Reusing Domain Ontologies on the Basis of Decision Analysis

Mari Carmen Suárez-Figueroa

Asunción Gómez-Pérez

Mariano Fernández-López

Antonio Jiménez

Alfonso Mateos

Work distributed under the license Creative Commons Attribution-Noncommercial-Share Alike 3.0

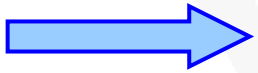
Date: 19/04/2012

Index

- Motivation
- State of the Art
  - Ontology Reuse Guidelines
  - Decision Analysis
- Ontology Reuse on the basis of Decision Analysis
- Conclusion
- Future Work

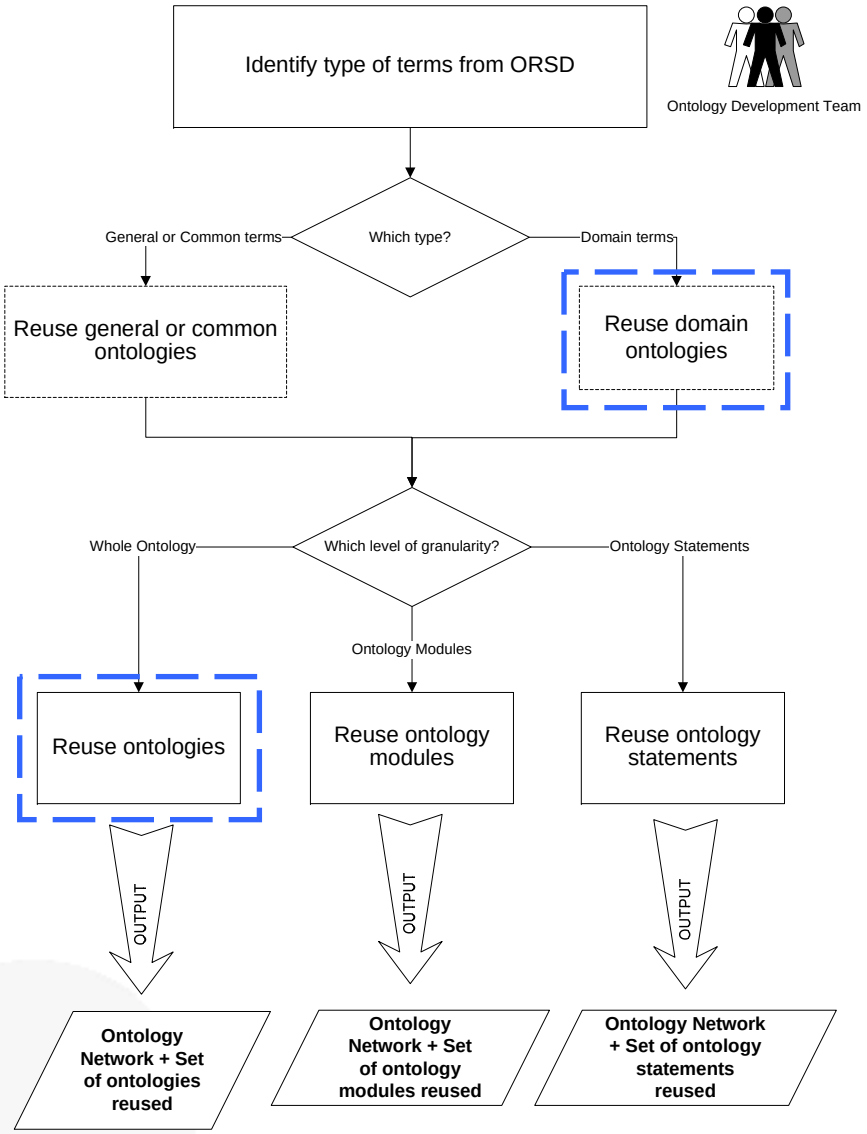
Motivation

- With the goal of speeding up the ontology development process, ontology practitioners are **reusing**
  - other ontologies [Simperl, 2009]
  - ontology modules [Cuenca-Grau et al., 2007]
  - ontology design patterns [Gangemi, 2007]
  - non-ontological resources [Jimeno-Yepes et al., 2009]
- **Different approaches for reusing ontologies** have been proposed [Gómez-Pérez et al., 2003; Uschold et al., 1998; Pinto and Martins, 2001; Paslaru and Mochol, 2005; Suárez-Figueroa, 2010]
- The selection of the most suitable domain ontologies for reuse in the development of a new ontology is a **complex decision-making problem**



The *selection of ontologies for reuse* in the development of a new ontology can be performed applying **the decision analysis methodology**.

Ontology Reuse Guidelines in the NeOn Methodology



Suárez-Figueroa, Mari Carmen (2010). **NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse**. PhD Thesis. UPM. 2010.

