



UNIVERSIDAD POLITÉCNICA DE MADRID
FACULTAD DE INFORMÁTICA

Technical Report

A SOFTWARE QUALITY MODEL FOR THE EVALUATION OF SEMANTIC TECHNOLOGIES

AUTHORS: FILIP RADULOVIĆ, RAÚL GARCÍA-CASTRO

July 2012

Contents

1	Semantic Technology Quality Model	1
1.1	Ontology Engineering Tools	1
1.1.1	Test Data	3
1.1.2	Basic Measures	3
1.1.3	Derived Measures	4
1.1.4	Quality Measures	6
1.1.5	ISO 9126 Quality Characteristics	8
1.1.6	Semantic Quality Characteristics	9
1.2	Reasoning Systems	10
1.2.1	Test Data	13
1.2.2	Basic Measures	14
1.2.3	Quality Measures	15
1.2.4	ISO 9126 Quality Characteristics	18
1.2.5	Semantic Quality Characteristics	18
1.3	Matching Tools	19
1.3.1	Test Data	21
1.3.2	Basic Measures	21
1.3.3	Derived Measures	21
1.3.4	Quality Measures	22
1.3.5	ISO 9126 Quality Characteristics	22
1.3.6	Semantic Quality Characteristics	23
1.4	Semantic Search Tools	25
1.4.1	Test Data	27
1.4.2	Basic Measures	27
1.4.3	Derived Measures	28
1.4.4	Quality Measures	29
1.4.5	ISO 9126 Quality Characteristics	32
1.4.6	Semantic Quality Characteristics	33
1.5	Semantic Web Service Tools	34
1.5.1	Test Data	36

1.5.2	Basic Measures	36
1.5.3	Derived Measures	36
1.5.4	Quality Measures	38
1.5.5	ISO 9126 Quality Characteristics	39
1.5.6	Semantic Quality Characteristics	40
1.6	Complete Overview of the Quality Model	41
Bibliography		43

List of Figures

1.1	Conformance scenario.	1
1.2	Interoperability scenario.	2
1.3	Scalability scenario.	2
1.4	Quality characteristics of ontology engineering tools.	10
1.5	Classification scenario.	11
1.6	Class satisfiability scenario.	11
1.7	Ontology satisfiability scenario.	12
1.8	Entailment scenario.	12
1.9	Non-entailment scenario.	13
1.10	Quality characteristics of reasoning systems.	20
1.11	Accuracy scenario.	20
1.12	Quality characteristics of ontology matching tools.	24
1.13	Automated scenario (I).	25
1.14	Automated scenario (II).	25
1.15	User-in-the-loop scenario (I).	26
1.16	User-in-the-loop scenario (II).	26
1.17	Internal and external quality characteristics of semantic search tools.	34
1.18	Quality in use of semantic search tools.	35
1.19	Semantic web service discovery scenario.	35
1.20	Quality characteristics of semantic web service tools.	41
1.21	External and internal quality characteristics of semantic technologies.	42

Chapter 1

Semantic Technology Quality Model

This chapter presents in detail the quality model obtained for all the evaluation campaigns performed in the SEALS project which covered: ontology engineering tools (Section 1.1), reasoning systems (Section 1.2), ontology matching tools (Section 1.3), semantic search tools (Section 1.4), and semantic web service tools (Section 1.5).

1.1 Ontology Engineering Tools

The evaluation campaign for ontology engineering tools contained three evaluation scenarios to evaluate the conformance (Fig 1.1), interoperability (Fig 1.2), and scalability (Fig 1.3) of these tools.

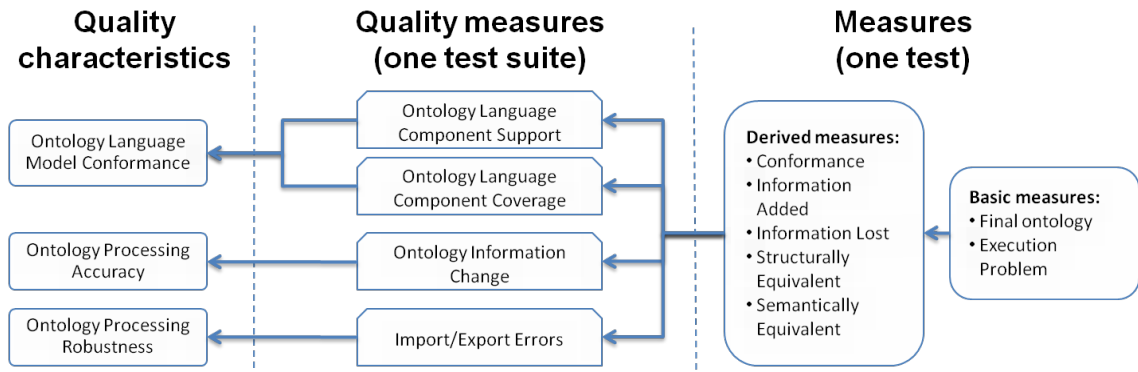


Figure 1.1: Conformance scenario.

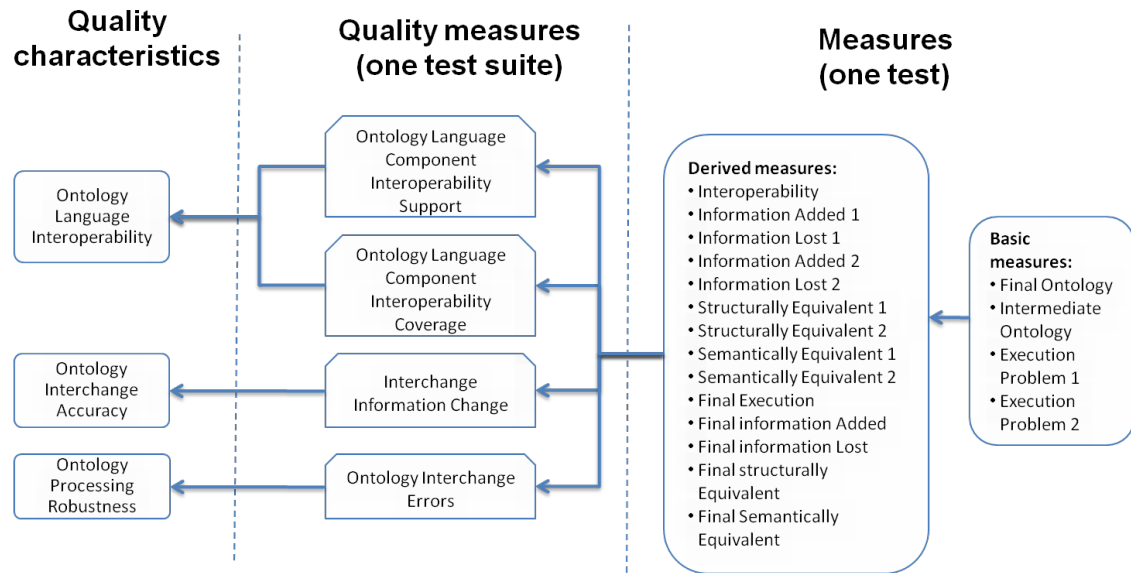


Figure 1.2: Interoperability scenario.

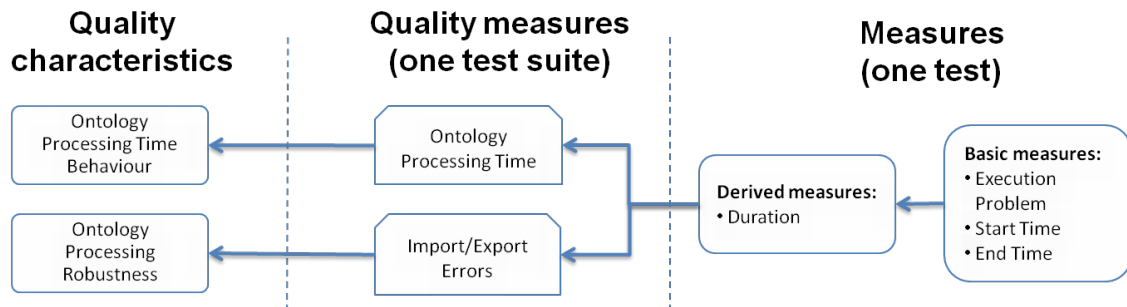


Figure 1.3: Scalability scenario.

1.1.1 Test Data

The test data used in the evaluation scenarios had the same structure. Different test suites were used, each containing several tests with one origin ontology each.

Origin Ontology - The ontology to be used as input

1.1.2 Basic Measures

Conformance Scenario

Final Ontology - The ontology that is produced by the tool when importing and exporting the origin ontology

Execution Problem - Whether there was any execution problem in the tool when importing and exporting the origin ontology. Possible values are *true*, and *false*

Interoperability Scenario

Intermediate Ontology - The ontology that is produced by the origin tool when importing and exporting the origin ontology

Final Ontology - The ontology that is produced by the tool when importing and exporting the intermediate ontology

Execution Problem 1 - Whether there was any execution problem in the tool when importing and exporting the origin ontology. Possible values are *true*, and *false*

Execution Problem 2 - Whether there was any execution problem in the tool when importing and exporting the intermediate ontology. Possible values are *true*, and *false*

Scalability Scenario

Execution Problem - Whether there was any execution problem in the tool when importing and exporting the origin ontology. Possible values are *true*, and *false*

Start Time - The time when the operation of importing and exporting the origin ontology starts

End Time - The time when the operation of importing and exporting the origin ontology ends

1.1.3 Derived Measures

Conformance Scenario

Information Added - The information added to the origin ontology (e.g., triples, axioms) after importing and exporting it

$$\text{final ontology} - \text{origin ontology}$$

Information Lost - The information lost from the origin ontology (e.g., triples, axioms) after importing and exporting it

$$\text{origin ontology} - \text{final ontology}$$

Structurally Equivalent - Whether the origin ontology and the final one are structurally equivalent. Possible values are *true*, and *false*

$$(\text{information added} = \text{null}) \wedge (\text{information lost} = \text{null})$$

