FONDAZIONE BRUNO KESSLER

SE SOFTWARE ENGINEERING

# Meta-Heuristic Generation of Robust XPath Locators for Web Testing


Maurizio Leotta

dibris

Andrea Stocco

dibris

Filippo Ricca

dibris

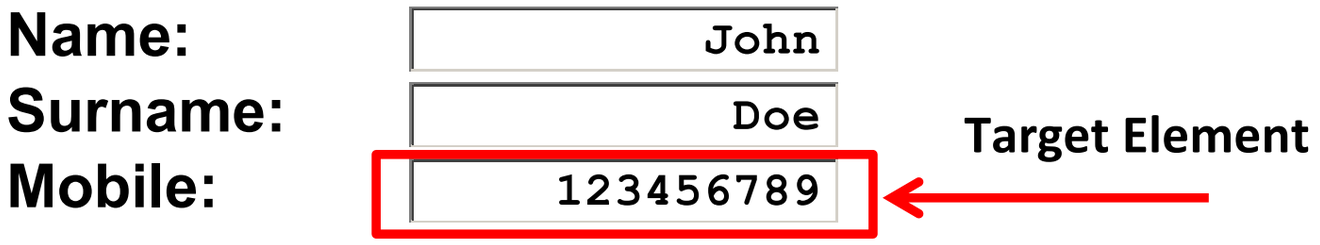
Paolo Tonella

FONDAZIONE BRUNO KESSLER

# Outline

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

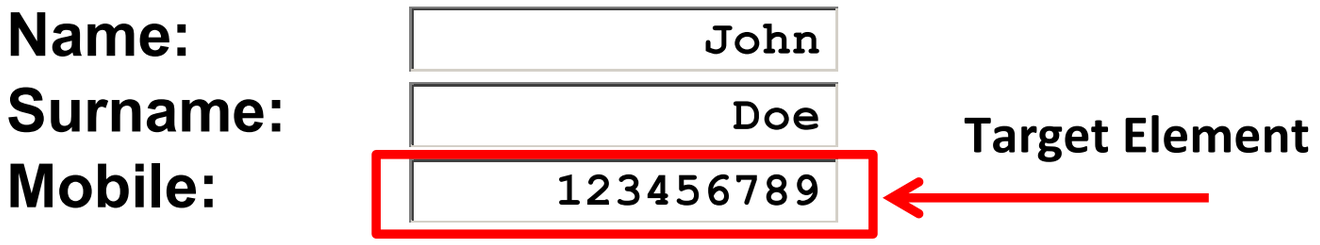
# Web testing




```
<html>
 <body>
  <table id="userInfo">
   <tr><td>Name:    </td><td title ="name">          John</td></tr>
   <tr><td>Surname:</td><td title ="surnanme">     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");
}
```

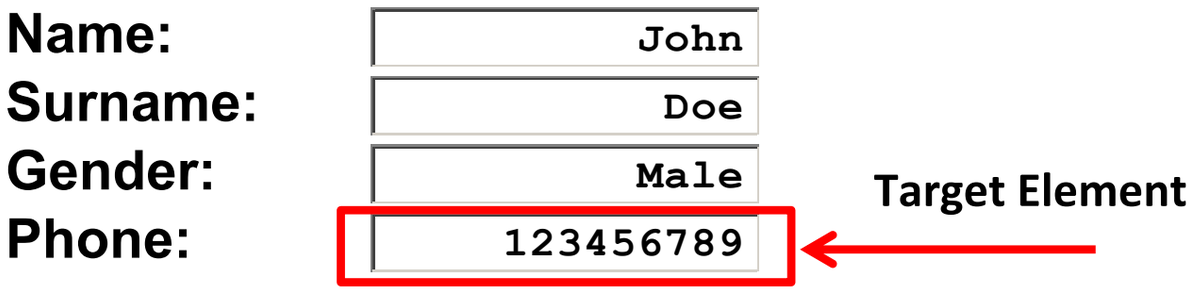
# Web element locators




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

| Tool | Kind | Generated XPath Locators for the Target Element |
|---|---|---|
| FirePath | abs | /html/body/table/tr[3]/td[2] |
| FirePath | rel | //*[@id="userInfo"]/tr[3]/td[2] |
| Chrome | rel | //*[@id="userInfo"]/tr[3]/td[2] |
| XPath Helper | abs | /html/body/table[@id="userInfo"]/tr[3]/td[@title="mobile"] |
| XPath Checker | rel | id('userInfo')/tr[3]/td[2] |
| ROBULA | rel | //td[@title="mobile"] |

