Metric-based Evaluation of Implemented Software Architectures

Eric Bouwers
Website

Accounts
Savings
Loans

Batch processing

Storage

External systems
ATAM, ARID, ALMA, ALPSM, SAAM, SAAMER, SAAMCS, SBAR, ESAAMI, SAEM, QASAR, .....
A first step...

- Abstraction
- Layering
- Logic in Database
- Module Inconsistency
- Module Size
- Source Grouping
- Technology Combination
- Textual Duplication

- Functional Duplication
- Libraries / Frameworks Usage
- Module Dependencies
- Module Functionality
- Relation Documentation and Implementation
- Technology Age
- Technology Usage

‘Criteria for the Evaluation of Implemented Architectures’ by E. Bouwers, J. Visser and A. van Deursen, ICSM 2009
... towards a structure, repeatable check.


28 questions
28 actions
5 topics
'A Cognitive Model for Software Architecture Complexity',
By E.Bouwers, C. Lilienthal, J. Visser, and A. van Deursen, ICPC 2010
Continuous Evaluation
Check >> Evaluation
Component Balance

Quantifying the analyzability of software architectures,
By E. Bouwers, J. Correia, A. van Deursen and J. Visser, WICS 2011
"Component co-evolution and component dependency: speculations and verifications."
By L. Yu, A. Mishra, and S. Ramaswamy, IET Software 2010
Correlating static and dynamic metrics

The graph illustrates the correlation between snapshot-based metrics and change-sets. Each change-set (cs0 to cs7) is marked with a specific metric value on the y-axis, ranging from 1 to 4. The change-sets are sequenced along the x-axis.
Step 1:
Determine the time-periods for which the snapshot-based metric is representative.

Step 2:
Calculate the evolutionary metric for these time-periods.

Step 3:
Calculate the correlation between the snapshot-based metric value and the evolutionary metric.
<table>
<thead>
<tr>
<th>Metric</th>
<th>Correlation</th>
<th>Corrected p-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCI</td>
<td>0.16</td>
<td>11.3</td>
<td>0.94</td>
</tr>
<tr>
<td>CCD</td>
<td>-0.27</td>
<td>0.13</td>
<td>0.01</td>
</tr>
<tr>
<td>ACD</td>
<td>-0.26</td>
<td>0.04</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>NCD</td>
<td>-0.19</td>
<td>0.59</td>
<td>0.05</td>
</tr>
<tr>
<td>CDI</td>
<td>0.32</td>
<td>11.94</td>
<td>1.00</td>
</tr>
<tr>
<td>PI</td>
<td>-0.30</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>RI</td>
<td>-0.31</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>IC</td>
<td>0.47</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>NBD</td>
<td>-0.22</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>CB</td>
<td>0.29</td>
<td>0.05</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>MSUI</td>
<td>-0.08</td>
<td>2.42</td>
<td>0.20</td>
</tr>
<tr>
<td>NC</td>
<td>-0.26</td>
<td>0.27</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Validity versus Usefulness

Does it measure what we want to measure?

What can we do with the metric?
A four step evaluation process

Embed

Data Gathering
- Observe
- Interview

Analyze
Measuring ISO 25010 maintainability

<table>
<thead>
<tr>
<th></th>
<th>Volume</th>
<th>Duplication</th>
<th>Unit Size</th>
<th>Unit Complexity</th>
<th>Unit Interfacing</th>
<th>Module Coupling</th>
<th>Component Independence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysability</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modifiability</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testability</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modularity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Reusability</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
Data gathering

*Memos and interviews*

49 memo’s
17 different consultants
11 different customers SUPpliers

30 minutes interviews with 11 consultants

Open discussion: ‘How do you use ....’

Closed questions (1-5 scale):
• How *useful* do you find the metric?
• Does it make your job *easier*?
Not easy to use  But useful
So are these metrics useful?

In the context of evaluating implemented architectures? YES!
Improve existing metrics

Develop more metrics

Improve the way metrics are used