Semantically Equivalent - Whether the origin ontology and the final one are semantically equivalent. Possible values are *true*, and *false*

$$\text{final ontology} \equiv \text{origin ontology}$$

Conformance - Whether the origin ontology has been imported and exported correctly with no addition or loss of information. Possible values are *true*, and *false*

$$\text{semantically equivalent} \wedge \neg(\text{execution problem})$$

Interoperability Scenario

Information Added 1 - The information added to the origin ontology after importing and exporting it

$$\text{intermediate ontology} - \text{origin ontology}$$

Information Lost 1 - The information lost from the origin ontology after importing and exporting it

$$\text{origin ontology} - \text{intermediate ontology}$$

Information Added 2 - The information added to the intermediate ontology after importing and exporting it

$$\text{final ontology} - \text{intermediate ontology}$$

Information Lost 2 - The information lost from the intermediate ontology after importing and exporting it

intermediate ontology – final ontology

Semantically Equivalent 1 - Whether the origin ontology and the intermediate one are semantically equivalent. Possible values are *true*, and *false*

intermediate ontology \equiv origin ontology

Semantically Equivalent 2 - Whether the intermediate ontology and the final one are semantically equivalent. Possible values are *true*, and *false*

final ontology \equiv intermediate ontology

Structurally Equivalent 1 - Whether the origin ontology and the intermediate one are structurally equivalent. Possible values are *true*, and *false*

(information added 1 = null) \wedge (information lost 1 = null)

Structurally Equivalent 2 - Whether the intermediate ontology and the final one are structurally equivalent. Possible values are *true*, and *false*

(information added 2 = null) \wedge (information lost 2 = null)

Final Information Added - The information added to the origin ontology after importing and exporting it by the first tool and then importing and exporting the intermediate ontology by the second tool

final ontology – origin ontology

Final Information Lost - The information lost from the origin ontology after importing and exporting it by the first tool and then importing and exporting the intermediate ontology by the second tool

origin ontology – final ontology

Final Structurally Equivalent - Whether the origin ontology and the final one are structurally equivalent. Possible values are *true*, and *false*

(final information added = null) \wedge (final information lost = null)

Final Semantically Equivalent - Whether the origin ontology and the final one are semantically equivalent. Possible values are *true*, and *false*

$$(\text{semantically equivalent 1}) \wedge (\text{semantically equivalent 2})$$

Final Execution - Whether there was any execution problem in the tool when importing and exporting the origin and intermediate ontology. Possible values are *true*, and *false*

$$(\text{execution problem 1}) \vee (\text{execution problem 2})$$

Interoperability - Whether the origin ontology has been interchanged correctly with no addition or loss of information. Possible values are *true*, and *false*

$$(\text{final semantically equivalent}) \wedge \neg(\text{execution final})$$

Scalability Scenario

Duration - The amount of time needed for importing and exporting the origin ontology

$$\text{end time} - \text{start time}$$

1.1.4 Quality Measures

Conformance Scenario

Ontology Language Component Support - Whether the tool fully supports an ontology language component

$$\frac{\# \text{ tests that contain the component where conformance} = \text{true}}{\# \text{ tests that contain the component}} = 1$$

Ontology Language Component Coverage - The ratio of ontology components that are shared by a tool internal model and an ontology language model

$$\frac{\# \text{ components in the ontology language where component support} = \text{true}}{\# \text{ components in the ontology language}} \times 100$$

Ontology Information Change - The ratio of information additions or losses when importing and exporting ontologies

$$\frac{\# \text{ tests where (information added} \neq \text{null or information lost} \neq \text{null)}}{\# \text{ tests}} \times 100$$

Import/Export Errors - The ratio of tool execution errors when importing and exporting ontologies

$$\frac{\# \text{ tests where execution problem} = \text{true}}{\# \text{ tests}} \times 100$$

Interoperability Scenario

Ontology Language Component Interoperability Support - Whether the tool fully supports an ontology language component interchange

$$\frac{\# \text{ tests that contain the component where interoperability} = \text{true}}{\# \text{ tests that contain the component}} = 1$$

Ontology Language Component Interoperability Coverage - The ratio of ontology components that can be interchanged with other tools

$$\frac{\# \text{ components in the ontology language where component interoperability support} = \text{true}}{\# \text{ components in the ontology language}} \times 100$$

Interchange Information Change - The ratio of information additions or losses when interchanging ontologies

$$\frac{\# \text{ tests where (final information added} \neq \text{null or final information lost} \neq \text{null)}}{\# \text{ tests}} \times 100$$

Ontology Interchange Errors - The ratio of tool execution errors when interchanging ontologies

$$\frac{\# \text{ tests where final execution} = \text{true}}{\# \text{ tests}} \times 100$$

Scalability Scenario

Ontology Processing Time - The average amount of time needed for importing and exporting ontologies

$$\frac{\sum_n \text{duration in the } n^{\text{th}} \text{ test}}{\# \text{ tests}}$$

Import/Export Errors - The ratio of tool execution errors when importing and

exporting ontologies

$$\frac{\# \text{ tests where execution problem} = \text{true}}{\# \text{ tests}} \times 100$$

1.1.5 ISO 9126 Quality Characteristics

Functional Suitability - The degree to which the software product provides functions that meet stated and implied needs when the software is used under specified conditions.

Functional Compliance - The degree to which the software product adheres to standards, conventions or regulations in laws and similar prescriptions relating to functional suitability. *Functional Compliance* is a sub-characteristic of *Functional Compliance*

Compatibility - The ability of two or more software components to exchange information and/or to perform their required functions while sharing the same hardware or software environment.

Interoperability - The degree to which the software product can be cooperatively operable with one or more other software products. *Interoperability* is a sub-characteristic of *Compatibility*

Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision. *Accuracy* is a sub-characteristic of *Functional Suitability*

Performance Efficiency - The degree to which the software product provides appropriate performance, relative to the amount of resources used, under stated conditions.

Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when performing its function, under stated conditions. *Time behaviour* is a sub-characteristic of *Performance Efficiency*

Reliability - The degree to which the software product can maintain a specified level of performance when used under specified conditions.

Robustness - The ability of the software product to function correctly in the

presence of invalid inputs or stressful environmental conditions. *Robustness* is a sub-characteristic of *Reliability*

1.1.6 Semantic Quality Characteristics

Ontology Language Model Conformance - The degree to which the knowledge representation model of the software product adheres to the knowledge representation model of an ontology language. *Ontology language model conformance* is a sub-characteristic of *Functional Compliance* and can be measured using:

- Ontology Language Component Coverage
- Ontology Language Component Support

Ontology Processing Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision when processing ontologies. *Ontology Processing Accuracy* is a sub-characteristic of *Accuracy* and can be measured using:

- Ontology Information Change

Ontology Language Interoperability - The degree to which the software product can interchange ontologies and use the ontologies that have been interchanged. *Ontology language interoperability* is a sub-characteristic of *Interoperability* and can be measured using:

- Ontology Language Component Interoperability Coverage
- Ontology Language Component Interoperability Support

Ontology Interchange Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision when interchanging the ontologies. *Ontology Interchange Accuracy* is the sub-characteristic of *Ontology Processing Accuracy* and can be measured using:

- Interchange Information Change

Ontology Processing Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when working with ontologies. *Ontology Processing Time Behaviour* is a sub-characteristic of *Time Behaviour* and can be measured using:

- Ontology Processing Time

Ontology Processing Robustness - The ability of the software product to process ontologies correctly in the presence of invalid inputs or stressful environmental conditions. *Ontology Processing Robustness* is a sub-characteristic of *Robustness* and it can be measured using:

- Import/Export Errors
- Ontology Interchange Errors

Figure 1.4 shows the part of the quality model obtained from the evaluation campaign for ontology engineering tools.

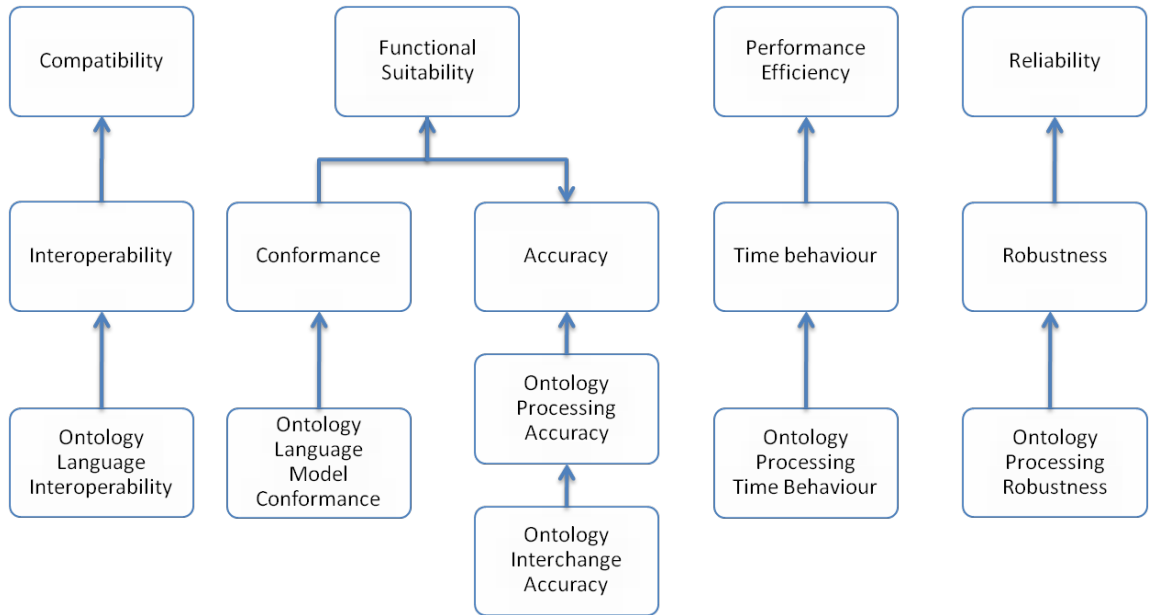


Figure 1.4: Quality characteristics of ontology engineering tools.

