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"

D5.6.2 Experimentation and Evaluation of the NeOn Methodology

Deliverable Co-ordinator: **Martin Dzbor**

Deliverable Co-ordinating Institution: **OU**

Other Authors: **Mari Carmen Suárez-Figueroa (UPM), Eva Blomqvist (CNR), Holger Lewen (UKARL), Mauricio Espinoza (UPM), Asunción Gómez-Pérez (UPM), Raul Palma (UPM)**

This is the first in the series of reports on the evaluation of the impact of the NeOn methodology in a range of selected categories on the performance of ontology designers. The purpose of this report is to provide an evaluative, user-based perspective on such aspects of the NeOn methodology as the use of ontology design patterns, selection of ontology network life cycle models, and the support for ontology specification and localization. In addition, we provide updated plans for more advanced testing of methods supporting ontology reuse in the social context.

Document Identifier:	NEON/2009/D5.6.2/v1.3	Date due:	February 28, 2009
Class Deliverable:	NEON EU-IST-2005-027595	Submission date:	February 28, 2009
Project start date:	March 1, 2006	Version:	v1.3
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:

OU, UPM, UKARL, CNR, FAO

Change Log

Version	Date	Amended by	Changes
0.1	15-01-2009	M Dzbor, Mari C Suárez-Figueroa	Outline & first draft planning
	06-02-2009	Holger Lewen	Initial input to ch. 8
0.2	06-02-2009	Eva Blomqvist	Initial input to ch. 2
	18-02-2009	Maurizio Espinoza	Initial inputs to ch. 4 and 7
0.3	18-02-2009	Raul Palma	Initial input to ch. 6
0.4	23-02-2009	Mari Carmen Suárez-Figueroa	Input to ch. 3 and 5
0.5	27-02-2009	Eva Blomqvist, Martin Dzbor	Update to ch. 2; alignment of terminology
0.6	02-03-2009	Martin Dzbor	Conclusions, analysis, plans
0.7	06-03-2009	M Dzbor, Mari C Suárez-Figueroa	Appendices & clean-up
0.8	10-03-2009	Eva Blomqvist, Martin Dzbor	Revision of ch.2
1.0	19-03-2009	Diana Maynard	Quality review, grammar check
1.1	03-04-2009	Martin Dzbor, Mari Carmen Suárez-Figueroa	Revisions of chapters 3-8 and appendices
1.2	14-04-2009	Eva Blomqvist	Revision of chapter 2 and appendix 1
1.3	15-04-2009	Martin Dzbor	Final cleanup, formatting, typo corrections

Executive Summary

The NeOn methodology designed and co-ordinated in WP5 has been designed in a structured and principled fashion by classifying a range of methods according to their position and place within an ontology design and use lifecycle. Due to this principled approach, the NeOn methodology goes beyond other, more narrowly focused methods/methodological guidelines, and as a consequence is being developed in steps. In order to obtain timely feedback we proposed in D5.6.1 to test partial aspects of the NeOn methodology in order to obtain a formative feedback to further refinement and effectiveness of our methodological proposals.

In the previous deliverable [6] we analyzed a range of methods available to us at that stage of the project and proposed several user studies and experiments with a subset of the methods. In the current deliverable we report on the actual execution and the results found. In particular, we report on the outcomes of studies with three user groups working with ontology design patterns, of a user study drawing upon the available support for ontology localization, ontology (requirement) specification and for the establishment of ontology lifecycles.

To respond to the evolving understanding of the project (and user) priorities, we are refining our proposals for experimenting with two critical areas of the project, namely, the role of ontology reuse and ontology mapping in the overall ontology development. In addition, we carried out two studies that were not explicitly planned in D5.6.1; however, we believe they are informative in terms of understanding the usability and the role of the NeOn technology. In particular, we will report on the studies with the Collaboration Server and the LabelTranslator.

Table of Contents

List of Tables

List of Figures

1. Introduction

One of the objectives of WP5 is to collate a coherent NeOn Methodology for building ontology networks comprising a range of concrete methods and guidelines to support activities involved in the typical ontology lifecycle and development process. In addition to the explicit objective of producing a "method book", WP5 also needs to provide qualitative and (where possible) quantitative evidence that using the NeOn Methodology leads to users being able to design ontologies faster and/or to better quality standards – in other words, to assess how effective NeOn Methodology is for the target user.