Domain Ontology Reuse Guidelines (I)

ORSD

INPUT

Ontology Development Team

Activity 1. Domain Ontology Search

Ontology Development Team

Activity 2. Domain Ontology Assessment

Ontology Development Team

Activity 3. Domain Ontology Select

OUTPUT

Set of selected domain ontologies

Ontology Development Team

Activity 4. Domain Ontology Integration

OUTPUT

Ontology network (interconnected ontologies + associated metadata) + set of domain ontologies

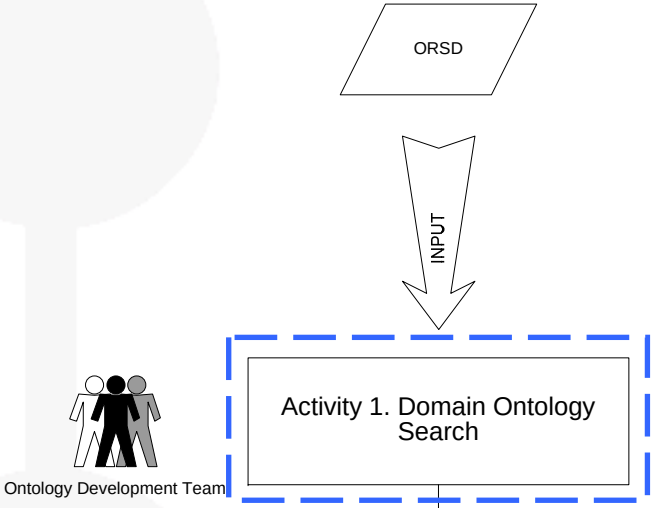
Domain Ontology Reuse Guidelines (II)

To search in libraries, repositories, and registries for candidate domain ontologies