# Robust locators




```
<html>
 <body>
  <table id="userInfo">
   <tr><td>Name:    </td><td title ="name">         John</td></tr>
   <tr><td>Surname:</td><td title ="surnanme">      Doe</td></tr>
   <tr><td>Gender: </td><td title ="gender">        Male</td></tr>
   <tr><td>Phone:   </td><td title ="mobile"> 123456789</td></tr>
  </table>
 </body>
</html>
```

| Tool | | XPath Locators Robustness (✓ robust ✗ broken) |
|---|---|---|
| FirePath | ✗ | /html/body/table/tr[~~3~~→4]/td[2] |
| FirePath | ✗ | //*[@id="userInfo"]/tr[~~3~~→4]/td[2] |
| Chrome | ✗ | //*[@id="userInfo"]/tr[~~3~~→4]/td[2] |
| XPath Helper | ✗ | /html/body/table[@id="userInfo"]/tr[~~3~~→4]/td[@title="mobile"] |
| XPath Checker | ✗ | id('userInfo')/tr[~~3~~→4]/td[2] |
| ROBULA | ✓ | //td[@title="mobile"] |

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

**e**: Target Web Element

**Full Absolute XPath locator** for the element **e:**

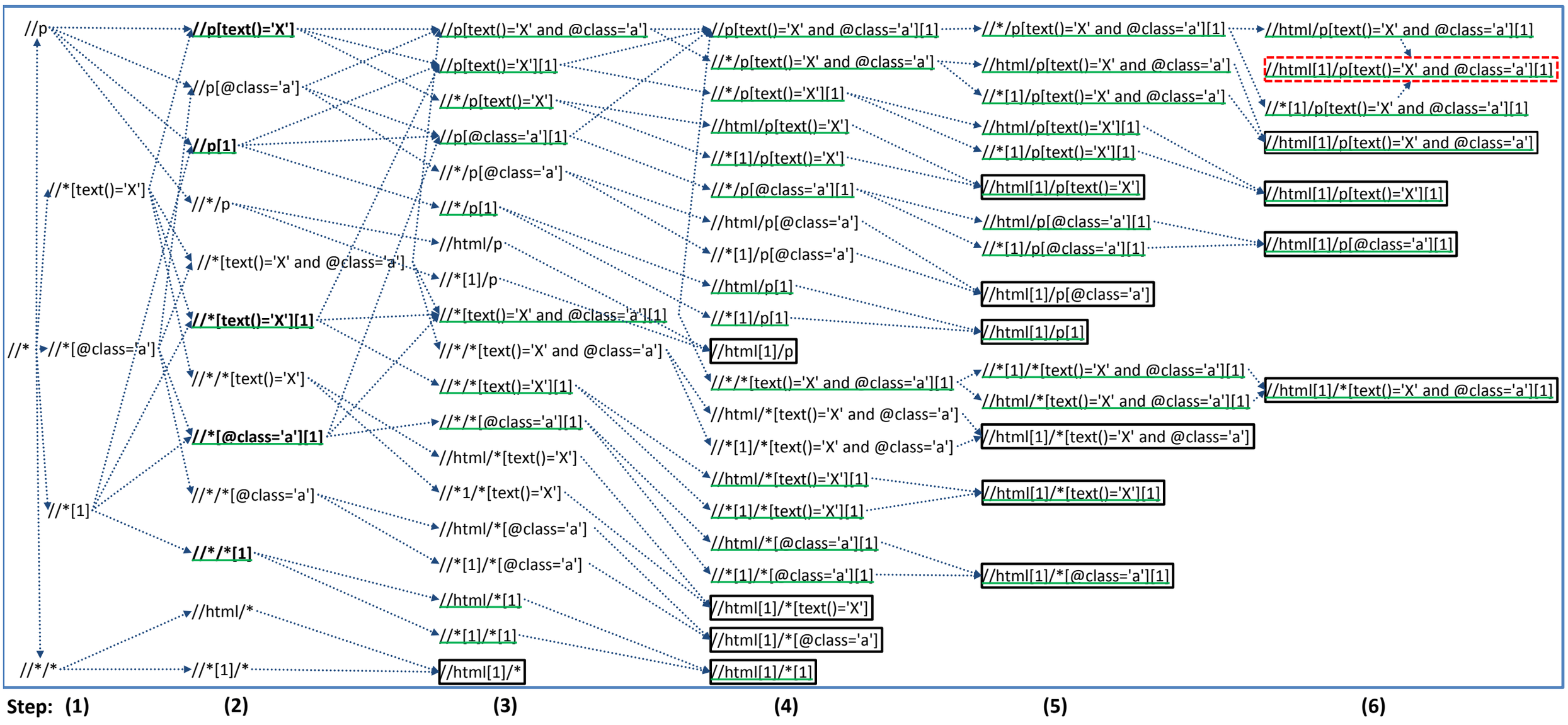
//html[1]/p[text()='X' and @class='a'][1]

