NeOn-project.org

NeOn: Lifecycle Support for Networked Ontologies

Integrated Project (IST-2005-027595)

Priority: IST-2004-2.4.7 – “Semantic-based knowledge and content systems”

D5.6.1 Experimentation with the NeOn methodologies and methods

---

Deliverable Co-ordinator: Raúl García-Castro

Deliverable Co-ordinating Institution: Universidad Politécnica de Madrid (UPM)

Other Authors: Mari Carmen Suárez-Figueroa (UPM); Mauricio Espinoza (UPM); Margherita Sini (FAO); Holger Lewen (UKARL); Eva Blomqvist (CNR)

One of the goals of WP5 is to provide qualitative and quantitative evidence that using the NeOn methodologies and methods ontologies and systems are built faster and better. This deliverable includes a first description of the experiments to be performed over the NeOn methodologies and methods during the following years.

Document Identifier:	NEON/2008/D5.6.1/v1.1	Date due:	February 29, 2008
Class Deliverable:	NEON EU-IST-2005-027595	Submission date:	March 31, 2008
Project start date:	March 1, 2006	Version:	V1.1
Project duration:	4 years	State:	Final
		Distribution:	Public

NeOn Consortium

This document is a part of the NeOn research project funded by the IST Programme of the Commission of the European Communities by the grant number IST-2005-027595. The following partners are involved in the project:

Open University (OU) – Coordinator Knowledge Media Institute – KMi Berrill Building, Walton Hall Milton Keynes, MK7 6AA United Kingdom Contact person: Martin Dzbor, Enrico Motta E-mail address: {m.dzbor, e.motta} @open.ac.uk	**Universität Karlsruhe – TH (UKARL)** Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB Englerstrasse 28 D-76128 Karlsruhe, Germany Contact person: Peter Haase E-mail address: pha@aifb.uni-karlsruhe.de
Universidad Politécnica de Madrid (UPM) Campus de Montegancedo 28660 Boadilla del Monte Spain Contact person: Asunción Gómez Pérez E-mail address: asun@fi.upm.es	**Software AG (SAG)** Uhlandstrasse 12 64297 Darmstadt Germany Contact person: Walter Waterfeld E-mail address: walter.waterfeld@softwareag.com
Intelligent Software Components S.A. (ISOCO) Calle de Pedro de Valdivia 10 28006 Madrid Spain Contact person: Jesús Contreras E-mail address: jcontreras@isoco.com	**Institut 'Jožef Stefan' (JSI)** Jamova 39 SI-1000 Ljubljana Slovenia Contact person: Marko Grobelnik E-mail address: marko.grobelnik@ijs.si
Institut National de Recherche en Informatique et en Automatique (INRIA) ZIRST – 655 avenue de l'Europe Montbonnot Saint Martin 38334 Saint-Ismier France Contact person: Jérôme Euzenat E-mail address: jerome.euzenat@inrialpes.fr	**University of Sheffield (USFD)** Dept. of Computer Science Regent Court 211 Portobello street S14DP Sheffield United Kingdom Contact person: Hamish Cunningham E-mail address: hamish@dcs.shef.ac.uk
Universität Koblenz-Landau (UKO-LD) Universitätsstrasse 1 56070 Koblenz Germany Contact person: Steffen Staab E-mail address: staab@uni-koblenz.de	**Consiglio Nazionale delle Ricerche (CNR)** Institute of cognitive sciences and technologies Via S. Martino della Battaglia, 44 - 00185 Roma-Lazio, Italy Contact person: Aldo Gangemi E-mail address: aldo.gangemi@istc.cnr.it
Ontoprise GmbH. (ONTO) Amalienbadstr. 36 (Raumfabrik 29) 76227 Karlsruhe Germany Contact person: Jürgen Angele E-mail address: angele@ontoprise.de	**Food and Agriculture Organization of the United Nations (FAO)** Viale delle Terme di Caracalla 1 00100 Rome Italy Contact person: Marta Iglesias E-mail address: marta.iglesias@fao.org
Atos Origin S.A. (ATOS) Calle de Albarracín, 25 28037 Madrid Spain Contact person: Tomás Pariente Lobo E-mail address: tomas.parientelobo@atosorigin.com	**Laboratorios KIN, S.A. (KIN)** C/Ciudad de Granada, 123 08018 Barcelona Spain Contact person: Antonio López E-mail address: alopez@kin.es

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed to the writing of this document or its parts:

UPM

FAO

UKARL

CNR

Change Log

Version	Date	Amended by	Changes
0.1	16-01-2008	Raúl García Castro	First draft of the document
0.2	25-01-2008	Raúl García Castro	Included experiment template
0.3	29-01-2008	Raúl García Castro	Included contributions from Holger Lewen, Margherita Sini and Eva Blomqvist
0.4	31-01-2008	Raúl García Castro	Updated reengineering and mapping experiments from Margherita Sini
0.5	01-02-2008	Raúl García Castro	Included ontology localization experiments from Mauricio Espinoza
0.6	05-02-2008	Raúl García Castro	Updated ontology design patterns experiments from Mari Carmen Suárez-Figueroa
0.7	09-02-2008	Raúl García Castro	Included ontology design pattern experiment results from Mari Carmen Suárez-Figueroa
0.8	15-02-2008	Raúl García Castro	Updated the lifecycle and specification experiments from Mari Carmen Suárez-Figueroa
0.9	14-03-2008	Raúl García Castro	Implemented some comments from Asunción Gómez-Pérez
0.10	27-03-2008	Raúl García Castro, Mari Carmen Suárez-Figueroa, Holger Lewen, Margherita Sini	Updated with the reviewer comments (Igor Mozetic)
1.0	03-04-2008	Valentina Presutti and Raúl García Castro	Updated with more reviewer comments
1.1	04-04-2008	Raúl García Castro	Final version with QA comments

Executive Summary

One of the goals of WP5 is to provide qualitative and quantitative evidence that using the NeOn methodologies and methods, ontologies and systems are built faster and better. This deliverable includes a first description of the experiments to be performed over the NeOn methodologies and methods during the following years.

This deliverable includes the planning of the experiments on:

- Ontology design patterns
- Ontology lifecycle
- Ontology reuse
- Ontology reengineering
- Ontology mapping
- Ontology localization
- Ontology specification

It also includes the results of some preliminary experiments on ontology design patterns.

Table of Contents

1. Introduction 7

2. Method followed to perform the experiments 8

3. Experiments on ontology design patterns 10

3.1. Experiment 1 – Experiment in the Artificial Intelligence course 11

3.2. Experiment 2 – Experiment in a Summer School 15

3.3. Experiment 3 – Problems solved by ontology patterns 20

3.4. Experiment 4 – Effects of content patterns on ontology design 23

3.5. Experiment 5 – Effects of content patterns on ontologies 27

3.6. Experiment 6 – Effects of pattern presentation on pattern selection 31

4. Experiment on the ontology network lifecycle 35

5. Experiment on ontology reuse 37

6. Experiment on ontology reengineering 40

6.1. Experiment 1 - From the AGROVOC and the ASFA thesauri to concept based models 40

6.2. Experiment 2 - From the current AGROVOC ontological model to the NeOn model (including the Lexical Informations - LIR) 44

7. Experiment on ontology mapping 49

8. Experiment on ontology localization 53

9. Experiment on ontology specification 56

10. Conclusions 58

11. References 59

APPENDIX I. Material for ontology design pattern experiment 1 60

APPENDIX II. Results of the ontology design pattern experiment 1 63

APPENDIX III. Material for ontology design pattern experiment 2 75

III.1. First experiment 75

III.2. First experiment with solutions 75

III.3. Second experiment 76

III.4. Second experiment with solutions 77

APPENDIX IV. Questionnaire for Experiments about Ontology Network Life Cycle 78

List of tables

Table 1. Experiments to be performed 7
Table 2. Actors in the experiments 44
Table 3. Steps and groups of people in the experiment 46
Table 4. Steps for each case 47
Table 5. Ontologies corpus statistics 55
Table 6. Results obtained for each metric 55

List of figures

Figure 1. General Results 15
Figure 2. General Results 63
Figure 3. General Results for Modelling Problem 1 63
Figure 4. Non Correct Pattern for Modelling Problem 1 64
Figure 5. General Results for Modelling Problem 2 64
Figure 6. Non Correct Pattern for Modelling Problem 2 65
Figure 7. General Results for Modelling Problem 3 65
Figure 8. Non Correct Pattern for Modelling Problem 3 66
Figure 9. General Results for Modelling Problem 4 66
Figure 10. Non Correct Pattern for Modelling Problem 4 67
Figure 11. General Results for Modelling Problem 5 67
Figure 12. Non Correct Pattern for Modelling Problem 5 68
Figure 13. General Results for Modelling Problem 6 68
Figure 14. General Results for Modelling Problem 7 69
Figure 15. Non Correct Pattern for Modelling Problem 7 69
Figure 16. General Results for Modelling Problem 8 70
Figure 17. Non Correct Pattern for Modelling Problem 8 70
Figure 18. General Results for Modelling Problem 9 71
Figure 19. Non Correct Pattern for Modelling Problem 9 71
Figure 20. General Results for Modelling Problem 10 72
Figure 21. Non Correct Pattern for Modelling Problem 10 72
Figure 22. General Results for Modelling Problem 11 73
Figure 23. General Results for Modelling Problem 12 73
Figure 24. General Results for Modelling Problem 13 74

1. Introduction

In the NeOn project, one methodology and different methods are being defined to support the different activities of the networked ontology lifecycle. One of the goals of WP5 is to provide qualitative and quantitative evidence that using the NeOn methodologies and methods, ontologies and systems are built faster and better.

Each of the experiments that will be performed in NeOn is composed of three consecutive phases:

- *Plan* phase, which includes the definition and design of the experiment.
- *Experiment* phase, which includes running the experiment.
- *Analysis* phase, which includes analysing the experiment results.

The experiments over the NeOn methodologies will take place in different moments of time. Table 1 shows the different experiments that will be performed in the NeOn project and the progress achieved in each of them. This table shows that there can be experiments of different types and, although this table shows 13 different experiments, further experiments will be performed in the NeOn project whenever new methods are defined.

		Plan	**Experiment**	**Analysis**
Ontology design patterns	E1	X	X	X
	E2	X	X	X
	E3	X		
	E4	X		
	E5	X		
	E6	X		
Ontology lifecycle	E1	X		
Ontology reuse	E1	X		
Ontology reengineering	E1	X		
	E2	X		
Ontology mapping	E1	X		
Ontology localization	E1	X		
Ontology specification	E1	X		

Table 1. Experiments to be performed

This deliverable includes a first description of the *Plan* phase of the experiments to be done during the following years. Regarding the *Experiment* and *Analysis* phases, at this moment there are only results for 2 experiments related to the use of ontology design patterns.

The rest of experiments will be executed in 2008 and future deliverables will include the description of how these experiments are executed and the analysis of their results.

2. Method followed to perform the experiments

As we mentioned before, each experiment is composed of three consecutive phases, following the phases considered in most software experimentation approaches [5, 6, 7]:

- *Plan* phase, which includes the definition and design of the experiment.
- *Experiment* phase, which includes running the experiment.
- *Analysis* phase, which includes analysing the experiment results.

In order to collect the descriptions of the *Plan* phase of the experiments, we defined a template to unify the format of the contributions and to ensure that no relevant topic about the experimentation was left out.

The template provided for defining the *Plan* phase of the experiments was available in the NeOn wiki[1] and is the following:

- **Experiment Definition:**
  - o Define experiment motivation. There is always a motivation either at personal or at organizational level for making experiments. Questions to be answered could be:
    - ▪ Why are you performing this experiment?
    - ▪ What do you want to know or learn?
  - o Define experiments goals. Which are the goals for the experiment?
    - ▪ Identify the beneficiaries of the experiments. The results of the experiment will benefit to some person, organization, community, etc. So, who is going to benefit from this experiment and in which way?
    - ▪ Identify experiment subject. Experiments must have a subject under study. So, which are the methods, techniques, tools, etc. to be studied in the experiment?
- **Experiment Design:**
  - o Identify the relevant characteristics. The subject selected for experimentation has different characteristics, but for the experiment only some of them are relevant. So, which attributes of the subject are going to be studied?
  - o Define the metrics and criteria. Metrics are the data that is needed to describe some attribute of the subject. Also, we need to know the criteria used for interpreting this data. Questions to be answered should be:
    - ▪ Which metrics will be used to measure these characteristics?
    - ▪ Which criteria will be used to interpret these measurements?
  - o Identify the variables (time, resources, etc.) that affect the measurement. Questions to be answered should be:
    - ▪ Which variables can affect the measurement and can be controlled?
    - ▪ Which variables can affect the measurement but cannot be controlled?
  - o Define the data collection process. Once we know which data we need, we have to define the procedure to follow for collecting this data (that is, which is the concrete

[1] http://www.neon-project.org/wiki/index.php?title=WP5D5.6.1

experiment to be carried out and how such experiment has to be done) and how this data will be stored. Questions to be answered could be:

- Which procedures will be followed to collect the measurements?
- How will the results be stored and documented?

o Define the requirements. Each experiment to perform has some requirements to carry it out. Questions could be the following:

- How many times each elementary experiment is to be repeated?
- How many people are needed to run the experiments?

o Define the analysis procedure. Once we have the data, we have to analyse it according to some procedure. Questions to be answered should be:

- How will the data be validated?
- How will the data be analysed?
- Will you compare the results of the experiment against a baseline?

o Define the experimentation plan. The experimentation plan includes the time and resources that will be used. Questions to be answered should be:

- Who will perform the experiments? (that is, who will carry out the experiment phase)
- When will the experiments be performed? (that is, when will the experiment phase carry out)
- Who will analyse the experiment results? (that is, who will carry out the analysis phase)
- When will the experiment results be analysed? (that is, when will the analysis phase carry out)

3. Experiments on ontology design patterns

The notion of ontology patterns (also called modelling components) has been described in detail in NeOn deliverable D5.1.1 [2]. Our concern here is to describe a set of experiments that can show the effects and benefits of using patterns in ontology engineering in a scientifically rigorous manner. There are several different aspects of patterns that need to be studied and several types of effects of pattern usage that needs to be defined and measured. So far no indisputable evidence has been put forward to support the benefits of using patterns. Only informal evidence such as subjective experience reports of ontology engineers can so far support the usefulness of ontology patterns. Logical and content patterns have for example been used for teaching, in tutorials and summer schools, and it has according to the tutors improved the way the students solve the tasks. Also, patterns have been used by ontology engineers when constructing and evaluating ontologies, but also in these cases it is only the personal subjective experience of the ontology engineers involved that suggest that patterns are indeed useful and provide benefits.

Patterns (when discussed generally, and specifically for ontology engineering) are commonly suggested to give three kinds of benefits: reuse benefits, guidance benefits and communication benefits. Reuse is here concerned with constructing “better” ontologies due to the use of patterns, guidance is referring to the assistance and learning aspects of using patterns and finally communication is concerned with patterns as a tool for describing existing ontologies, situations and problems with ontologies. The proposed set of experiments below intend to in the long term address all of these issues for the different kinds of patterns, but as a first set of experiments only a few have been selected focusing mainly on showing that some specific kinds of patterns do have both reuse and guidance benefits.

Initially two preliminary experiments involving students, at UPM and at a summer school on ontologies, were designed and conducted during 2007. These two experiments and the analysis of their results are described in sections 3.1 and 3.2. The experiments gave some very important initial results and insights on how subjects really perceive, understand, and use ontology design patterns. Additionally, these experiments were an excellent opportunity to test the settings and discover good ways to experiment with ontology design patterns. In this sense the experiments have also been used as a starting point, providing valuable experiences for designing a more complete set of experiments on ontology design patterns.

- The goal of the first experiment at UPM is to test if, given a modelling problem expressed in natural language, people modelling ontologies are able to identify the correct pattern for such problem.

- The goal of the second experiment at a summer school on ontologies is to test if a subset of logical and content design patterns described in NeOn deliverable D5.1.1 [2] is well explained and if such patterns are easy to understand and to apply in particular modelling problems.

Further proposed experiments that still need to be performed are the following, divided in two categories concerning the motivation of patterns and the actual usage of pattern-based ontology design:

- Motivation and characteristics of pattern-based design
    - Problems solved by ontology patterns (1 exploratory study)
    - The effects of patterns on the ontology development process (2 controlled experiments)
    - The effects of patterns on constructed ontologies (4 controlled experiments)
- Requirements and characteristics of support for pattern-based ontology design
    - Requirements on ontology construction environment for pattern-based design (1 exploratory study)
    - The effects of pattern representation and presentation on manual pattern selection (1 controlled experiment)

- o The effects of pattern abstraction level on manual selection and automatic usage (2 controlled experiments)
- o Requirements on pattern repository (1 exploratory study)
- o Pattern extraction (1 controlled experiment)

To motivate the use of patterns the first further experiment (an exploratory study) intends to show that patterns actually exist in real-world ontologies and that patterns solve common problems. This would indicate that reuse is possible and thereby motivate and support the general notion of ontology patterns. Next the guidance benefit would be studied through two controlled experiments (using first content patterns and then architecture patterns), in order to show that pattern usage can actually guide the ontology engineers to construct correct and suitable ontologies tentatively both easier and faster. The reuse benefit would then be further studied in the next four controlled experiments, focusing on comparing the "quality" of ontologies constructed using patterns with ontologies constructed without using patterns (three experiments using a manual design process and logical, content and architecture patterns respectively, and one automatic process using content patterns).

The second set of further experiments naturally follows from the first set, assuming that it has been proven that patterns really do give benefits. This set of experiments focus on what tools and functionalities are needed to support pattern-based design and how patterns should be constructed, presented and stored. So far these two sets of proposed experiments mainly focus on the reuse and guidance benefits of ontology patterns in ontology design, but of course even more experiments could be envisioned focusing on communication benefits when for example using patterns to evaluate ontologies, document ontologies or communicate between developers.

Initially four of the above experiments have been selected to be included in WP5. The following four experiments have been selected and will be described in further detail below:

1. Exploratory study on problems solved by logical and content ontology patterns in "real-world" ontologies.
2. Controlled experiment concerning the effects of content patterns on the manual ontology design process.
3. Controlled experiment concerning the effects of content patterns on the "quality" of ontologies constructed manually using pattern-based design.
4. Controlled experiment concerning the effects of patterns presentation on manual pattern selection.

The first planning of these experiments suggested that results should already be present in time for this deliverable, but during the further planning of these experiments it has become clear that the experiment results will benefit from somewhat delaying the experiments and from the results of the two preliminary experiments included in this chapter. This is due to the fact that the second version of the NeOn pattern catalogue (first version presented in D5.1.1 [2]) is only now being developed, and thereby no complete pattern repository has so far been available (see also D2.5.1 [9] for details on the repository). Since the suggested experiments rely heavily on the availability of patterns, as the actual object of study, a decision was made to delay the experiments until all patterns are available.

3.1. Experiment 1 – Experiment in the Artificial Intelligence course

3.1.1. Experiment Definition

Experiment motivation.

The main idea is to know if people modelling ontologies in a certain domain identify the correct ontology design patterns for solving particular modelling problems.

Experiments goals.

The goal of the experiment is to test if, given a modelling problem expressed in natural language, people modelling ontologies are able to identify the correct pattern for such problem. In the case people are not able to identify the correct pattern, the objective is to know if typically a wrong pattern is used for the modelling problem instead of the correct one.

Beneficiaries of the experiments.

Partners involved in research about modelling ontologies and ontology design patterns will obtain a benefit of this experiment that will serve them to improve their research.

Experiment subject.

A set of modelling problems involved the following ontology design patterns, from those presented in D5.1.1 [2], including the following ones: subclass-of relation (LP-SC-01), disjoint classes (LP-Di-01), exhaustive classes (LP-EC-01), object property (LP-OP-01), individual (LP-In-01), defining n-ary relations: using a list for arguments (LP-NR-01), simple part-whole relations: part-whole relation (CP-PW-01), simple part-whole relations: part-whole class hierarchy (CP-PW-02), datatype property (LP-DP-01), representing specified values in OWL: values as sets of individuals (LP-SV-01), and representing specified values in OWL: values as subclasses (LP-SV-02).

3.1.2. Experiment Design

Relevant characteristics.

ODPs have different characteristics as solution proposed, problem addressed, formal and graphical representation, etc. In this experiment, we are going to study if people modelling ontologies when they have a modelling problem give as a solution the correct ontology design pattern, described using the template defined in [2].

Metrics and criteria.

- Which metrics will be used to measure these characteristics?

  In this experiment we propose different modelling problems, expressed in natural language, whose solution involved particular ontology design patterns from those described in [2].

- Which criteria will be used to interpret these measurements?

  For interpreting the measurements, we propose to compare the solution for the modelling problems with the solutions or answers obtained after carrying out the experiment.

  Correct solutions imply that the correct pattern has been applied, and thus the pattern has been well identified from the modelling problems expressed in natural language. Wrong solutions can imply either the wrong pattern has been applied or a solution without using patterns has been applied. In both cases, it indicates that the pattern is not well identified.

Variables (time, resources, etc.) that affect the measurement.

- Which variables can affect the measurement and can be controlled?