1.2 Reasoning Systems

The evaluation campaign for reasoning systems contained evaluation scenarios to evaluate the accuracy, scalability, and robustness of these tools. Those scenarios include: classification scenario (Fig 1.5), class satisfiability scenario (Fig 1.6), ontology satisfiability scenario (Fig 1.7), entailment scenario (Fig 1.8), and non-entailment scenario (Fig 1.9).

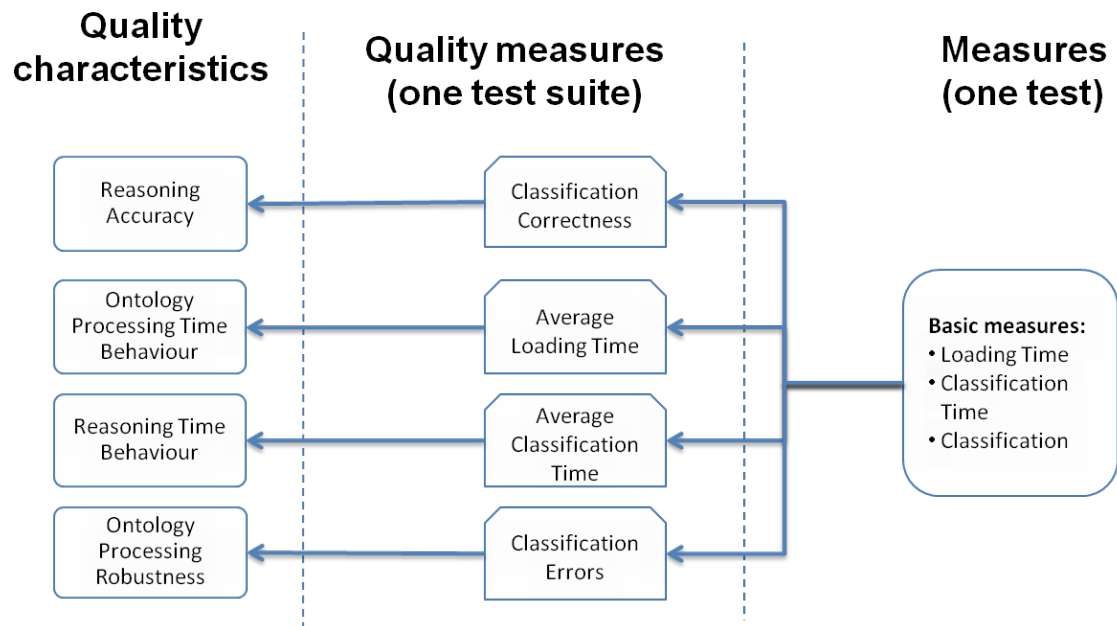


Figure 1.5: Classification scenario.

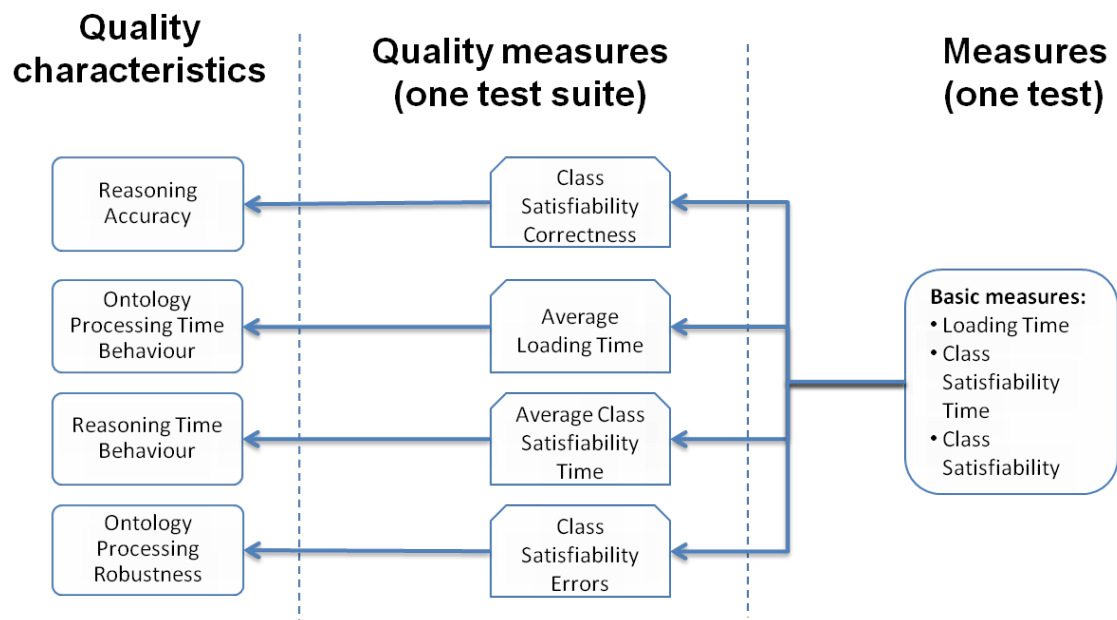


Figure 1.6: Class satisfiability scenario.

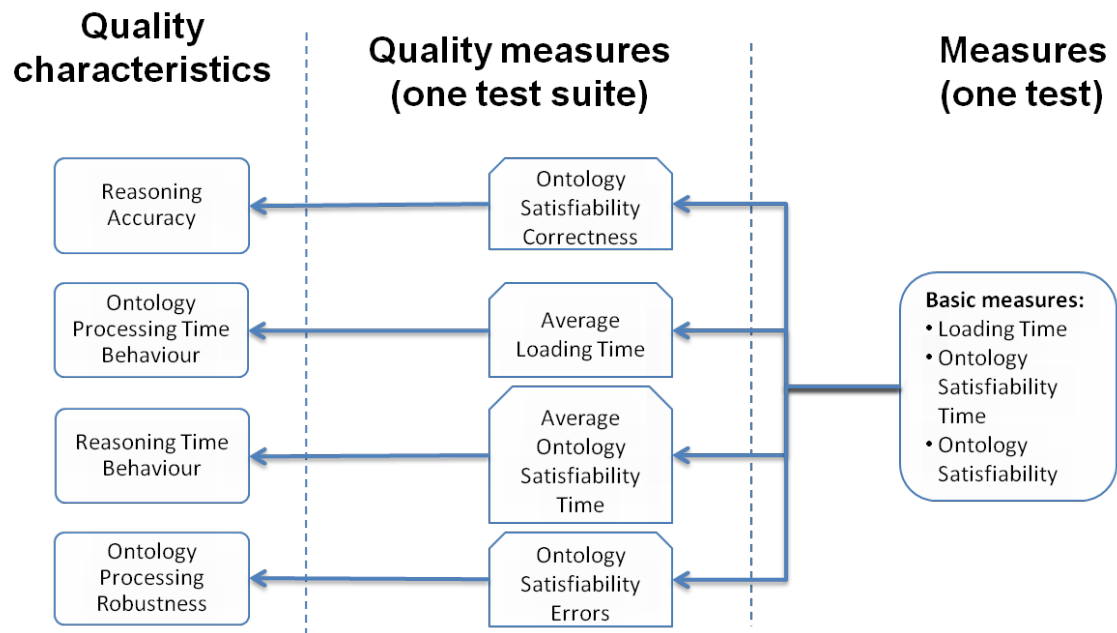


Figure 1.7: Ontology satisfiability scenario.

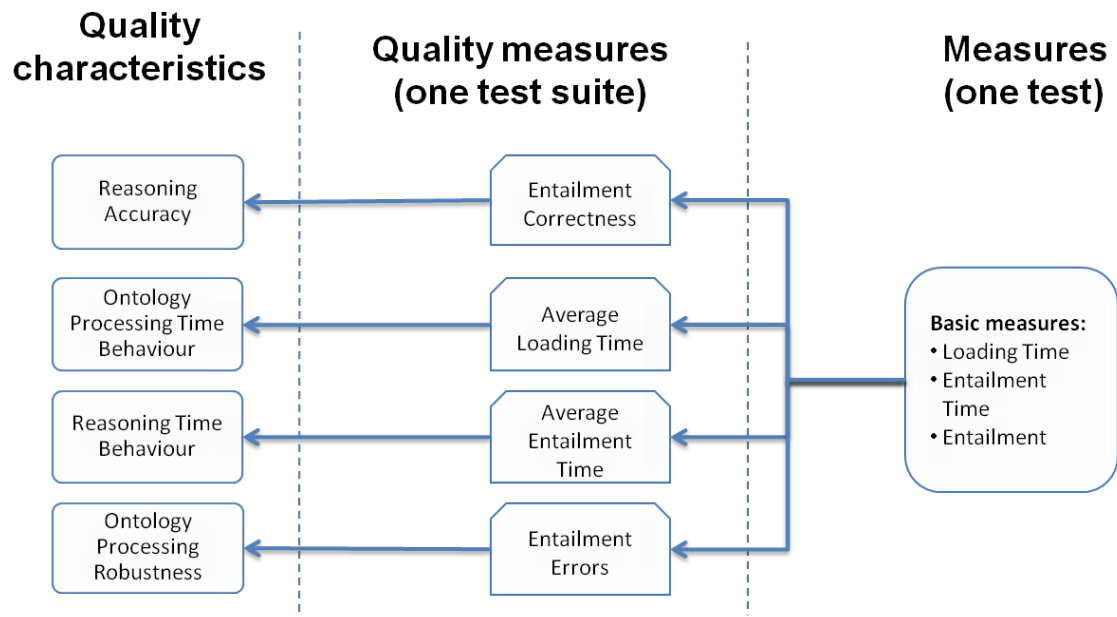


Figure 1.8: Entailment scenario.