see step (7)

**Info for $G_e$:**

| | | |
|---|---|---|
| 80 | N of Vertices (XPaths) in $G_e$ | 64 of length 2 + 16 of length 1 |
| 1 | N of Vertices with InDegree = 0 | i.e., only //* |
| 16 | N of Vertices with OutDegree = 0 | boxed |
| 5 | N of Boundary Locators | **bold – underlined in green** |
| 49 | N of Locators | underlined in green |

**Graph $G_e$**




***Unique locators*** of *e* are underlined in green.

# XPath fragility

```
transfAddName
   //*/td -> //tr/td                                   FC += W_tag

transfAddPredicate
   //tr/td -> //tr[@name='data']/td                FC += W(@name)
   //tr/td -> //tr[2]/td                              FC += W_pos

transfAddLevel
   //tr/td -> //*/tr/td                               FC += W_lev
```

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

**e**: Target Web Element

**Full Absolute XPath locator** for the element **e:**

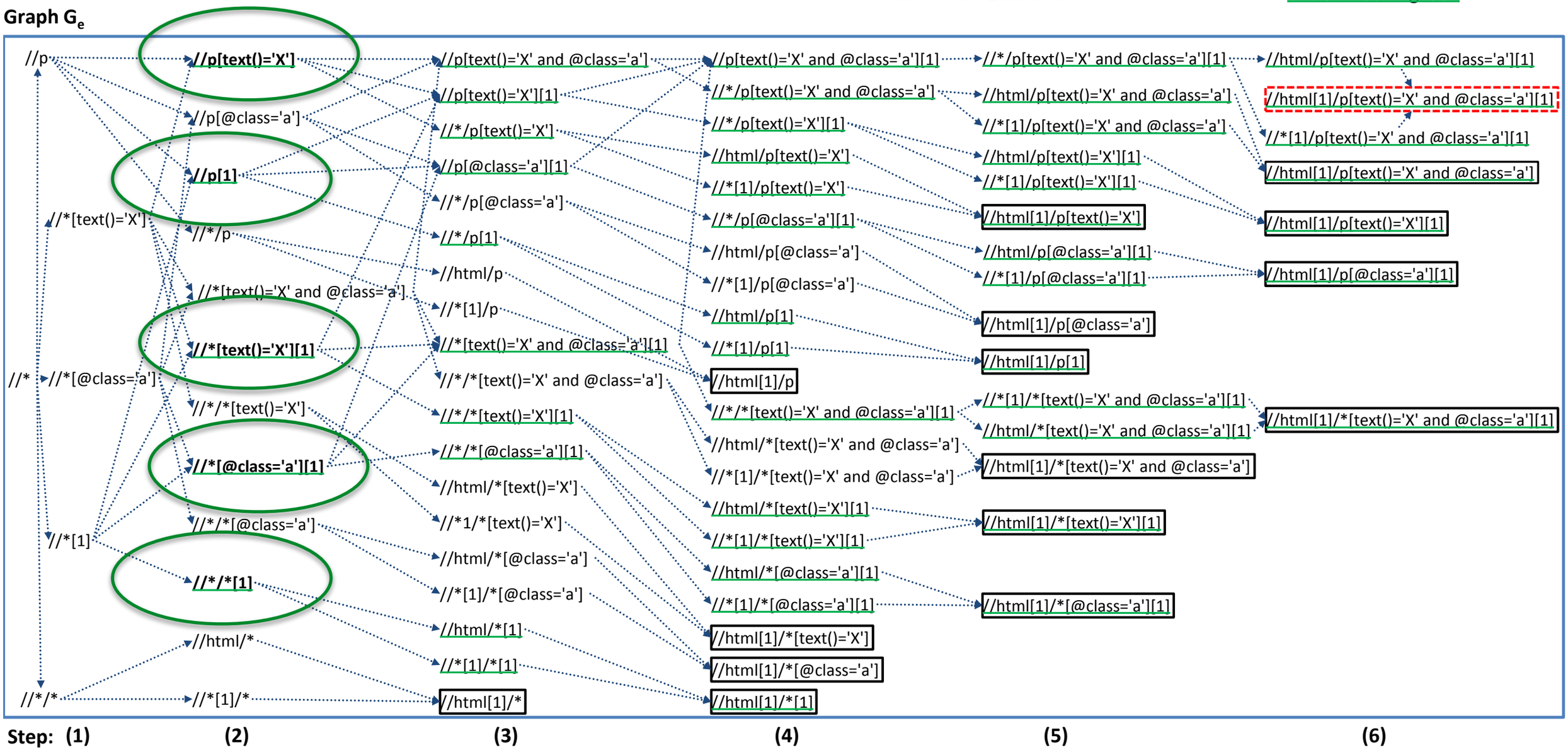
//html[1]/p[text()='X' and @class='a'][1]