  The experience level and background of people carrying out the experiment. In this case, subjects carrying out the experiments have different experience in modelling (databases, software, etc.) but no extensive practical experience in ontology modelling.

  The complexity of the proposed problems is held low, but incremental throughout the experiment.

- Which variables cannot be controlled?

  Subjects can be familiar with the domain of the proposed problems, but it is not reasonable to measure this variable.

Data collection process.

This experiment is carried out within the "Artificial Intelligence (AI)" course at Facultad de Informática (Universidad Politécnica de Madrid) with undergraduated students, having background in database modelling (e.g., entity relationship, etc.), software engineering modelling (e.g., UML), and AI modelling (e.g., frames, semantic networks, description logics).

The experiment is divided in the following phases:

1. Lectures provide to students 13 modelling problems with incremental degree of difficulty. Such modelling problems have as solution models involving particular ontology design patterns (logical and content ones) described in [2]. The modelling problems presented to the students are included in Appendix I.

2. Students individually solve the modelling problems in 45 minutes. It is expected here that they use their previous knowledge about ontology design patterns for solving the proposed modelling problems, because they are not provided with the catalog of patterns during the experiment execution. Students solve the problems in white papers.

The solutions provided for each student will be both in the paper given by the student and in an excel file comparing the right solution with the solution provided by the student.

Requirements.

The experiment about the use of ontology design patterns during modelling ontologies should be carried out one time by at least 65 students from the "Artificial Intelligence" course at Facultad de Informática.

Analysis procedure.

- How will the data be validated?

  The data will be validated by means of reviewing the solution given by students and of checking that such solutions are models.

- How will the data be analysed?

  The solutions given by students will be compared with the correct solutions (involved ontology design patterns described in [2]) for each proposed modelling problem.

  Correct solutions to the proposed modelling problems are included in Appendix I.

Conclusions about if the correct pattern was used or not, about if normally a wrong solution instead of a pattern is used (that is, use of anti-patterns), etc., will be provided.

Experimentation plan.

The experiment is carried out on June 2007 by at least 65 undergraduate students attending the “Artificial Intelligence” course during 2007 at Facultad de Informática.

The results of carrying out the experiment are analysed by UPM people (Jose Angel Ramos Gargantilla and Mari Carmen Suárez-Figueroa). Such analysis is carried out during December 2007 and January 2008.

3.1.3. Experiment Phase

The experiment is carried out on June 2007 by 69 undergraduate students attending the “Artificial Intelligence” course during 2007 at Facultad de Informática. Students are in the same room during 45 minutes to carry out the experiment. They are provided with the 13 modelling problems and only white papers for solving the problems.

The results of the experiments are collected and validated after the students finish modelling the 13 problems.

3.1.4. Analysis Phase

The results of the experiment are stored in an excel file to facilitate the analysis of the results and the extraction of statistics from such results. In this section, we provide some conclusions extracted from the analysis of the experiment results.

Part of the statistics extracted from the experiment results are included in Appendix II.

The experiment includes 13 modelling problems solved by 69 students, and in general as **Figure 1** shows 59% of the solutions provided by the students corresponded with the correct ontology design pattern. 22% of the solutions included a wrong pattern. 11% of the solutions were empty and 8% of the answers were wrong but without involving any ontology design pattern. As a general conclusion we can say that more than half of the solutions provided include the correct ontology design pattern; however, there is almost a quarter of the solutions provided that includes a wrong pattern. One of the objectives of the research on ontology design patterns should be to reduce such percentage, and for knowing how to reduce such number some other experiments are proposed in this deliverable.

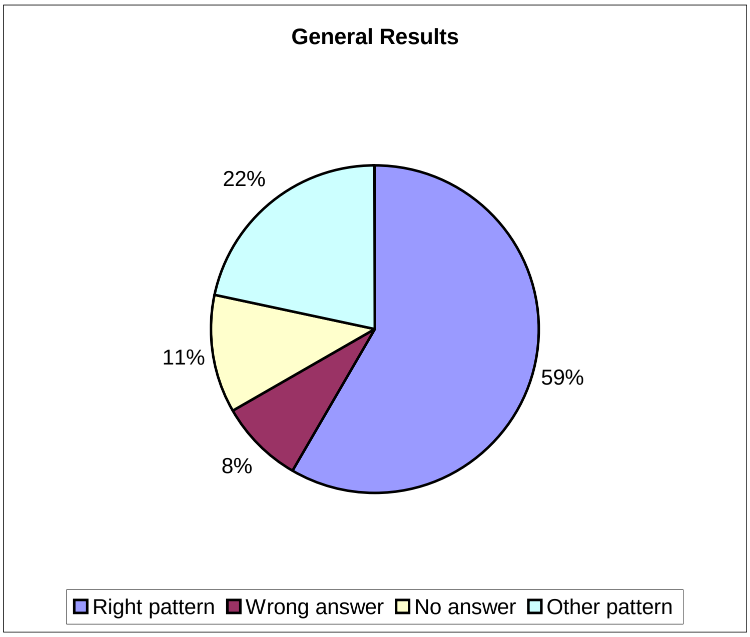

Figure 1. General Results

- More specific conclusions are: Half of the students identified correctly the “subclass-of” relation (LP-SC-01) from natural modelling problems proposed. However, more than a quarter of the students confused the “subclass-of” relation with other patterns; in most of the cases with “exhaustive classes” and “disjoint classes”.
- Most of the students identified correctly the “disjoint classes” (LP-Di-01) from the natural modelling problem proposed.
- Almost 40% of the students identified correctly the “exhaustive classes” (LP-EC-01) from the natural modelling problem proposed. However, the same percentage of students confused the ““exhaustive classes” with other patterns; in most of the cases with “subclass-of“ relation and in 23% of the cases with “disjoint classes”.
- Most of the students identified correctly “object properties” (LP-OP-01) from the natural modelling problem proposed.
- The majority of the students confused “defining n-ary relations: using a list for arguments” with “object properties”.
- A considerable number of students did not identify correctly the “part-whole” relation, confunding it with the “subclass-of” relation.

3.2. Experiment 2 – Experiment in a Summer School

3.2.1. Experiment Definition

Experiment motivation.

The main idea is to learn about the understandability and correct identification of ontology design patterns for a particular problem when modelling ontologies.

Experiments goals.

The goal of the experiment is to test if a subset of logical and content design patterns described in NeOn deliverable D5.1.1 [2] is well explained and if such patterns are easy to understand and to apply in particular modelling problems.

Beneficiaries of the experiments.

Partners working on ontology modelling and ontology design patterns will obtain a benefit of this experiment that will serve them to improve the research on such issues.

Experiment subject.

A subset of the inventory of ontology design patterns presented in D5.1.1 [2], including the following logical and content patterns: subclass-of relation (LP-SC-01), disjoint classes (LP-Di-01), exhaustive classes (LP-EC-01), object property (LP-OP-01), universal restriction (LP-UR-01), union of classes (LP-UO-01), individual (LP-In-01), defining n-ary relations: using a list for arguments (LP-NR-01), defining n-ary relations: using a list for arguments (LP-NR-02), simple part-whole relations: part-whole relation (CP-PW-01), simple part-whole relations: part-whole class hierarchy (CP-PW-02), existential restriction (LP-ER-01), representing specified values in OWL: values as sets of individuals (LP-SV-01), representing specified values in OWL: values as subclasses (LP-SV-02), and role-task (CP-RT-01).

3.2.2. Experiment Design

Relevant characteristics.

ODPs have different characteristics as solution proposed, problem addressed, formal and graphical representation, etc. In [2] ODPs are described using a template with different slots informing about such characteristics. In this experiment, we are going to study if the ODPs as they are described in [2] following the template are well understood, and if such template helps in the task of identifying an ontology design pattern for a particular modelling problem.

Metrics and criteria.

- Which metrics will be used to measure these characteristics?

  In this experiment we propose different modelling problems, expressed in natural language, whose solution involved particular ontology design patterns

- Which criteria will be used to interpret these measurements?

  For interpreting the measurements, we propose compare the correct solution for the modelling problems with the solutions obtain after carrying out the experiment.

  Correct solutions implies that the correct pattern has been applied, and thus the pattern has been well identified from the modelling problems expressed in natural language. Wrong solutions can imply either the wrong pattern has been applied or a solution without using patterns has been applied. In both cases, it indicates that the pattern is not well identified.

Variables (time, resources, etc.) that affect the measurement.

- Which variables can affect the measurement and can be controlled?

The experience level and background of people carrying out the experiment. In this case, subjects carrying out the experiments have different experience in modelling (databases, software, etc.) but no extensive practical experience in ontology modelling.

The complexity of the proposed problems is held low, but incremental throughout the experiment.

- Which variables cannot be controlled?

  Subjects can be familiar with the domain of the proposed problems, but it is not reasonable to measure this variable.

Data collection process.

This experiment is carried out with Master and PhD students attending the 5th Knowledge Web summer school on Ontological Engineering and the Semantic Web. Students have different background in modelling. Students will be asked about previous courses taken on database modelling (e.g., entity relationship, etc.), software engineering modelling (e.g., UML), AI modelling (e.g., frames, semantic networks, description logics), etc.

The experiment is divided in the following phases:

1. Summer School tutors will provide to students three modelling problems with incremental degree of difficulty. Such modelling problems have as solution models involved particular ontology design patterns (logical and content ones) described in [2]. The modelling problems presented to the students are included in Appendix II.

2. Students will individually solve in papers the modelling problems in 45 minutes.

3. Summer School tutors will explain what an ontological design pattern is, and will give the students a collection of ontology design patterns (taken from NeOn deliverable D5.1.1 [2]). Tutors will not explain the collection of patterns.

4. Tutors will propose new modelling problems. The modelling problems must replicate the initial three problems but in another domain. The modelling problems presented to the students are included in Appendix II.

5. Students will individually solve in papers the new modelling problems in 45 minutes using the collection of ontology design patterns provided by tutors.

The solutions provided for each student will be both in the paper given by the student and in an excel file comparing the right solution with the solution provided by the student.

Requirements.

The experiment about the use of ontology design patterns during modelling ontologies should be carried out one time by 50 Master and PhD students attending the 5th Knowledge Web summer school on Ontological Engineering and the Semantic Web.

Analysis procedure.

- How will the data be validated?

The data from the first set of modelling problems will be validated by means of reviewing the solution given by students and of checking that such solutions are models (graphics, written in an ontology language, etc.).

The data from the second set of modelling problems will be validated by means of reviewing the solution given by students and of checking that such solutions correspond to name or identifier of ontology design patterns from the provided collection.

- How will the data be analysed?

  The solutions given by students for the first and the second set of modelling problems will be compared with the correct solutions (involved ontology design patterns) for each proposed modelling problem.

  Correct solutions to the proposed modelling problems are included in Appendix II.

  Conclusions about if the correct pattern was used or not, about if some patterns are misunderstanding, about if normally a wrong solution instead of a pattern is used (that is, use of anti-patterns), about if after providing the collection of pattern the use of them for modelling improves, etc., will be provided.

Experimentation plan.

The experiment is carried out on July 2007 by 50 Master and PhD students, with different background in modelling, participating in the 5th Knowledge Web summer school on Ontological Engineering and the Semantic Web.

The results of carrying out the experiment are being analysed by UPM people (Jose Angel Ramos Gargantilla and Mari Carmen Suárez-Figueroa). Such analysis is carried out during December 2007 and January 2008.

3.2.3. Experiment Phase

The experiment is carried out on July 2007 by 50 Master and PhD students, with different background in modelling, participating in the 5th Knowledge Web summer school on Ontological Engineering and the Semantic Web. Students are in the same room during 45 minutes to carry out the first part of the experiment and 45 minutes to carry out the second part of the experiment. They are provided with the 10 modelling problems for the first part and with other 10 modelling problems for the second part. They are provided with white papers for solving the problems, and with the catalog of ontology design patterns in the second part of the experiment.

The results of the experiments are collected and validated after the students finish modelling the provided problems.

3.2.4. Analysis Phase

The results of the experiments are stored in an excel file to facilitate the analysis of the results and the extraction of statistics from the results. In this section, we provide some conclusions extracted from the analysis of the experiment results.

From the first part of the experiment, in which students carry out the modelling problems without the collection of ontology design patterns, we can mention as conclusions the following:

- 88% of the students correctly identified the “subclass-of” relation, and 72% of the students correctly modelled the disjoint knowledge.

- 90% of the students correctly identified “individual”.
- 70% of the students correctly identified “object property”; however, 22% of the students confused with “part-whole”.
- 66% of the students did not identify the “exhaustive classes”; such students provided solutions included a wrong pattern, mainly “object property” and “subclass-of”.
- 74% of the students did not correctly identified “n-ary relation”, confusing such pattern with “object property” in most of the cases.
- A considerable number of students confused “part-whole” with “subclass-of” or with “object property”.

From the second part of the experiment, in which students carry out the modelling problems using the collection of ontology design patterns included in D5.1.1 [2], we can mention as conclusions the following:

- 52% of the students mainly confused “subclass-of” with “disjoint classes”.
- 62% of the students mainly confused “exhaustive classes” with “subclass-of”.
- 66% of the students mainly confused “disjoint classes” with “exhaustive classes”.
- 54% of the students mainly confused “n-ary relation” with “object property” or with “datatype property”.
- A considerable number of students confused “part-whole” with “object property” or with “subclass-of” or with “subproperty-of”.
- 48% of the students mainly confused “specific values” with “object property” or with “exhaustive classes”.
- 38% of the students mainly confused “role-task” with “plan-execution” or with “specific values”.

As general conclusion of the first part of the experiment, we can say that students had problems with the identification of the following ontology design patterns: “exhaustive classes”, “n-ary relation”, and “part-whole”.

As general conclusion of the second part of the experiment, we can say that students had problems in the identification of the following ontology design patterns: “subclass-of”, “exhaustive classes”, “disjoint classes”, “n-ary relation”, “part-whole”, “specific values”, and “role-task”.

Due to in the first part of the experiment, students had no too many difficulties on identifying “subclass-of” and “disjoint classes”, we can conclude that in both cases ontology design patterns in the collection should be improved, and also methodological guidelines should be provided to easily differentiate between “subclass-of”, “disjoint classes”, and “exhaustive classes”.

In the cases of “exhaustive classes”, “n-ary relation”, and “part-whole”, students had problems both in the first part and in the second part (when they used the collection of ontology design patterns). Therefore, these ontology design patterns should be better explained in the collection, and methodological guidelines for distinguishing between such ontology design patterns should be also provided.

Regarding “specific values” and “role-task”, it is also clear that these ontology design patterns should be improved in the collection in order to allow their better understanding.

3.3. Experiment 3 – Problems solved by ontology patterns

3.3.1. Experiment definition

Experiment motivation

To determine what problems are actually solved (or could have been solved) by logical and content patterns in “real” ontologies that are used today. To show that patterns solve “real” problems will serve as a general motivation to why patterns are useful.

Experiment goals

To show that existing ontologies either implement a recurrent design solution in order to solve a certain modelling problem, or that they have some undesirable effect i.e. they do not address a certain problem, and would benefit from the application of the pattern. Ontology design patterns emerge, therefore such experiments wants on one hand to show that they can be extracted once and reuse instead of re-inventing the wheel. On the other hand, we also want to show that there are common mistakes that patterns could have helped to resolve. Some simple statistics how many ontologies (out of the test set) are actually using the selected patterns (consciously or without knowing about patterns) for solving common problems.

Beneficiaries of the experiments

The pattern research community, through having a set of problems that patterns are known to solve (a motivation for the patterns) and showing that patterns are actually already used (i.e. they are really some kind of best-practices).

Experiment subject

Logical and content patterns and the problems they can solve.

3.3.2. Experiment design

Relevant characteristics

The problems that can be (and are in reality) solved by a set of logical and content patterns.

Metrics and criteria

Pattern presence will be measured through manual inspection and classified either into the category “complete usage” or “partial usage” and the total number of complete or partial identifications of a pattern in an ontology will be summed up. In addition for each pattern the number of “places” in the ontology where the pattern could have been used but is not, resulting in some kind of problem arising, will be assessed manually. In addition these “places” will be stored as evidence (examples) showing a problem that arises when not using patterns.

Statistics of the pattern usage will be the result of summing up the usage counts for each ontology and the total set of ontologies. The percentage of ontologies using a specific pattern will be computed and interpreted as an indication of the support the pattern has in “real-world” ontologies. Additional statistics will be derived from the potential pattern usage counts, but no general conclusions can be drawn since the selection of the ontologies was not conducted completely randomly, the results has to be seen only as an indication on the actual usage of patterns. The wrong design solutions will be collected and stored as examples of “anti-patterns” that can be used in the description of the *correct* pattern.

There is no baseline to which we can compare results because the patterns are something new. What the experts are asked to do with this experiments is to evaluate existing ontologies against

NeOn ontology design patterns and possibly show that when there are problems they can be solved by patterns. Furthermore, the validation is subjective because it is based on personal expertise and experience. We want to show that there are recurrent solutions in existing ontologies and that they are encoded in NeOn patterns, therefore they can be reused instead of *re-inventing the wheel*, favouring interoperability and saving resources.

Variables that affect the measurement

The **selection of ontologies** can in theory be controlled, so that it is done randomly and without a bias, but in practice probably there will not be so many ontologies available that are large and complex enough to be relevant, and still come with documentation about its intended usage and requirements. What **patterns** are searched for, this should be a fixed catalogue of well-established patterns determined before the experiment starts. The **ontology representation language** could affect the analysis, therefore the language is set to be OWL DL for the ontologies (and in cases where the patterns are language specific this should also hold for the patterns). The **tool** used by the experts for visualising the ontologies might influence what pattern and problems are easy or hard to detect. On one hand the experts will perform better using a tool they are familiar with, but on the other hand the tool might constitute a bias towards certain types of patterns or types of errors. If possible (with respect to tool availability) the experts should use the same tool in order for the results to be comparable (even though certain results might be missed due to this).

The **selection of ontologies** can in practice only partly be controlled, as mentioned above. Another variable is the **expertise and judgment** of the “experts” performing the experiment. Preferably there should be several “experts” (at least 2) performing the same tasks and then the agreement of the experts can be used as the result.

Data collection process

First the ontologies have to be selected. The ontologies have to satisfy some criteria to make them relevant for the study. They need to be large enough to have the possibility to display the use of several patterns, we set the limit at a minimum of 20 concepts. They also need to display some level of complexity in order to include patterns, so we set the requirement that at least the ontology should display something more than a simple taxonomy and all concepts should be used in some relation. Additionally, to determine the presence of “anti-patterns” the requirements and intended usage of the ontology need to be known, enough documentation to determine if a certain construct in the ontology meets its requirements or not is needed. With these criteria in mind ontologies will be retrieved from known projects and additionally searched for on the Internet. All ontologies found matching the above criteria will be stored initially. The study needs at least 20 ontologies, therefore the initial search should retrieve at least 40 ontologies (or as many as can be found that fulfil the requirements). If less than 20 ontologies are found the size criteria has to be reduced and additionally the complexity criteria. When more than 40 ontologies have been found half of them are randomly selected for use in the experiment.

The retrieval and selection of ontologies will be conducted by a person familiar with ontologies and ontology engineering (to be able to assess the ontologies easily). This should not be the same person as the experiment supervisor or the pattern experts. The patterns will be selected from the NeOn modelling components (logical patterns and content patterns) described in deliverable D5.1.1. The selection should result in 5 patterns of each kind (logical and content) based on the complexity of the patterns, so that a more complex pattern is preferred over a simpler one. The intuition behind this selection is to avoid including patterns that are so simple that they will always be present in all ontologies. The patterns are selected by the experiment supervisor, who will also write the instructions for the experts, submit the experiment material and collect and analyse the results. Additionally at least two pattern experts need to be found in order to run the experiment. The pattern experts should have practical experience in ontology engineering and in the use of patterns, acquired prior to this experiment (and should not be the same persons as the two experiment supervisors mentioned above).

3.3.3. Conducting the experiment

The set of ontologies is then distributed to the two experts. They may perform the study independently and submit the results when they are finished, only a general time-limit of one week is set in order to finish the experiment on schedule. The experts are instructed to base their decisions only on the material presented and not on any additional knowledge about the origin of the ontologies, their current usage etc. For each ontology the experts fill out a form where they note down, for each pattern, how many complete and partial occurrences were found in that specific ontology and where in the ontology those were found. Additionally all the "places" where the pattern could have been used are described in detail (where it occurred, the context of the occurrence, the problem it gives rise to in that specific ontology with respect to the requirements, how the pattern could have solved the problem etc.) This is done for all the ontologies and the results are sent to the experiment supervisor specified in the instructions.

For each ontology and each pattern the experts fill out a form and describe the occurrences of the pattern in that ontology. Additionally the experts produce a list of detected "problem areas", a description of the problem, its location and how a certain pattern could have solved it, for each ontology and each pattern.

Requirements

The analysis is performed by each expert for each ontology combined with each pattern. There have to be at least 20 ontologies and there are 10 patterns.

One person is needed to collect the ontologies and another person to prepare and supervise the rest of the experiment and analyse the results, then two pattern experts are needed to perform the actual experiment. This sums up to 4 people in total, where two are pattern experts and two are at least familiar with ontology engineering.