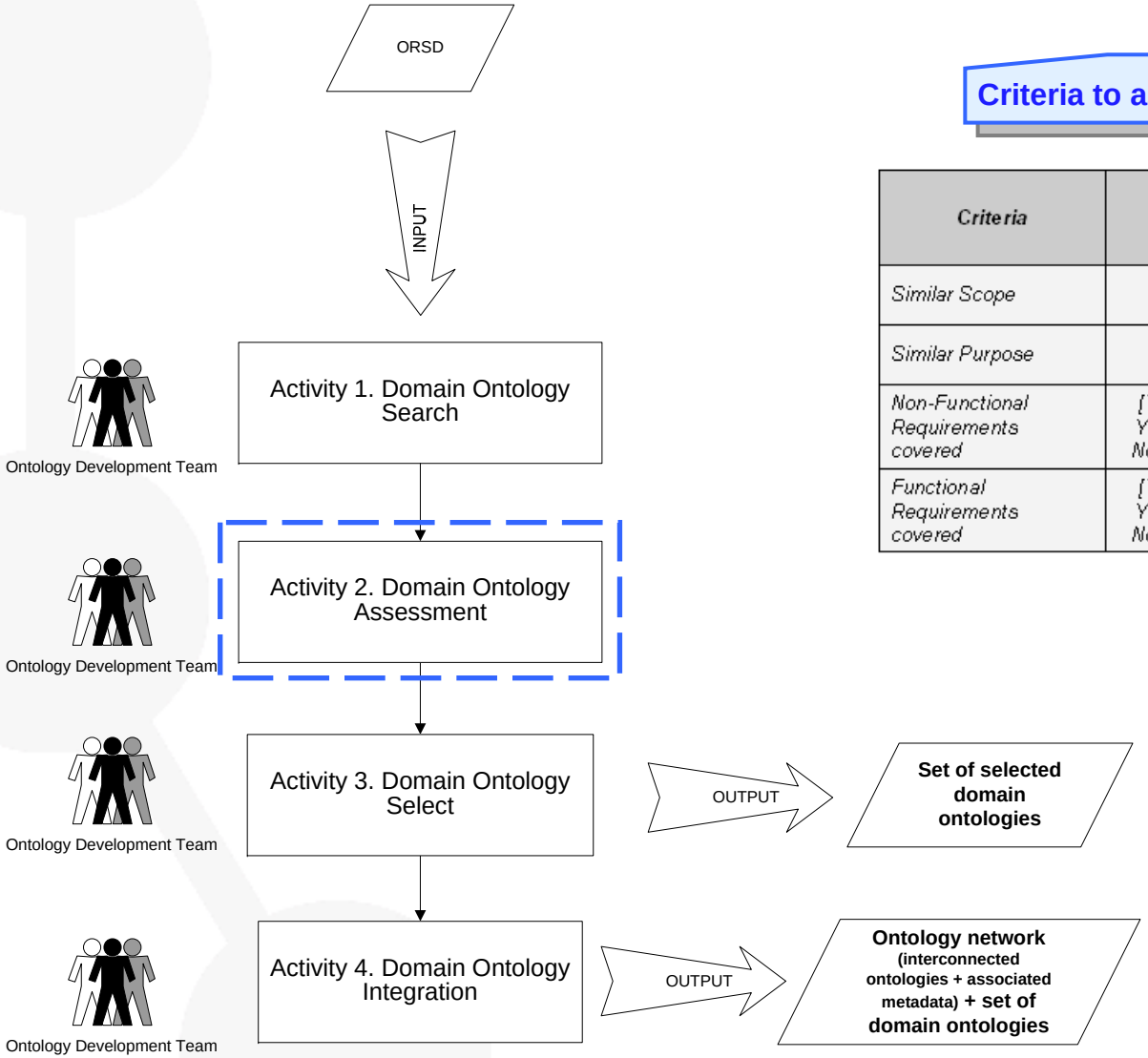
watson exploring the semantic web

Swoogle semantic web search 2007

Linked Open Vocabularies



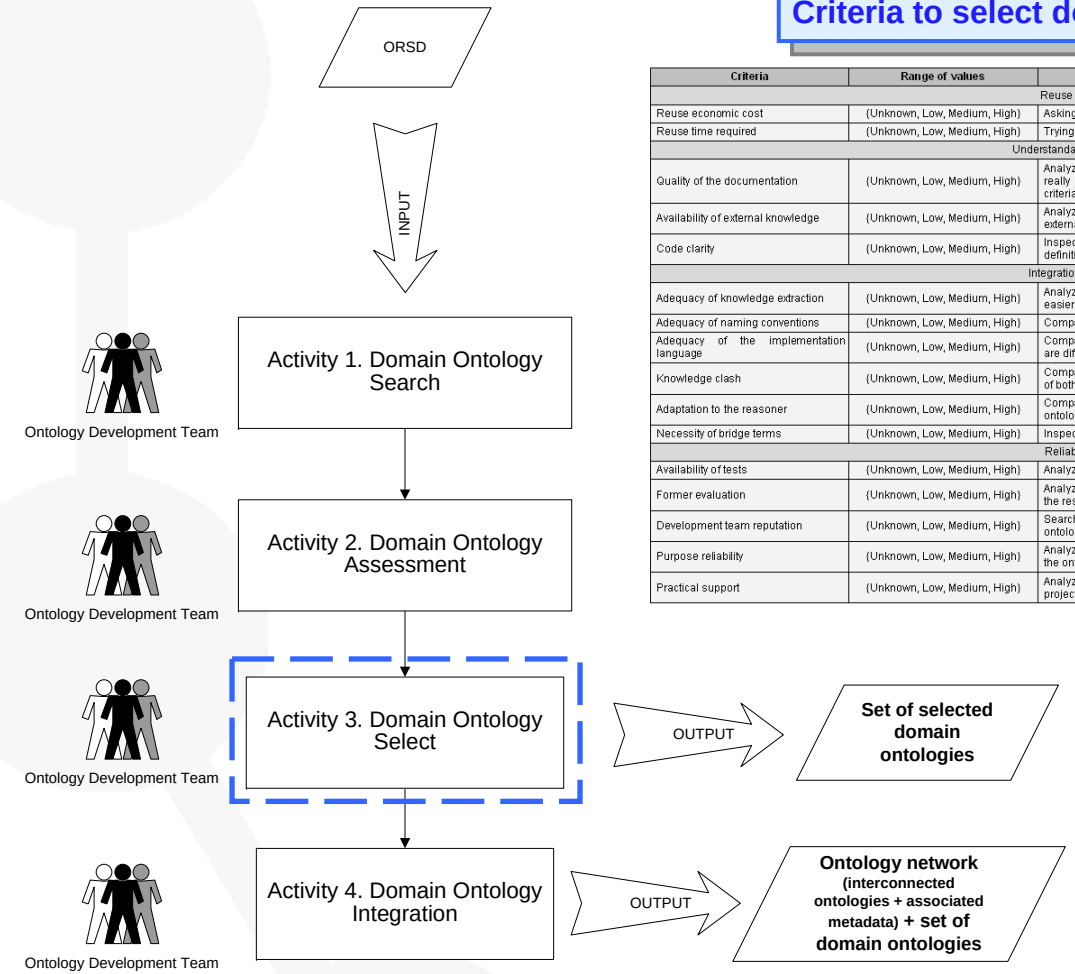
Domain Ontology Reuse Guidelines (III)



Criteria to assess candidate domain ontologies

Criteria	Range of Values	Candidate Domain Ontologies		
		Ontology 1	Ontology 2	Ontology 3
Similar Scope	[Yes, No, Unknown]			
Similar Purpose	[Yes, No, Unknown]			
Non-Functional Requirements covered	[Yes-Totally, Yes-Partially, No, Unknown]			
Functional Requirements covered	[Yes-Totally, Yes-Partially, No, Unknown]			

Domain Ontology Reuse Guidelines (IV)



