OUQUARE
A SQUaRE-based Quality Evaluation Framework for Ontologies

Astrid Duque-Ramos
Jesualdo Tomás Fernández-Breis
Robert Stevens
Nathalie Aussenac-Gilles

Ontology Summit 2013: Ontology Evaluation Across the Ontology Lifecycle

January, 31th, 2013
SQuaRE: Standard for software product Quality Requirements and Evaluation (ISO 25000)

- Comprehensive specification and evaluation model
- Common language for specifying user requirements
- It is based on observation
- It makes quality evaluation reproducible
<table>
<thead>
<tr>
<th>ISO/IEC 25003n Quality Requirements</th>
<th>ISO/IEC 25001n Quality Model</th>
<th>ISO/IEC 25004n Quality Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO/IEC 25000n Product Quality General division</td>
<td>ISO/IEC 25002n Quality Metrics</td>
<td></td>
</tr>
</tbody>
</table>
OUR GOAL

• Adapting SQuaRE to ontology evaluation
  – Identifying strengths and flaws of the ontologies
  – Supporting users and developers in making informed decisions
Outline of Quality Model and Quality Metrics

Quality Model

Characteristic 1

- Subcharacteristic 1
- Subcharacteristic P

Characteristic K

- Subcharacteristic J
- Subcharacteristic M

Quality Metrics

- Metric 1
- Metric 2
- Metric 3
- Metric 4
- Metric 5
- Metric N-1
- Metric N
Detailed information: [http://miuras.inf.um.es/evaluation/oquare](http://miuras.inf.um.es/evaluation/oquare)
Origin of the quality metrics

• Adaptation from software metrics
  – Coupling between objects
  – Weighted method count

• Reuse of ontology metrics
  – Cohesion (Yao, Orme, and Etzkorn (2005))
  – Metrics from Tartir and Arpinar (2007)
• Notation

• C1; C2; ...; Cn: Classes in the ontology.
• R_{C1}; R_{C2}; ...; R_{Ck}: Relationships of each class Ci.
• P_{C1}; P_{C2}; ...; P_{Cz}: Properties of each class Ci.
• I_{C1}; I_{C2}; ...; I_{Cm}: Individuals of each class Ci.
• Sup_{C1}; Sup_{C2}; ...; Sup_{Cm}: Direct superclasses of a given class C.
• Thing: Root class of the ontology.
Some adapted software metrics

Coupling Between Objects: \[ CBO_{Onto} = \frac{\sum | Sup_{Ci} |}{\sum (| C_i | - | R_{Thing} |)} \]

Depth of Inheritance Tree: \[ DIT_{Onto} = \text{Max} \sum | D_{Ci} | \]

Weighted Method Count: \[ WMCO_{Onto} = \left( \frac{\sum | R_{Ci} | + \sum | R_{Ci} |}{\sum | C_i |} \right) \]

Number of Ancestor Classes: \[ NAC_{Onto} = \frac{\sum | Sup_{leaf_i} |}{\sum | C(leaf)_{i} |} \]
Some reused ontology metrics

Attributes Richness: \( AROnto = \sum |Att_{Ci}|/\sum |C_i| \)

Relationships per class: \( INROnto = \sum |R_{Ci}|/\sum |C_i| \)

Number of properties: \( NOMOnto = (\sum |P_{Ci}|)/\sum |C_i| \)

Annotations Richness: \( AnOnto = (\sum |An_{Ci}|)/\sum |C_i| \)
Association between Quality Model and Quality Metrics

**Quality Model**

- Structural
  - Formal relations support
  - Redundancy

- Functional adequacy
  - Schema and value reconciliation
  - Knowledge acquisition-representation

- Maintainability
  - Modularity

- Reliability
  - Recoverability

**Quality Metrics**

- RROnto
- ANOnto
- AROnto
- Formal degree
- Consistency degree
- NOMOnto
- WMCONt0
- CBOnto
- DITOnto
- LCOMOnto
Getting scores from the value of the metrics

• SQuaRE scores are in the range \([1,5]\)
  – 1: not acceptable
  – 3: minimally acceptable
  – 5: exceeds the requirements

• Values of the metrics are mapped onto \([1,5]\)

• Higher values do not always mean higher quality
Definition of the value mappings following best practices