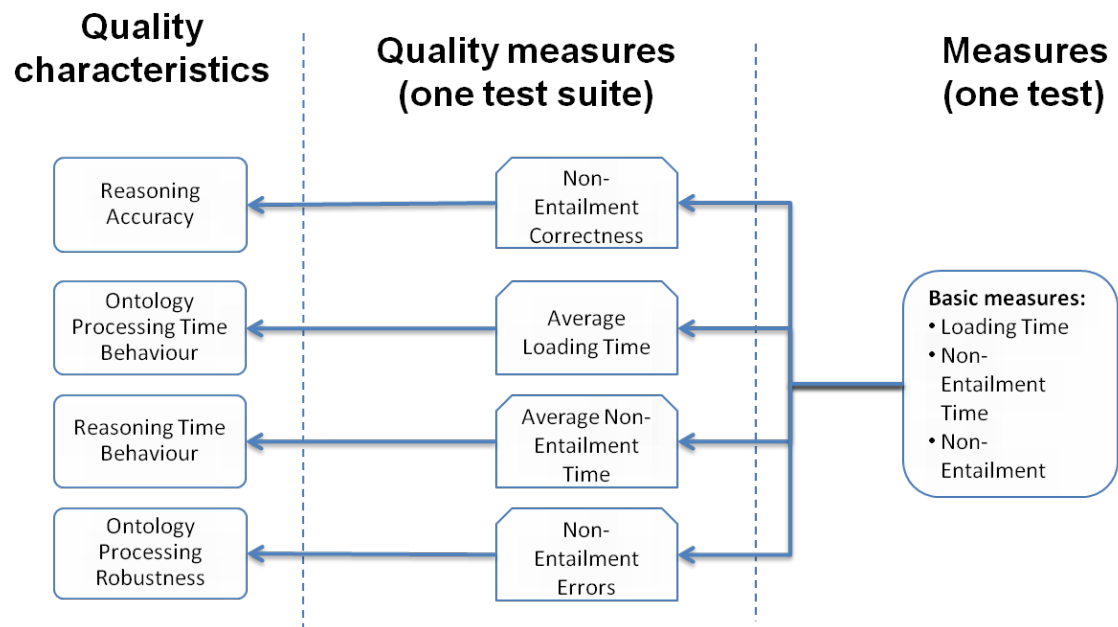


Figure 1.9: Non-entailment scenario.

1.2.1 Test Data

Classification Scenario

Ontology - The ontology to be classified

Class Satisfiability Scenario

Ontology - The ontology to be used in the class satisfiability task

Class URIs - URIs of the classes to be checked for satisfiability

Ontology Satisfiability Scenario

Ontology - The ontology to be checked for satisfiability

Entailment Scenario

Premise Ontology - The ontology to be used in the entailment scenario

Conclusion Ontology - The ontology to be checked for being logically entailed by the premise ontology

Non-Entailment Scenario

Premise Ontology - The ontology to be used in the non-entailment scenario

Conclusion Ontology - The ontology to be checked for not being logically entailed by the premise ontology

1.2.2 Basic Measures

Classification Scenario

Loading time - The amount of time needed to load the ontology

Classification Time - The amount of time needed to perform a classification operation

Classification - Whether the classification of the ontology is performed correctly. Possible values are *true*, *false*, *unknown*, and *error*

Class Satisfiability Scenario

Loading time - The amount of time needed to load the ontology

Class Satisfiability Time - The amount of time needed to perform a class satisfiability operation

Class Satisfiability - Whether a specific class from the ontology is satisfiable. Possible values are *true*, *false*, *unknown*, and *error*

Ontology Satisfiability Scenario

Loading time - The amount of time needed to load the ontology

Ontology Satisfiability Time - The amount of time needed to perform an ontology satisfiability operation

Ontology Satisfiability - Whether the origin ontology is satisfiable. Possible values are *true*, *false*, *unknown*, and *error*

Entailment Scenario

Loading time - The amount of time needed to load the ontology

Entailment Time - The amount of time needed to perform entailment operation

Entailment - Whether the conclusion ontology is being logically entailed by the premise ontology. Possible values are *true*, *false*, *unknown*, and *error*

Non-Entailment Scenario

Loading time - The amount of time needed to load the ontology

Non-Entailment Time - The amount of time needed to perform non-entailment

operation

Non-Entailment - Whether the conclusion ontology is not being logically entailed by the premise ontology. Possible values are *true*, *false*, *unknown*, and *error*

1.2.3 Quality Measures

Classification Scenario

Average Loading Time - The average time needed for the tool to load an ontology

$$\frac{\sum_n \text{loading time for the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Classification Time - The average time needed for the tool to perform a classification task

$$\frac{\sum_n \text{classification time for the } n^{th} \text{ test}}{\# \text{ tests}}$$

Classification Errors - The ratio of tool execution errors when performing a classification task

$$\frac{\# \text{ tests where classification} = \text{error}}{\# \text{ tests}} \times 100$$

Classification Correctness The ratio of correctly performed classification tasks

$$\frac{\# \text{ tests where classification} = \text{true}}{\# \text{ tests}} \times 100$$

Class Satisfiability Scenario

Average Loading Time - The average time needed for the tool to load an ontology

$$\frac{\sum_n \text{loading time for the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Class Satisfiability Time - The average time needed for the tool to perform a class satisfiability task

$$\frac{\sum_n \text{class satisfiability time for the } n^{th} \text{ test}}{\# \text{ tests}}$$

Class Satisfiability Errors - The ratio of tool execution errors when performing class satisfiability tasks

$$\frac{\# \text{ tests where class satisfiability} = \text{error}}{\# \text{ tests}} \times 100$$

Class Satisfiability Correctness The ratio of correctly performed class satisfiability tasks

$$\frac{\# \text{ tests where class satisfiability} = \text{true}}{\# \text{ tests}} \times 100$$

Ontology Satisfiability Scenario

Average Loading Time - The average time needed for the tool to load an ontology

$$\frac{\sum_n \text{loading time for the } n^{\text{th}} \text{ test}}{\# \text{ tests}}$$

Average Ontology Satisfiability Time - The average time needed for the tool to perform an ontology satisfiability task

$$\frac{\sum_n \text{ontology satisfiability time for the } n^{\text{th}} \text{ test}}{\# \text{ tests}}$$

Ontology Satisfiability Errors - The ratio of tool execution errors when performing an ontology satisfiability task

$$\frac{\# \text{ tests where ontology satisfiability} = \text{error}}{\# \text{ tests}} \times 100$$

Ontology Satisfiability Correctness The ratio of correctly performed ontology satisfiability tasks

$$\frac{\# \text{ tests where ontology satisfiability} = \text{true}}{\# \text{ tests}} \times 100$$

Entailment Scenario

Average Loading Time - The average time needed for the tool to load an ontology

$$\frac{\sum_n \text{loading time for the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Entailment Time - The average time needed for the tool to perform an entailment task

$$\frac{\sum_n \text{entailment time for the } n^{th} \text{ test}}{\# \text{ tests}}$$

Entailment Errors - The ratio of tool execution errors when performing an entailment task

$$\frac{\# \text{ tests where entailment} = \text{error}}{\# \text{ tests}} \times 100$$

Entailment Correctness The ratio of correctly performed entailment tasks

$$\frac{\# \text{ tests where entailment} = \text{true}}{\# \text{ tests}} \times 100$$

Non-Entailment Scenario

Average Loading Time - The average time needed for the tool to load an ontology

$$\frac{\sum_n \text{loading time for the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Non-Entailment Time - The average time needed for the tool to perform a non-entailment task

$$\frac{\sum_n \text{non-entailment time for the } n^{th} \text{ test}}{\# \text{ tests}}$$

Non-Entailment Errors - The ratio of tool execution errors when performing a non-entailment task

$$\frac{\# \text{ tests where non-entailment} = \text{error}}{\# \text{ tests}} \times 100$$

Non-Entailment Correctness The ratio of correctly performed non-entailment

tasks

$$\frac{\# \text{ tests where non-entailment} = \text{true}}{\# \text{ tests}} \times 100$$

1.2.4 ISO 9126 Quality Characteristics

Functional Suitability - The degree to which the software product provides functions that meet stated and implied needs when the software is used under specified conditions.

Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision. *Accuracy* is a sub-characteristic of *Functional Suitability*

Performance Efficiency - The degree to which the software product provides appropriate performance, relative to the amount of resources used, under stated conditions.

Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when performing its function, under stated conditions. *Time behaviour* is a sub-characteristic of *Performance Efficiency*

Reliability - The capability of the software product to maintain a specified level of performance when used under specified conditions

Robustness - The ability of the software product to function correctly in the presence of invalid inputs or stressful environmental conditions. *Robustness* is a sub-characteristic of *Reliability*

1.2.5 Semantic Quality Characteristics

Ontology Processing Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision when processing ontologies. *Ontology Processing Accuracy* is a sub-characteristic of *Accuracy*

Reasoning Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision when performing the reasoning task. *Reasoning Accuracy* is a sub-characteristic of *Ontology Processing Accuracy* and can be measured using:

- Class Satisfiability Correctness
- Ontology Satisfiability Correctness
- Classification Correctness
- Entailment Correctness

- Non-Entailment Correctness

Ontology Processing Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when working with ontologies. *Ontology Processing Time Behaviour* is a sub-characteristic of *Time Behaviour* and can be measured using:

- Average Loading time

Reasoning Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when performing reasoning tasks. *Reasoning Time Behaviour* is a sub-characteristic of *Ontology Processing Time Behaviour* and can be measured using:

- Average Classification Time
- Average Class Satisfiability Time
- Average Ontology Satisfiability Time
- Average Entailment Time
- Average Non-Entailment Time

Ontology Processing Robustness - The ability of the software product to process ontologies correctly in the presence of invalid inputs or stressful environmental conditions. *Ontology Processing Robustness* is a sub-characteristic of *Robustness* and it can be measured using:

- Classification Errors
- Class Satisfiability Errors
- Ontology Satisfiability Errors
- Entailment Errors
- Non-Entailment Errors

Figure 1.10 shows the part of the quality model obtained from the evaluation campaign for reasoning tools.

1.3 Matching Tools

The evaluation campaign for matching tools contained one evaluation scenario to evaluate accuracy (Fig 1.11).

