Using ontologies for verification and validation of workflow-based experiments

Tomasz Miksa
Andreas Rauber

TU Wien & SBA Research, Austria
miksa@ifs.tuwien.ac.at
Replicability

- Can I reuse her workflow?
- Do the results match?
- Have I done it correctly?
Replicability

- Current studies show very low reproducibility in
  - medicine
  - economy
  - computer science

- Reproducibility requires
  - well documented research workflows
  - precise information on the experiment's environment

http://dx.doi.org/10.1016/j.jbi.2016.10.011
Context Model

- OWL ontology
- Modular architecture
  - Domain Independent Ontology
  - Domain Specific Ontologies
- Process preservation
Context Model

Validation metrics

Workflow definition

File formats

Software dependencies

Provenance
VFramework - Run static analysis

- Transformation requires
  - mapping between concepts
  - development of conversion tools

Script

```
#!/bin/bash
# fetch data
java -jar gestBarragensWSClientIQData.jar
unzip -o IQData.zip
# fix encoding
iconv -f LATIN1 -t UTF-8 iq.r > iq_utf8.r
# generate references
R --vanilla < iq_utf8.r > IQout.txt
# create pdf
pdflatex iq.tex
pdflatex iq.tex
```

Taverna Workflow

Archimate model

Context Model (OWL ontology)
VFramework - Run dynamic analysis
Local dependencies

- Process Migration Framework (PMF) [1]
  - designed for automatic redeployments into virtual machines
  - uses `strace` to monitor system calls
  - complete log of all accessed resources
  - creates context model

Validation report for the WeatherExample

Evaluation result: There are 2 not fulfilled metrics. Please see tables below for details.
Comparison performed using following workflow execution traces

Original Workflow
ID: 37b4d2fb-e71c-4b67-b7b3-1788ee82977
Timestamp: 2015-10-14 18:58:06.475

Compared Workflow
ID: ede04b87-5f68-4a89-b0c3-e179957cbad0
Timestamp: 2015-11-13 13:59:56.443

Table 1: Overview of requirements

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Description</th>
<th>Is Fulfilled</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>The inputs to the workflow are the same</td>
<td>true</td>
</tr>
<tr>
<td>R2</td>
<td>The outputs of the workflow are the same</td>
<td>false</td>
</tr>
<tr>
<td>R3</td>
<td>The workflow step ExtractTemperature must have identical outputs</td>
<td>true</td>
</tr>
<tr>
<td>R4</td>
<td>The workflow step GetWeatherData must have identical outputs</td>
<td>true</td>
</tr>
<tr>
<td>R5</td>
<td>The workflow step MakeDecision must have identical outputs</td>
<td>true</td>
</tr>
<tr>
<td>R6</td>
<td>The workflow step ExtractWeatherType must have identical outputs</td>
<td>true</td>
</tr>
<tr>
<td>R7</td>
<td>The workflow step VisualiseTemperature must have identical outputs</td>
<td>false</td>
</tr>
<tr>
<td>R8</td>
<td>Execution duration of each of the workflow steps shall be similar</td>
<td>true</td>
</tr>
</tbody>
</table>

Table 2: List of requirements and metrics that failed.

<table>
<thead>
<tr>
<th>Req</th>
<th>Sub-req</th>
<th>Sub-requirement description</th>
<th>Measurement point</th>
<th>Metric</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td>R7.1</td>
<td>The output plot of workflow step VisualiseTemperature must be identical</td>
<td>plot</td>
<td>ImageFingerprintEquality</td>
<td>false</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ImageResolutionEquality</td>
<td>true</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AbsoluteErrorCount</td>
<td>false</td>
</tr>
<tr>
<td>R7</td>
<td>R7.1</td>
<td>The output plot of workflow step VisualiseTemperature must be identical</td>
<td>plot</td>
<td>ImageFingerprintEquality</td>
<td>false</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ImageResolutionEquality</td>
<td>true</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AbsoluteErrorCount</td>
<td>false</td>
</tr>
</tbody>
</table>
Temperature in 322690

Temperature

Location

Temperature in 322690

Temperature

Location
ADDED

NOT USED
Taverna workflow analysis
- Limited validation for external services (Rserve scripts)
  - global variables
  - only final result of workflow computation is validated
    • black box testing
- Data created through shell calls
  - such files are not a part of provenance traces in Taverna
- ‘real’ workflow outputs
  - provenance traces can contain paths only, but not the content
  - workflows can create files not linked to any output
Results – validation metrics

- metadata included in data
  - generation timestamp
- file format comparison improved the results
  - ZIP archives

---

% latex table generated in R 3.0.2 by xtable 1.7-1 package

% Tue Apr 21 13:37:05 2015
\begin{table}
\begin{tabular}{rrrrrrr}
\hline
& Min & 1Q & Mediana & Media & 3Q & Max \\
\hline Residuos (mm) & -2.42 & -1.23 & -0.52 & -0.00 & 0.45 & 5.52 \\
\hline
\end{tabular}
\caption{Residuos do modelo}
\end{table}
Recommendations

- Analyse dependencies and evade shell calls
  - e.g. use scripting mechanisms provided by the workflow engine

- Write code that runs on all platforms
  - e.g. do not encode specific paths

- Publish experiment setup and context
  - e.g. exact versions of tools used

- Publish validation data
  - e.g. provenance but also other files created during execution

- Test the replicability on your own
  - e.g. try rerunning your experiment in a clean virtual machine
Conclusions

I have repeated her experiment in the same way!
I got the same results!
I can reuse any part of it!

VFramework + Context Model

Original experiment

Re-executed experiment
Publications