see step (7)

**Info for $G_e$:**

| | | |
|---|---|---|
| 80 | N of Vertices (XPaths) in $G_e$ | 64 of length 2 + 16 of length 1 |
| 1 | N of Vertices with InDegree = 0 | i.e., only //* |
| 16 | N of Vertices with OutDegree = 0 | boxed |
| 5 | N of Boundary Locators | **bold – underlined in green** |
| 49 | N of Locators | underlined in green |

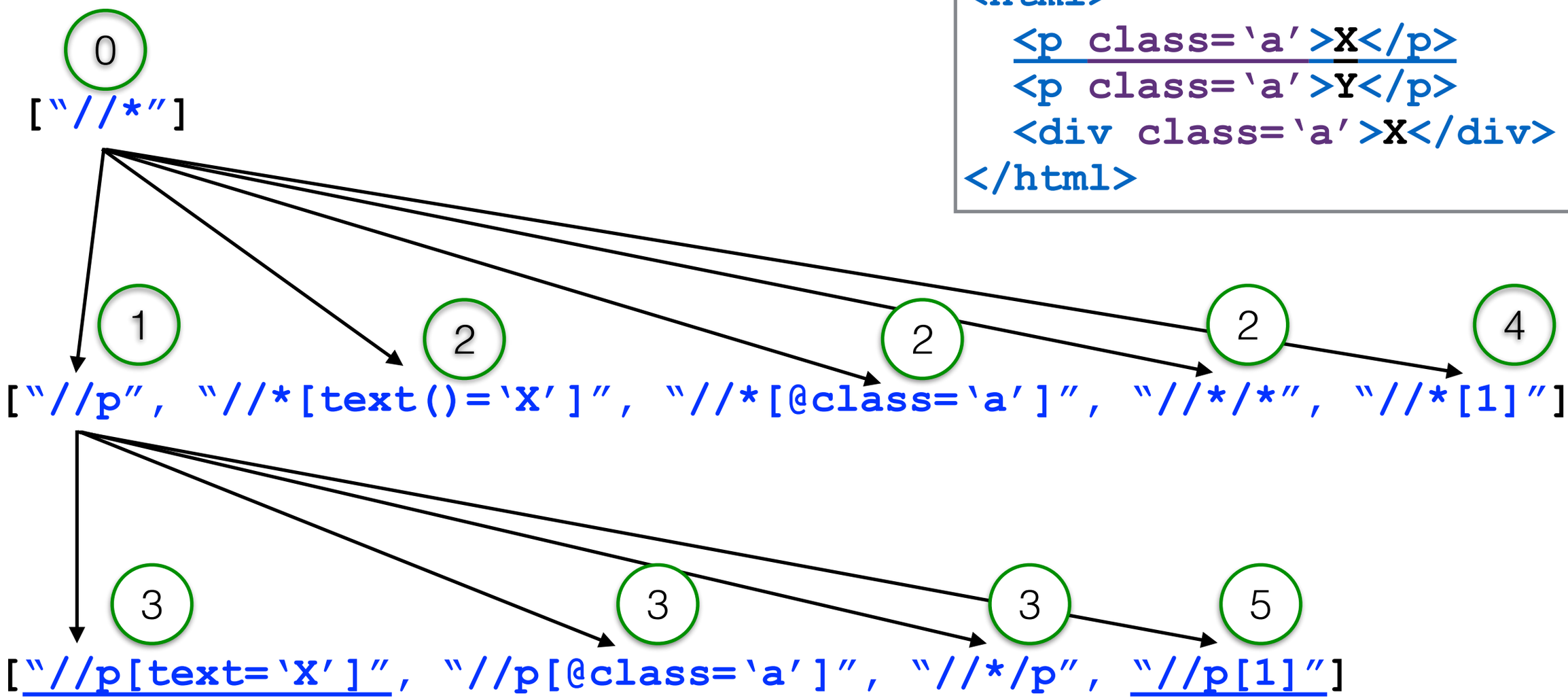
**Graph $G_e$**




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

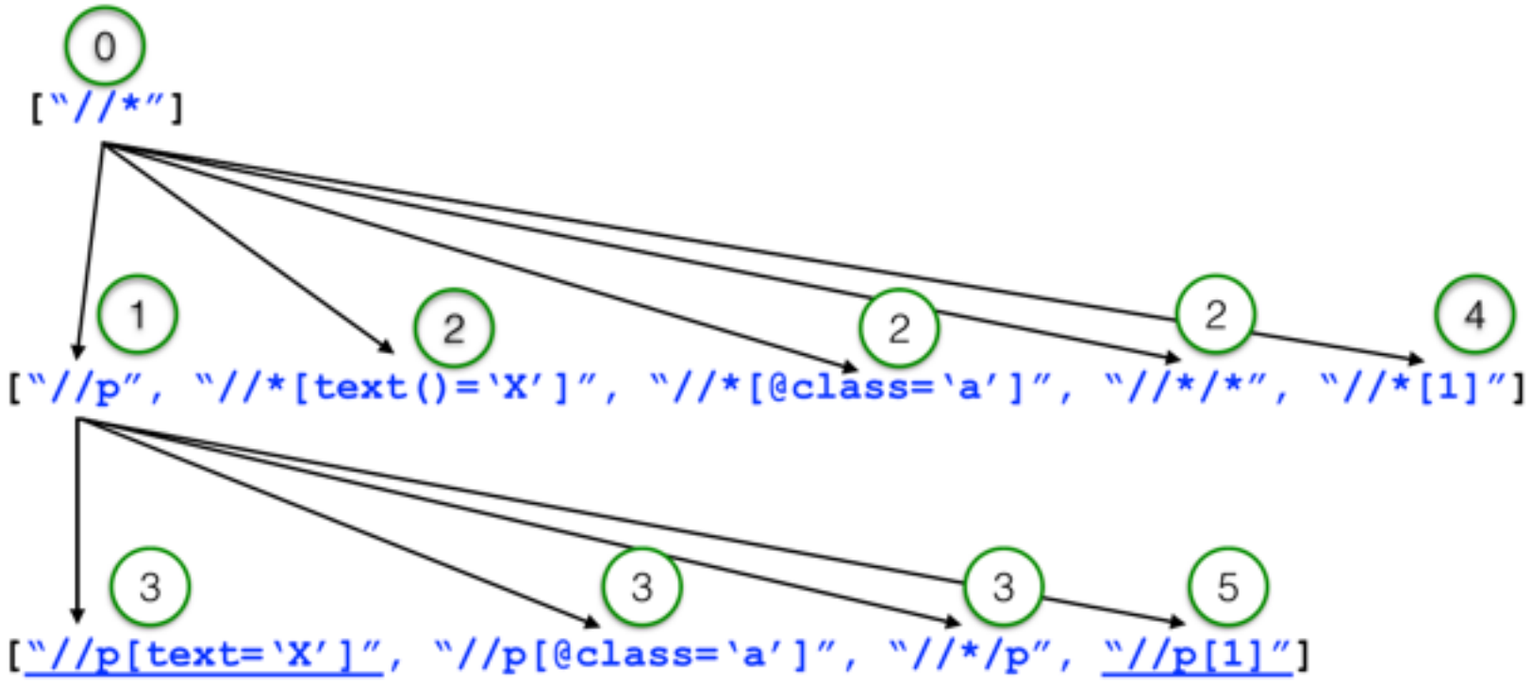
# Greedy algorithm

```
<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

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

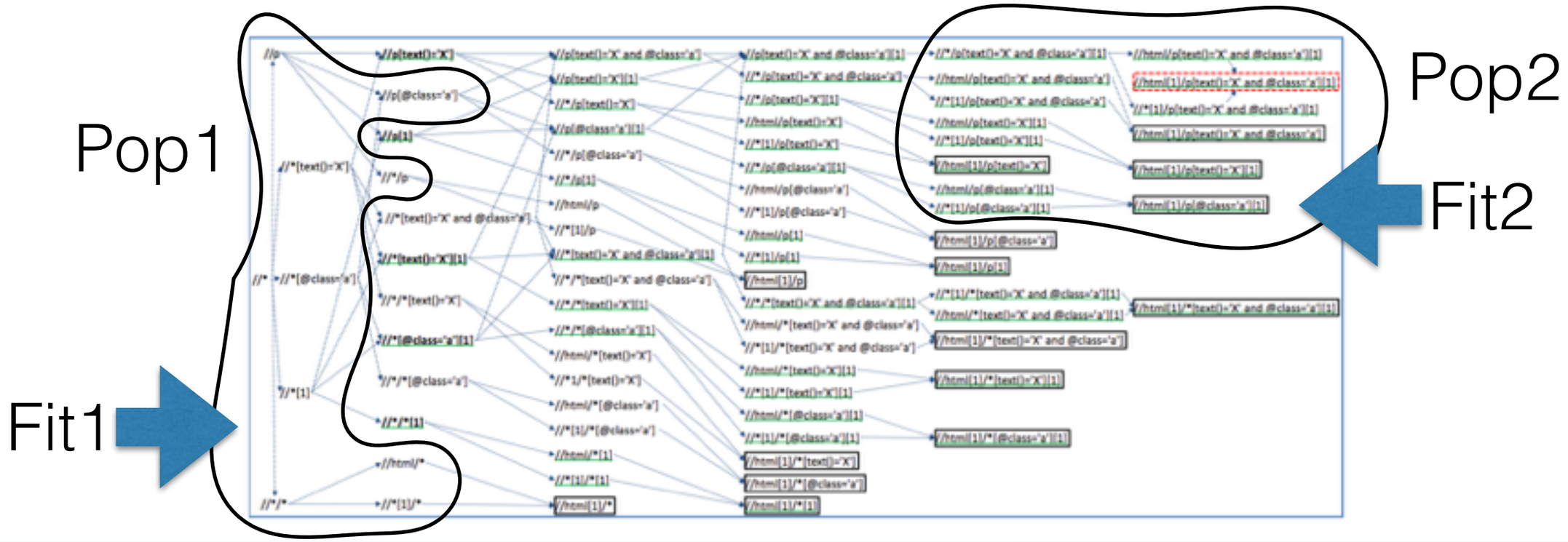