In the previous deliverable [6] we identified several areas where (i) methodological support was available at a sufficient level of detail, and (ii) user studies could be performed between months M27-M36:

- Ontology design patterns
- Ontology lifecycle
- Ontology reuse
- Ontology reengineering
- Ontology mapping
- Ontology localization
- Ontology (requirement) specification

In the remainder of this section we summarize the content of this document by saying a few words about each of the planned, executed and refined studies.

1.1 Overview of the document

When the original plan of user studies was compiled in M24, several assumptions were made in terms of functional priorities pursued by the NeOn project, in terms of deployment of a particular tool and/or method in practice, and in terms of availability of target user groups. Since that time, several "deviations" from the original plan were made, which are explained and justified in the remainder of the chapter. In order to structure the deliverable, we divide its content into three broad categories that correspond to the status of the respective study with respect to the plans in D5.6.1:

1. Executed, planned user studies … in section 1.2 we list the user studies carried out in line with the plans made in the previous deliverable or only with minor procedural variations;
2. Additional user studies … in section 1.3 we present several studies of tools that emerged after month M24 – this category does not cover all tools and techniques arising from the NeOn effort; on the contrary, we only include the studies with the tools addressing one of the primary needs either of a use case or of the project vision;
3. Refined plans for studies … in section 1.4 we highlight two major changes to our plans: In our plans from D5.6.1 we included a generic intention to experiment with ontology reuse and with the ontology mapping – these were analyzed in more depth between months M25-M36 and the conclusion has been reached to better embed these activities into a concrete, broader ontology development problem and thus refine the respective user study.

1.2 Summary of executed, planned user studies

The first set of findings concerns those user studies that were carried out in accordance with the plans made in deliverable D5.6.1. We dedicate a separate chapter to each study; thus the purpose of this summary is to refresh the reader's knowledge of our proposals and to cross-reference to the subsequent parts of this document.

First, the notion of ontology patterns (also called modelling components) has been described in detail in NeOn deliverable D5.1.1 [11], and more recently in deliverable D2.5.1 [19]. The concern in the user studies is to show the effects and benefits of using patterns in ontology engineering in a scientifically rigorous manner. There are several different aspects of patterns that may be studied and several types of pattern usage effects that may be explored and analyzed. Patterns in design and thus in ontology engineering are suggested to give three kinds of benefits: reuse benefits, guidance benefits and communication benefits. Reuse is concerned with constructing "better" ontologies due to the use of proved modelling chunks, fragments; the guidance is referring to the assistance and learning opportunities a structured pattern offers to the user, and finally, the communication is concerned with patterns as a tool for describing existing ontologies, modelling situations in a standard, shareable manner.

The proposed set of experiments in D5.6.1 intends to address all of these issues for the different kinds of patterns in the long term, but as a first pass, only a few have been selected focusing mainly on showing that specific kinds of patterns do have both reuse and guidance benefits. Two preliminary experiments involving students were designed and conducted during 2007. These two experiments and the corresponding analysis of their results were described in D5.6.1. The experiments gave some very important initial results and insights on how subjects really perceive, understand, and use ontology design patterns. The goal of the experiment was to test if, given a modelling problem expressed in natural language, people modelling ontologies were able to identify the correct pattern for such problem. The auxiliary goals were to test if a subset of logical and content design patterns described in D5.1.1 [11] was well explained and if such patterns were easy to understand and to apply in particular modelling problems.

Thus, in assessing the benefits of patterns, we have taken the depth-first approach in order to ensure the main expectations about the benefits of patterns are indeed explicitly identifiable in the design process. We believe this focus on the primary role of patterns is important and carried out a similar set of studies with several user groups to ensure broader validity of our conclusions. Further particulars and detailed findings from the **user studies with ontology design patterns are presented in chapter 2**.