Analysis procedure

The results are validated through the scale and threshold described below, sorting out unreliable judgments from the experts before the final analysis is performed. Validation of the problem examples are done through manual assessment by the person analysing the results.

When the results have been collected by the experiment supervisor that person is also responsible for the aggregation and analysis of the results. The pattern identifications are re-weighted on a scale from "one expert found it partially", "one expert found it complete", "one found it complete and one partially", "both found it partially" to "both found it complete". This gives a "confidence scale" (from 1 to 5) that shows how reliable we can consider the statement that this pattern actually is present in the ontology in this specific position. The intuition is that if only one of the experts found it and it was only partially used then it is quite uncertain that the pattern was actually used, while if both experts identified the pattern in this position it is much more certain that this is actually a correct judgment. Depending on how it is desirable to present the results a threshold can be set on this four level scale for when to consider that a pattern has actually been found. A strict interpretation would be to only consider the pattern as present if both experts have found it completely or partially, a slight relaxation would be to also consider the case where both found it but one expert considers it to be only partially present etc.

Then the statistics over pattern presence can be computed as, for each pattern, the number (and fraction) of ontologies where it was present and the total number of occurrences of the pattern found. Additionally the collection of problems should be analysed for general categories of "mistakes" that each pattern could have solved, so that if several examples exist for a pattern these examples can be grouped into problem categories and each group generalised into a kind of "anti-pattern".

The intention of the experiment is not to show any improvements or compare the result to something, only to point at the presence of patterns in "real-world" ontologies and collect typical situations where they can be/have been used.

Experimentation plan

Experiment preparation, pattern selection, supervision: Eva Blomqvist

Collection of the ontologies: Valentina Presutti

Pattern experts: Aldo Gangemi and Enrico Motta

The experiments will be performed during 2008 (schedule to be determined in February 2008). The experiment results will be analysed by Eva Blomqvist and Valentina Presutti during 2008 (schedule to be determined in February 2008).

3.4. Experiment 4 – Effects of content patterns on ontology design

3.4.1. Experiment definition

Experiment motivation

The motivation is to determine if patterns really make manual ontology design "easier" and faster.

Experiment goals

By testing the effects of content pattern usage on the ontology construction time and the actual and perceived "assistance" given by the patterns we want to learn what the actual benefits of using content patterns are, in terms of process benefits.

Beneficiaries of the experiments

In the end ontology engineers will benefit from knowing what methods help them construct ontologies faster and easier. Also, the ontology pattern community will benefit from having evidence showing benefits and drawbacks with pattern-based methods.

Experiment subject

Ontology content patterns.

3.4.2. Experiment design

Relevant characteristics

The ability of content patterns to assist users when manually constructing ontologies.

Metrics and criteria

Perceived (subjective) "assistance" by the ontology developers will be measured through questionnaires given to the subjects, including both a structured part (user assesses the process by selecting answers from a Likert-scale[2]) and an unstructured part where the user can reflect on the process in free text. Actual (objective) "assistance" will be measured in terms of construction time for each task versus number of clear errors in the result (assessed manually), coverage of

[2] A likert scale is a type of psychometric response scale developed by Rensis Likert often used in questionnaires, and is the most widely used scale in survey research. Likert items have responses on a continuum and response categories such as "strongly agree," "agree," "disagree," and "strongly disagree".

result with respect to the task description (assessed manually on a Likert-scale), and the “quality” of the output ontology with respect to a “gold-standard” developed beforehand.

The questionnaires will give statistics on how the subjects perceive the process from a set of viewpoints, which can then be used to draw conclusions on the overall perceived assistance and satisfaction of the subjects. In addition the free text answers will be used to help motivate details in the statistical results, as well as giving insight into the preferences and intuitions of the subjects. The construction times will be used to compute the completion time differences between the group using patterns and the control group not using patterns. A shorter construction time in combination with lower (or the same) error rate and good coverage and quality would indicate a high level of actual assistance through the use of patterns. A shorter construction time but more errors, less coverage or poor quality indicates that the patterns are used but are used incorrectly. Less coverage using patterns could also indicate that patterns restrict the subjects and that the patterns available did not cover the task completely. Longer construction time would generally indicate that patterns do not give so much assistance or that they are hard to understand, but this might be somewhat outweighed by the perceived assistance or explained by the answers to the questionnaires.

Variables that affect the measurement

The **experience level and background** of the subjects can be controlled. Subjects should be chosen from two major groups, inexperienced subjects and experienced subjects. Inexperienced indicating experience in modelling but no extensive practical experience in ontology modelling, such as master students or PhD students not directly involved with ontologies but who have taken some course on ontologies or tried some simple exercises. All these subjects should be given a short tutorial on ontology modelling in OWL to leverage their previous knowledge (ontology patterns should only be mentioned briefly). Experienced users should have at least some practical experience in “real-world” ontology modelling (more than “toy examples”) using OWL and extensive theoretical knowledge of ontologies. They should be given a short tutorial on ontology content patterns to leverage their knowledge of patterns. There should be approximately equal number of experienced and inexperienced subjects included in the experiment. **Complexity of the problems** can affect the result, this will not be used as a factor of the experiment so the complexity will be held low and (approximately) constant in the two tasks given. **Pattern presentation** is another variable that could affect the result, but this will not be considered as a factor either, all patterns will be presented equally throughout the experiment (using the same template and the same kind of representation in OWL).

Subjects’ **familiarity with the domain and the tool used** for solving the problems presented cannot be controlled. It is not reasonable to do a complete “background check” on the subjects, but the domain and problem types should be varied in the two tasks to reduce the effect of this, and the tool is kept the same for all subjects. Subjects’ own **speed of analysing** the problem. The effect of this will be reduced by studying the time difference over the complete groups and not absolute times of individual subjects. The **order of the tasks**, design with or without patterns, is a factor that may influence the results (if patterns were used first then the subject may have learned some good practices through the first task). This variable will be taken into account when analysing the results. The actual assistance will be primarily measures during the first round, when the groups have a clear distinction between using patterns and not using patterns, but still it is worth using two tasks and switching groups in order to get more valuable feedback in the questionnaires after the second task.

Data collection process

The preparation of the experiment in terms of writing the problem descriptions, developing the “gold-standard” ontology, selecting patterns and preparing the experiment environment will be done by one person. The problems will be designed to be solvable by an inexperienced subject in about two hours. Additionally the questionnaires and the tutorials to be given have to be prepared. Several questionnaires should be prepared for each group. Prior to the experiment (some day in advance) all the subjects get a short questionnaire asking about their background and level of experience using the tool and designing ontologies. This is to assist the experiment supervisor to divide the subjects into even groups, with respect to experience etc. Next there is a questionnaire to be answered by the subjects after the first round of modelling. This questionnaire is prepared in two versions, one for those who used patterns and one for those who did not. For those who used patterns the important questions are what problems they had, how and what they used the patterns for and if the patterns helped, while to the ones that did not use patterns the questions focus on perceived problems (in order to see differences in perceived problems between the groups). Finally, there is a questionnaire at the end of the task, also in two versions with again slightly different questions depending on the groups.

3.4.3. Conducting the experiment

The subjects are divided into two groups, one which will use patterns in the first round and one group that will not. The groups should be divided as equally as possible, with equal number of experienced and inexperienced subjects in each group. Before the experiment the subjects will answer a small questionnaire about their background, then they will be divided into pairs based on equal kinds of background, from each pair the two persons will end up in different groups. Immediately before the experiment also the tutorials (on OWL modelling for the inexperienced group and on patterns for the experienced group) will be given to the subjects. The tutorials will include a short introduction to the tool that is to be used. The tutorials will be only a “lecture” and some reading material, no exercises. The subjects are then given the first task (same for all), a problem description covering the intended domain of the ontology and a few competency questions as the basis for modelling. One group additionally has a set of content patterns available, both as OWL models and described with additional information and examples (in the template used in NeOn D5.1.1).

The content patterns are given to this group a few minutes before both groups get the first task (approximately 2 minutes per pattern given to read and understand the patterns). When the subjects finish the task the time for completion is recorded and the resulting ontology stored. The subject is also given an intermediate questionnaire to record their initial perception of the experience. A maximum time of 2 hours is set for completing the first task. When all have finished the groups are switched, so that the group using patterns now will work without patterns and the other group gets the patterns instead (the same procedure with this group getting acquainted to the patterns through reading the information is applied, approximately 2 minutes per pattern). A new task is presented and the subjects start another modelling round. A maximum time of 2 hours is set for completing also the second task. After this round (when completion time and ontologies have again been stored) another questionnaire is given to each subject, with questions about the last part of the experiment but also asking the subjects to compare the two parts and reflect on their experience of using patterns.

The data (task completion times, resulting ontologies and answered questionnaires) will be stored for each subject with a connection to an ID-number for that subject, in order to be able to trace the set of results to each user but without connecting it to the person’s identity. Task completion times are noted on the set of questionnaires handed in, and the filenames of the ontology files are written on these papers as well.

Requirements

As described above the experiment is run twice, after “switching” the setting (with respect to patterns) in between, and in each round the groups will contain at least 10 people each (approximately 5 experienced users and 5 inexperienced).

At a minimum the experiment needs 10 inexperienced and 10 experienced users, but it would be desirable to have as much as 40 subjects in total. In addition the experiment is to be prepared by two people and lead by at least one person for each site (if distributed over several places) also responsible for giving the tutorials before the session. After the experiment at least one “expert” is needed to manually evaluate the resulting ontologies, and one person is needed to summarise and analyse the results.

Analysis procedure

The time measurements will be validated through every subject additionally noting down the time when they finish the tasks themselves. The resulting ontologies will be classified in two categories, finished tasks and unfinished tasks (in the case that some subject does not manage to perform the task at all, leaves ahead of time, or is not finished when the maximum time is reached). Measures of ontology errors, coverage and quality are subjective decisions by an “expert” and cannot be validated.

The resulting ontologies will be assessed by and “expert” to determine the number of clear errors in each ontology (with respect to the problem description). A clear error is something that contradicts the description, only leaving something out or interpreting it differently but still consistent with the description is not counted as an error. For each ontology the number of clear errors is entered in a form. Next, the expert will assess the coverage of ontologies. Since this is a subjective assessment, that will only be used to check that the coverage holds an acceptable level, a mark on a Likert scale and not a number will be used as the result of this assessment. The expert should check for each statement in the problem description and each competency question if these are realised in the ontology. The result is entered in the same form as above, for each ontology.

The quality of the ontologies will be measured with respect to a “gold-standard” ontology constructed before the experiment, representing a somewhat “optimal” way of describing the problem (“optimal” in the sense that it uses no unnecessary concept, only covers the precise domain of the problem description but still solves all competency questions). The details of ontology quality are not considered in this experiment, since the focus is on the process of construction, this quality assessment attempts to discriminate ontologies that are too large and extensively detailed, modelled in a complicated way etc. To do this, the number of concepts are counted in the constructed ontologies and then divided by the number of concept present in the “gold-standard”. The closer to 1 the better ontology quality is the intuition. Additionally the same is done for the number of properties. Finally the ontologies are additionally assessed for “connectedness”, it is investigated if the ontologies can be shown as a connected graph or not.

All this information is entered into the form for each ontology. The difference in completion time between the experienced users and the inexperienced users, as well as the group using patterns and the group not using patterns will be computed and compared for each round. This information is then analysed together with the error rates, coverage assessments an ontology quality that have been manually assessed as described above. Ideally the quality is expected to be the same or better for the pattern-based ontologies while they are constructed faster, the quality assessments are needed to check that the ontologies are really equal or better if faster times are measured for the group using patterns.

The baseline for comparison is the group in the first round not using patterns.

Experimentation plan

The experiment may be performed in different locations, as long as it can be ensured that the experiment conditions are equal in all settings.

Questionnaire preparation: Eva Blomqvist.

Problem descriptions, pattern selection, tool selection and tutorial preparation: Valentina Presutti.

Subject groups:

Jönköping University – 3 experienced subjects, 3 inexperienced subjects (experiment supervisor: Eva Blomqvist).

CNR Rome – 7 experienced subjects, 7 inexperienced subjects (experiment supervisor: Valentina Presutti).

The experiments will be performed during 2008 (schedule to be determined in February 2008).

The analysis of the results will take place during 2008 (schedule to be determined in February 2008).

Expert assessments (errors, coverage and quality): Valentina Presutti.

Result collection and analysis: Eva Blomqvist.

3.5. Experiment 5 – Effects of content patterns on ontologies

3.5.1. Experiment definition

Experiment motivation

To determine what properties are imposed on the designed ontology based on the use of content patterns, and to determine if these properties really indicate that the ontology is "better" in some sense.

Experiment goals

A set of characteristics that ontologies constructed based on content pattern commonly exhibit, compared to manually constructed ontologies without using patterns (as a motivation to why someone should use content patterns).

Beneficiaries of the experiments

Ontology engineers will benefit from knowing the characteristics related to a specific method of ontology construction and the pattern community will benefit from experimental results supporting pattern-based design as a good ontology design method.

Experiment subject

Ontology content patterns.

3.5.2. Experiment design

Relevant characteristics

The properties imposed on the resulting ontology when patterns are used in the design process.

Metrics and criteria

The construction time will be measured for each subject, which will be the time from when the subjects get the problem description until the ontology is in their opinion finished. The error rate is measured through an "expert" finding clear errors in the modelling with respect to the problem description. The coverage of the ontology with respect to the input problem description, assessed again by "expert" judgment using a manual count of included parts (a manually determined precision and recall measure). Characteristics of the ontology, measured through the following set of ontology evaluation measures: average depth, maximal breadth, average density, tangledness, class/property ratio, number of root classes, number of leaf classes, and individual/concept ratio.

The construction time will only be used to explain exceptionally error prone or low-covering ontologies that in addition took the subject exceptionally short time to construct. This is most likely due to low motivation of the subject to participate in the experiment and to complete the task. Due to the simple nature of the problems a low error rate is expected, in combination with a high coverage. Higher error rates and/or lower coverage (in combination with longer completion time) would indicate that the subject found the task difficult. Such differences will be analysed and compared between the groups of subjects and between the group using patterns and the group not using patterns. The error rate and coverage differences can also indicate the "quality" of the constructed ontologies. The ontology characteristics will be studied to find differences between the ontologies constructed with and without using patterns, in terms of general "qualities" such as cognitive ergonomics, transparency, reusability, compliance to expertise (that can be derived from the combinations of the individual assessments).

Variables that affect the measurement

The **experience level and background** of the subjects can be controlled. Subjects should be chosen from two major groups, inexperienced subjects and experienced subjects. Inexperienced indicating experience in modelling but no extensive practical experience in ontology modelling, such as master students or PhD students not directly involved with ontologies but who have taken some course on ontologies or tried some simple exercises. All these subjects should be given a tutorial on ontology modelling in OWL to leverage their previous knowledge (ontology patterns should only be mentioned briefly). Experienced users should have at least some practical experience in "real-world" ontology modelling (more than "toy examples") using OWL and extensive theoretical knowledge of ontologies. They should be given a short tutorial on ontology design patterns to leverage their knowledge of patterns. There should be approximately equal number of experienced and inexperienced subjects selected for the experiment. **Pattern presentation** is another variable that could affect the result, but this will not be considered a factor, all patterns will be presented equally throughout the experiment (using the template used in NeOn and the same kind of representation in OWL). The **tool** used for the experiment will also be the same for all subjects.

Complexity of the problem can affect the result and cannot be controlled, this will not be used as a factor of the experiment, since due to time and subject availability restrictions the experiment can only be performed once. Subjects' **familiarity with the domain and the tool used** for solving the problems presented. It is not reasonable to do a complete "background check" on the subjects, but by using a sufficiently large number of subjects the effects on the result can be reduced. The tool is kept the same for all subjects. Subjects' own **speed of analysing** the problem affects the time to complete the task, but in this case time is only used to explain users that did not take the task seriously. The subjects' own **style of modelling** could affect some of the quality results, since some subjects may be more or less inclined to model in a "pattern"-like way even without the patterns and some may have other experiences that affects the way they model. The effects will be reduced through a sufficient number of subjects.

Data collection process

The preparation of the experiment in terms of preparing the tutorials, writing the problem description, selecting patterns and preparing the experiment environment can be done by one person. A suitable tool should be selected, which gives sufficient support for pattern-based design and which is available on all experiment sites. The problem will be designed to be solvable by an inexperienced subject in less than four hours (without patterns). A small questionnaire should also be prepared and sent to the subjects prior to the experiment (some day in advance) asking about their background and level of experience using the tool and designing ontologies. This is to assist the experiment supervisor to divide the subjects into even groups, with respect to experience etc.

3.5.3. Conducting the experiment

The subjects are divided into two groups, one which will use patterns and one group that will not. The groups should be divided as equally as possible, with equal number of experienced and inexperienced subjects in each group. Based on the answers to the initial questionnaire they will be divided into pairs based on equal kinds of background, from each pair the two persons will end up in different groups. Before the experiment also the tutorials (on OWL modelling for the inexperienced subjects and on patterns for the experienced subjects) will be given. The tutorials will be in the form of a lecture and some reading material (no exercises), and a brief introduction to the tool to be used.

The subjects are then given the task, a problem description covering the intended domain of the ontology and a few competency questions as the basis for modelling. One group additionally gets a set of content patterns, both as OWL models and described with additional information (e.g., in the template used in NeOn D5.1.1), this group is given some time (about 2 minutes per pattern in the catalogue) to read through the information about the patterns before both groups are given the task. When the subjects finish the modelling task the time for completion is recorded and the resulting ontology stored. The maximum time to complete the task is set to 4 hours.

Every subject gets a unique ID so the results will be stored for each user and identified by the ID-number. The ontologies are stored as OWL files, and the completion times are written down in a list.

Requirements

The experiment is conducted once for each subject.

As a minimum the experiment should have 30 subjects, for each type of subject i.e., inexperienced and experienced subjects. Possibly we will not have a unique group of 30 people. In that case we will replicate the task with smaller groups. If we will realize that the results would need more subjects we will iterate the replication on new groups. In addition to the subjects, at least one person is required to supervise the experiment, give the tutorials, prepare the experiment and collect results and analyse them. After the experiment at least two “experts” are needed in order to evaluate the ontologies according to the measurements.

Analysis procedure

The time measurements will be validated through every subject additionally themselves noting down the time on paper when they finish the tasks. The resulting ontologies will be classified in two categories, finished tasks and unfinished tasks (in the case that some subject does not manage to perform the task at all, leaves ahead of time or does not finish within 4 hours) the unfinished results will not be used for the analysis but reported as failures in the result of the study. The judgments of ontology errors and coverage are subjective decisions by an “expert” and cannot be validated, other than through the agreement of the “experts” (see below). Individual measures of ontology quality factors can be checked for erroneous calculations.

The construction time will only be used to explain error prone or low-covering ontologies that in addition took the subject a short time to construct, as low subject motivation to complete the task. Due to the simple nature of the problems a low error rate is expected, in combination with a high coverage, especially among the experienced subjects. The “quality” and characteristics of the ontologies will be assessed through measuring number of clear errors in the ontologies, the coverage of the ontologies over the problem description, and a set of characteristics commonly used for ontology evaluation (as mentioned previously). Together the results of these measures can be analysed to derive general characteristics of the ontologies, such as transparency etc. mentioned previously, for the ontologies constructed by the different subject groups.

First, the resulting ontologies will be assessed by the “experts” to determine the number of clear errors in each ontology (with respect to the problem description and competency questions). A clear error is something that contradicts the description, only leaving something out or interpreting it differently but still consistent with the description is not counted as an error. For each ontology the number of clear errors is entered in a form. Next, the expert will assess the coverage of the ontologies. Beforehand the problem description has been divided into “primitive” problems, by the person designing the problem description. A “primitive problem” is for example every concept in the text that should somehow be part of the resulting ontology, every relation or property described in the text that should somehow be realised in the ontology, every other axiom detectable in the text etc. For example, the sentence “There are only two kinds of animals, mammals and fish, where the fish generally live in the sea.” contains at least 8 “primitives” (concepts needed: “animal”, “mammal”, “fish” and “sea”, relations needed: “mammal is an animal”, “fish is an animal” and “fish live in sea”, plus possibly a constraint that mammals and fish exhaustively divide the concept animals). The experts evaluating the coverage of the ontologies should take this list of “primitives” needed to model the complete problem, and go through the resulting ontologies connecting parts of the ontologies to the “primitives”.

The result of this should be for each ontology a measure of how many primitives of the total number stated in the list (based on the problem description) were actually included and how many additional concepts, properties and other axioms were additionally included in the ontology. From this the precision and recall of the ontology can be computed as follows: *precision = number of realised primitives from the list / total number of primitives in the ontology* and *recall = number of realised primitives from the list / number of primitives on the list*. The result is entered in the same form as above, for each ontology.

Finally, the characteristics of the ontology are measured through the following set of ontology evaluation measures:

Number of root concepts = number of concepts only one step from the taxonomic root.

Number of leaf concepts = number of concepts with no child in the taxonomic hierarchy.