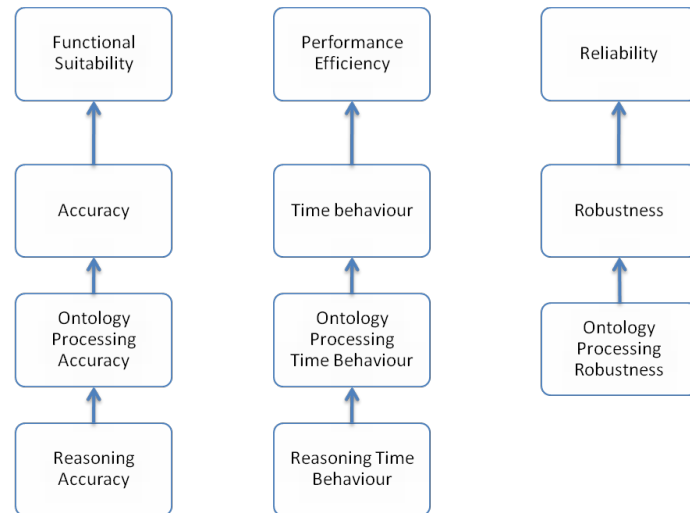


Figure 1.10: Quality characteristics of reasoning systems.

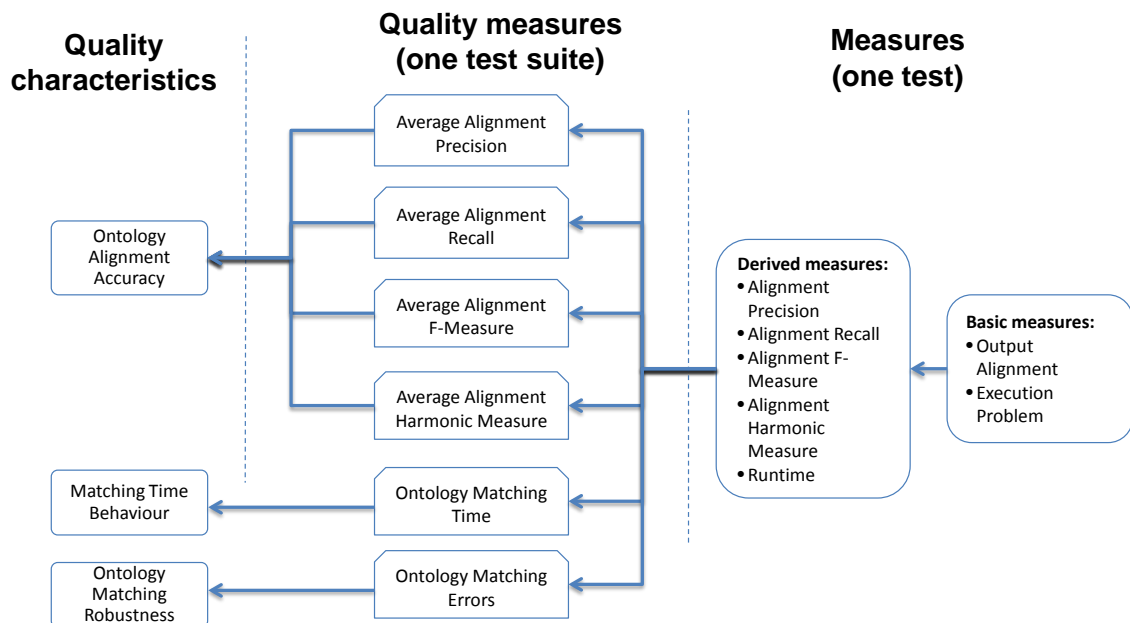


Figure 1.11: Accuracy scenario.

1.3.1 Test Data

Ontology 1 - One ontology to be used as input for an alignment task

Ontology 2 - Another ontology to be used as input for an alignment task

Expected Alignment - The expected alignment to be obtained when aligning the two input ontologies

1.3.2 Basic Measures

Output Alignment - The alignment that is produced by the tool when aligning ontology 1 and ontology 2

Execution Problem - Whether there was any execution problem in the tool when performing matching task. Possible values are *true*, and *false*

1.3.3 Derived Measures

Alignment Precision - The precision of the alignment task

$$\frac{\# \text{ output alignments that match expected alignments}}{\# \text{ output alignments}}$$

Alignment Recall - The recall of the alignment task

$$\frac{\# \text{ output alignments that match expected alignments}}{\# \text{ expected alignments}}$$

Alignment F-Measure - The aggregation measure of alignment precision and alignment recall

$$\frac{\text{alignment precision} \times \text{alignment recall}}{(1-\alpha) \times \text{alignment precision} + \alpha \times \text{alignment recall}}, \alpha = [0..1]$$

Alignment Harmonic Measure - The aggregation measure of alignment precision and alignment recall (value of alignment F-Measure when $\alpha = 0.5$)

$$\frac{\text{alignment precision} \times \text{alignment recall}}{\text{alignment precision} + \text{alignment recall}} \times 2$$

Runtime - The amount of time needed for the matching task

1.3.4 Quality Measures

Average Alignment Precision - The average precision of the ontology matching tool

$$\frac{\sum_n \text{alignment precision of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Alignment Recall - The average recall of the ontology matching tool

$$\frac{\sum_n \text{alignment recall of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Alignment F-Measure - The average F-Measure of the ontology matching tool

$$\frac{\sum_n \text{alignment F-Measure of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Alignment Harmonic Measure - The average harmonic measure of the ontology matching tool

$$\frac{\sum_n \text{alignment harmonic measure of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Ontology Matching Time - The average amount of time needed for the matching task

$$\frac{\sum_n \text{runtime in the } n^{th} \text{ test}}{\# \text{ tests}}$$

Ontology Matching Errors - The ratio of tool execution errors when performing matching tasks

$$\frac{\# \text{ tests where execution problem} = \text{true}}{\# \text{ tests}} \times 100$$

1.3.5 ISO 9126 Quality Characteristics

Functional Suitability - The degree to which the software product provides functions that meet stated and implied needs when the software is used under specified conditions.

Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision. *Accuracy* is a sub-characteristic of

Functional Suitability

Performance Efficiency - The degree to which the software product provides appropriate performance, relative to the amount of resources used, under stated conditions.

Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when performing its function, under stated conditions. *Time behaviour* is a sub-characteristic of *Performance Efficiency*

Reliability - The degree to which the software product can maintain a specified level of performance when used under specified conditions.

Robustness - The ability of the software product to function correctly in the presence of invalid inputs or stressful environmental conditions. *Robustness* is a sub-characteristic of *Reliability*

1.3.6 Semantic Quality Characteristics

Ontology Processing Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision when processing ontologies. *Ontology Processing Accuracy* is a sub-characteristic of *Accuracy*

Ontology Processing Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when working with ontologies. *Ontology Processing Time Behaviour* is a sub-characteristic of *Time Behaviour*.

Ontology Processing Robustness - The ability of the software product to process ontologies correctly in the presence of invalid inputs or stressful environmental conditions. *Ontology Processing Robustness* is a sub-characteristic of *Robustness*.

Ontology Alignment Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision when performing the alignment task. *Ontology alignment accuracy* is a sub-characteristic of *Ontology Processing Accuracy* and can be measured using:

- Average Alignment Precision
- Average Alignment Recall

- Average Alignment F-measure
- Average Alignment Harmonic Measure

Matching Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when performing the alignment task. *Matching time Behaviour* is a sub-characteristic of *Ontology Processing Time Behaviour* and can be measured using:

- Ontology Matching time

Ontology Matching Robustness - The ability of the software product to perform alignment task correctly in the presence of invalid inputs or stressful environmental conditions. *Ontology Matching Robustness* is a sub-characteristic of *Ontology Processing Robustness* and can be measured using:

- Ontology Matching Errors

Figure 1.12 shows the part of the quality model obtained from the evaluation campaign for matching tools.

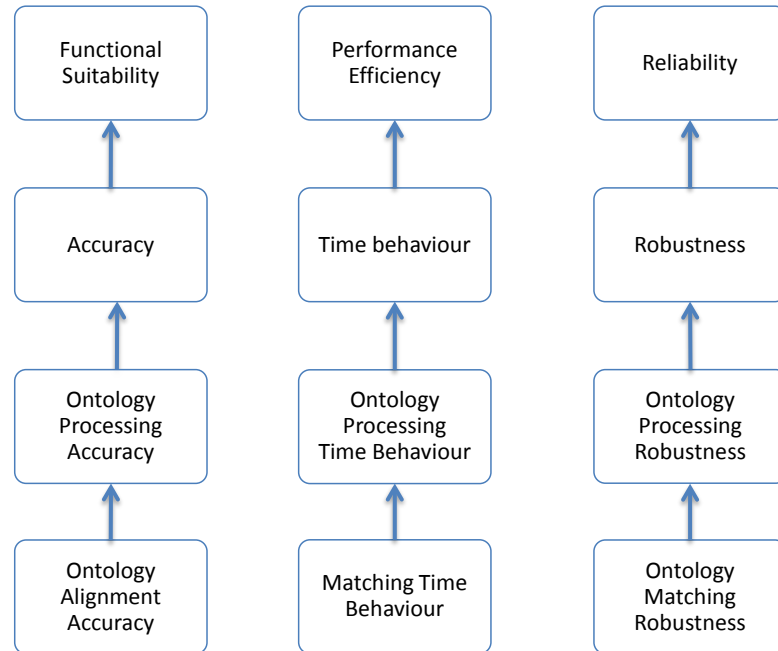


Figure 1.12: Quality characteristics of ontology matching tools.

1.4 Semantic Search Tools

The evaluation campaign for semantic search tools contained evaluation scenarios to evaluate accuracy, performance and usability. The evaluation was divided into two separate phases, an automated phase (Fig 1.13 and Fig 1.14) and a user-in-the-loop phase (Fig 1.15 and Fig 1.16).