Second, the next study planned concerns the support for ontology (requirement) specification activity. The goal of this experiment is to test the benefits of using the proposed methodological guidelines for obtaining the ontology requirement specification document (ORSD) as an output of the ontology (requirement) specification activity. The ORSD acts as an input to the whole ontology development. The main motivation for this study is to learn about and assess the ease of comprehension and the usability of the proposed methodological guidelines for carrying out the ontology (requirement) specification activity. The study focuses on the roles of software developers, analysts and various knowledge engineering practitioners who capture requirements for ontology development, and the expected outcome of the study is the improvement of the guidelines. Further particulars and detailed findings from the **user studies with ontology (requirement) specification methodological guidelines are presented in chapter 3**.

The third study in the category of planned and executed concerns ontology localization. The ontology localization activity comprises the adaptation of an ontology designed in a particular natural language (e.g., English) to another concrete language and culture community (e.g., Spanish), as defined in [12]. The main goal of the experiment was to test the benefits of using the proposed guidelines for obtaining a multilingual ontology as output of the ontology localization activity. Further particulars and detailed findings from the **user studies with ontology localization are presented in chapter 4**.

The fourth study planned and executed is about the establishment of ontology network lifecycle, in which the main idea was to learn about the understandability and usability of the proposed guidelines for helping software developers and ontology practitioners to decide which ontology network lifecycle model is the most appropriate for their ontology network and which concrete activities should be carried out in their ontology network lifecycle. The main goal of the experiment was to test the benefits of using the proposed guidelines for obtaining the ontology network life cycle. Further particulars and detailed findings from the **user studies with ontology lifecycle are presented in chapter 5**.

1.3 Summary of additional user studies

The next set of experimentation and findings concerns those user studies that were carried out with two technologies deployed as NeOn Toolkit plug-ins. These studies focused on testing the overall performance and design suitability of the implemented techniques, rather than the effectiveness of the methodological components. In particular, two such elements were tested: (i) the Collaboration Server and (ii) LabelTranslator.

Thus, the sixth reported study refers to testing the adequacy and usability of the Collaboration Server in the context of supporting distributed teams working together on ontology development and maintenance activities. The main motivation of this experiment was to evaluate the models and strategies proposed for the management of ontology changes to support the development and maintenance of ontologies in a collaborative scenario in FAO. An identified FAO workflow model distinguishing ontology engineers, editors, validators and users was followed, and each category of users was queried and observed whilst using the proprietary infrastructure of the Collaboration Server provided by Ontoprise, GmbH. Further details on **studying collaborative support for ontology development in the NeOn Toolkit are available in chapter 6**.

The seventh reported study in this deliverable concerns the tests of the quality of the new translation-ranking algorithm and usability of the LabelTranslator plug-in. The goal of first experiment was to re-evaluate the quality of the new translation-ranking algorithm implemented in the second version of our ontology localization system. On the other hand, the goal of the second user study was to assess the user satisfaction of the LabelTranslator system for carrying out the ontology localization activity. A formal Software Usability Measurement Inventory (SUMI) method was followed, using questionnaires with 50 asked-about items for which the user selects one of three responses (“agree", “don’t know", “disagree"). The questionnaires were designed to measure the affect, efficiency, capability to learn, helpfulness and overall control of the LabelTranslator environment. Further details on **LabelTranslator studies are in chapter 7**.

1.4 Summary of refined plans for user studies

The third set of contributions made in this deliverable contains the refinements to our plans for the core activity in the context of networked ontologies: reuse of ontological resources. Although an experiment on this topic was proposed earlier, new developments in NeOn between M25-M36 brought substantial advances in both tool support and our methodological understanding of the roles of these activities in a broader ontology lifecycle.

In particular, the original proposition to study ontology reuse was based upon a very loose and ad-hoc connectivity between Watson as a part of NeOn infrastructure and Open Rating System (ORS). The relationship between Watson, ORS and ontology design has been since formalized in tightened to form Cupboard – a comprehensive support framework for accessing, rating, submitting, and organizing ontologies in a way that would reflect organizational focus, rather than the earlier “one size fits all” approach. Hence, the refined experiment will not study ontology reuse as an abstract activity, but as an activity that is inherently social (i.e., dependent on previous ratings), conceptually circumscribed to the interests of a particular organization or design team (i.e., reusing a number of approved models rather than anything that can be found on the web), and wrapped in one, user-friendly engineering framework (i.e., the Cupboard platform [1]). Further particulars and more detailed proposals for **the ontology reuse studies are presented in chapter 8**.