Average depth (in the taxonomic hierarchy) = sum of number of steps from the root for each primitive concept / number of primitive concepts.

Maximal breadth (in the taxonomic hierarchy) = max (the number of concepts with path length x from the root).

Concept/relation ratio = number of concepts / number of relations.

Average density = sum (sum for each concept the number of relations (taxonomic and other properties) connected to (going to or from) the concept and number of siblings in the taxonomy of the concept) / number of concepts.

Tangledness (of the taxonomic hierarchy) = total number of primitive concepts / number of primitive concept with more than one direct parent in the hierarchy.

Individual/concept ratio = number of individuals / number of concepts.

As a first step in the analysis all the measures are recomputed as an average over the results received from each of the experts. Next, the differences of all these measures between the experienced users and the inexperienced users, as well as the group using patterns and the group not using patterns will be computed. This information is then analysed in order to try and generalize the characteristics and draw conclusions about more general properties of the ontologies constructed using patterns.

The baseline to be used is in this case is the ontologies constructed without using patterns.

Experimentation plan

The experiment may be performed in different locations, as long as it can be ensured that the experiment conditions are equal in all settings.

Questionnaire and forms for the expert evaluations: Eva Blomqvist.

Problem descriptions, tutorials, pattern selection and tool selection: Valentina Presutti.

Subject groups:

Jönköping University – 3 experienced subjects, 3 inexperienced subjects (experiment supervisor: Eva Blomqvist).

CNR Rome – 7 experienced subjects, 7 inexperienced subjects (experiment supervisor: Valentina Presutti).

The experiments will be performed during 2008 (schedule to be determined in February 2008).

The results will be analysed during 2008 (schedule to be determined in February 2008).

Expert assessments (errors, coverage and characteristics):

Valentina Presutti and Aldo Gangemi

Result collection and analysis: Eva Blomqvist.

3.6. Experiment 6 – Effects of pattern presentation on pattern selection

3.6.1. Experiment definition

Experiment motivation

To determine the effects of different pattern presentation formats and explanations on user understanding, to verify what parts of the currently used pattern template are most important to the user and if something is missing or superfluous.

Experiment goals

How important different parts of the pattern description template are for the user understanding, what parts make a big difference and what parts are not so important to all subjects. Additionally to get ideas on more content from the subjects.

Beneficiaries of the experiments

Pattern users by getting more useful descriptions of patterns, and the pattern community by getting empirical evidence supporting the template.

Experiment subject

The template parts used to describe ontology patterns, specifically content patterns.

3.6.2. Experiment design

Relevant characteristics

The ability of a certain part of the pattern template information to support user understanding.

Metrics and criteria

Number of correctly classified problems for each pattern template tested, and the time differences when solving the different classification tasks for each subject.

A correctly classified problem indicates that the user has understood at least some of the patterns correctly. If the time for classification is short this increases the evidence that the patterns were easy to understand, on the other hand if the time is longer the patterns apparently required more time to be studied and understood. A wrong classification would indicate that the patterns were hard to understand, especially if the time is also long. If the time is short and the answer wrong it could indicate that the user was not motivated to perform the task or misread the problem or pattern description. One of the cases are slightly different from the others, when the subjects get all the information in the template, in this case the time for reading it will be considerably longer than in the other cases and the absolute time can therefore not be compared to the others.

Variables that affect the measurement

The **experience level and background** of the subjects can be controlled. Subjects should be chosen from two major groups, inexperienced subjects and experienced subjects. Inexperienced indicating experience in modelling but no extensive practical experience in ontology modelling. Experienced users should have practical experience in ontology modelling using OWL. Both groups should be given a very short tutorial on patterns in general before the experiment (as a self-study, no exercises). There should be approximately equal number of experienced and inexperienced subjects selected for the experiment. **Complexity of the problems** can affect the result, this will not be used as a factor of the experiment so the complexity will be held low and (approximately) constant throughout the classification tasks.

Subjects' **familiarity with the domain** of the problems presented cannot be controlled. It is not reasonable to do a complete "background check" on the subjects, but the domain and problem types should be varied across the extent of the experiment to reduce the effect of such pre-existing familiarity. Subjects' own **speed of reading and analysing** the problem. The effect of this will be reduced by not considering absolute times, but only looking at time differences between the classifications of each subject. The **order of the tasks** have an impact on how fast the user is when selecting the correct pattern, if the same pattern has been investigated before, the user might be faster in ruling it out or selecting it. This cannot be controlled in our setting, but the effect is minimised through presenting the tasks in random order to all participants.

Data collection process

The experiment will be set up as a "quiz" accessible on the web. Preparations include the preparation of questions to determine the subjects' backgrounds, a short tutorial text on ontology content patterns and a set of problem descriptions and pattern descriptions (and their combinations into "tasks", the correct classification is set in advance for each such problem). As a preparation for checking the questions before the actual experiment, two pattern experts and two completely inexperienced subjects (not familiar with ontologies) are selected to complete it under supervision from the experiment designers. This is done in order to discover problems with the experiment that requires redesign of some parts, and to validate that the time limits are reasonable. Additionally a questionnaire is prepared to be answered by the subjects after completing the classification tasks. The subjective views of the users on what parts of the pattern template were most useful, and why, will be collected.

3.6.3. Conducting the experiment

Invitations to potential participants will be sent out via e-mail. The "quiz" starts with a survey containing a few questions about the subjects' background and experience, to afterwards being able to divide the subjects into groups. Next, the subjects are instructed to read a short introduction about ontologies and content pattern (the tutorial), including a few examples (with patterns not used in the actual experiment). Then the subject is presented with a set of problems, one at a time, including a problem description and a set of patterns described using a subset of the pattern template. The subject has to select one pattern that he believes solves the problem before pressing the submit button to get the next problem. When the submit button is pressed, the result is evaluated automatically and stored in a database together with the time it took the user to solve the task. Also the order of the tasks is stored for each subject. All subjects will do the same set of tasks but in a random order.

There will be 7 different kinds of content pattern descriptions used (1 – only the formal OWL notation, 2 – OWL+name, 3 – OWL+use-case, 4 – OWL+informal, 5 – OWL+graphical, 6 – OWL+relationships, 7 – all information from the template used in NeOn D5.1.1) that should be tested 3 times each for each user. This indicates that each user will get 21 tasks to solve where each task lets the users select between patterns all described using only one of the 7 alternatives. The maximum time to solve each classification problem is set to 4 minutes, which gives a total maximum time of 84 minutes for completing the quiz (not more than 120 minutes for the complete survey including initial questions and tutorial). The subjects should be instructed to set aside 2 hours without interruption to do the experiment. At the end the subjects fill in a short questionnaire about their experience and give their subjective view on what parts of the template they found most useful and why.

The result is a set of completion times, registration of correct/incorrect answers (connected to their respective task and the order of the tasks), and questionnaire answers stored in a database.

Requirements

There are 21 tasks to be completed by each subject, all tasks are to be completed by all subjects (at least 20 people).

The experiment should be performed with a minimum of two groups of 30 people (experienced and inexperienced). The experiment will be performed distributed over the web. Before the experiment is performed with the subjects, two pattern experts perform it, as well as two novice subjects that are not at all accustomed to ontologies and have got no introduction. Also at least one person is needed afterwards to analyse the results.

An issue in this experiment and in the previous one might be to have the availability of expert users. If we do not reach a critical mass, the experiments will be conducted only with inexperienced users (possibly trained students in the experiment 5).

Analysis procedure

The completion times cannot be validated, except by checking that they are below the maximum time limit. The correct classifications have been determined beforehand, the correction of the tasks is automatically performed based on the answers at runtime of the experiment. The results are also compared to the four “extreme” subjects, used in the preparation of the experiment, and should be within the range between these extremes, otherwise there might be some error that needs to be analysed further.

Time differences will be computed for each subject as additional time compared to the fastest classification of that particular subject. So for example if a subject completes task 3 in his set of tasks in 1 minute and 10 seconds, and this is the fastest time of this subject, the time difference for task 3 is set to 0 seconds. If task 1 was completed in 1 minute and 54 seconds, the time difference for this task will be +44 seconds. For each task of each subject a set of result values are attached, consisting of the ID of the subject, the time difference (as described above), the ID of the task, the position of this task in the subject’s individual random sequence of tasks, and the boolean variable correct/incorrect answer.

At the time of analysis the average time differences and error rates are computed for the subject groups of inexperienced and experienced subjects, and for each of the 7 categories of pattern presentation. The position of the tasks in the sequence is only used to explain exceptional results. The questionnaire results are summed up and compared for the different subject groups, to discover differences and generalize over inexperienced and experienced subjects. Additionally free text suggestions are collected.

Two pattern experts perform the classification beforehand, to determine a probable “lower bound” on the classification times and error rate, and additionally two complete novice users perform the classification to determine an “upper bound”, but these are only used to validate the results.

Experimentation plan

The actual experiment is performed on the web, an invitation will be sent to approximately twice as many people as actually needed.

Initial and final questionnaires: Eva Blomqvist.

Problem descriptions, pattern selection, tutorial text and survey construction and setup: Valentina Presutti.

Pattern experts for initial testing: Aldo Gangemi and Valentina Presutti.

Novice users for initial testing: 2 PhD students from Jönköping University.

Subject groups:

Jönköping University – 3 experienced subjects, 3 inexperienced subjects get the invitation.

NeOn project members and additionally people invited through ontology-related mailing lists.

The experiment will be performed during 2008 (schedule to be determined in February 2008).

The results will be analysed during 2008 (schedule to be determined in February 2008).

Result collection and analysis: Eva Blomqvist and Valentina Presutti.

4. Experiment on the ontology network lifecycle

4.1.1. Experiment Definition:

Experiment motivation.

The idea is to learn about the understandability and usability of the proposed guidelines for helping software developers and ontology practitioners to decide which ontology network life cycle model is the most appropriate for their ontology network and which concrete activities should be carried out in their ontology network life cycle.

Experiments goals.

The goal of the experiment is to test the benefits of using the proposed guidelines for obtaining the ontology network life cycle.

Beneficiaries of the experiments.

Software developers and ontology practitioners involved in developing ontologies will obtain a benefit of this experiment that will serve us to improve the proposed guidelines.

Experiment subject.

Proposed guidelines for helping software developers and ontology practitioners to decide which ontology network life cycle model is the most appropriate for their ontology network and which concrete activities should be carried out in their ontology network life cycle. These guidelines are included in deliverable D5.3.1 [3].

Also, additional material like collection of ontology network life cycle model, NeOn Glossary of Activities and Table of Required-If Applicable Activities. This material is also included in deliverable D5.3.1.

4.1.2. Experiment Design:

Relevant characteristics.

In this experiment, we are going to study the usability and understandability of the guidelines and additional material included in D5.3.1 [3].

Metrics and criteria.

- Which metrics will be used to measure these characteristics?

In this experiment we propose a questionnaire (included in Appendix IV) about the guidelines, in order to be answered by people carry out the experiment.

- Which criteria will be used to interpret these measurements?

For interpreting the measurements, we propose to analyze the answered questionnaire and extract statistics and conclusions.

Variables (time, resources, etc.) that affect the measurement.

- Which variables can affect the measurement and can be controlled?

The experience level and background of people carrying out the experiment. In this case, subjects carrying out the experiments have different experience (in databases, software engineering, etc.) but no extensive practical experience in ontology engineering.

Data collection process.

This experiment is carried out within the "Artificial Intelligence (AI)" master course at Facultad de Informática (Universidad Politécnica de Madrid) with master students, having background in databases, software engineering, and artificial intelligence.

The experiment is divided in the following phases:

1. Lecture will provide to students the proposed guidelines, explanations and additional documents (Collection of ontology network life cycle models, NeOn Glossary of Activities and Table of Required-If Applicable Activities).
2. Students will by pairs follow the guidelines to establish the ontology network life cycle for their problem.
3. Students will document each step followed.
4. Students will fill a questionnaire about the proposed guidelines.

The result document and the questionnaire will be in electronic format (doc files, pdf files, etc.)

Requirements.

The experiment about the establishment of a particular ontology network life cycle should be carried out one time by at least 30 students from the "Artificial Intelligence" master course at Facultad de Informática.

Analysis procedure.

- How will the data be validated?

The data will be validated by means of reviewing the answers in the questionnaire and the output of the experiment.

- How will the data be analysed?

Questionnaire and output will be analysed to extract the conclusions.

Experimentation plan.

The experiment will be carried out during 2007 and 2008 by at least 30 master students attending the "Artificial Intelligence" master course at Facultad de Informática. Students have two weeks for carrying out the experiment using the provided material.

The results of the experiment will be collected and validated after the students finish the experiment.

5. Experiment on ontology reuse

The task: Build an ontology according to a given specification. Reuse knowledge (ontologies) already available on the WWW as widely as possible. Do all this while following the NeOn Reuse Methodology, described in D5.4.1 [8].

The different participant groups:

Group1: Access to the NeOn toolkit and the Internet.

Group2: Access to the NeOn toolkit, the Internet and the Watson Plugin

Group3: Access to the NeOn toolkit, the Internet, the Watson Plugin and Ontology Rating information

Expected outcome:

Group1 will likely take the longest time completing the task, because they have to turn to the Internet for finding potential ontologies. Since they do not have access to the Ontology Rating information, they will have to find a different way to evaluate potentially relevant ontologies. Also they do not have an easy way to integrate found knowledge into the NeOn Toolkit and will have to import it by hand.

Group2 will likely benefit from the integration of Watson and the NeOn toolkit. They can search for reusable knowledge directly from within the toolkit, and also integration is as easy as the click on a button. Only the rating information is not available, so they have to find a way to evaluate whether an ontology is suited for reuse or not.

Group3 should outperform the other groups, because not only do they have integrated access to the Watson semantic gateway, they will also get ontologies already ranked in a way that maximized usefulness. Using the information from the rating system, ontology evaluation is taken out of the hands of the user.

It is likely that all three groups will find the proposed NeOn reuse methodology useful.

5.1.1. Experiment definition

Experiment motivation:

The goal of the experiment is twofold: On the one hand the NeOn reuse methodology is evaluated, and on the other hand the benefit of NeOn technology for facilitating reuse is measured.

Experiments goals:

The first goal is to evaluate whether users think the proposed NeOn reuse methodology is intuitive and easy to follow. Most importantly, data is gathered on how users having access to NeOn technologies perform in a reuse scenario against those with limited access to NeOn tools. The resulting data can be used for scientific publications and as justification for NeOn review meetings.

5.1.2. Experiment design

Relevant characteristics:

Attributes to be studied are "ease of use", "acceptance by user" for both the proposed methodology and the NeOn technologies. "Time used to perform a task" is also captured to compare different user groups with access to different levels of NeOn technologies.

Metrics and criteria:

The characteristics will be measured as: time needed to complete the tasks as well as results of questionnaires filled out by the users. A comparison of results from the three test groups in terms of time taken to complete task, quality of resulting ontology and answers to questionnaires can be used to interpret the measurements. For example, it is better if the task was performed faster, or if users gave more positive feedback.

Variables affecting the measurement:

Because the time is not limited, the complexity of the task will likely affect the time the experiment takes, in case we do not stop the experiment after a predetermined period of time. By deciding on the complexity of the task, the duration of the experiment can be influenced.

A variable that cannot be controlled is the motivation of the test participants. If people are not willing to put effort into the experiment, results could be affected. Also participants could give up if the task seems to be too cumbersome.

Data collection process:

As already used in the WP4 experiments, The Camtasia Screen Capturing Software will be used to record the actions of the users. The process of recording the screen has no overhead during the conduction of the experiment. It will be used to measure the time taken to perform each task and subtask. We will also be able to gain important information about the usability of the underlying NeOn tools. For example whether a user has problem finding certain functionality or spends more time than expected performing a task.

Thus time can be measured easily. Also the resulting ontologies will be saved. A questionnaire will gather the impressions of the candidates.

The Results will then be stored on server. Questionnaires can be Word documents or scans of handwritten paper and stored electronically as well.

Requirements:

The experiment may not be repeated per participant. After one run the participant will already know the solution and might affect the results of the experiment if participating again.

One person supervising the experiment and setting up the machine beforehand can be enough to run the experiment. Depending on time and test candidates / supervisors available, the experiments can be performed in parallel, in small groups or individually. Each participant will be assigned to one group and given access to NeOn technology according to the rules described above. Since the experiments are all independent, it is not necessary that multiple users or the whole test group perform the experiment at once. Potentially better results can be expected when one supervisor interacts with one participant and can provide help along the way. In case of multiple groups performing the experiment in parallel, it would be good to have at least one supervisor per group.

Analysis procedure:

The data is implicitly validated by having multiple groups and multiple participants in each group. Having multiple participants per group evens out personal dependencies, and multiple test group should enable us to interpret the results in correlation with the availability of NeOn technologies. Also usual statistical preprocessing steps can be applied to ensure against outliers.

The data will be analyzed in the following way:

The time taken to perform a subtask (a step in the methodology) will be calculated by analyzing the Camtasia video, which provides the timeline of the experiment. We will then know how long it takes to complete a subtask in the methodology in correlation with availability of NeOn technologies.

The questionnaire will be analyzed using the current best practices in questionnaire evaluation.

In our case the baseline will be group1 which will likely take the longest time to complete the task. Speedups resulting of additional NeOn technology available can then be calculated as difference to that baseline.

Additionally we hope to gain knowledge about users interacting with our tools when performing the analysis of the video.

5.1.3. Experimentation plan

Since missing reuse is one of the current pressing problems in ontology engineering, our potential impact on the community if we prove that we can increase or facilitate reuse by means of our technology is huge. Therefore all academic partners as well as the case study partners should have at least 3 people participating (one per group 1-3). We also need 1 supervisor in each institution.

Given availability of our technology and the rating information the experiments should be performed in September 2008 (or earlier). Given the possibility, it would be good to also get results from students at the 6th European Summer School on Ontological Engineering and the Semantic Web (SSSW-2008)

UKARL will lead the analysis of the experimentation results. Hopefully other institutions can provide support.

The analysis of the experiment will be time consuming, but will be performed within two month after the last dataset is provided by the participating institution.

5.1.4. Experiment phase

In this phase the experiment is carried out.

Some of the technologies needed are already available, the rest is currently under development. Before we can start with the experiment, we have to develop a feasible task based on technology available. The task will be the same for all the groups. Instruction along with a potential solution will be sent to the supervisors at the respective participating institutions.

After the experiments are carried out, the results will be sent to UKARL for analyzes, where they will also be validated.

5.1.5. Analysis Phase

In this phase the experiment results are analyzed. The obtained data is examined and analyzed. We will use current best practices for the statistical analysis of the experiments and the questionnaire. After the results are available, they will be included in an upcoming deliverable and published to a scientific conference.

6. Experiment on ontology reengineering

Ontology Reengineering refers to the process of retrieving and transforming a conceptual model of an existing and implemented ontology into a new, more correct and more complete conceptual model which is re-implemented.

By implementation we mean generating computable models according to the syntax of a formal representation language (e.g., RDF(S), OWL, FLogic).

With this in mind, it is possible for example to:

- transform existing Knowledge Organization Systems (KOS), such as thesauri, glossaries, database schemas, subject directories, etc., into ontologies, by means of semi-automatic methods;
- transform an ontology by reorganizing all its elements (concepts, instances, attributes, etc.) maintaining the same representation language.

Thus, we are proposing to carry out two different types of reengineering experiments:

1. from non-ontological resources to a more defined ontology: from the AGROVOC and the ASFA thesauri to concept based models;
2. from an existing ontology to a different ontological model: from the current AGROVOC ontological model to the NeOn model.

6.1. Experiment 1 - From the AGROVOC and the ASFA thesauri to concept based models

The AGROVOC Thesaurus has been developed by FAO over the years in many languages (today is available online in 18 languages). It has been developed following the traditional ISO standard for the development of monolingual and multilingual thesauri. It is currently stored in a MySQL database[3] and available online at http://www.fao.org/aims/ag_intro.htm.

Since many years several efforts in FAO have been carried out to represent the content of this thesaurus in a more extensive way, in order not only to represent the minimum semantic of a thesaurus, but to make it more closer to a full fledge ontology.

Thus, a new system, called the AGROVOC Concept Server has been identified which reorganize the elements of the thesaurus into a concept based model[4].

ASFA[5] is a specific monolingual English thesaurus for the fisheries domain. It is available in ISO2709 format and XML. It is also available in OWL format (created with Protégé with OWL Plugin 2.0 beta, Build 250). However this file has been created two years ago, prior to the NeOn project and needs to be recreated.

6.1.1. Experiment definition

The objective of this experiment is to identify the best method for converting non-ontological data (from a relational database or XML) into an ontology represented in RDFS or OWL. Several methods will be executed and evaluated.

---

[3] The structure of the database can be seen here: ftp://ftp.fao.org/gi/gil/gilws/aims/kos/agrovoc_formats/db_format.doc

[4] The basic description of the AGROVOC Concept Server OWL model is available here ftp://ftp.fao.org/docrep/fao/009/ah801e/ah801e00.pdf

[5] http://www4.fao.org/asfa/asfa.htm

Experiment motivation

In the context of the Agricultural Ontology Service (AOS) initiative[6], FAO has moved steps forwards on remodelling a thesaurus into an ontology (the AGROVOC Concept Server). The thesaurus is stored in a relational database, but the final ontology should be represented with the OWL language.