Criteria to select domain ontologies

Criteria	Range of values	How to measure it	Weight	
Reuse cost				
Reuse economic cost	{Unknown, Low, Medium, High}	Asking the owner for an estimate.	(-)	9
Reuse time required	{Unknown, Low, Medium, High}	Trying the connection to the server.	(-)	7
Understandability effort				
Quality of the documentation	{Unknown, Low, Medium, High}	Analyzing if the ontology has documentation, and if such documentation really explains the domain and the ontology itself, as well as modelling criteria using during the ontology development.	(+)	8
Availability of external knowledge	{Unknown, Low, Medium, High}	Analyzing if in the ontology documentation there is any reference to external sources that could be used to better understand the ontology.	(+)	7
Code clarity	{Unknown, Low, Medium, High}	Inspecting the ontology code by analyzing the complexity of the definitions (and axioms) implemented the ontology.	(+)	8
Integration effort				
Adequacy of knowledge extraction	{Unknown, Low, Medium, High}	Analyzing if the ontology is modularized, or if it can be modularized in an easier way.	(+)	9
Adequacy of naming conventions	{Unknown, Low, Medium, High}	Comparing the naming conventions of both ontologies.	(+)	5
Adequacy of the implementation language	{Unknown, Low, Medium, High}	Comparing the ontology language of both ontologies. If both languages are different, analyzing the loss of knowledge in the translation.	(+)	7
Knowledge clash	{Unknown, Low, Medium, High}	Comparing modelling decisions and knowledge representation decision of both ontologies.	(-)	7
Adaptation to the reasoner	{Unknown, Low, Medium, High}	Comparing the reasoners related to the ontology language of both ontologies.	(+)	7
Necessity of bridge terms	{Unknown, Low, Medium, High}	Inspecting the ontology code.	(-)	6
Reliability				
Availability of tests	{Unknown, Low, Medium, High}	Analyzing if the ontology documentation refers to existing unit tests.	(+)	8
Former evaluation	{Unknown, Low, Medium, High}	Analyzing if the ontology documentation refers to existing unit tests and the results of such tests.	(+)	8
Development team reputation	{Unknown, Low, Medium, High}	Searching for information about the ontology development team (other ontologies developed, papers published, etc.).	(+)	8
Purpose reliability	{Unknown, Low, Medium, High}	Analyzing if the ontology documentation refers to the purpose for which the ontology was developed.	(+)	3
Practical support	{Unknown, Low, Medium, High}	Analyzing if the ontology documentation refers to other ontologies and/or projects reusing the ontology.	(+)	7

16 criteria organized in 4 dimensions

Method to calculate the score of each candidate domain ontology

1. To transform the different values into the same scale [0,3]
2. To calculate the score of the different candidate domain ontologies

$$Score_{i(+)} = \sum_{j(+)} Value_{T_{i,j}} \times \frac{Weight_j}{\sum_j Weight_j} \qquad Score_{i(-)} = \sum_{j(-)} Value_{T_{i,j}} \times \frac{Weight_j}{\sum_j Weight_j}$$

$$Score_i = Score_{i,(+)} - Score_{i,(-)}$$

Decision Analysis (I)

- **Decision analysis** (DA) is aimed at structuring and simplifying the task of making hard decisions
- DA is based on the assumption that the alternatives will appeal to experts depending on
  - how likely the possible performances of each alternative are
    - Existing information, collected data, models and professional judgments are used to quantify the likelihoods of a range of performances
  - what preferences experts have for the possible performances
    - Utility theory is used to quantify preferences
- **DA Methodology** can be divided into four steps:
  1. structure the problem;
  2. identify feasible strategies, their impact and uncertainty;
  3. quantify preferences; and
  4. evaluate alternatives and analyze sensitivity.

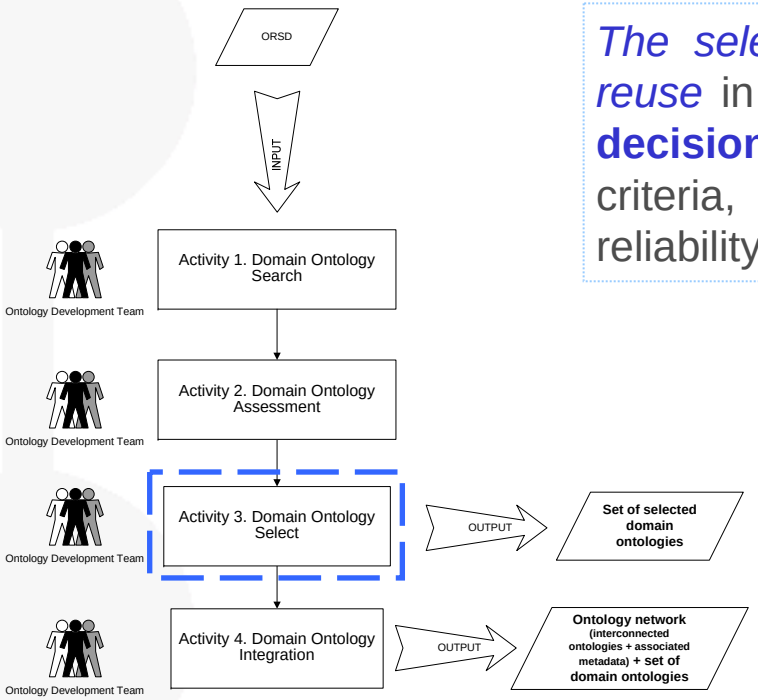
Decision Analysis (II)