2. Using ontology design patterns

Ontology design patterns are an important part of the NeOn methodology, and experiments showing the benefits and drawbacks of using such patterns are equally important. So far one set of the proposed experiments (see [6]) has been performed, the focus has been to first show that patterns are in fact perceived useful and give real benefits, i.e., we take a depth first approach and first show the benefits of using patterns, before going in depth on their presentation and identification (as proposed in the original experiment plan [6]). At this early stage of the development of a community using and developing patterns it is important to identify requirements of tools and methodologies for using patterns. Important questions include how patterns are in fact used and how this can be supported?

2.1 Overview and objectives

Before analyzing and discussing specific findings we remind the reader of the motivation stated for this experiment in D5.6.1 [6]:

The main purpose and aim of the study is to determine if patterns really make manual ontology design 'easier' and faster.

At this point, it is also worthwhile to remind the reader that the focus for this particular study was set to content patterns for the first version of this experiment. The experiment settings were slightly changed from the description in D5.6.1 [6] due to practical reasons, and availability of experiment subjects, but the major intention still holds. The most significant change was in the setup of the sessions. Instead of having two groups (one working with patterns and one working without patterns) working in parallel on the same task we had to settle for only one group. The reason was that too few participants were available for the study. To leverage the effects of this we introduced an initial task, in order to record their modelling abilities before introducing them to patterns, and then compared the results of this initial task to the results after the pattern training and the final task where patterns was used. This introduces slightly more uncertainty into the interpretation of the results as will be discussed further in section 2.4.2.

By studying participants modelling ontologies, with and without patterns, we wish to determine both how the patterns are perceived by their users, the participants in our study, and how they in fact affect the modelling results. Several questions can be asked, and some of them we will try to answer during the presentation of this study:

- Are the content patterns perceived as useful by the participants?
- Are the ontologies constructed using content patterns 'better' in some modelling quality sense, than the ontologies constructed without patterns?
- Are the tasks given to the participants solved faster when using patterns?
- How do participants use the patterns provided, and what support for pattern selection and usage would be beneficial?
- What common problems in modelling solutions can be identified, both when not using patterns and when using the available content patterns?

Not all of these questions will receive a clear answer based on this study, but the questions will be revisited at the end of the chapter in order to draw some conclusions.

2.2 Assumptions and user study setup

Data collection has been focused on two major types of data:

- Subjective user opinions, and
- Objective measurements.

The subjective user opinions have been collected through questionnaires given to the participants during the study, while the objective measurements are based on studying the backgrounds of the participating users, the study setting itself, and the resulting ontologies produced by the participants. The two types of data are presented together in this study and conclusions will be drawn from the combination of both kinds of data.

The setup of this study was to divide the experiment into three sessions, each with a slightly different setup and slightly different type of participant. The main idea of each setting was to let a set of participants construct ontologies: first, without using patterns and then with the help of a set of pre-selected content patterns. All participants have been given appropriate training, in order to leverage their previous knowledge in the experiment. First, one part of the training addressing the modelling of ontologies using OWL in general; another part focused on introducing ontology patterns in general and, in particular, content patterns. In some of the sessions, practical training was used, while in one session only theoretical information was given before the actual experiment tasks were given. All settings used the same tool for modelling, in our case, TopBraid Composer[1], and all settings used the same two tasks for recording the participants' modelling abilities, the first set within the music industry domain and the second one within the domain of hospitals as work places. The main reason for choosing TopBraid Composer for the user study was the timing of these training sessions and the release plan of the NeOn Toolkit – at the time participants were available, we did not have a stable release of the toolkit supporting the functionality needed. The possibility to work with several ontologies and import ontologies through an intuitive 'drag and drop' interface was the primary factor in favour of TopBraid Composer. We felt that such basic support for handling the patterns as building blocks was essential to receive useful suggestions on additional functionality needed, beyond the basics.

Before the sessions all participants filled out the same background questionnaire, and after each task was completed their experience was recorded through a new questionnaire. The total number of participating users was 45, and the distribution of this population over the three sessions can be seen in Figure 1. Below, the three settings are described in detail.