$$fit(x) = \begin{cases} |\text{query}(x, D)| & \text{query}(x, D) \neq \{e\} \\ f_c(x) & \text{query}(x, D) = \{e\} \end{cases}$$

["//html/p", "//html[1]/p[text()='X'][1]", "//*/*]

mutation

crossover

["//html[1]/p", "//*/p", "//*/p[text()='X'][1]", "//html[1]/*]



# Conclusions

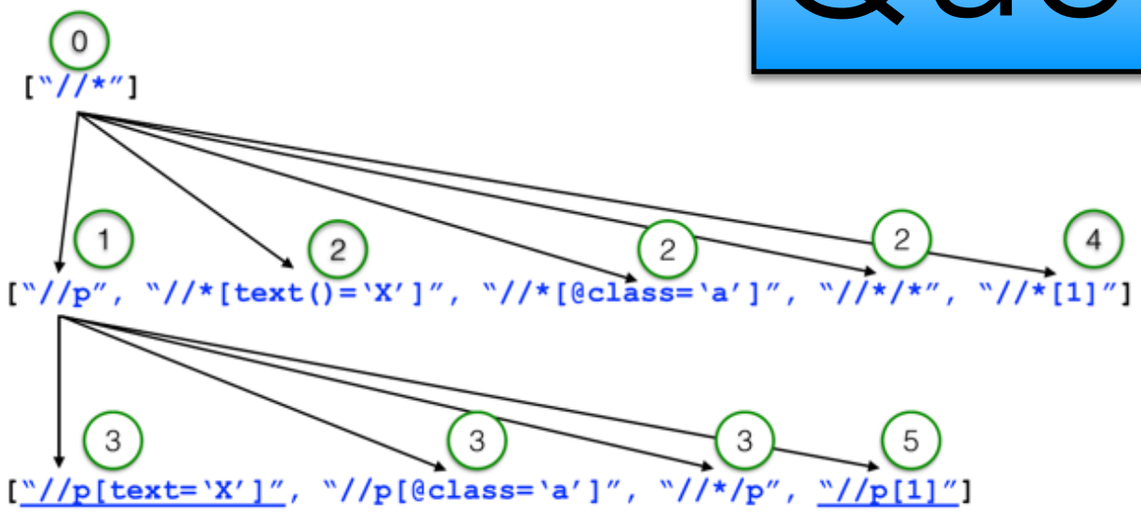
**Name:** John
**Surname:** Doe
**Gender:** Male
**Phone:** 123456789

Target Element