- The **Generic Multi-Attribute Analysis System (GMAA)** is a decision support system that is intended to allay the operational difficulties involved in the DA methodology.
- **GMAA**
  - is based on a multi-attribute additive utility model
  - accounts for
    - uncertainty about alternative performances
    - incomplete information about decision-maker (DM) preferences

http://www.dia.fi.upm.es/~ajimenez/GMAA

- A. Jiménez, S. Ríos-Insua and A. Mateos. **A decision support system for multiattribute utility evaluation based on imprecise assignments.** Decis Sup Syst 36(2003) 65-79.
- A. Jiménez, S. Ríos-Insua and A. Mateos. **A generic multi-attribute analysis system.** Comput Oper Res 33(2006) 1081-1101.

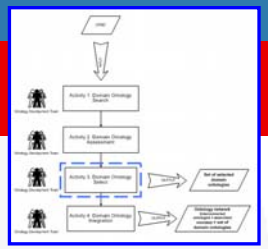
Ontology Reuse on the basis of Decision Analysis (I)



The selection of the most suitable domain ontologies for reuse in the development of a new ontology is a **complex decision-making problem** where different conflicting criteria, like understandability effort, integration effort, and reliability, have to be taken into account simultaneously.

DA Methodology Steps:

1. structure the problem;
2. identify feasible strategies, their impact and uncertainty;
3. quantify preferences; and
4. evaluate alternatives and analyze sensitivity



Ontology Reuse on the basis of Decision Analysis (II)

1. Structuring the problem

Several conflicting criteria have to be taken into account simultaneously *when selecting domain ontologies for reuse*

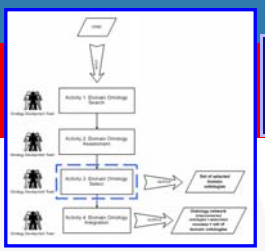
16 criteria organized in 4 dimensions

Criteria	Range of values	How to measure it	Weight	
Reuse cost				
Reuse economic cost	{Unknown, Low, Medium, High}	Asking the owner for an estimate.	(-)	9
Reuse time required	{Unknown, Low, Medium, High}	Trying the connection to the server.	(-)	7
Understandability effort				
Quality of the documentation	{Unknown, Low, Medium, High}	Analyzing if the ontology has documentation, and if such documentation really explains the domain and the ontology itself, as well as modelling criteria using during the ontology development.	(+)	8
Availability of external knowledge	{Unknown, Low, Medium, High}	Analyzing if in the ontology documentation there is any reference to external sources that could be used to better understand the ontology.	(+)	7
Code clarity	{Unknown, Low, Medium, High}	Inspecting the ontology code by analyzing the complexity of the definitions (and axioms) implemented the ontology.	(+)	8
Integration effort				
Adequacy of knowledge extraction	{Unknown, Low, Medium, High}	Analyzing if the ontology is modularized, or if it can be modularized in an easier way.	(+)	9
Adequacy of naming conventions	{Unknown, Low, Medium, High}	Comparing the naming conventions of both ontologies.	(+)	5
Adequacy of the implementation language	{Unknown, Low, Medium, High}	Comparing the ontology language of both ontologies. If both languages are different, analyzing the loss of knowledge in the translation.	(+)	7
Knowledge clash	{Unknown, Low, Medium, High}	Comparing modelling decisions and knowledge representation decision of both ontologies.	(-)	7
Adaptation to the reasoner	{Unknown, Low, Medium, High}	Comparing the reasoners related to the ontology language of both ontologies.	(+)	7
Necessity of bridge terms	{Unknown, Low, Medium, High}	Inspecting the ontology code.	(-)	6
Reliability				
Availability of tests	{Unknown, Low, Medium, High}	Analyzing if the ontology documentation refers to existing unit tests.	(+)	8
Former evaluation	{Unknown, Low, Medium, High}	Analyzing if the ontology documentation refers to existing unit tests and the results of such tests.	(+)	8
Development team reputation	{Unknown, Low, Medium, High}	Searching for information about the ontology development team (other ontologies developed, papers published, etc.).	(+)	8
Purpose reliability	{Unknown, Low, Medium, High}	Analyzing if the ontology documentation refers to the purpose for which the ontology was developed.	(+)	3
Practical support	{Unknown, Low, Medium, High}	Analyzing if the ontology documentation refers to other ontologies and/or projects reusing the ontology.	(+)	7