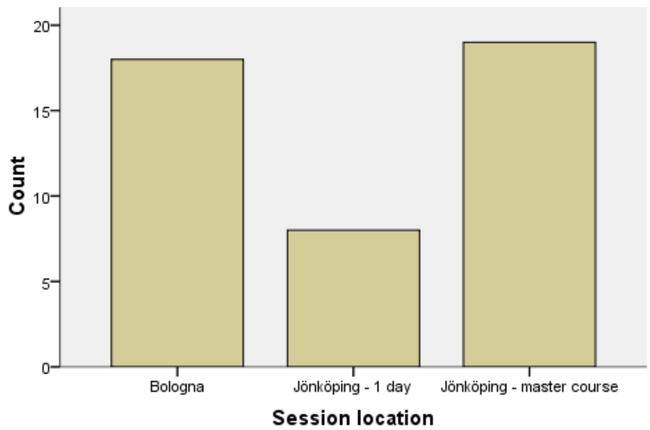

Figure 1. Distribution of participants over the three settings.

1 Available from http://www.topquadrant.com

Not only were the questionnaires of all individual participants recorded, in addition the ontologies they constructed were stored and later analysed. In two out of three of the sessions the ontologies were constructed in pairs of two participants, i.e. the number of ontologies stored was then half of the number of participants. The evaluations performed are closely related to general ontology evaluation methods, therefore we first give a brief introduction to commonly used evaluation strategies. In NeOn deliverables such as D2.2.1 [20] ontology evaluation is treated in detail, the interested reader is referred to these deliverables and their referred literature. In brief, ontology evaluation can be divided into three dimensions (or levels):

- Structural evaluations
- Functional evaluations
- Usability profiling evaluations

Structural evaluations treat the ontology as an information object, studying the syntax and formal semantics of the ontology definitions. Functional evaluations additionally study the intended conceptualization and view the ontology in its intended context. Usability evaluations focus on the ontology as a semiotic object, trying to convey some sort of meaning. Structural evaluations are commonly performed through simple measures of ontology characteristics, such as depth, breadth, and other measures, and through checking definitions and axioms for correctness and consistency. Functional evaluations commonly compare the ontology to some external source of knowledge, such as a ‘gold standard’ ontology of the domain, or measure the ontology efficacy and performance in its intended task environment. Usability profiling evaluations are concerned with user satisfaction, understandability and re-usability.

Within this study, the focus of the evaluations has been primarily on the functional and usability profiling level, although some of the methods used additionally cover the structural level to some extent. The ontologies were analysed with respect to four different aspects that we name as:

- Coverage of problem requirements
- Usability profiling
- Modelling mistakes/unresolved problems
- Patterns used

Since the tasks were rather compact, it was possible to construct a gold standard ontology representing a reasonably good solution to the problem, given the requirements of the task. Based on this solution and the competency questions (CQs) derivable from the task descriptions the essential elements of the ‘gold standard’ were listed. Hence, for example, in the first, music industry task there were bands and musicians playing different instruments described, whereby concepts such as ‘band, ‘person’ or ‘musical instrument’ would be seen as essential parts of the ontology. Also, properties such as the ‘release date’ of an album or a track in an album realising a specific song could be identified and were considered as essential. Using this list of elements the ontologies were then analysed to determine the coverage of the ontology over the problem stated in the task description. The coverage of the ontology can be expressed as a percentage of the essential elements that were included. An element did not have to be realised in precisely the same manner as in the ‘gold standard’ but it needed to be in some way included in the modelled ontology, i.e., the general problem it represents had to be solved somehow, in order to be counted as covered. In this way all the ontologies received a percentage representing their coverage of the task, how much of the task the participants had managed to solve. In the sense of functional ontology evaluations this represents the completeness of the conceptualization to what extent the ontology solves the intended task.

With respect to usability, the ontologies were analysed regarding their clarity and understandability, i.e., whether concepts were properly defined and axioms included were appropriate, whether names were clear and followed some naming convention, and whether the ontology contained also labels and comments. More specifically, two types of measures were used; one set representing usability profiling measures and one set representing structural measures that we believe also contribute to the clarity of the ontology semantics. The first set contains the following measures:

- Presence of naming conventions – an assessment if naming conventions were used and in that case for what elements of the ontology.
- Labels - the fraction of concepts, properties and instances with labels specified.
- Comments - the fraction of concepts and properties with comments specified.