The reengineering of a thesaurus into a more formalized entity is necessary in order to allow users of the AGROVOC thesaurus to be able to represent more semantics. In addition to that the final ontology may be maintained in a collaborative and distribute environment, by multiple users worldwide, and therefore relieving FAO from the current centralized maintenance of AGROVOC.

ASFA is one of the main thesaurus used by FAO experts, and is a fundamental component in the new ontology-based infrastructure envisaged for the Fisheries Department in FAO. It is therefore necessary to convert ASFA into a well structured ontology.

Experiments goals

The objective of the experiment is to test and describe methods for reengineering data from thesauri into an ontology.

In this experiment we study several methods for converting non-ontological data to an ontology. The methods under study are:

1. transform the AGROVOC and ASFA thesauri into an RDFS or OWL[7] model using OntoLift[8];
2. transform the AGROVOC and ASFA thesauri into an RDFS or OWL[9] model using ODEMapster[10];
3. express the ASFA thesaurus with the SKOS[11] formal syntax.

No particular order should be respected in performing these tests.

It will be better if at least the first two experiments will be carried out possibly by the same people, so that comparison of the two methods can be done.

The resulting model in the first two experiments should be the one proposed by the AGROVOC Concept Server (the ontology model is given and defined in a paper which will be provided together with the database).

The results of the experiment will be highly beneficial for FAO and for the AGROVOC and ASFA communities, because the final model will be used as backbone of the workbench system, developed by AOS partners (FAO, Kasetsart University - KU), for the collaborative maintenance of the AGROVOC Concept Server.

6.1.2. Experiment design

[6] http://www.fao.org/aims/agrovoccs.jsp

[7] (the ontology model is given and defined in a paper which will be provided together with the database)

[8] http://wonderweb.semanticweb.org/deliverables/documents/D11.pdf and http://wonderweb.semanticweb.org/deliverables/documents/D12.pdf

[9] (the ontology model is given and defined in a paper which will be provided together with the database)

[10] http://www2006.org/programme/files/pdf/p160-slides.pdf

[11] http://www.w3.org/2004/02/skos/

Relevant characteristics

Each conversion should produce a validated RDFS or OWL model. In the first two cases, the resulting ontology in OWL should be identical and should be ready for import in the workbench tool[12] (the tool for maintaining the AGROVOC Concept Server).

The OWL models should validate with an OWL validator (several are already available online).

The SKOS generated model should be validated with the SKOS validator[13].

Metrics and criteria

For the first two methods, the resulting models should have the same number of concepts, instances and properties organized in the same way.

Success criteria will be the validation of the ontology with an OWL validator, the possibility to be edited by Protégé and by the NeOn toolkit, and the compliancy of the resulting model with the workbench tool.

Variables

In the resulting models, the number of concepts, properties and instances can be calculated easily. In addition the number of assigned relationships should be monitored.

No variables are available that can affect the measurement but cannot be controlled.

Data collection process for AGROVOC

The original data are in a MySQL database. There are 3 main tables necessary for the conversion:

1. *agrovocterm*: contains all the terms of the thesaurus in all languages. Contains also indication if a term is a descriptor (statusid greater than 0 and lower than 61) or if it is a non-descriptor (status greater than 61). The status of a concept (field name = statusid) equal to 0 identify deleted terms (in the ontological module they should be marked as deprecated).
2. *termlink*: contains information about relationships between concepts and terms. The definitions of the type of relationships are included in the table *linktype*.
3. *termtag*: contains notes for concepts and terms.

The data transformation should consider that:

- Every concept is identified by every descriptor in the thesaurus.
- Every concept should have a unique URI (generated from the termcode of the corresponding descriptor).
- Every term in every language, independently by their status, should be instantiated as an instance of a class called "lexicalization". The URI of these instances can be composed by merging the language code and the term code.
- Every term in every language representing a concept (all languages of a descriptor plus all the other terms related by the "UF"/"UF+" relationships), instantiated as instances of the class lexicalization (see previous bullet), should be linked to they corresponding concept.
- Concepts should be organized in a hierarchy following the BT/NT relationships between the thesaurus descriptors.
- Concepts should be put in relationships (with a generic "related_concept" object type property) if the corresponding descriptors in the thesaurus are connected by the "RT" relationship.

---

[12] http://naist.cpe.ku.ac.th/agrovoc/

[13] http://esw.w3.org/topic/SkosValidator

- Scope notes, definitions and other types of notes from the *termtag* table should be appropriately linked to corresponding concepts.
- Each language in the same termcode will be considered the exact translations of these terms.
- Each non descriptor can be considered synonym of corresponding descriptors.
- The scopeid field can be used to identify terms which are acronyms, geographical terms or scientific terms.

Data collection process for ASFA

The ASFA thesaurus is available in XML or OWL format. The structure is similar to the one mentioned above for the AGROVOC Thesaurus.

In order to collect the measurements (number of concept, number of relationships, etc.) we would need to analyze and compare the RDFS/OWL models resulting from the conversions.

Requirements

Each elementary experiment (corresponding to a specific method) need to be repeated only once.

It is enough that one person only run the experiments. However, whenever more resources are available, more than one person can execute every single experiment with one method.

Analysis procedure

The analysis of the results of this experiment corresponds to analyse the performances for getting the results and to analyze the resulting models:

- How long took to execute the experiment?
- How complex was to set it up?
- Is the result validated by a specific ontology validator? By multiple validators?
- Is the result compliant with the AGROVOC Concept Server model or the SKOS requirements?
- Can the resulting model be edited with an RDFS/OWL editor? Which one?
- Can the resulting model be edited with the NeOn Toolkit?
- If the model is in the OWL format can it be loaded in the AGROVOC Concept Server Workbench tool?

Experimentation plan

The experiment should follow these steps:

1. The first step is to analyze the MySQL databases or XML files which constitutes the source of the experiment. The experimenters need to be familiar with its structure and content.
2. On the 3 cases (OntoLift, ODEMapster, SKOS) the experimenters needs to be familiar with the characteristics of these methods. Literature and web sites are available on these methods.
3. The experimenters need to understand how to transform each element of the original data into final model elements. Any further request on this topic can be addressed to FAO (Margherita.sini@fao.org or FAO-AGRIS-CARIS@fao.org).
4. The experimenters need to identify the methodology for running the experiment, for example identifying how to run corresponding tools or to develop specific needed tools.
5. Run the experiment.
6. Document and describe results.

There is no need of concurrent process (the three methods can be executed in different moments or at the same time). We assume that every method for every source should not require more than one working week.

The following table identifies actors in the experiments:

Experiment name	Experimenters	Testers and Validator	Analysis of the experiment results
1. OntoLift	To be decided	To be decided	To be decided
2. ODEMapster	To be decided	To be decided	FAO
3. SKOS	To be decided	To be decided	FAO

Table 2. Actors in the experiments

The experiments can be executed prior to month M36 (February 2009).

6.1.3. Experiment phase

In this phase the experiment is carried out, described, and the results will be identified.

Necessary tools are already available, but customization or adaptation may be needed. If no tools are available, they will be specifically developed if needed.

It will be necessary to validate the obtained results with a validator. If results are not validated the methods needs to be adjusted and the experiment re-executed.

Keep track of how long the experiment took.

6.1.4. Analysis phase

In this phase the results will be examined:

- Is the final OWL model validated?
- How big is the model?
- Can be loaded with ontology tools?

Is it possible to compare two methods and identify the most efficient one?

This analysis phase will involve domain experts from FAO fishery department and the domain experts that usually maintain AGROVOC; however the evaluation of the mentioned methods requires Information and Knowledge experts.

6.2. Experiment 2 - From the current AGROVOC ontological model to the NeOn model (including the Lexical Informations - LIR)

As mentioned, AGROVOC is a thesaurus which is currently being restructured as an Ontology. It exists already as OWL file, with basic distinction between terms and concepts.

In the context of NeOn, all lexical information related to concepts have been organized with a specific model and stored into an ad-hoc repository.

It would be interesting to re-model the AGROVOC ontology as an ontology following NeOn indications and lexical model structures.

6.2.1. Experiment definition

The objective of this experiment is to use the ONIONS[14] method to generate a large ontology with a specific structure (the NeOn + LIR structure) starting from an ontology derived by a thesaurus related to the agricultural domain.

Experiment motivation

The AGROVOC thesaurus is of high relevance within and outside FAO. In light with the new semantic technologies, FAO is interested in representing AGROVOC as an ontology. This ontology may also be maintained by FAO with the NeOn toolkit.

The objective of this experiment is to test the ONIONS mechanism for the representation of the AGROVOC ontology as a NeOn ontology: it would be needed to re-model the AGROVOC current ontology following the NeOn model (including the Lexical Information Repository) in order to be sure that the ontology support multilinguality, that the ontology can be used in a networked environment, and that the ontology can be edited by the NeOn toolkit.

Experiments goals

The objective of this experiment is to test the ONIONS methodology by converting an existing basic ontology (the AGROVOC thesaurus available in OWL format), into a NeOn model that has been identified as better for a multilingual, dynamic, multi-collaborative environment such as FAO.

The beneficiaries of the experiments will be the AGROVOC editors because it will be easier for them to maintain the resource, the FAO users because could access data online and eventually to contribute to their maintenance, all AGROVOC users all over the world because would be able to add lexicalizations in their language, FAO tools and systems because could access a dynamic resource making sure that legacy systems will work. In addition, the obtained AGROVOC ontology may interact with other FAO ontologies (e.g., biological species, ASFA).

6.2.2. Experiment design

Relevant characteristics

In this experiment is under study the ONIONS method for converting one ontology model to another more complete model, which will make use o the Lexical Information Repository (LIR) as identified in WP2 of the NeOn project, and which allow the definition of other characteristics such as versioning, mappings, etc.

Metrics and criteria

In order to provide a correct evaluation of the ONIONS methodology, the experiment will need to be carried out twice: one time following the methodology and another time without following the methodology.

The measurable indicators for this experiment are:

1. the time spent on producing the resulting model (with and without the methodology);
2. the correctness of the final resulting model.

The final model will need to be validated with an OWL validator (for the lexical syntax) and be edited with the NeOn toolkit (for the structural validation). It should also interact easily with other ontologies may be available in the net.

[14] http://www.loa-cnr.it/Papers/dke.pdf

Success is obtained if we get validation of the restructured ontology. We will have a failure if the validation of the new ontology fails or if the new ontology is incomplete or cannot be easily maintained.

Variables

Time is an issue: it is question to determine if, when restructuring an ontology, the current proposed ONIONS methods could help to have a result in a faster way, ensuring correctness.

The validation of the model and the ability to represent all AGROVOC data in correct way are key factor for success.

Data collection process

The AGROVOC ontology is available in OWL format. This file needs to be studied to analyze which elements needs to be added or converted into specific elements of the NeOn model.

The measurements of the results involve an OWL validator, the NeOn toolkit, including plug-ins.

The final AGROVOC ontology will be documented and published in the FAO AIMS web site, in Watson, as well as in the most known ontology repository (schemaweb.org, froogle, etc.).

In general we identify that:

- Every term (descriptor) in AGROVOC can become a concept in the new ontology model. The URI will need to be determined for every concept.
- If language variances are available in the original data (e.g., for English or Spanish) these should be represented in the LIR model.
- Relationships between descriptors in the AGROVOC thesaurus will be transformed in object type relationships.
- Non descriptors should be properly represented in the LIR model or as synonyms or other variances for descriptors.

Requirements

The experiment needs to be repeated twice: once with the use of the ONIONS method and once without consider this methodology. Two groups of people will be needed for this task:

	Step 1	Step 2
Group A	Execute ontology transformation without following any specific method	Execute the ontology transformation following the ONIONS method.
Group B	Execute the ontology transformation following the ONIONS method.	Execute ontology transformation without following any specific method

Table 3. Steps and groups of people in the experiment

The competences of the two groups should be ontology design (e.g., they can be ontology experts or engineers).

FAO will be made available original data to start from.

Analysis procedure

The final data will be validated with an OWL validator and the NeOn toolkit. The validation should also verify the possibility of editing of the ontology, the possibility of versions and ensure compatibility with legacy systems.

The data will be analyzed manually (no applications are available to verify automatically consistency between the final model and the original data).

The results of the 2 steps of the experiment should be compared each other by both groups of people. The two groups should also confront each other in order to see what has been done and how with and without the NeOn methodology.

Experimentation plan

The two cases of this experiment will need to perform several steps:

	Case description	Steps
Case 1	Execute ontology transformation without following any specific method	1. Identify the original data 2. Identify the model of the ontology that should be produced, including the LIR section (linguistic information) 3. Represent the old ontology based on model identified in point 2.
Case 2	Execute the ontology transformation following the ONIONS method.	1. Identify the original data 2. Identify the model of the ontology that should be produced, including the LIR section (linguistic information) 3. Study the ONIONS method and identify steps to do to follow this methodology 4. Represent the old ontology based on model identified in point 2 following the steps analyzed in point 3.

Table 4. Steps for each case

The experiment can be performed prior to month M36 (February 2009).

Some tools will need to be available (e.g., OWL validator, NeOn toolkit).

6.2.3. Experiment phase

In this phase the experiment is carried out, described, and the results will be identified.

Necessary tools are already available, but the ONIONS method needs to be analyzed.

It will be necessary to validate the obtained results with a validator. If results are not validated the methods needs to be adjusted and the experiment re-executed.

The experts of the two groups need to keep track of how long the experiment took with and without the methodology.

6.2.4. Analysis Phase

In this phase the experiment results are analyzed. The obtained data are examined and compared.

Questions to be answered will be:

- Are the results obtained without using the ONIONS method validated?
- Are the results obtained using the ONIONS method validated?
- How long took to produce a result without using the ONIONS method?
- How long took to produce a result using the ONIONS method?
- Was the ONIONS method useful to perform the task?
- Could the results obtained without using the ONIONS method be loaded with ontology tools?
- Could the results obtained using the ONIONS method be loaded with ontology tools?
- Was the ONIONS method useful to perform the task?

7. Experiment on ontology mapping

Mapping (also known as Ontology Alignment) refers to finding the correspondences between entities of two or more ontologies or ontology modules and storing them for later use.

For FAO this is a crucial activity in particular for resources that may contain overlapping information, such as the AGROVOC multilingual thesaurus, the FAO Terminology system, and the Biological Species Ontology.

It would be of high utility for FAO users, to create a mapping between these resources. Therefore it is proposed to execute tests of the mapping experiments on these FAO real data.

Currently there are several methods for mapping ontologies. Some of the most known are[15]:

- COMA++ (Sabine Massmann, massmann@informatik.uni-leipzig.de)
- Falcon-AO (Wei Hu, whu@seu.edu.cn)
- HMatch (Alfio Ferrara, alfio@balsi-informatica.it)
- PRIOR (Ming Mao, berylmm@gmail.com)
- RiMOM (Yi Li, ly@keg.cs.tsinghua.edu.cn)

Other methods exist (such as SAMBO and AOAS), but they are more domain-dependent and therefore not applicable for our purposes.

One of the best identified between these methods is recognized to be Falcon-AO and therefore the one suggested to be used within this experiment. Alternatively another method could be used (e.g., the Alignment server by INRIA).

7.1.1. Experiment definition

For the mapping experiment we will consider the following resources:

- The AGROVOC thesaurus: existing in more than 18 languages contains around 30000 concepts organized hierarchically. Associative relationships between concepts are available.
- The FAOTERM terminology database: exists as XML repository containing over 70000 entities in 5 or sometimes 6 languages. No relationships are available between the entities except for terminological relationships between terms of the same concept (e.g., translations, synonymies, spelling variances).
- The Biological Species Ontology16.

Objective of the experiment is to test the Falcon-AO methodology or the Alignment server methods for mapping these huge resources.

Experiment motivation

All these resources (AGROVOC, FAOTERM and the Biological Species Ontology) contain overlapping information, such as the concepts identifying biological entities (in AGROVOC and in the Ontology organized in a taxonomy: genus-family-species).

[15] See also ftp://ftp.inrialpes.fr/pub/exmo/publications/euzenat2007g.pdf

[16] http:/www.fao.org/aims/aos/fi/species_v1.0.owl

The Biological Species ontology manages reference data about biological species needed for fisheries fact sheets and statistical information, among other resources. Species items are organized and maintained in the Aquatic Science and Fisheries Information System (ASFIS) and currently includes nearly 11.000 species items related to Fisheries and Aquaculture.

In FAO, it is needed to map the concepts from these 3 mentioned resources.

We would like to map these resources and evaluate the mapping method and the result of the mapping, in order to identify it possible later reuse in other mapping projects. In case the mapping method is not yet incorporate in the NeOn Toolkit as plug-ins, and in case of successful mapping results, the experiment will identify this method as a possible resource for the Toolkit.

Experiments goals

FAO is aiming to create a connection between AGROVOC, FAOTERM and the Biological Species Ontology. The experiment will be beneficial for FAO as one of the results will be the mapping between these resources. However, the real motivation for the experiment is to test the mentioned mapping technique between ontologies so that, if of an acceptable quality, this technique may be incorporated in the NeOn toolkit.

The result of the mappings will be made available to the entire agricultural community through the registry of mappings of the AIMS[17] web site. All users already using these resources will benefits from the mapping: they can access a more complete resource, by taking information from the other mapped resources.

Experiment subject

Under study with this experiment is the Falcon-AO mapping method (other methods could be also used). The results of this method will be compared to other mappings executed in FAO, or by other partners, without following this method.

7.1.2. Experiment design

Relevant characteristics

Mapping is done at the concept level. We do not consider other mappings.

The mapping will be represented in OWL format putting together the URI of the mapped concepts and the identifier of the mapping relationship. Alternatively a set of triple identifying the mapped concepts and the mapping relationships will be enough (e.g., represented using an excel sheet).

Metrics and criteria

To analyze the results, we should consider the total number of objects mapped between the mentioned resources. The success criterion is that all concepts from the biological species will be mapped to both elements in AGROVOC and FAOTERM. Not all AGROVOC and FAOTERM concepts can be mapped to the biological species ontology because these 2 sources are much more general and cover other domains and not only biological entities.

- All concepts from Biological Species Ontology should be mapped to AGROVOC and FAOTERM. If some of them are not mapped is because they do not exists in AGROVOC and FAOTERM and in this case we needs to mark them as non-mapped.
- All concepts representing fishery biological species from AGROVOC should be mapped to the Biological Species Ontology and to the corresponding concepts in FAOTERM.

[17] http://www.fao.org/aims/kos_intro.htm

- All concepts representing fishery biological species from FAOTERM should be mapped to the Biological Species Ontology and to the corresponding concepts in AGROVOC.

Variables

AGROVOC contains currently 39635 terms in English. Some of these (around 13000) cover fishery species names.

All FAOTERM terms related to fisheries biological species can be obtained by filter using the subject "Fish Names and Species": around 24565 terms. These may contains scientific names and common names in multiple languages.

The Biological Species Ontology contains 11571 species.

The experiments will need ontology-aware people or IT experts who will need to manage large schemas.

In order to evaluate the quality of the mappings, one or more domain expert(s) may be consulted.

Time is not crucial as the mapping will be executed only once. However, for the purpose of the experiment, if there is a time constraint problem, the experiment can be performed on a subset of the data.

The equipment available for the experiment should be compliant with the size of the resource used in the experiment in order not to effect the computational time.

If tools are used to implement the mapping method, a user-friendly interface may speed up the process.

Data collection process

The resources should be represented in a form suitable for this method (possibly RDFS, SKOS or OWL). FAO team will take care to provide the resources in the required format.

Based on the deadlines and the available time the experiment can be performed with a subset of the original data.

In order to analyze the results will be important to evaluate:

- how long the mapping took compared to resource size;
- the quality of the mapping;
- the type of output presentation.

Requirements

The mapping between the FAO resources will be carried out following the Falcon-AO method (or analogous) and evaluated over other mappings already executed in FAO using a manual approach.

The experiment can be executed only once. There is no need to perform the same mapping for multiple times.

Analysis procedure

Once the mapping will be done there will be the need of evaluating the quality of the mapping. This may require a domain expert. This evaluation needs to be done by hand because no automatic tools exist for this task.

The final evaluation will need to consider:

- How fast was the execution of the mapping?
- How reliable?
- Were all the possible mapping relationships used?
- Were any parameters in the execution of the methodology affecting the quality of the result?

Experimentation plan

To execute this mapping we need at least a person who is aware of the following concepts:

- what is an ontology;
- what is a mapping and what types of mapping relations exist;
- be aware of mapping methodologies;
- be aware of tools to execute mappings;
- be aware of mapping serialization (e.g., OWL)

The experiment can be performed prior to month M36 (February 2009).

FAO will analyse the results of the experiments just after they have been performed.

7.1.3. Experiment phase

In this phase the experiment is carried out, described, and the results will be identified. There should be 3 results:

- the mapping between AGROVOC and FAOTERM;
- the mapping between AGROVOC and the Biological Species Ontology;
- the mapping between FAOTERM and the Biological Species Ontology;

Necessary tools will be identified prior the execution of the experiment.