Criteria = Objectives

Associated Attributes

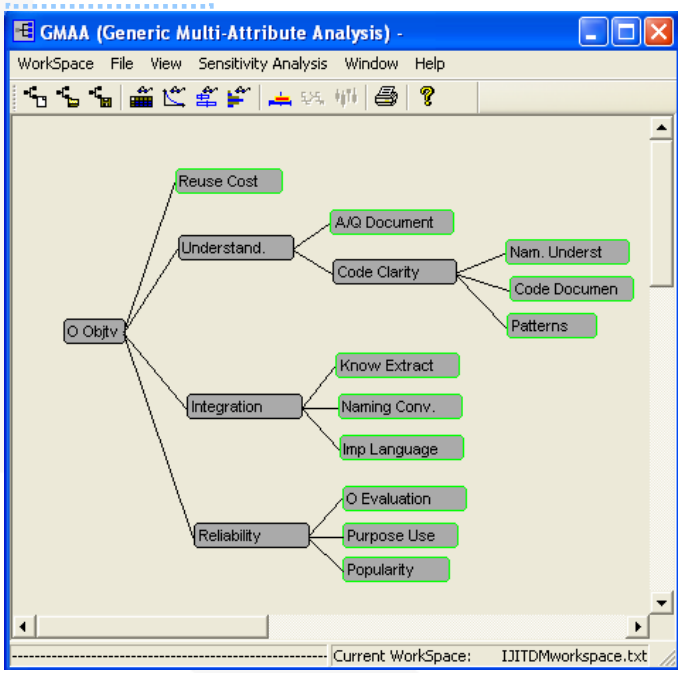
11 criteria and their attributes



Ontology Reuse on the basis of Decision Analysis (III)

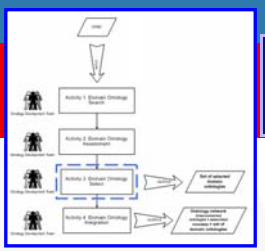
1. Structuring the problem

11 criteria



11 criteria and their objective attributes

Criteria		Associated Attributes
Reuse Cost		
Reuse Cost		Continuous Attribute
Understandability Effort		
Documentation Availability and Quality		Discrete Attribute
Code Clarity	Naming Understandability	Binary Attribute
	Code Documentation	Continuous Attribute
	Pattern Conformance	Discrete Attribute
Integration Effort		
Adequacy of Knowledge Extraction		Binary Attribute
Adequacy of Naming Conventions		Binary Attribute
Adequacy of Implementation Language		Discrete Attribute
Reliability		
Ontology Evaluation		Discrete Attribute
Purpose of Use		Discrete Attribute
Popularity		Discrete Attribute



Ontology Reuse on the basis of Decision Analysis (IV)

e.g.

1. Structuring the problem

Criteria		Associated Attributes
Reuse Cost		
Reuse Cost		Continuous Attribute
Understandability Effort		
Documentation Availability and Quality		Discrete Attribute
Code Clarity	Naming Understandability	Binary Attribute
	Code Documentation	Continuous Attribute
	Pattern Conformance	Discrete Attribute
Integration Effort		
Adequacy of Knowledge Extraction		Binary Attribute
Adequacy of Naming Conventions		Binary Attribute
Adequacy of Implementation Language		Discrete Attribute
Reliability		
Ontology Evaluation		Discrete Attribute
Purpose of Use		Discrete Attribute
Popularity		Discrete Attribute

Documentation availability and quality (A/Q Document) refers to whether there is any communicable material used to describe or explain different aspects of the candidate ontology, as well as the enacted development process.

A ***discrete attribute*** with three possible values has been established :

- low, when the candidate ontology has no documentation;
- medium, when the candidate ontology has documentation (in the form of web page, wiki, and/or paper) detailing the ontology; and
- high, when the candidate ontology has documentation (in the form of web page, wiki, and/or paper) detailing both the candidate ontology and its development process (ORSD, modeling decisions, etc.).

Code documentation (Code Documen) refers to whether the ontology code is documented.

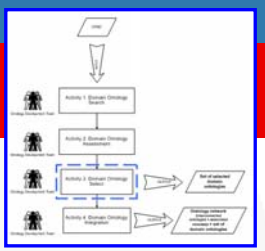
A ***continuous attribute*** ([0,100]) was used: $\frac{\#CommentedOntologyElements}{\#OntologyElements} \times 100,$

where #CommentedOntologyElements is the number of elements commented in the ontology (rated by searching the code for the "#rdfs:comment" string) and #OntologyElements is the number of elements in the ontology (#OntologyElements = 1+ #Classes + #ObjectProperties + #DatatypeProperties + #Instances).

Adequacy of naming conventions (Naming Conv) refers to whether the two ontologies (the candidate ontology and the ontology under development) follow the same rules for naming the different ontology components.

A ***binary attribute***

- value 1 when the candidate ontology and the ontology under development follow the same naming conventions
- value 0 when they do not.

