded.
XPath generation graph

DOM D of the Web Page:
```html
<html>
  <p class='a'>X</p>
  <p class='a'>Y</p>
  <div class='a'>X</div>
</html>
```

Full Absolute XPath locator for the element e:
```
/html[1]/p[text()='X' and @class='a'][1]
```

Info for $G_e$:
- 80 N of Vertices (XPaths) in $G_e$
- 1 N of Vertices with InDegree = 0
- 16 N of Vertices with OutDegree = 0
- 5 N of Boundary Locators
- 49 N of Locators

64 of length 2 + 16 of length 1

i.e., only `/**` boxed

**bold – underlined in green**

underlined in green

Unique locators of e are underlined in green.
### XPath fragility

<table>
<thead>
<tr>
<th>Transformation</th>
<th>XPath Pattern</th>
<th>Fragility Count (FC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>transfAddName</td>
<td><code>//*</code>/td -&gt; <code>//tr/td</code></td>
<td>FC += W_tag</td>
</tr>
<tr>
<td>transfAddPredicate</td>
<td><code>//tr/td</code> -&gt; <code>//tr[@name='data']/td</code></td>
<td>FC += W(@name)</td>
</tr>
<tr>
<td></td>
<td><code>//tr/td</code> -&gt; <code>//tr[2]/td</code></td>
<td>FC += W_pos</td>
</tr>
<tr>
<td>transfAddLevel</td>
<td><code>//tr/td</code> -&gt; <code>//*;/tr/td</code></td>
<td>FC += W_lev</td>
</tr>
</tbody>
</table>

**Fragility Count (FC)** is zero for ``//*`; it is incremented whenever edges are added to the XPath generation graph.
XPath fragility

DOM D of the Web Page:
```html
<html>
<p class='a'>X</p>
<p class='a'>Y</p>
<div class='a'>X</div>
</html>
```

Graph $G_e$

Full Absolute XPath locator for the element $e$:
```
/html[1]/p[text()='X' and @class='a'][1]
```

Info for $G_e$:
- 80 N of Vertices (XPaths) in $G_e$
- 1 N of Vertices with InDegree = 0
- 16 N of Vertices with OutDegree = 0
- 5 N of Boundary Locators
- 49 N of Locators

Step:
1. //html[1]/p[text()='X' and @class='a'][1]
2. //p[1]
3. //*[text()='X'][1]
4. //*[@class='a'][1]
5. //html/p
6. //*[1]/p
7. //*[text()='X' and @class='a'][1]
8. //*[1]/*[text()='X' and @class='a'][1]
9. //*[1]/*[1]/*[text()='X' and @class='a'][1]
10. //*[1]/*[1]/*[1]/*[text()='X' and @class='a'][1]
11. //*[1]/*[1]/*[1]/*[1]/*[text()='X' and @class='a'][1]
12. //*[1]/*[1]/*[1]/*[1]/*[1]/*[text()='X' and @class='a'][1]
13. //*[1]/*[1]/*[1]/*[1]/*[1]/*[1]/*[text()='X' and @class='a'][1]

Minimum FC locators are at the boundary between unique and non unique locators.
Greedy algorithm

```
<html>
  <p class='a'>X</p>
  <p class='a'>Y</p>
  <div class='a'>X</div>
</html>
```

Globally optimal locator: `//p[text='X']`
Greedy algorithm

**Termination:** the algorithm is ensured to terminate, since in the worst case it returns the absolute XPath.

**Correctness:** the algorithm returns the global optimum because FC is monotonically increasing for successively explored locators.

**Complexity:** the algorithm is exponential in the number of predicates and levels:

\[
|V| = \sum_{i=1}^{h} 2^{(\sum_{k=1}^{i} |P_k|) + i} = |X_e|
\]
Genetic algorithm

\[ fit(x) = \begin{cases} 
|\text{query}(x, D)| & \text{query}(x, D) \neq \{e\} \\
\text{fc}(x) & \text{query}(x, D) = \{e\} 
\end{cases} \]
Conclusions

<table>
<thead>
<tr>
<th>Tool</th>
<th>XPath Locators Robustness</th>
<th>Robust</th>
<th>Broken</th>
</tr>
</thead>
<tbody>
<tr>
<td>FirePath</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FirePath</td>
<td>/[id='userInfo'][3-4]/td[2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chrome</td>
<td>/[id='userInfo'][3-4]/td[2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XPath Helper</td>
<td>/html/body/table[@id='userInfo'][3-4]/td[@title='mobile']</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XPath Checker</td>
<td>id('userInfo')[3-4]/td[2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROBULA</td>
<td>/td[@title='mobile']</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Questions?**

**Locator generation**

**Greedy optimal**

**GA sub-optimal**