7.1.4. Analysis Phase

During this phase the validity of the mappings method will be analyzed. Evaluators will need to identify:

- How many concepts have been automatically mapped?
- How many concepts have not been mapped?
- What is the quality of the automatically mapped concepts? An analysis of the quality of the mappings will be needed.
- How long took to execute the mapping using this methodology?
- How good is this mapping method compared to other mappings executed in FAO not using this method?

The replies to these questions will be documented and reported to the NeOn partners.

8. Experiment on ontology localization

Typically ontologies are described in a determined natural language only. Thus, in order to achieve a more generally applicable ontology, it is necessary to guarantee that the same knowledge be recognizable among different natural languages. LabelTranslator, a plug-in for the NeOn Toolkit automatically localizes ontologies in English, Spanish and German. The Ontology Localization Activity (OLA) consists of adapting an ontology to a concrete language and culture community, as defined in [3]. LabelTranslator takes as input an ontology whose labels are expressed in a source natural language, and obtains the most probable translation of each ontology label into a target natural language. We expect that the automatization of the translation process reduces the human efforts to localize manually the ontology. For the experiment, we have defined some metrics to evaluate the quality of the ontology translations. This is done on the base of the comparison of the translations provided by an expert ("gold standard") in front of the translations provided by the ranking algorithm used in our plug-in. A set of interfaces has been implemented to capture the subjective scores and parameters about the translations of each ontology label.

8.1.1. Experiment Definition

The goal of this experiment is to evaluate some aspects of the translation ranking method (used by LabelTranslator) which tries to select the most appropriate translation for each ontology label. In particular, we evaluate two aspects of the ranking method: the *obtained output using manual and automatic operation*, and the *quality of translation*. Based on these aspects we define some metrics to see whether the plug-in facilities the automatic localization of ontologies among different natural languages.

In the case that the results of the experiments show that LabelTranslator plug-in facilitates the automatic localization of ontologies the main beneficiaries will be the Neon users and UPM team. We hope additionally that the obtained results enable us to verify the impact of the different lexical resources used to obtain the translations of each ontology label.

8.1.2. Experiment Design

For this experiment we decide to apply a manual evaluation on two relevant characteristic of our plug-in: 1) quality of translation and, 2) manual and automatic operation of the translation algorithm. The manual evaluation was done by Spanish speakers with a good level of English. One of the evaluators had a fluid level of German. In all the experiments a reference translation ("gold standard") provided by the evaluators was used. The gold standard allows users to compare the quality of the translations provided by an expert in front of the translations provided by the algorithm. In the following we give a short overview of the metrics used to evaluate the relevant characteristics described above.

Metrics

Accuracy: In order to evaluate the quality of the output of the ranking method in automatic operation mode we propose a measure of accuracy. The accuracy measures the capacity of the algorithm of translation to get in an automatical way a correct translation according to context. To measure the accuracy of the algorithm, we counted the number of times where the first translation was correct.

$$\text{Accuracy} = \frac{\text{number of times where the first translation is correct}}{\text{number of labels of the ontology}}$$

Precision and Recall: The previous measure does not allow checking the completeness of the translations since it does not observe the behaviour of all the translated labels. Thus, we have measured *precision* as the number of correct translations of all the translations provided by the system divided by the total number of translations provided by the system. In other words the *precision* measures the capacity of the translation algorithm to provide correct translations according to the context. *Recall* measures the capability of the translation algorithm to provide all correct translations according to the context. Recall measure is calculated dividing the number of correct translations of all the translations provided by the system for the number of correct translations (provided by the gold standard). To calculate both measures each evaluator identifies for each ontology label which one is a correct translation.

$$\text{Precision} = \frac{\text{number of correct translations of all the provided for the system}}{\text{total number of translations of all the provided for the system}}$$

$$\text{Recall} = \frac{\text{number of correct translations provided for the system}}{\text{number of correct translations}}$$

Adequacy and Fluency: In order to measure the quality of the translation of compound labels we propose a subjective 1-5 score for adequacy and fluency. The adequacy measures the capability of the translation algorithm to determinate the quantity in that the meaning of a correct translation is preserved. On the other hand, the fluency measures the capability of the algorithm to determine how good the corresponding language is. In this experiment, each evaluator assigned fluency and adequacy ratings for each translated label. Each score ranges from one to five (with one being the poorest grade and five the highest).

$$\text{Adequacy} = \text{percentage of obtained values in each category of adequacy}$$

$$\text{Fluency} = \text{percentage of obtained values in each category of fluency}$$

The used criteria to interpret the measures above described are: for accuracy, precision, and recall measures we expect to obtain values near one. For adequacy and fluency the values ranges of one to five (with one being the poorest grade and five the highest). In our initial tests we have detected at least two variables that can affect the measurement: the required time to perform the evaluation and the motivation of the evaluators. In the first case, the number of concepts, attributes and relationships of an ontology can increment the time required to perform the experiment. We propose using a sample set of ontological terms to improve this problem. The second variable (motivation) cannot be controlled, thus the results could be affected.

Data Collection Process

In order to collect the measurements a set of interfaces has been implemented. These interfaces allow gathering the subjective scores and parameters about the translations of each ontology label. All data are stored in memory and the results exported to a file.

Each elementary experiment is performed only one time per participant. Also, in order to run the experiment is necessary one person to supervise. With the purpose of the validate the data, a specific experiment will be performed for different evaluators, the scores of two evaluators for each ontology label will be average together, and an overall average score will be calculated for each evaluated ontology. Some statistical operations will be performed on data.

Experimentation Plan

An initial experiment evaluation has been performed by four members of the SID group of the Zaragoza University. The ontology corpus used for this evaluation was selected from the set of Knowledge Web [4] ontologies used to manage EU projects. All ontologies are described in English. The corpus statistics are given in Table 5. In order to evaluate English-Spanish translations, all ontologies were used. However, in order to evaluate English-German translations, we used the last three ontologies only. The data of each evaluator was analyzed for the developer group of LabelTranslator.

Ontology Domain	Number of Ontological Terms			% Compound labels	
	concepts	**attributes**	**relationships**	**≤ 3 tokens**	**> 3 tokens**
Documentation&Meeting	42	61	22	44%	25.6%
Person&Project	25	18	12	47.2%	10.9%
Organization	10	7	11	46.4%	7.1%
Office	20	12	8	12.5%	0%
University	30	10	12	17.3%	0%

Table 5. Ontologies corpus statistics

Our experimental results (see Analysis Phase section) encourage us to tackle further improvements and tests to our ranking algorithm. Thus, we are planning to extend the evaluation using as evaluators the master students of the course "Ontology and Semantic Web" from UPM. In this case each student performs the evaluation on the ontology created during the course. These ontologies could be described in English or Spanish language. All evaluators are Spanish speakers with a good level of English.

8.1.3. Analysis Phase

In Table 6 we show the results achieved by our prototype in each experiment. The values are organized by target language. All percentages of adequacy and fluency shown in this table correspond to those translations punctuated with a value greater than 4. The obtained results reveal several results. First, the worst value of accuracy was obtained for the documentation ontology. This is not surprising, because this ontology has a 25.6% of composed labels with more than three tokens, which cannot be managed in this version. Second, the percentages of adequacy and fluency obtained for English-German composed label translations are lower than the percentages of the English-Spanish ones. Our explication is that a major effort was put (in the current version) for the learning of templates between English-Spanish languages.

Ontology Domain	Spanish					German				
	Accu.	***Prec.***	***Rec.***	***Adeq.***	***Flu.***	***Accu.***	***Prec.***	***Rec.***	***Adeq.***	***Flu.***
Documentation	0.59	0.57	0.84	68%	91%	-	-	-	-	-
Person&Project	0.73	0.71	0.81	89%	93%	-	-	-	-	-
Organization	0.81	0.79	0.78	87%	95%	0.73	0.71	0.64	73%	69%
Office	0.79	0.77	0.88	93%	95%	0.78	0.76	0.74	67%	76%
University	0.80	0.78	0.87	96%	93%	0.71	0.69	0.71	69%	73%

Table 6. Results obtained for each metric

9. Experiment on ontology specification

In this section we provide a preliminary definition and design of this experiment; more detailed definition and design will be included in the next version of this deliverable.

9.1.1. Experiment Definition

Experiment motivation.

The main idea is to learn about the understandability and usability of the proposed methodological guidelines for carrying out the ontology (requirement) specification activity.

Experiments goals.

The goal of the experiment is to test the benefits of using the proposed methodological guidelines for obtaining the ontology requirement specification document (ORSD) as output of the ontology (requirement) specification activity.

Beneficiaries of the experiments.

Software developers and ontology practitioners involved in developing ontologies will obtain a benefit of this experiment that will serve us to improve the proposed methodological guidelines.

Experiment subject.

Proposed guidelines for carrying out the ontology (requirements) specification activity, included in deliverable D5.4.1 [8].

9.1.2. Experiment Design

Relevant characteristics.

In this experiment, we are going to study the usability and understandability of the methodological guidelines included in D5.4.1 [8] for carrying out the ontology (requirement) specification activity.

Metrics and criteria.

- Which metrics will be used to measure these characteristics?

In this experiment we propose a questionnaire about the methodological guidelines, in order to be answered by people carry out the experiment.

- Which criteria will be used to interpret these measurements?

For interpreting the measurements, we propose to analyze the answered questionnaire and extract statistics and conclusions.

Variables (time, resources, etc.) that affect the measurement.

- Which variables can affect the measurement and can be controlled?

The experience level and background of people carrying out the experiment. In this case, subjects carrying out the experiments have different experience in databases, software engineering, etc.) but no extensive practical experience in ontology engineering.

Data collection process.

This experiment will be carried out within the "Artificial Intelligence (AI)" master course at Facultad de Informática (Universidad Politécnica de Madrid) with master students, having background in databases, software engineering, and artificial intelligence.

The attending students will be divided into two groups and each group will follow a different version of the methodological guidelines (Guidelines-1 and Guidelines-2).

The experiment is divided in the following phases:

1. Lecture will provide to students the proposed guidelines (1 or 2).
2. Student groups will follow the methodological guidelines proposed to carry out the ontology (requirements) specification activity.
3. Students will document in detail each task proposed in the methodological guidelines.
4. Students will fill a questionnaire about the proposed methodological guidelines.

Requirements.

The experiment about the ontology (requirement) specification activity should be carried out one time by at least 30 students from the "Artificial Intelligence" master course at Facultad de Informática.

Analysis procedure.

- How will the data be validated?

The data will be validated by means of reviewing the answers in the questionnaire and the output of the experiment.

- How will the data be analysed?

Questionnaire and output will be analysed to extract the conclusions.

Experimentation plan.

The experiment will be carried out during 2007 and 2008 by at least 30 master students attending the "Artificial Intelligence" master course at Facultad de Informática. Students have two weeks for carrying out the experiment using the provided material.

The results of the experiment will be collected and validated after the students finish the experiment.

10. Conclusions

This deliverable has presented the first planning of the experiments to be performed over the NeOn methodologies and methods. We have obtained results of two of these experiments related to ontology design patterns. The rest of the experiments have been planned and defined, will be performed in 2008, and their results will appear in the next version of this deliverable.

In the future, more experiments will be proposed for the rest of the NeOn methodologies and methods. Examples of future experiments are to perform experiments over different methods of networked ontology evaluation (either developed inside the NeOn project or outside it) or experiments to evaluate the methodology for the development of large scale Semantic Web applications that will be defined in WP5.

As the NeOn Toolkit must provide support to the methods defined in NeOn, and this support will be one of the key points in the adoption of both the NeOn Toolkit and the NeOn methodologies and methods, future experiments will also involve a deep evaluation of the NeOn Toolkit support to the NeOn methodology for developing networked ontologies.

11. References

1. P. Haase, S. Rudolph, Y. Wang, S. Brockmans, R. Palma, and J. Euzenat, M. d'Aquin. *NeOn Deliverable D1.1.1 Networked Ontology Model*. November 2006.

2. M. C. Suárez-Figueroa, S. Brockmans, A. Gangemi, A. Gómez-Pérez, J. Lehmann, H. Lewen, V. Presutti, M. Sabou. NeOn Deliverable D5.1.1. NeOn Modelling Components. NeOn Project (http://www.neon-project.org). February, 2007.

3. Suarez-Figueroa, M.C. (coordinator). NeOn Development Process and Ontology Life Cycle. NeOn Project Deliverable 5.3.1 (2007).

4. O. Corcho, A. López-Cima, and A. Gómez-Pérez. The ODESeW 2.0 semantic web application framework. In *Proceedings of the 15th International Conference on World Wide Web (Edinburgh, Scotland, May 23 - 26, 2006). WWW '06*, pages 1049-1050. ACM Press, New York, NY, 2006.

5. V.R. Basili, R.W. Selby, and D.H. Hutchens. Experimentation in software engineering. IEEE Transactions on Software Engineering 12, (7):733–743, January 1986.

6. S.L. Pfleeger. Experimental design and analysis in software engineering, part 2: How to set up and experiment. ACM SIGSOFT Software Engineering Notes, 20(1):22–26, January 1995.

7. B.A. Kitchenham, S.L. Pfleeger, L.M. Pickard, P.W. Jones, D.C. Hoaglin, K. El-Emam, and J. Rosenberg. Preliminary guidelines for empirical research in software engineering. IEEE Transactions on Software Engineering 28, (8):721–734, 2002.

8. M.C. Suárez-Figueroa, K. Dellschaft, E. Montiel-Ponsoda, B. Villazón-Terrazas, Z. Yufei, G. Aguado, A. García, M. Fernández-López, A. Gómez-Pérez, M. Espinoza, M. Sabou. D5.4.1 NeOn Methodology for Building Contextualized Ontology Networks. March 2008.

9. V. Presutti, A. Gangemi, S. David, G. Aguado de Cea, M. C. Suárez-Figueroa, E. Montiel-Ponsoda, M. Poveda. NeOn Deliverable D2.5.1. A Library of Ontology Design Patterns: reusable solutions for collaborative design of networked ontologies. NeOn Project. February 2008.

APPENDIX I. Material for ontology design pattern experiment 1

Modelling problems

Suppose that some one in a company wants to model tasks involved in a plan. Types of tasks are management tasks, financial tasks, marketing tasks and control tasks. Plans can be also classified within the same tasks categories.

We know that:

- Plans and tasks are disjoints.
- Business plans have business tasks, apart from other tasks.
- Control tasks can be only begin task, end task or sequential task.
- A concrete task can only be classified in one of these categories.
- A research plan is part of a research project and that a research plan is composed by a theoretical plan and an experimental plan.
- Plans could be in one of the following stages: accepted, non-accepted, in process of revision.

The modelling problems proposed are the following ones:

1. Model that tasks are management tasks, financial tasks, marketing tasks and control tasks.
2. Model that plans are management plans, financial plans, marketing plans and control plans.
3. Model that plans and tasks are disjoints.
4. Model that control tasks can be only begin task, end task or sequential task; and that a concrete task can only be classified in one of these categories.
5. Model that business plans have business tasks, apart from other tasks.
6. Model that Plan-1 is a concrete business plan, Task-1 is a concrete business task, and that Plan-1 has Task-1.
7. Model that a financial department carries out a financial plan for the marketing department.
8. Model that research plan is composed by a theoretical plan and an experimental plan.
9. Model that a research plan is part of a research project.
10. Model that experimental plans and theoretical plans are part of the research project assuming that part-of is transitive.
11. Model that experimental plans and theoretical plans are part of the research project assuming that part-of is not transitive.
12. Model the stage of a plan as a property of the plan.

13. Model the stage of a plan as a class related to plans.

Solutions to modelling problems

Suppose that some one in a company wants to model tasks involved in a plan. Types of tasks are management tasks, financial tasks, marketing tasks and control tasks. Plans can be also classified within the same tasks categories.

We know that:

- Plans and tasks are disjoints.
- Business plans have business tasks, apart from other tasks.
- Control tasks can be only begin task, end task or sequential task.
- A concrete task can only be classified in one of these categories.
- A research plan is part of a research project and that a research plan is composed by a theoretical plan and an experimental plan.
- Plans could be in one of the following stages: accepted, non-accepted, in process of revision.

The modelling problems proposed are the following ones:

1. Model that tasks are management tasks, financial tasks, marketing tasks and control tasks (LP-SC-01).
2. Model that plans are management plans, financial plans, marketing plans and control plans (LP-SC-01).
3. Model that plans and tasks are disjoints (LP-DI-01).
4. Model that control tasks can be only begin task, end task or sequential task; and that a concrete task can only be classified in one of these categories. (LP-EC-01)
5. Model that business plans have business tasks, apart from other tasks (LP-OP-01).
6. Model that Plan-1 is a concrete business plan, Task-1 is a concrete business task, and that Plan-1 has Task-1 (LP-IN-01).
7. Model that a financial department carries out a financial plan for the marketing department. (LP-NR-01 | LP-NR-02)
8. Model that research plan is composed by a theoretical plan and an experimental plan. (CP-PW-01 | CP-PW-02)
9. Model that a research plan is part of a research project. (CP-PW-01 | CP-PW-02)
10. Model that experimental plans and theoretical plans are part of the research project assuming that part-of is transitive. (CP-PW-01 | CP-PW-02)
11. Model that experimental plans and theoretical plans are part of the research project assuming that part-of is not transitive. (CP-PW-01 | CP-PW-02)