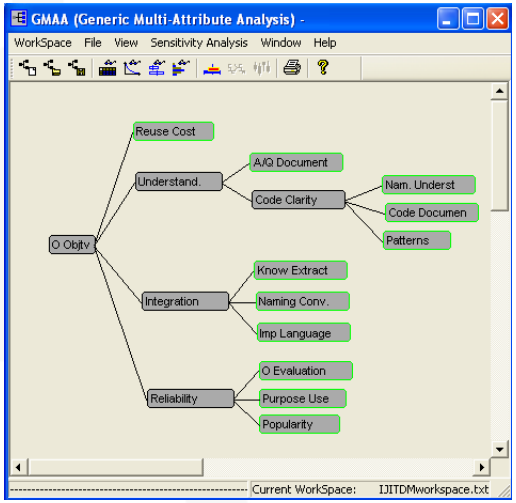


Ontology Reuse on the basis of Decision Analysis (V)

2. Identifying feasible strategies, their impact and uncertainty

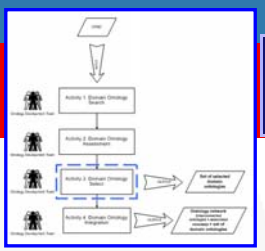
Feasible alternatives are those *candidate domain ontologies obtained in Activity 2* (Domain Ontology Assessment)

Alternative performances should be established in terms of attributes associated with the lowest level objectives in the hierarchy



Example(s)

	Athlete	*Baseball*	*OSO*	*OntoSem*	*Rissen*	*Soccer*	*Tsinaraki*	*SoccerV2*	*UNSPSC*
Document A/Q	Low	Medium	Low	Medium	Low	Low	Low	Medium	Medium
Nam. Underst	1	1	1	0	1	1	0	1	0
Code Documen	0	0	[35, 45]	0	[80, 90]	0	[65, 75]	0	0
Patterns	0	0	0	0	0	0	CP	0	0
Know Extract	0	0	0	0	0	0	1	0	0
Naming Conv.	1	1	0	0	0	0	1	0	0
Imp Language	High	High	High	High	High	Medium	High	Medium	Medium
O Evaluation	2	2	6	[0, 21]	2	0	0	0	[0, 21]
Purpose Rel	Low	Low	Low	High	Low	Low	Low	Medium	High
Popularity	Low	Medium	Low	Medium	Low	Medium	Low	Medium	Medium

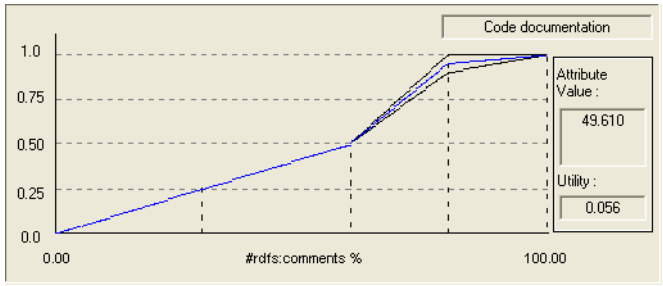


Ontology Reuse on the basis of Decision Analysis (VI)

3. Quantifying preferences

3.1. Quantifying preferences involves assessing *single attribute utilities* that represent DM preferences concerning the possible ontology performances.

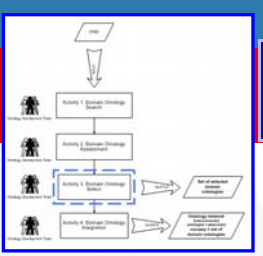
Utility functions for continuous attributes



Utility intervals for discrete attributes

Imprecise values

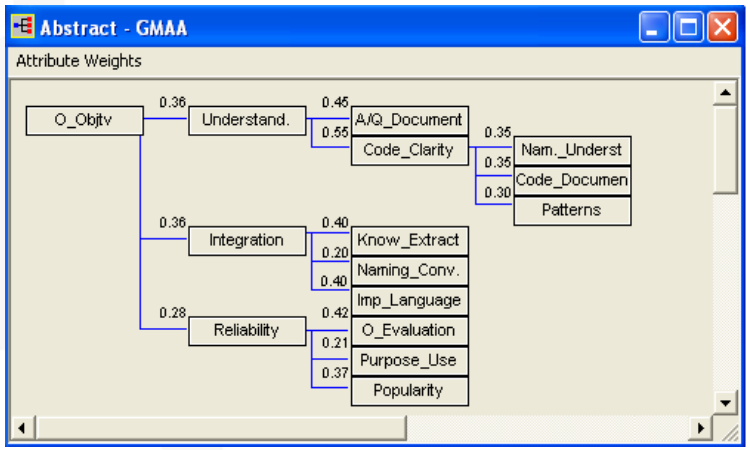
	Low	*Medium*	*High*	*Very High*	*0*	*1*
A/Q Document	[0, 0.2]	[0.3, 0.6]	[0.7, 1]	-	-	-
Nam. Underst	-	-	-	-	[0.0, 0.2]	[0.3, 1]
Know Extract	-	-	-	-	0	1
Naming Conv.	-	-	-	-	[0.0, 0.3]	[0.4, 1]
Imp Language	0	[0.4, 0.6]	1	-	-	-
Purpose of use	[0, 0.2]	[0.3, 0.6]	[0.7, 1]	-	-	-
Popularity	0	[0.3, 0.4]	[0.7, 0.8]	[0.9, 1]	-	-