The second set, i.e. structural aspects that contribute to clarity of the semantics, include the following measures:

- Inverse relations – the fraction of object properties that have inverses defined.
- Disjointness axioms – the fraction of concepts that have disjointness axioms defined.
- Level of axiomatization – the fraction of concepts that have a 'formal' definition.

The intention was to see if good practices had been followed with respect to making the ontology understandable and reusable by others. The result of this part of the analysis was a list of usability problems for each ontology with respect to the above-mentioned observable characteristics. The reason for including the structural measures under the usability heading was that the focus is not actually on the details and correctness on those structures but rather on the presence of the structures. Imagine a case where the ontology is to be understood or even reused by a person not involved in its initial development. In such a case to have properly defined concepts and suitable axioms present makes assumptions and intentions explicit. Thereby the ontology will be easier to understand and to reuse.

The third aspect is concerned with problems, or 'mistakes' in the designed ontologies. Even an ontology that covers the CQs of its requirements might have some inherent problems that would for example hamper its usage in a system or make it less flexible, not easy to extend or update. Such mistakes are not easily discovered, but are often closely related to incorrect modelling choices that make the model incoherent with respect to the reality, or that make the model 'static' in some way. An example of such an issue is to model 'musician' as a specialization (sub-class) of 'person'. At a first glance, this seems to be a valid approach – musicians are in fact people. But are musicians really a *specific kind of* person? What about children that become musicians later in their lives? What about a musician that stops playing and becomes a taxi driver? In both cases, the individuals should retain their 'person' belongingness, but may lose or acquire their 'musician' status. These problems arise because a musician is not really a kind of person; it is *a role* that any person may take in some specific setting and during some specific period of time. Such differences in meaning and subtle modelling errors have been discussed and discovered using methodologies such as OntoClean [22] and different methods of taxonomic evaluation (see [21]). In this particular experiment the ontologies are small enough to be grasped without using any formal methodology, nevertheless discovering the same issues and problems. This was the focus of the third aspect of analysis. As a result, a list of problems was compiled for each ontology. These problems can be both on the structural or functional level.

Finally, the pattern aspect was analysed. For the ontologies constructed based on patterns it was of course interesting to see how many patterns, and which ones, had been used, but also for the ontologies constructed without patterns this dimension was analysed. In this case the intention was to see if any patterns were used 'unintentionally', i.e. without knowing about patterns, or if some problems could be discovered that could have been solved by using an existing pattern in the available catalogue. The result of this analysis was (i) a list of patterns actually reused (where applicable), (ii) patterns used unintentionally, and (iii) problems that patterns could have solved. Since the focus of this study is on content patterns, the patterns solve specific issues with respect to conceptualisation, and therefore this evaluation can be viewed as mainly a functional evaluation, however the problems observed can also be both structural and of a usability character.

2.2.1 User study #1: PhD Course in Bologna

The first session took place at a 20 hour (four day) PhD course in Bologna (IT). The course outline can be seen in Figure 2. The participants were PhD students and other researchers in computer science and related fields that had applied to the PhD course. Before coming to the course the participants filled in the background questionnaires and sent their responses to the tutors. On the first day the participants had some lectures on ontologies in general and modelling with OWL (Lecture 1 in Figure 2). Then they did a first modelling exercise, in order to practise the basics of OWL (Exercise 1 in Figure 2). All exercises were in this course done in pairs of two students (although questionnaires were still answered individually), and the participants were suggested to follow a methodology called eXtreme Design that was introduced during the lectures. The methodology introduces a divide-and-conquer way of addressing the set problem, solving the modelling problem piece by piece. It fits well with the use of patterns but does not assume patterns to be present in the general case. After this first exercise there was no questionnaire, but results in the form of the ontologies were collected in order to provide background information on the modelling abilities of the participants. The second day contained some more lectures on ontologies and ontology design (Lecture 2 in Figure 2), but content patterns were not mentioned, thereafter the participants got a different modelling task than the day before. This was the first task of the actual experiment, Task 1 in Figure 2, resulting in the set of ontologies O_1 in Figure 2. The time given for completing the task was 2 hours. After this task the participants answered a questionnaire about their experiences, resulting in a set of answers to be analysed (illustrated as Q_1 in Figure 2).