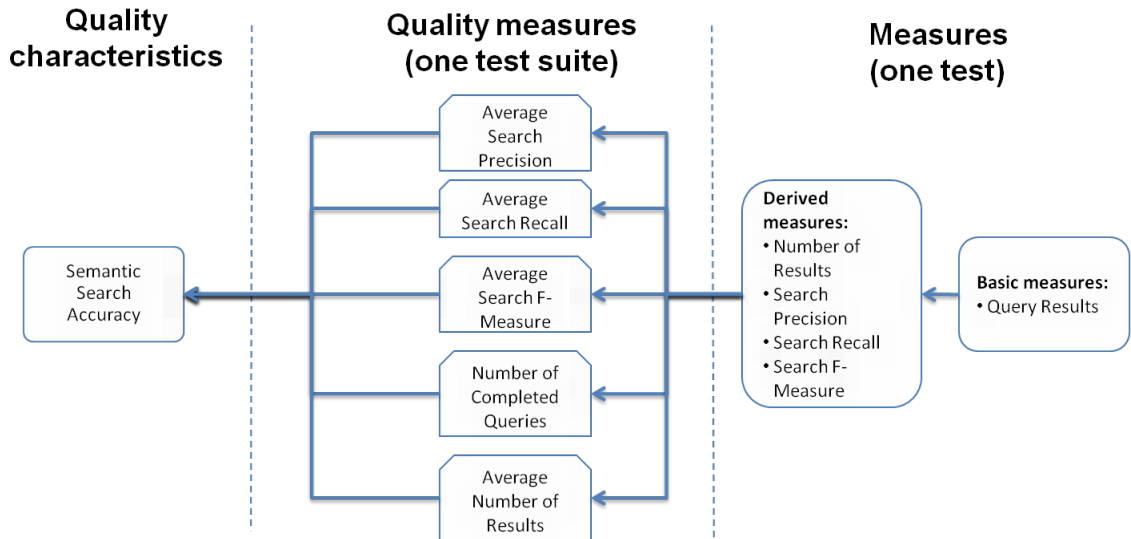


Figure 1.13: Automated scenario (I).

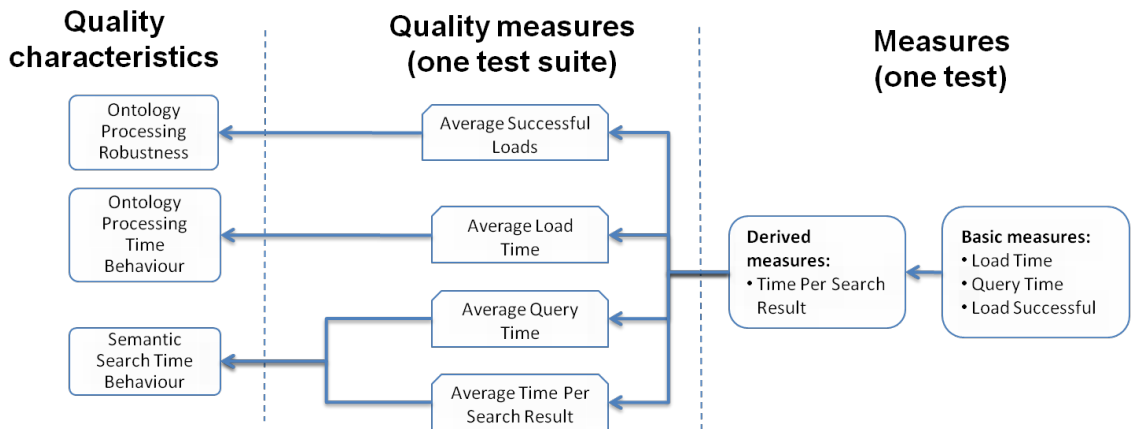


Figure 1.14: Automated scenario (II).

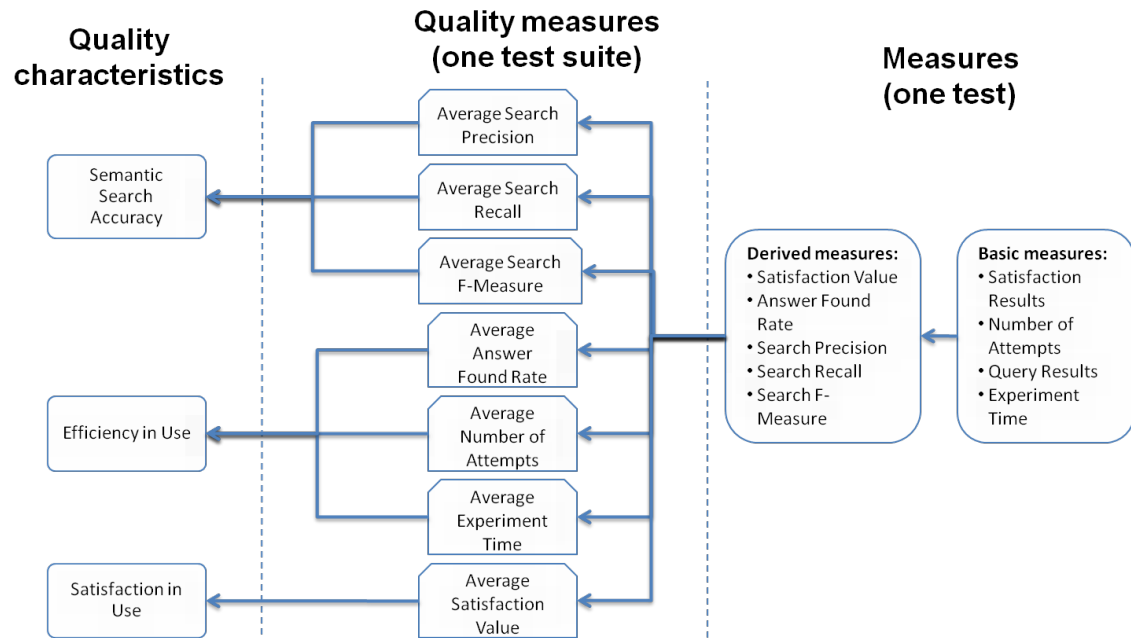


Figure 1.15: User-in-the-loop scenario (I).

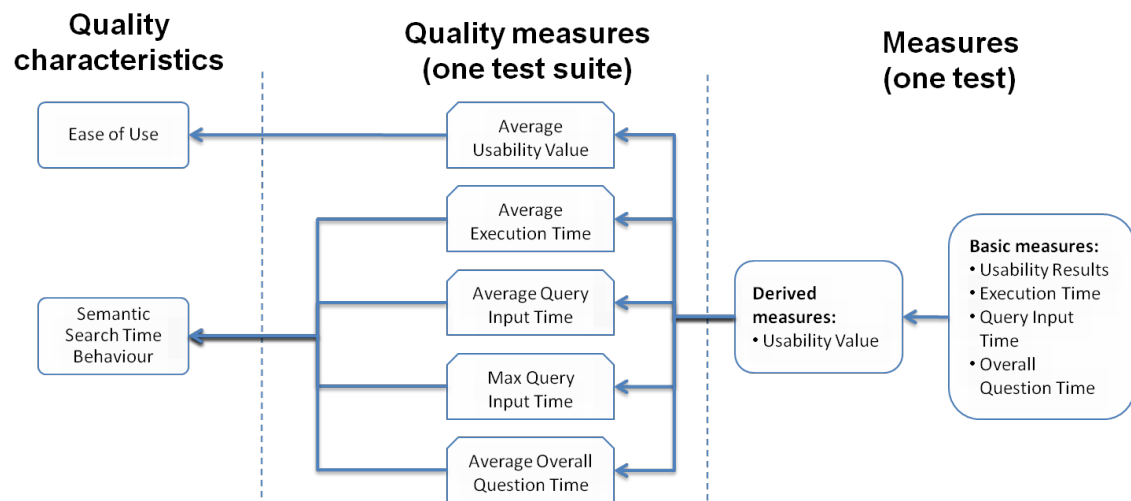


Figure 1.16: User-in-the-loop scenario (II).

1.4.1 Test Data

Automated Scenario

Origin Ontology - The ontology to be used as input for the search task

Input Question - The question to be used as input for the search task

Expected Answer - The expected answer to be obtained from the origin ontology for the input question

User-In-The-Loop Scenario

Origin Ontology - The ontology to be used as input for the search task

Input Question - The question to be used as input for the search task

Expected Answer - The expected answer to be obtained from the origin ontology for the input question

Usability Questionnaire - Standardized Usability Test by Brooke [4] containing ten standardized questions to be given to the users

Satisfaction Questionnaire - User satisfaction questionnaire to be given to the users

Demographics Questionnaire - Questionnaire about the characteristics of a population to be given to the users

1.4.2 Basic Measures

Automated Scenario

Query Result - The results returned by the tool when executing the query

Load Successful - Whether the ontology has been successfully loaded. Possible values are *true* or *false*

Load Time - The amount of time needed to load the ontology

Query Time - The amount of time it takes for the query to execute

User-In-The-Loop Scenario

Usability Result - The results of the usability questionnaire filled by the user

Satisfaction Result - The results of the satisfaction questionnaire filled by the user

Number of Attempts - The number of times the user had to reformulate the query using the tool interface in order to obtain satisfactory answers

Query Result - The result of the executed query

Experiment Time - The amount of time needed for one user to complete the whole experiment

Query Input Time - The amount of time the user spends for formulating a query

using the tool's interface

Execution time - The amount of time needed to execute the query

Overall Question Time - The amount of time needed for answering one question

1.4.3 Derived Measures

Automated Scenario

Number of Results - The number of results obtained for the query

$$\# \text{ query result}$$

Time Per Search Result - The amount of time needed to obtain one result

$$\frac{\text{query time}}{\# \text{ results}}$$

Search Precision - The ratio between the relevant answers and all answers returned by the tool

$$\frac{\# \text{ relevant answers returned by the tool}}{\# \text{ answers returned by the tool}}$$

Search Recall - The ration between the relevant answers returned by the tool and answers in groundtruth

$$\frac{\# \text{ relevant answers returned by the tool}}{\# \text{ answers in groundtruth}}$$

Search F-Measure - The aggregation measure of search precision and search recall

$$\frac{\text{search precision} \times \text{search recall}}{\text{search precision} + \text{search recall}} \times 2$$

User-In-The-Loop Scenario

Usability Value - The value obtained from the usability questionnaire

Satisfaction Value - The value obtained from satisfaction questionnaire

Answer Found Rate - The ratio of finding the appropriate answer after a number of attempts and the user giving up after a number of attempts

$$\frac{\text{number of attempts}}{\# \text{ of attempts before giving up}}$$

Search Precision - The ratio between the relevant answers and all answers returned by the tool

$$\frac{\# \text{ relevant answers returned by the tool}}{\# \text{ answers returned by the tool}}$$