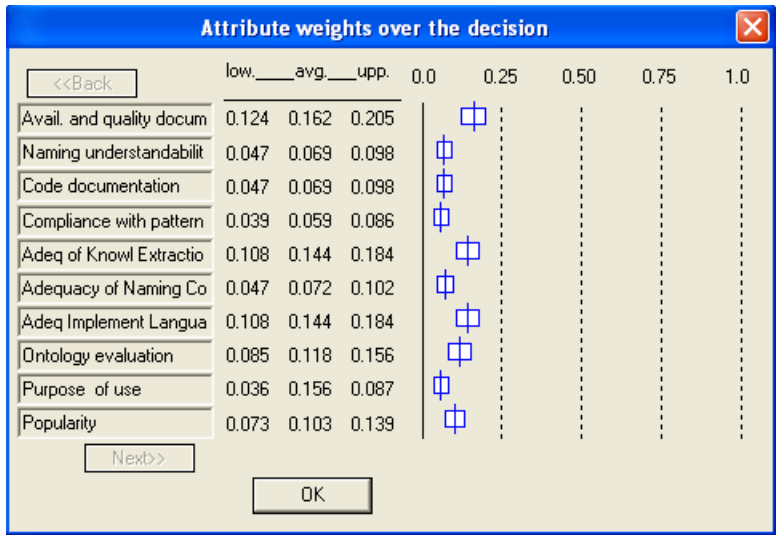


Ontology Reuse on the basis of Decision Analysis (VII)

3. Quantifying preferences

3.2. Quantifying preferences involves eliciting weights that represent the *relative importance of criteria*.





Documentation availability and quality is the most important criteria, followed by adequacy of knowledge extractions and adequacy of implementation language and ontology evaluation, respectively.

Ontology Reuse on the basis of Decision Analysis (VIII)

4. Evaluating alternatives

The *additive model* is

$$u(O_i) = \sum_{j=1}^{n} w_j u(x_{ij}),$$

where xij is the performance of ontology Oi for attribute Xj; uj(xij) is the utility associated with value xij, and wj are the weights of each attribute.

The *additive model* is used to assess average overall utilities and minimum and maximum overall utilities.

The *ranking of alternatives* is based on the average overall utilities

the minimum and maximum overall utilities provide further insight into the robustness of the ranking

Example(s)

Alternative Classification

Alternatives :	Overall Utilities Min:	Avg:	Max:	Rank:
Tsinaraki	0.2856	0.4516	0.6848	1
OntoSem	0.2317	0.4393	0.7428	2
Baseball	0.2469	0.4249	0.6823	3
UNSPSC	0.1666	0.3673	0.6690	4
Rissen	0.2111	0.3593	0.5714	5
OSO	0.1989	0.3381	0.5397	6
Athlete	0.1879	0.3322	0.5448	7
SoccerV2	0.1660	0.3228	0.5566	8
Soccer	0.1179	0.2454	0.4399	9

Conclusion

- **GMAA system**
  - used to apply the decision analysis (DA) methodology
  - has proved to be a useful tool for complex decision-making problems including different conflicting objectives (such as the **domain ontology selection for reuse**)
  - helps DMs to think about the problem in more depth, and accounts for imprecision in the inputs
  - makes the process less stressful for experts and suitable for group decision-making
- **Updated set of criteria** for ontology reuse has been created
  - **Importance of criteria** has been established by means of weights.
- **New attributes** for measuring the performances of the candidate domain ontologies for the different criteria have been identified
  - most are objective rather subjective attributes
  - easier for experts to measure

- This approach (*mixing ontology engineering and decision analysis*) has been used in  in the following domains
  - Multimedia
  - Sport
  - News

Future Work

- **Questionnaires** will be created and sent to ontology experts asking for
  - feedback about current criteria and attributes
  - feedback about utilities for each attribute
  - feedback about importance of each criteria

  } ***Preferences***

- **Criteria and associated attributes** for reusing domain ontologies will be updated

- **Global preferences** will be obtained by applying *aggregation processes*

Reusing Domain Ontologies on the Basis of Decision Analysis

Mari Carmen Suárez-Figueroa

Asunción Gómez-Pérez

Mariano Fernández-López

Antonio Jiménez

Alfonso Mateos

Work distributed under the license Creative Commons Attribution-Noncommercial-Share Alike 3.0

Date: 19/04/2012