12. Model the stage of a plan as a property of the plan. (LP-DP-01)

13. Model the stage of a plan as a class related to plans. (LP-SV-01 | LP-SV-02).

APPENDIX II. Results of the ontology design pattern experiment 1

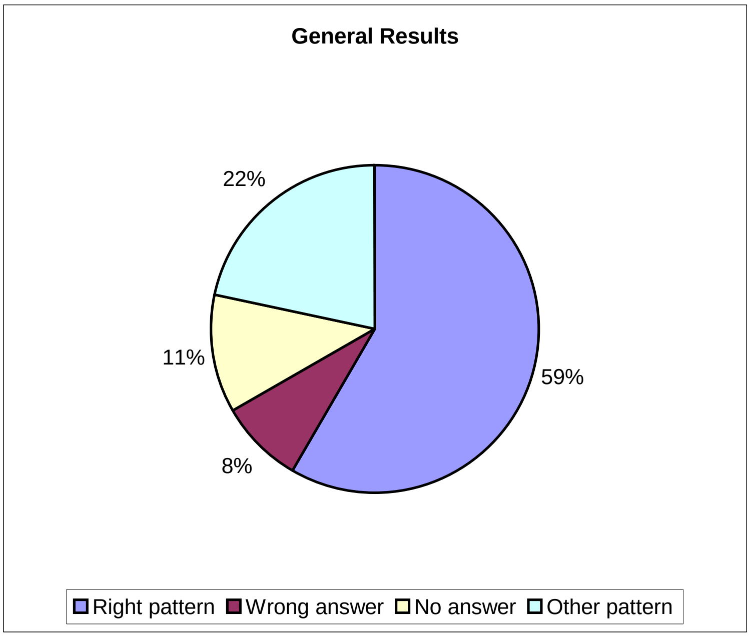

Figure 2. General Results

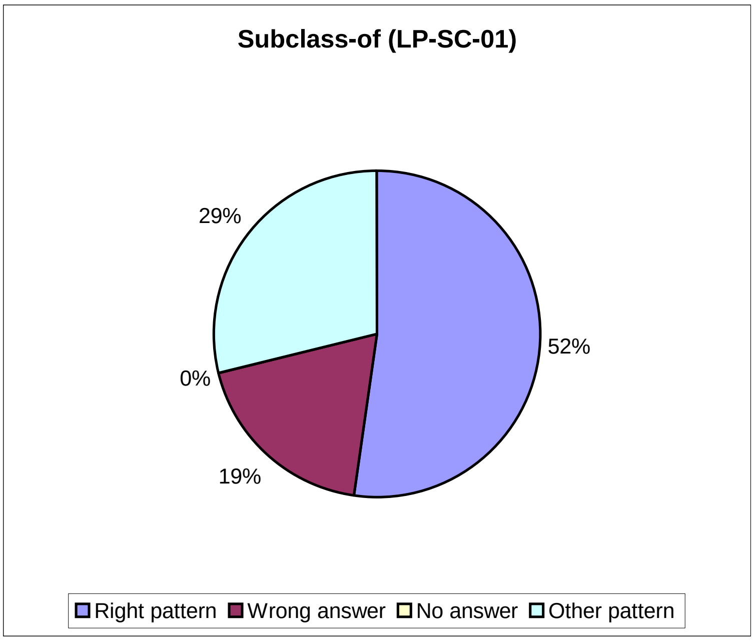

Figure 3. General Results for Modelling Problem 1

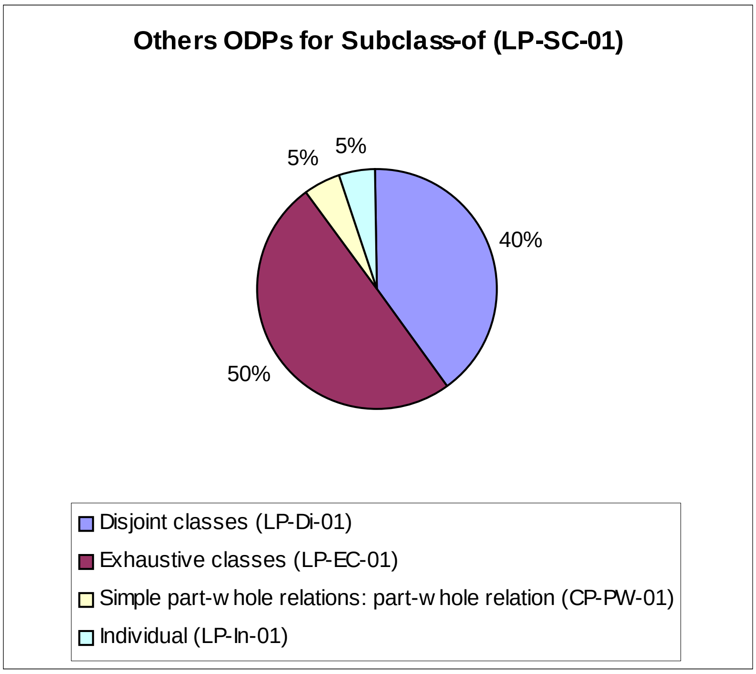

Figure 4. Non Correct Pattern for Modelling Problem 1

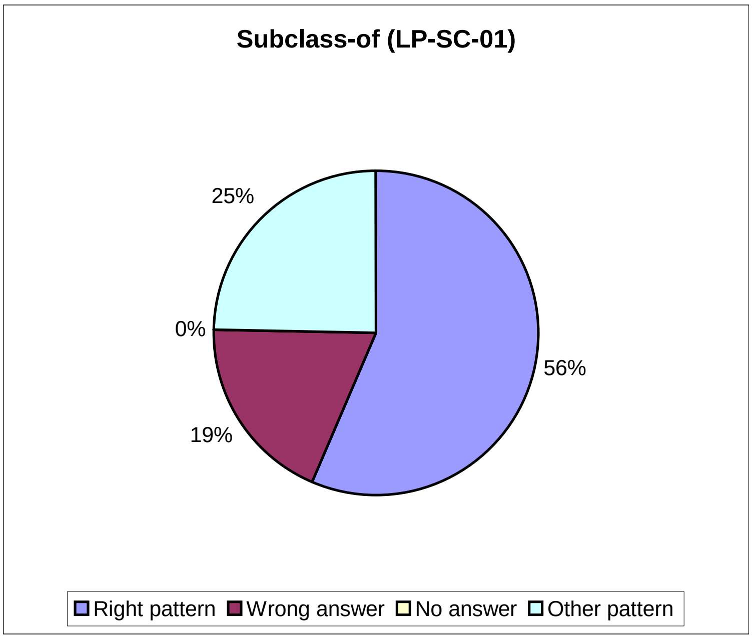

Figure 5. General Results for Modelling Problem 2

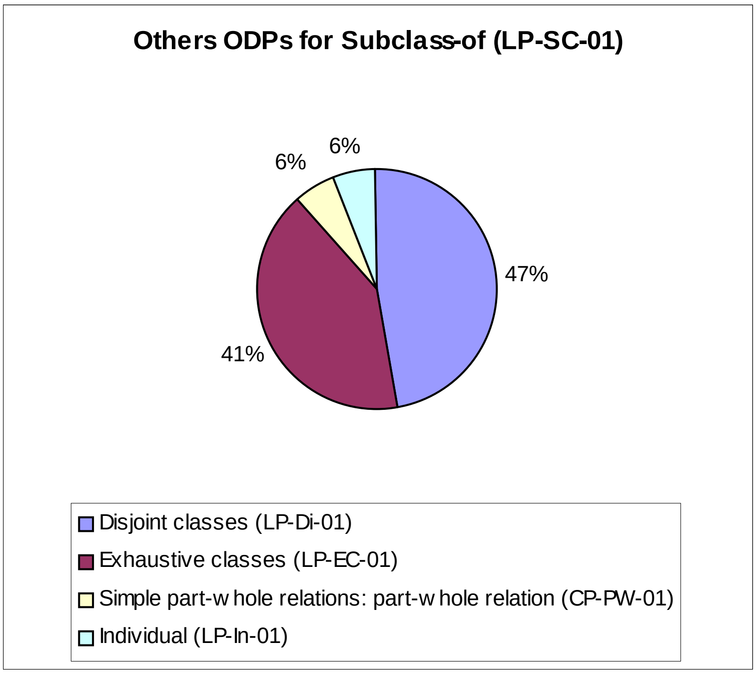

Figure 6. Non Correct Pattern for Modelling Problem 2

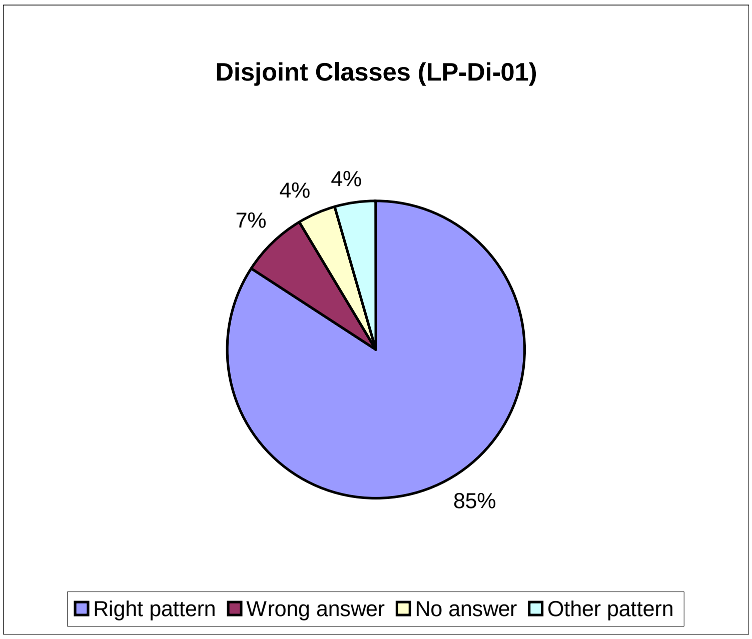

Figure 7. General Results for Modelling Problem 3

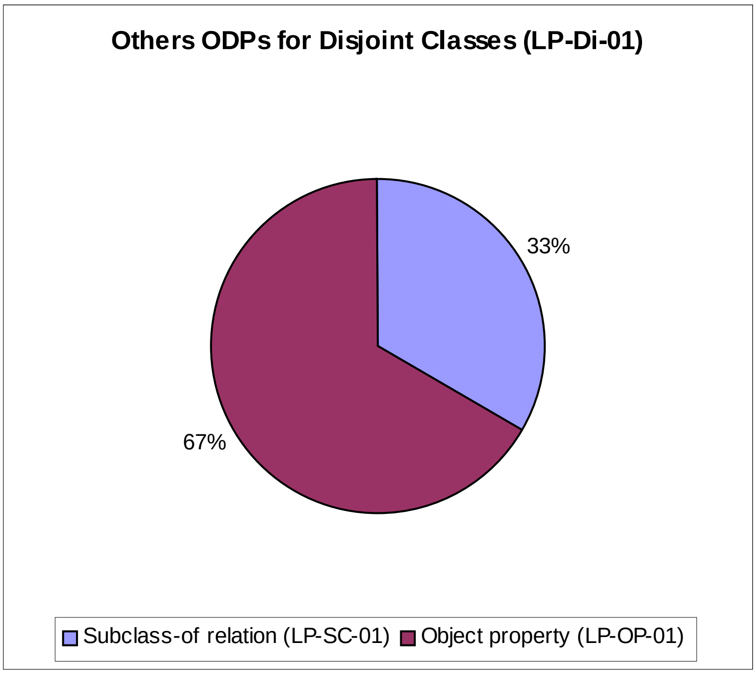

Figure 8. Non Correct Pattern for Modelling Problem 3

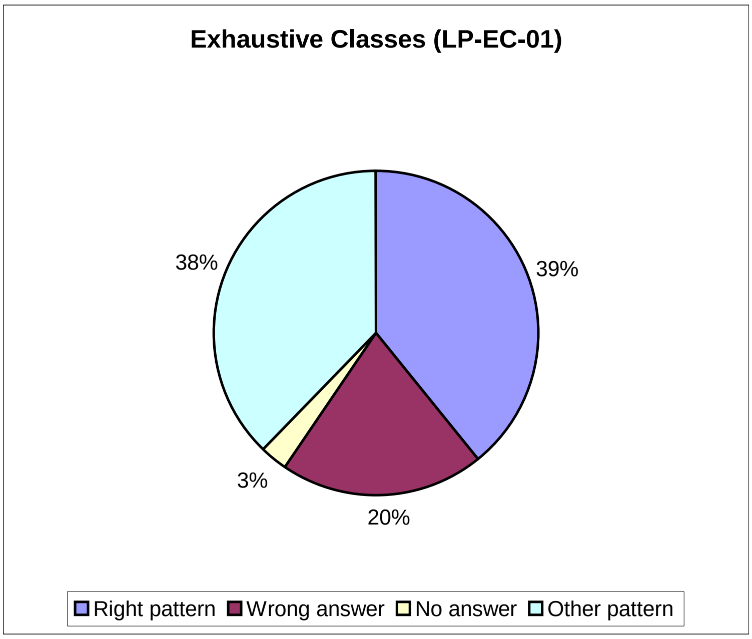

Figure 9. General Results for Modelling Problem 4

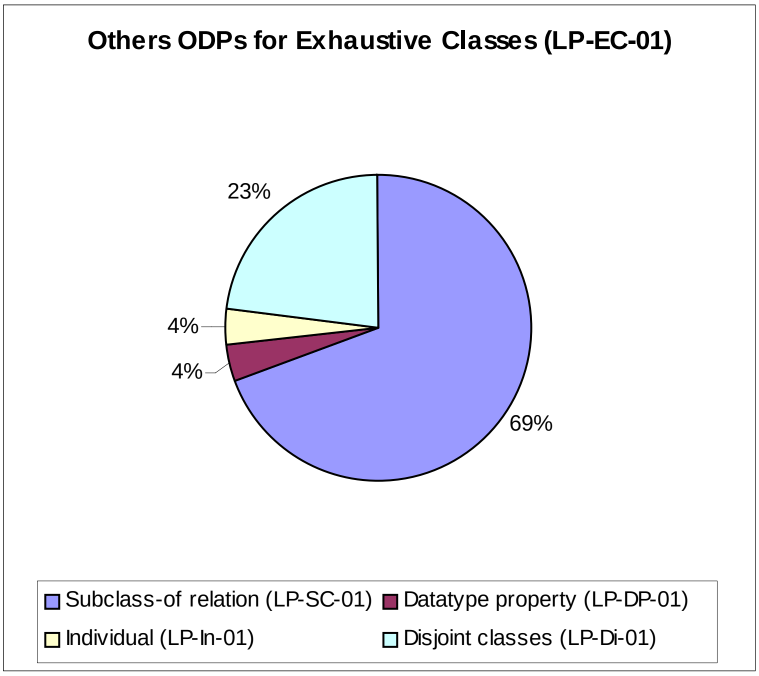

Figure 10. Non Correct Pattern for Modelling Problem 4

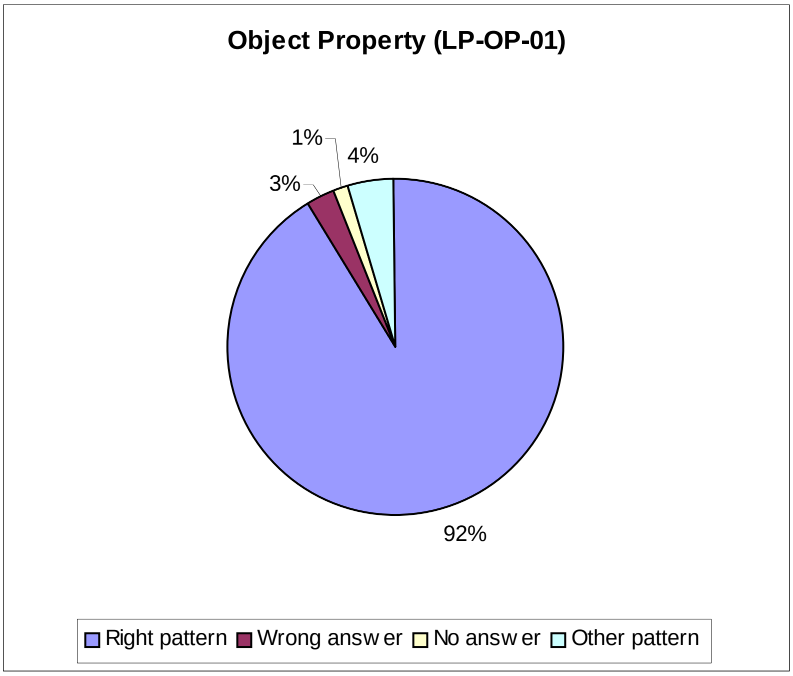

Figure 11. General Results for Modelling Problem 5

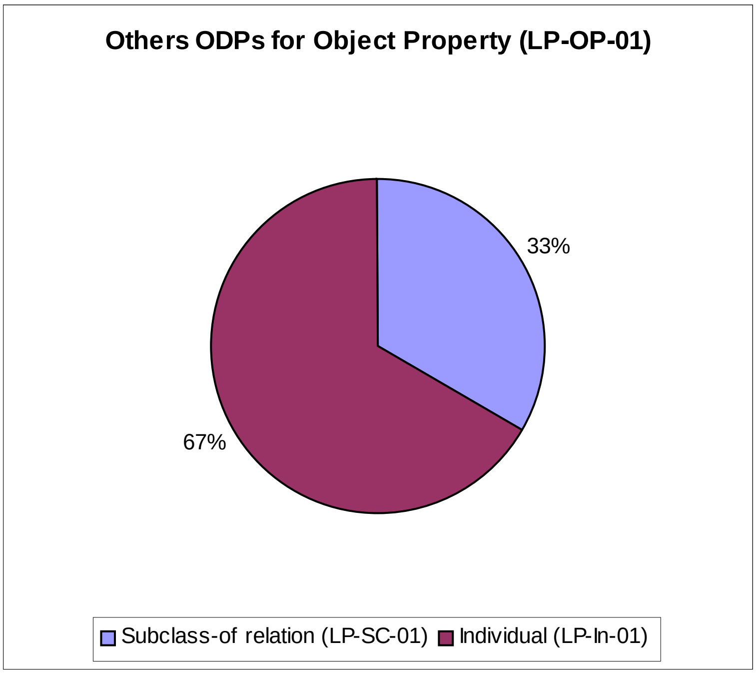

Figure 12. Non Correct Pattern for Modelling Problem 5

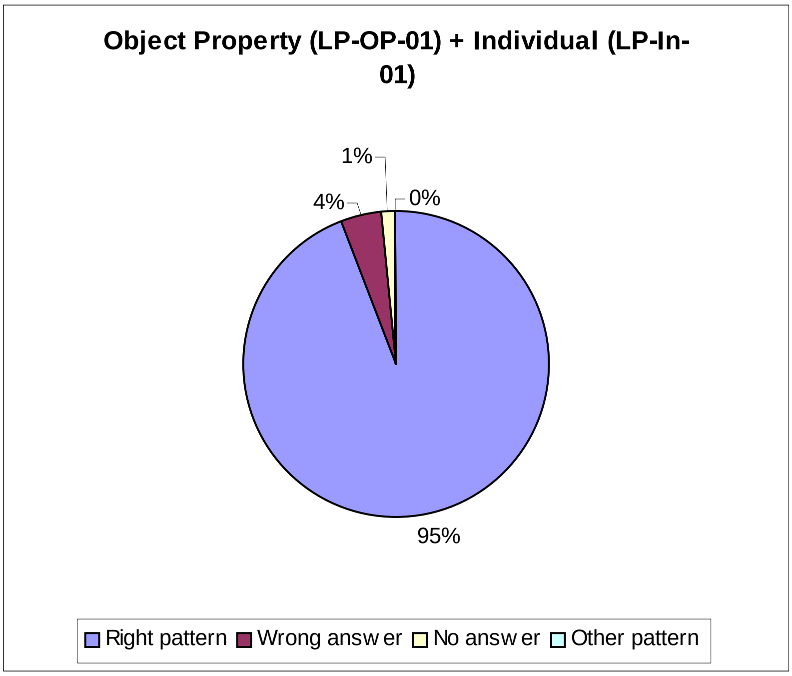

Figure 13. General Results for Modelling Problem 6

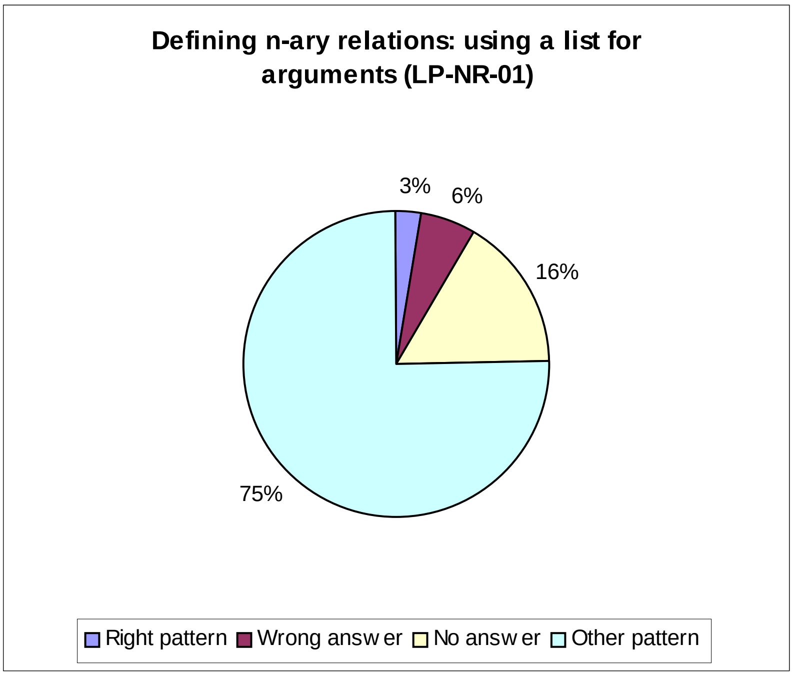

Figure 14. General Results for Modelling Problem 7

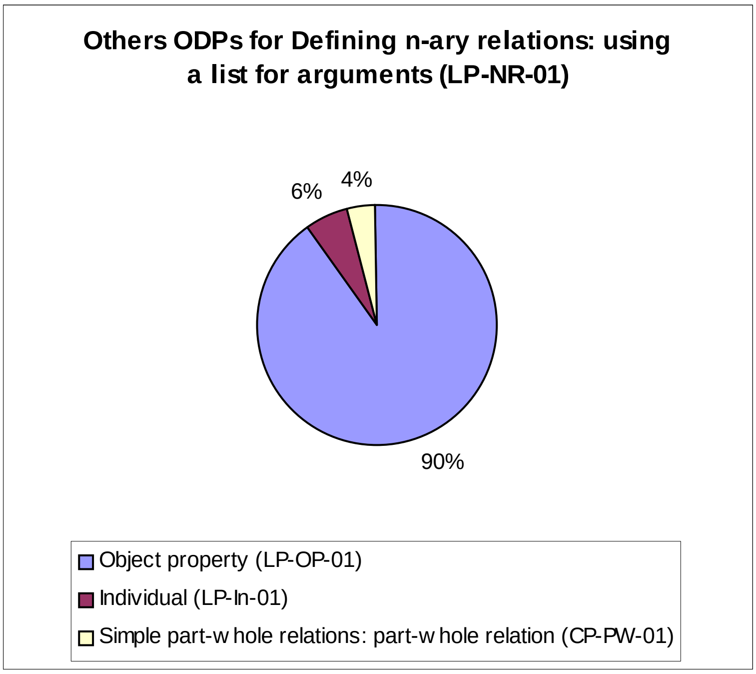

Figure 15. Non Correct Pattern for Modelling Problem 7

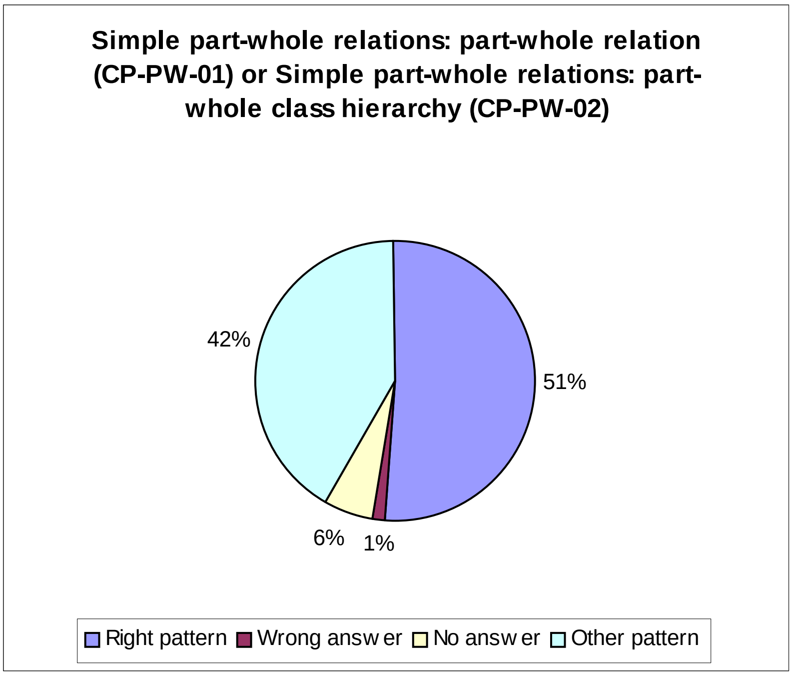

Figure 16. General Results for Modelling Problem 8

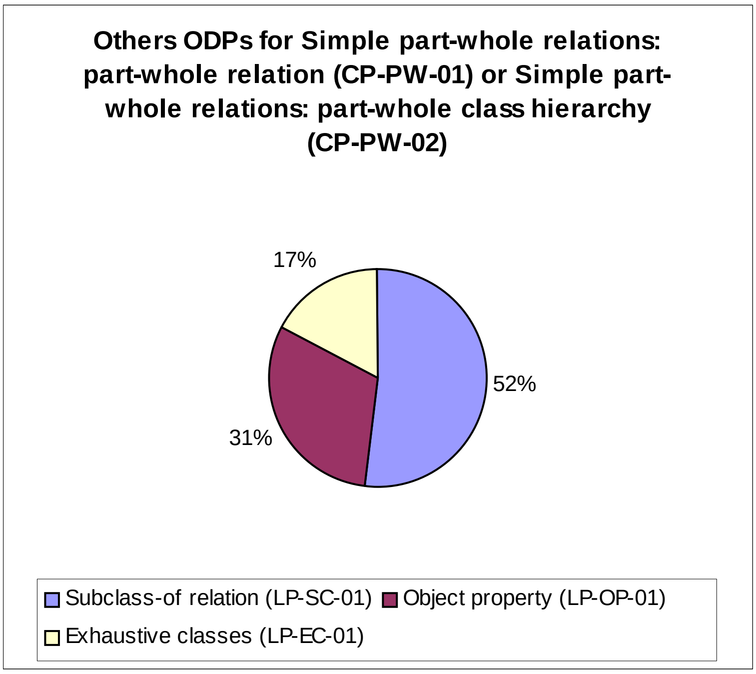

Figure 17. Non Correct Pattern for Modelling Problem 8

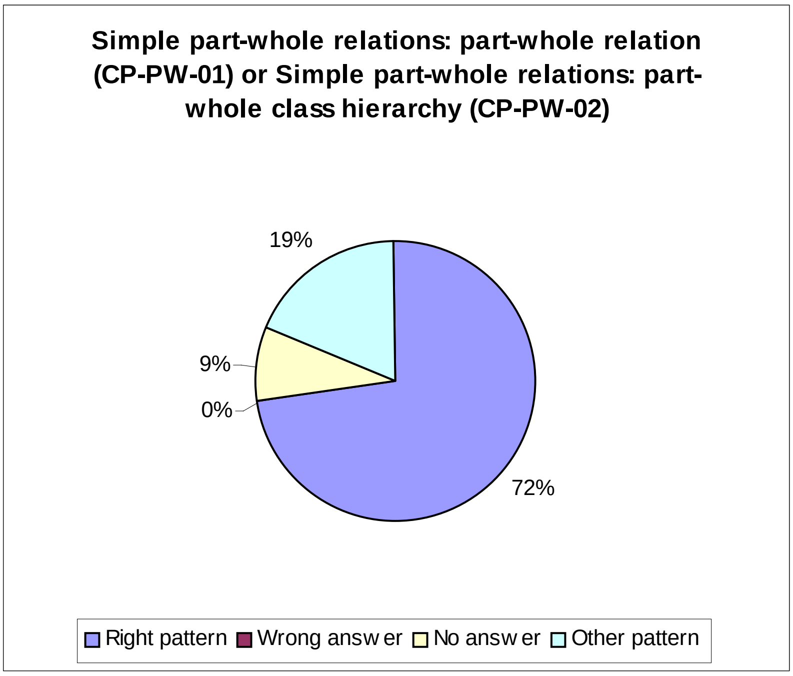

Figure 18. General Results for Modelling Problem 9

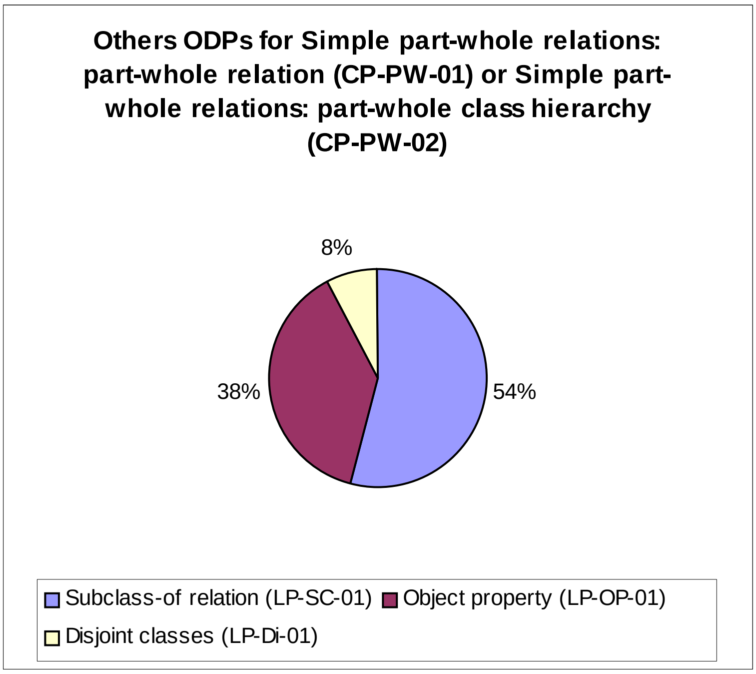

Figure 19. Non Correct Pattern for Modelling Problem 9

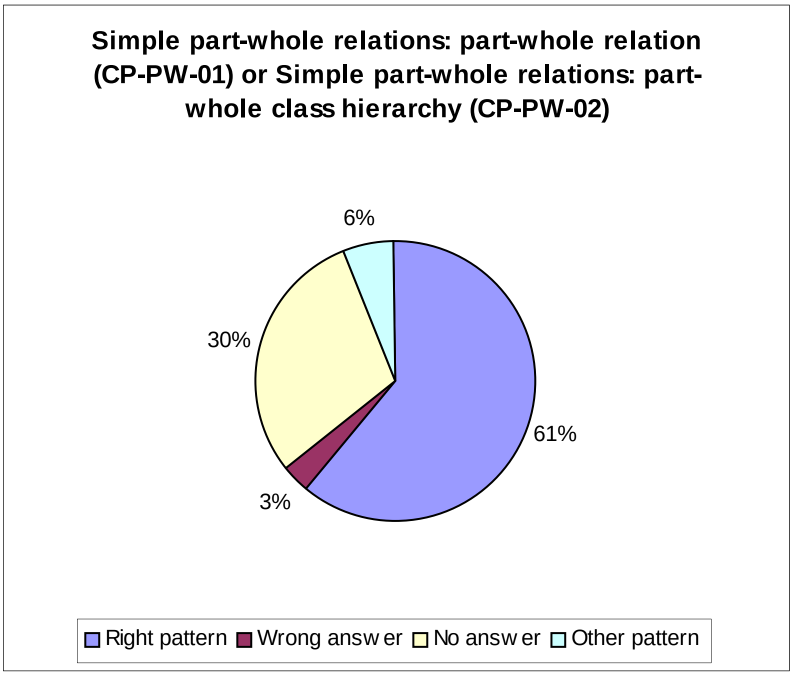

Figure 20. General Results for Modelling Problem 10

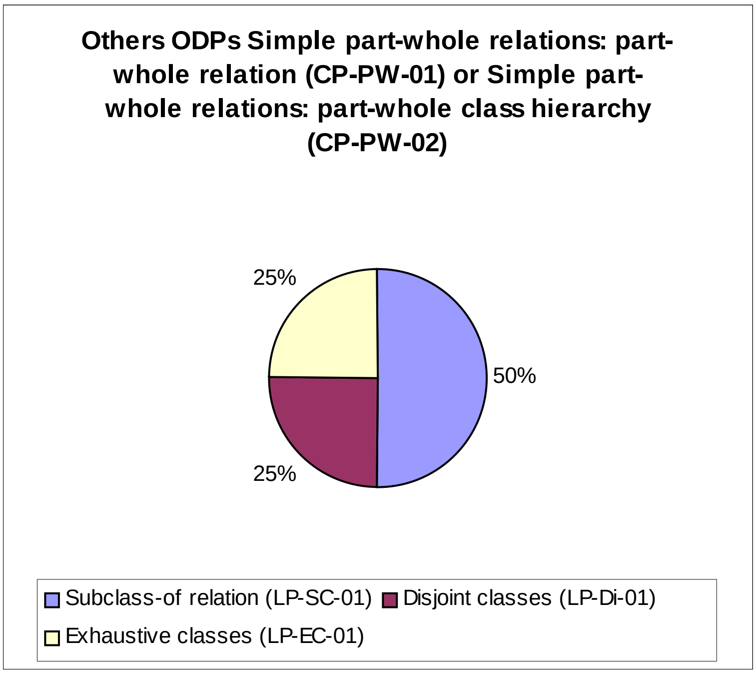

Figure 21. Non Correct Pattern for Modelling Problem 10

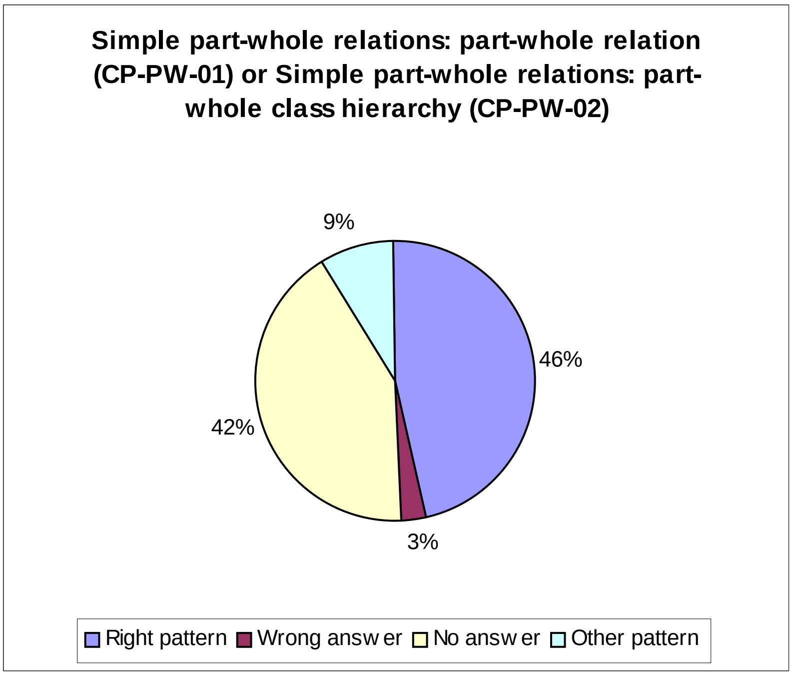

Figure 22. General Results for Modelling Problem 11

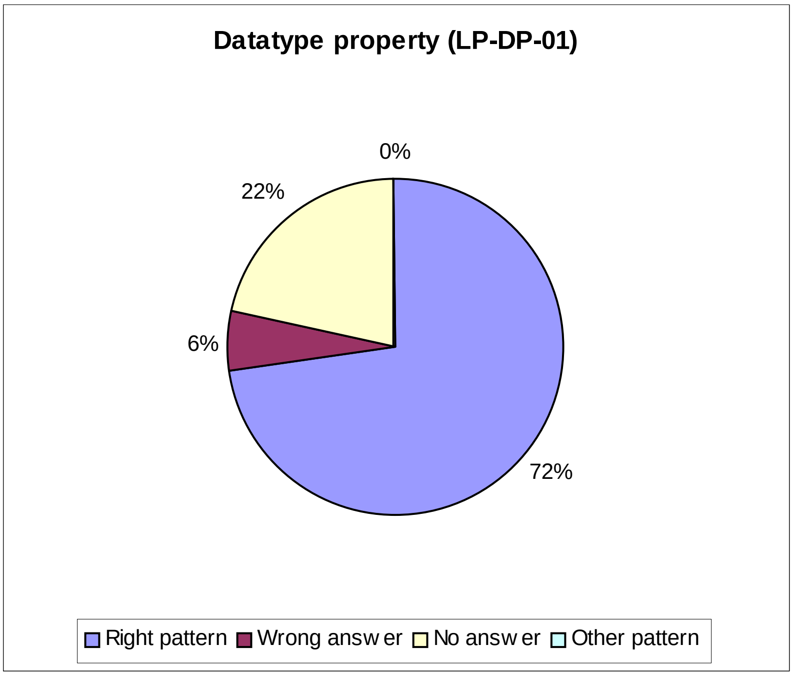

Figure 23. General Results for Modelling Problem 12

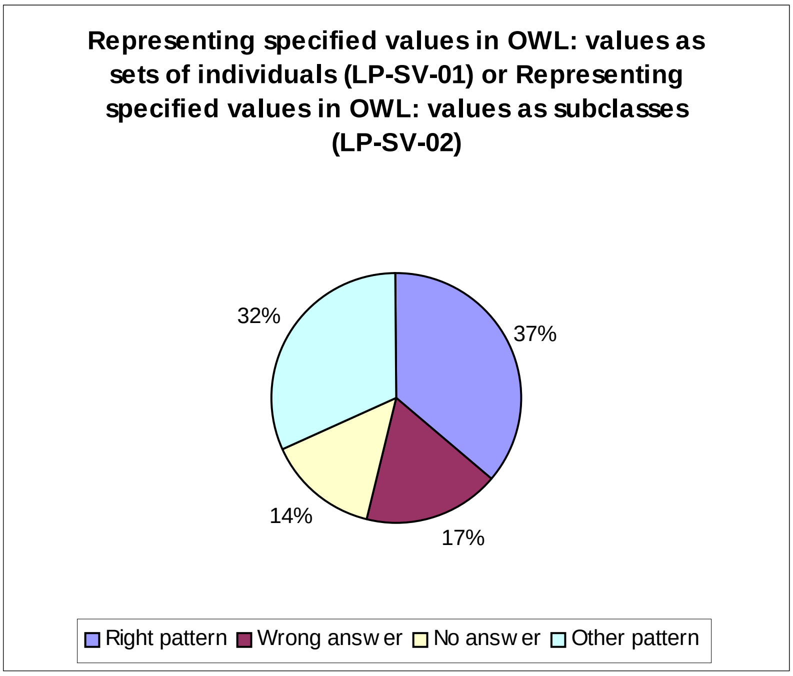

Figure 24. General Results for Modelling Problem 13

APPENDIX III. Material for ontology design pattern experiment 2

III.1. First experiment

We propose three modelling problems to be solved by students.

Requirement set 1:

The knowledge manager from a company wants to model tasks involved in a business plan. Types of tasks include management tasks, financial tasks, marketing tasks and control tasks. We also know that control tasks are used to declare begin, end, branching, and synchronization of the flow within a plan.

Solve the following modelling problems inspired by the previous requirements:

1. Tasks can be management tasks, financial tasks, marketing tasks, or control tasks.
2. Plans cannot be tasks, and vice-versa.
3. Control tasks can be used to declare begin, end, branching, and synchronization of the flow within a plan.
4. Business plans have financial tasks.
5. Plan-1 is a business plan, Task-1 is a financial task, and Plan-1 has Task-1.

Requirement set 2:

1. Financial departments are typically in charge of executing financial tasks on behalf of their marketing departments.

Requirement set 3:

A research plan is part of a research project, and a research plan is composed of a theoretical plan, and an experimental plan.

Solve the following modelling problems inspired by the previous requirements:

1. Research plans are part of some research project.
2. Research plans are composed of a theoretical plan, and an experimental plan.
3. Experimental plans and theoretical plans are part of some research plan as well as of a research project.
4. Experimental plans and theoretical plans are direct components of some research plan, but not necessarily components of a research project.

III.2. First experiment with solutions

We propose three modelling problems to be solved by students.

Requirement set 1:

The knowledge manager from a company wants to model tasks involved in a business plan. Types of tasks include management tasks, financial tasks, marketing tasks and control tasks. We also know that control tasks are used to declare begin, end, branching, and synchronization of the flow within a plan.

Solve the following modelling problems inspired by the previous requirements:

1. Tasks can be management tasks, financial tasks, marketing tasks, or control tasks (LP-SC-01)

2. Plans cannot be tasks, and vice-versa (LP-DI-01)

3. Control tasks can be used to declare begin, end, branching, and synchronization of the flow within a plan (LP-EC-01)

4. Business plans have financial tasks (LP-OP-01)