Search Recall - The ration between the relevant answers returned by the tool and answers in groundtruth

$$\frac{\# \text{ relevant answers returned by the tool}}{\# \text{ answers in groundtruth}}$$

Search F-Measure - The aggregation measure of search precision and search recall

$$\frac{\text{search precision} \times \text{search recall}}{\text{search precision} + \text{search recall}} \times 2$$

1.4.4 Quality Measures

Automated Scenario

Average Successful Loads - The ratio of successful ontology loads

$$\frac{\# \text{ tests where load successful} = \text{true}}{\# \text{ tests}} \times 100$$

Average Load Time - The average amount of time it takes for the ontology to load

$$\frac{\sum_n \text{load time of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Query Time - The average amount of time it takes for a query to execute

$$\frac{\sum_n \text{query time of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Number of Results - The average number of results obtained for the query

$$\frac{\sum_n \text{number of results of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Number of Completed Queries - The number of queries that produced expected

answer

tests where expected answer = query result

Average Time Per Search Result - The average amount of time needed to obtain one result

$$\frac{\sum_n \text{time per search result of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Search Precision - The average precision of search

$$\frac{\sum_n \text{search precision of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Search Recall - The average recall of search tool

$$\frac{\sum_n \text{search recall of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Search F-Measure - The average F-Measure of search tool

$$\frac{\sum_n \text{search F-Measure of the } n^{th} \text{ test}}{\# \text{ tests}}$$

User-In-The-Loop Scenario

Average Experiment Time - The average time needed for the experiment to complete

$$\frac{\sum_n \text{experiment time of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Usability Value - The average value of results obtained from usability questionnaire

$$\frac{\sum_n \text{usability value of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Satisfaction Value - The average value obtained from satisfaction ques-

tionnaire

$$\frac{\sum_n \text{satisfaction value of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Number of Attempts - The average number of attempts before the user is happy with the results

$$\frac{\sum_n \text{number of attempts of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Answer Found Rate - The ratio of finding the appropriate answer after a number of attempts and the user giving up after a number of attempts

$$\frac{\sum_n \text{answer found rate of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Execution Time - The average time needed to execute the question

$$\frac{\sum_n \text{execution time of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Query Input Time - The average time needed for the query to be formulated

$$\frac{\sum_n \text{input time of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Max Query Input Time - The maximum time needed for the query to be inputted

$$\max(\text{input time of the } n^{th} \text{ test})$$

Average Overall Question Time - The average amount of time needed for answering one question

$$\frac{\sum_n \text{overall question time of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Search Precision - The average precision of the search tool

$$\frac{\sum_n \text{search precision of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Search Recall - The average recall of the search tool

$$\frac{\sum_n \text{search recall of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Search F-Measure - The average F-Measure of the search tool

$$\frac{\sum_n \text{search F-Measure of the } n^{th} \text{ test}}{\# \text{ tests}}$$

1.4.5 ISO 9126 Quality Characteristics

Functional Suitability - The degree to which the software product provides functions that meet stated and implied needs when the software is used under specified conditions.

Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision. *Accuracy* is a sub-characteristic of *Functional Suitability*

Performance Efficiency - The degree to which the software product provides appropriate performance, relative to the amount of resources used, under stated conditions.

Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when performing its function, under stated conditions. *Time behaviour* is a sub-characteristic of *Performance Efficiency*

Reliability - The capability of the software product to maintain a specified level of performance when used under specified conditions

Robustness - The ability of the software product to function correctly in the presence of invalid inputs or stressful environmental conditions. *Robustness* is a sub-characteristic of *Reliability*

Operability - The degree to which the software product can be understood, learned, used and attractive to the user, when used under specified conditions.

Ease of Use - The degree to which the software product makes it easy for users to operate and control it. *Ease of Use* is a sub-characteristic of *Operability*, and can be measured using:

- Average Usability Value

Usability in Use - The degree to which specified users can achieve specified goals with effectiveness in use, efficiency in use and satisfaction in use in a specified context of use.

Efficiency in Use - The degree to which specified users expend appropriate amounts of resources in relation to the effectiveness achieved in a specified context of use. *Efficiency in Use* is a sub-characteristic of *Usability in Use*, and can be measured using

- Average Number of Attempts
- Average Answer Found Rate
- Average Experiment Time

Satisfaction in Use - The degree to which users are satisfied in a specified context of use. *Satisfaction in Use* is a sub-characteristic of *Usability in Use*, and can be measured using:

- Average Satisfaction Value

1.4.6 Semantic Quality Characteristics

Semantic Search Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision when performing the search task. *Semantic search accuracy* is a sub-characteristic of *Accuracy* and can be measured using

- Average Search Precision
- Average Search Recall
- Average Search F-measure
- Number of Completed Queries
- Average Number of Results

Ontology Processing Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when working with ontologies. *Ontology Processing Time Behaviour* is a sub-characteristic of *Time Behaviour* and can be measured using:

- Average Load Time

Ontology Processing Robustness - The ability of the software product to process ontologies correctly in the presence of invalid inputs or stressful environmental conditions. *Ontology Processing Robustness* is a sub-characteristic of *Robustness* and it can be measured using:

- Average Successful Loads

Semantic Search Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when performing the search tasks. *Semantic Search Time Behaviour* is a sub-characteristic of *Time Behaviour* and can be measured using:

- Average Query Time
- Average Time Per Search Result
- Average Execution Time
- Average Query Input Time
- Max Query Input Time
- Average Overall Question Time

Figure 1.17 shows the part of the quality model internal and external characteristics obtained from the evaluation campaign for semantic search tools. Fig 1.18 shows the quality in use.

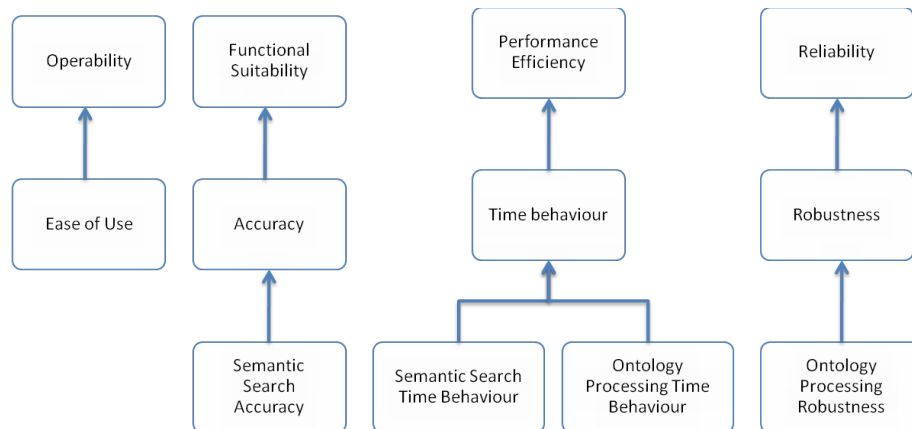


Figure 1.17: Internal and external quality characteristics of semantic search tools.

1.5 Semantic Web Service Tools

The evaluation campaign for semantic web services contained one evaluation scenario to evaluate discovery (Fig 1.19).

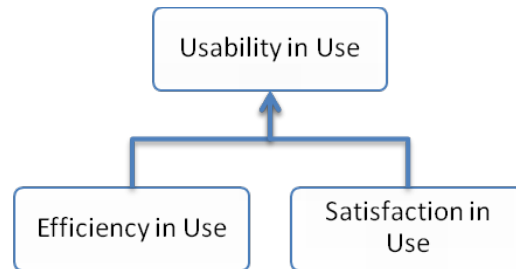


Figure 1.18: Quality in use of semantic search tools.

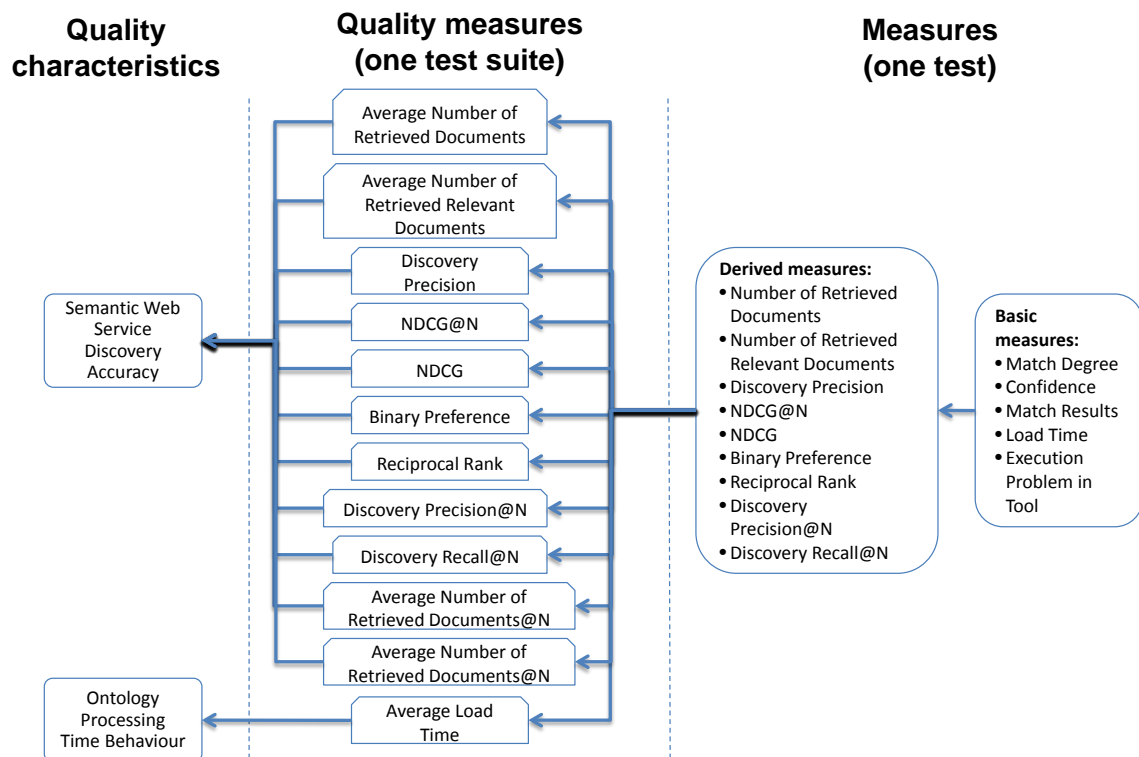


Figure 1.19: Semantic web service discovery scenario.