```
<html>
 <body>
  <table id="userInfo">
   <tr><td>Name:    </td><td title ="name">         John</td></tr>
   <tr><td>Surname:</td><td title ="surnanme">      Doe</td></tr>
   <tr><td>Gender: </td><td title ="gender">       Male</td></tr>
   <tr><td>Phone:  </td><td title ="mobile"> 123456789</td></tr>
  </table>
 </body>
</html>
```

| Tool | | XPath Locators Robustness ✓ robust × broken |
|---|---|---|
| FirePath | × | /html/body/table/tr[~~3~~→4]/td[2] |
| FirePath | × | //*[@id="userInfo"]/tr[~~3~~→4]/td[2] |
| Chrome | × | //*[@id="userInfo"]/tr[~~3~~→4]/td[2] |
| XPath Helper | × | /html/body/table[@id="userInfo"]/tr[~~3~~→4]/td[@title="mobile"] |
| XPath Checker | × | id('userInfo')/tr[~~3~~→4]/td[2] |
| ROBULA | ✓ | //td[@title="mobile"] |

## Robust locators




## Greedy optimal

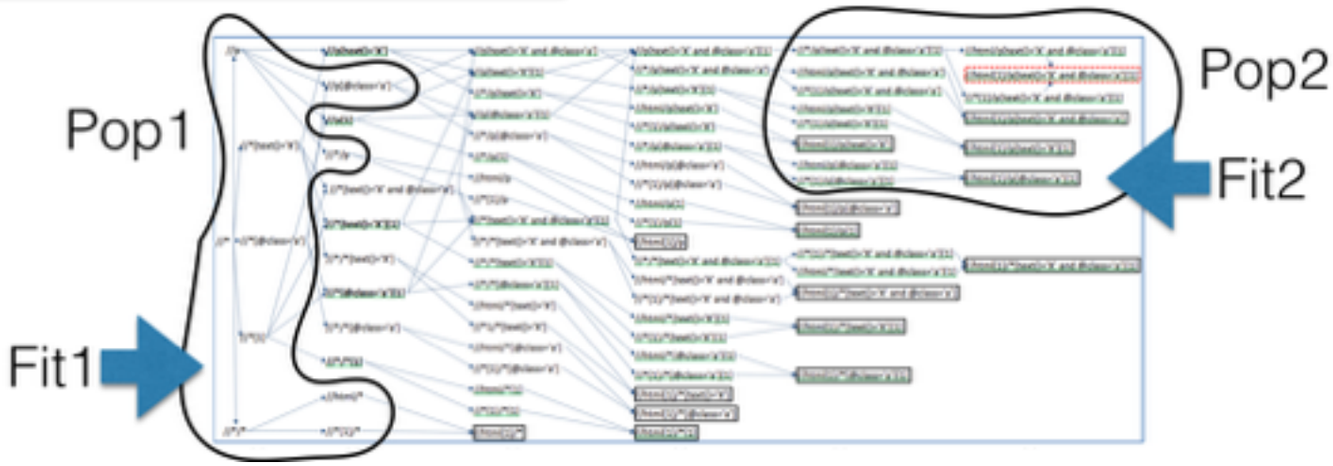
```
transfAddName
    //*/td -> //tr/td

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

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

## Locator generation

# Questions?




## GA sub-optimal