5. Plan-1 is a business plan, Task-1 is a financial task, and Plan-1 has Task-1 (LP-IN-01)

Requirement set 2:

1. Financial departments are typically in charge of executing financial tasks on behalf of their marketing departments (LP-NR-01 | LP-NR-02)

Requirement set 3:

A research plan is part of a research project, and a research plan is composed of a theoretical plan, and an experimental plan.

Solve the following modelling problems inspired by the previous requirements:

1. Research plans are part of some research project. (CP-PW-01 | CP-PW-02)

2. Research plans are composed of a theoretical plan, and an experimental plan. (CP-PW-01 | CP-PW-02)

3. Experimental plans and theoretical plans are part of some research plan as well as of a research project. (CP-PW-01 | CP-PW-02, LP-ER-01)

4. Experimental plans and theoretical plans are direct components of some research plan, but not necessarily components of a research project. (CP-PW-01 | CP-PW-02)

III.3. Second experiment

We propose the following modelling problems to be solved:

1. Mammals, birds and reptiles are different types of animals.

2. Animals are primarily categorized into two groups: vertebrates and invertebrates.

3. A courier company can be situated either in a city or in a town.

4. Customers are either regular customer or sporadic customer.

5. The “hepatitis patients” clinical guideline is targeted to “hepatitis A”.

6. Packages are sent by courier companies at a specific date, and from an origin to a destination.

7. Regions of the human body include head, trunk/torso, and limbs/extremities.

8. Horse legs are regions from their bodies, and have hoofs.

9. A train trip has a departure period, which can be White, Yellow or Blue.

10. A research project is executed by a number of persons, which can play the following roles: director, project manager, researcher, and developer.

III.4. Second experiment with solutions

We propose the following modelling problems to be solved:

1. Mammals, birds and reptiles are different types of animals. (LP-SC-01)

2. Animals are primarily categorized into two groups: vertebrates and invertebrates. (LP-EC-01)

3. A courier company can be situated either in a city or in a town. (LP-OP-01, LP-UR-01, LP-UO-01)

4. Customers are either regular customer or sporadic customer. (LP-DI-01)

5. The “hepatitis patients” clinical guideline is targeted to “hepatitis A”. (LP-IN-01)

6. Packages are sent by courier companies at a specific date, and from an origin to a destination. (LP-NR-01 | LP-NR-02)

7. Regions of the human body include head, trunk/torso, and limbs/extremities. (CP-PW-01 | CP-PW-02)

8. Horse legs are regions from their bodies, and have hoofs. (CP-PW-01 | CP-PW-02)

9. A train trip has a departure period, which can be White, Yellow or Blue. (LP-SV-01 | LP-SV-02)

10. A research project is executed by a number of persons, which can play the following roles: director, project manager, researcher, and developer. (CP-RT-01)

APPENDIX IV. Questionnaire for Experiments about Ontology Network Life Cycle

Questionnaire about the proposed guidelines to decide which ontology network life cycle model is the most appropriate for their ontology network and which concrete activities should be carried out in their ontology network life cycle.

Please answer each question with detail and be honest! Thank you very much.

General issues:

1. Are the proposed guidelines well explained?

2. Is more detail needed in the guidelines?

3. In your opinion, are this guidelines complete? If not, what is missing?

Guidelines. Step 1:

1. How difficult was to establish the requirements for your ontology?

Guidelines. Step 2:

1. How difficult was to select the ontology network life cycle model (ONLCM)?

2. How useful was the proposed decision tree?

3. If you needed to define a new ONLCM (not included in the current collection), please explain why.

4. Was the collection of ONLCM enough explained?

   Is more detail needed in the explanation of each model?

Guidelines. Step 3:

1. If you have developed more than five ontologies before this experiment, can you easily identified the activities needed for your project based on the Required-If Applicable activities?

   If not, please explain why.

2. If you have not developed more than five ontologies before this experiment, was useful the set of natural language questions for identified the activities needed for your project?

   If not, please explain why.

3. Is the NeOn Glossary of Activities well explained?

   Is something missing?

Guidelines. Step 4:

1. Which kind of detailed explanation you expected in this step?

Guidelines. Step 5:

1. Which kind of detailed explanation you expected in this step?

Final Comments:

1. How we can improve the proposed guidelines?

How the activity of establishing the life cycle for a concrete ontology network could be carried out faster?