1.5.1 Test Data

Goal Document - The semantic web service document containing users' request description to be used in discovery test

Service Document - The semantic web service document containing service offer to be used in discovery test

Queries - The query to be used in discovery test

1.5.2 Basic Measures

Match Degree - Information about the match degree. Possible values are *Non*, *Plugin*, *Match*, *Exact*, *Subsumption*

Confidence - The confidence value of the produced match

Match Results - The results that are produced by the tool in the process of discovery

Load Time - The amount of time needed to load the query

Execution Problem in Tool - Whether there were any execution problems in the tool while performing the discovery process. Possible values are *true*, and *false*

1.5.3 Derived Measures

Number of Retrieved Documents - The number of retrieved documents

retrieved documents

Number of Retrieved Relevant Documents - Number of retrieved documents that were judged relevant

retrieved documents where (match degree = match \vee exact)

Discovery Precision - Precision of the semantic web service discovery

$$\frac{\# \text{ relevant documents retrieved}}{\# \text{ retrieved documents}}$$

Normalized Discounted Cumulative Gain (NDCG) - A normalized measure of the effectiveness of the search algorithm according to the relevance of a search result

$$rel_1 + \sum_{i=2} \frac{rel_i}{\log_2 i}$$

Normalized Discounted Cumulative Gain @N - Normalized discounted cumulative gain at a given number of documents retrieved

$$N \frac{\sum_i (2^{r(i)} - 1)}{\log(1 + i)}$$

Binary Preference - Measures the relevant documents retrieved when dealing with incomplete information

$$\frac{1}{R} \sum_r 1 - \frac{|nrankedgreaterthanr|}{R}$$

Reciprocal Rank - Returns the reciprocal of the rank of the first relevant document retrieved, or zero if no relevant documents were retrieved

$$\frac{1}{\text{rank of the first relevant document retrieved}}$$

Discovery Precision@N - The precision of the retrieval at a given number of documents retrieved

$$\frac{\# \text{ relevant documents retrieved}}{\# \text{ retrieved documents}}$$

Discovery Recall@N - The recall of the retrieval at a given number of documents retrieved

$$\frac{\# \text{ relevant documents retrieved}}{\# \text{ relevant documents in groundtruth}}$$

Number of Retrieved Documents@N - The number of retrieved documents at cutoff point

$$\# \text{ retrieved documents at cutoff point}$$

Precision at Relevant Retrieved Cutoff Point - The precision of the retrieval at a number of relevant documents retrieved

$$\frac{\# \text{ relevant documents retrieved}}{\# \text{ retrieved documents}}$$

1.5.4 Quality Measures

Average Number Retrieved Documents - The average number of retrieved documents for all tests

$$\frac{\sum_n \text{average number of retrieved documents of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Relevant Documents Retrieved - The average number of retrieved documents that were judged relevant

$$\frac{\sum_n \text{number of retrieved relevant documents of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Discovery Precision - The average precision of semantic web service discovery

$$\frac{\sum_n \text{discovery precision of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average NDCG@N - The average NDCG@N for all tests

$$\frac{\sum_n \text{NDCG@N of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average NDCG - The average NDCG for all tests

$$\frac{\sum_n \text{NDCG of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Binary Preference - The average binary preference for all tests

$$\frac{\sum_n \text{binary preference of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Reciprocal Rank - The average reciprocal rank for all tests

$$\frac{\sum_n \text{reciprocal rank of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Discovery Precision@N - The average Precision@N for all tests

$$\frac{\sum_n \text{discovery precision@N of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Discovery Recall@N - The average Recall@N for all tests

$$\frac{\sum_n \text{discovery recall@N of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Number of Retrieved Documents@N - The average number of retrieved documents at cutoff point

$$\frac{\sum_n \text{number of retrieved documents @N of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Precision at Relevant Retrieved Cutoff Point - The average precision of the retrieval at a number of relevant documents retrieved

$$\frac{\sum_n \text{precision at relevant retrieved cutoff point of the } n^{th} \text{ test}}{\# \text{ tests}}$$

Average Load Time - The average amount of time needed to load the service

$$\frac{\sum_n \text{load time of the } n^{th} \text{ test}}{\# \text{ tests}}$$

1.5.5 ISO 9126 Quality Characteristics

Functional Suitability - The degree to which the software product provides functions that meet stated and implied needs when the software is used under specified conditions.

Accuracy - The degree to which the software product provides the right or specified results with the needed degree of precision. *Accuracy* is a sub-characteristic of *Functional Suitability*

Performance Efficiency - The degree to which the software product provides appropriate performance, relative to the amount of resources used, under stated conditions.

Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when performing its function,

under stated conditions. *Time behaviour* is a sub-characteristic of *Performance Efficiency*

1.5.6 Semantic Quality Characteristics

Ontology Processing Time Behaviour - The degree to which the software product provides appropriate response and processing times and throughput rates when working with ontologies. *Ontology Processing Time Behaviour* is a sub-characteristic of *Time Behaviour* and can be measured using:

- Average Load time

Semantic Web Service Discovery Accuracy - The accuracy of the process of finding services that can be used to fulfill a given requirement from the service requester. *Semantic web service discovery accuracy* is a sub-characteristic of *Accuracy* and can be measured using:

- Average Number Retrieved Documents
- Average Relevant Documents Retrieved
- Average Precision
- Average NDCG@N
- Average NDCG
- Average Binary Preference
- Average Reciprocal Rank
- Average Precision@N
- Average Recall@N
- Average Number of Retrieved Documents@N
- Average Precision at Relevant Retrieved Cutoff Point

Figure 1.20 shows the part of the quality model obtained from the evaluation campaign for semantic web services.

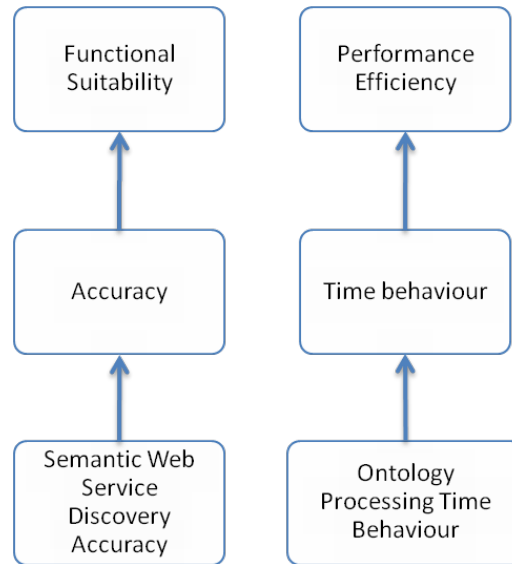


Figure 1.20: Quality characteristics of semantic web service tools.

1.6 Complete Overview of the Quality Model

Fig. 1.21 shows a complete overview of the quality model for semantic technologies that was obtained after analyzing all the evaluation scenarios and results from the SEALS evaluation campaigns.

In the final version of the model, we have also taken into account the analysis presented in the state of the art (Section ??). Two characteristics that were not obtained from the SEALS evaluation campaigns appear in the analysis, and were also included in the model. Those are:

- *Semantic web service time behaviour*. The capability of the software product to provide appropriate response and processing times when performing semantic web service discovery tasks.
- *Matching time behaviour*. The capability of the software product to provide appropriate response and processing times when performing matching tasks.

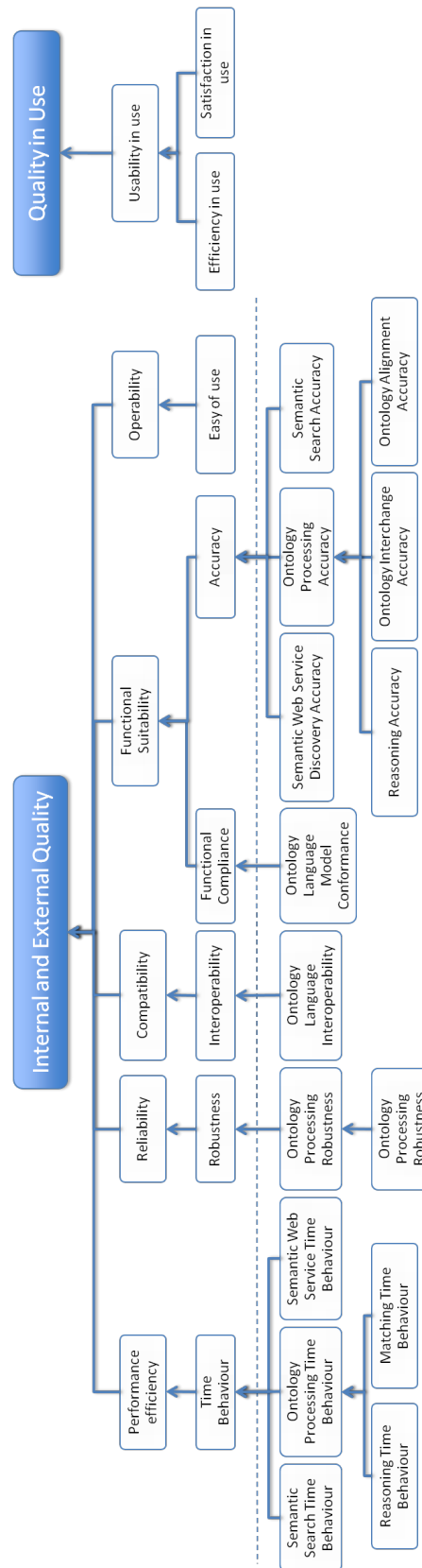


Figure 1.21: External and internal quality characteristics of semantic technologies.

Bibliography