<table>
<thead>
<tr>
<th>Scale / Metric</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCOMOnto</td>
<td>&gt; 8</td>
<td>(6-8]</td>
<td>(4,6]</td>
<td>(2, 4]</td>
<td>&lt;=2</td>
</tr>
<tr>
<td>DITOnto</td>
<td>&gt; 8</td>
<td>(6-8]</td>
<td>(4,6]</td>
<td>(2, 4]</td>
<td>[1,2]</td>
</tr>
<tr>
<td>NACOnto</td>
<td>&gt; 12</td>
<td>(8-12]</td>
<td>(6,8]</td>
<td>(3,6]</td>
<td>[1,3]</td>
</tr>
<tr>
<td>RROnto</td>
<td>[0,20]%</td>
<td>(20-40]%</td>
<td>(40-60]%</td>
<td>(60-80]%</td>
<td>&gt; 80%</td>
</tr>
<tr>
<td>AROnto</td>
<td>[0,20]%</td>
<td>(20-40]%</td>
<td>(40-60]%</td>
<td>(60-80]%</td>
<td>&gt; 80%</td>
</tr>
<tr>
<td>INROnto</td>
<td>[0,20]%</td>
<td>(20-40]%</td>
<td>(40-60]%</td>
<td>(60-80]%</td>
<td>&gt; 80%</td>
</tr>
</tbody>
</table>
• Cell type ontology: Two versions of CTO
• Units of Measurements: Ten Unit of Measurements Ontologies

Main results of each case study available at http://miuras.inf.um.es/evaluation/oquare
CASE STUDIES
Manual evaluation of Unit Measurement

Ontologies of unit of measurements manual Comparison

- Structural
- Maintainability
- Functional Adequacy
- Operability
- Reliability
- Average
CASE STUDIES
Automatic evaluation of Unit Measurement

Ontologies of unit of measurements comparison

- Structural
- Maintainability
- Functional Adequacy
- Operability
- Reliability
- Transferability
- Compatibility
- Average
• Similar scores in both evaluations

• Findings about the ontologies
  – Highest score has been obtained for the structural and functional adequacy characteristics
  – Lowest score for reliability and operability.
  – Ontologies require more effort to be used, understood and learnt
• Preliminary assessment on OQuaRE:
  – Positive and negative aspects of the current version of OQuaRE
  – Completeness and usefulness of the quality metrics
  – Independent experts on biomedical ontologies: Stefan Schulz, Michel Dumontier, Mikel Egaña
EVALUATION OF THE FRAMEWORK
The Process

• Step 1) Manual evaluation
  • Difficulty in understanding some subcharacteristics
  • Need for knowing the intended context of use

• Step 2) Manual evaluation with the support of OQuaRE metrics
  • Difficulty in understanding some metrics because of their definition in an OWL-independent way.
  • Metrics provided additional information to the experts contributing to a more precise understanding of the subcharacteristics.
EVALUATION OF THE FRAMEWORK
Findings about the quality model

Subcharacteristic Vs Appropriateness-Difficulty

- High Aprop and Low Diff.: 35%
- High Aprop and Media Diff.: 17%
- High Aprop and High Diff.: 13%
- Media Aprop and Low Diff.: 11%
- Media Aprop and Media Diff.: 8%
- Media Aprop and High Diff.: 4%
- Low Aprop and Low Diff.: 8%
- Low Aprop and Media Diff.: 4%
- Low Aprop and High Diff.: 0%
• To define new metrics or new associations between metrics and subcharacteristics

• To have a limited number of metrics per subcharacteristic.

• To distinguish between context dependent and independent metrics

• To divide the structural accuracy into subcharacteristics.
COMMUNITY ACTIONS

• Agreement on subcharacteristics and metrics

• Agreement on interpretation of values of metrics

• Need for metrics for some quality subcharacteristics

• Contributions are welcome
  • http://miuras.inf.um.es/oquarewiki/
  • https://docs.google.com/spreadsheet/viewform?fromEmail=true&formkey=dFlsZGpkbjlHaEk1d0l0RU1oYzVwN0E6MQ
• Adjusting OQuaRE by increasing interaction with the ontology engineering community

• Definition of the quality requirements module for determining potential contexts of use

• Finishing the development of an online tool for ontology evaluation based on OQuaRE
OQUARE
A SQUaRE-based Quality Evaluation Framework for Ontologies

Astrid Duque-Ramos astrid.duque@um.es
Jesualdo Tomás Fernández-Breis jfermand@um.es
Robert Stevens robert.stevens@manchester.ac.uk
Nathalie Aussenac-Gilles aussenac@irit.fr

Ontology Summit 2013: Ontology Evaluation Across the Ontology Lifecycle

January, 31th, 2013