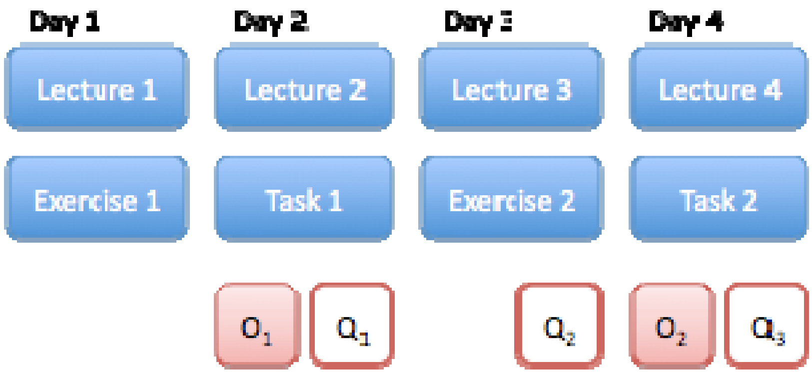

Figure 2. Activities and analysed results during the four days.

Next, on the third day they were given lectures on ontology patterns and pattern-based design (Lecture 3 in Figure 2), and then they had to redo the same exercise as the day before but now with a pattern catalogue available (Exercise 2 in Figure 2). The intention of this was to practise using patterns, in order not to record only the “initial confusion”, when introducing a new concept and performing the actual experiment task. Again there was a questionnaire after this exercise, precisely to record this “initial confusion” that was foreseen (the set of answers recorded is denoted Q_2 in Figure 2). Finally, on the last day some additional lectures on reengineering and the semantic web were given (Lecture 4 in Figure 2) and the subjects then got a new exercise to solve (Task 2 in Figure 2), with the same pattern catalogue as the day before. This was the second task of the actual experiment, resulting in the set of ontologies denoted as O_2 in Figure 2. The time given for completing the task was again 2 hours. The session ended with a concluding questionnaire (resulting responses denoted as Q_3), also asking the participants to compare their experiences over the four days of lectures and exercises.

Both the tasks (for days 2-3 and day 4) contained approximately the same modelling issues but in different domains. Additionally, the final exercise was not equally well specified as the others, thereby reducing the amount of help given to the participants. More in detail, the task was on the same level of difficulty, but the requirements were less explicit and not divided into simple sentences beforehand, as was the case for the first experiment task. The reason for this was to avoid introducing any unfair advantage, in favour of the task where patterns were used, with respect to the level of difficulty of solving the task. Since the participants had now had some more training, the requirements were presented just as a small 'story', without any additional clarification.

2.2.2 User study #2: Dedicated session in Jönköping

The second session was conducted in Jönköping (SE) as a tutorial in less than one day, on ontology design and patterns. This was a dedicated experiment session in the sense that the participants were recruited specifically for participating in the experiment, i.e., the session was not part of any course as the other two sessions. The participants were PhD students and researchers in computer science and related fields, invited based on having some basic knowledge of, or interest in, ontologies. The setup from Bologna was compressed into two lectures and only the two tasks that were actually part of the experiment, i.e. no intermediate exercises. The first lecture gave the basics of ontologies and OWL (see Lecture 1/2 in Figure 3), and then the users were directly given the task from day 2 of the Bologna course (Task 1 in Figure 3), with the subsequent questionnaire (results recorded in the set of responses Q_1 and set of ontologies O_1, as illustrated in Figure 3).

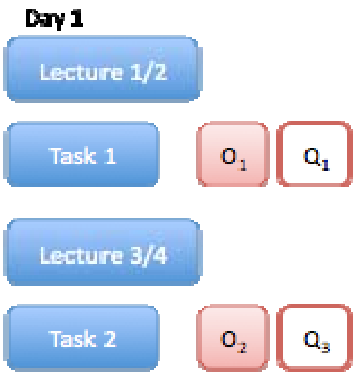

Figure 3. Activities and analysed results during the session.

The same methodology for problem solving as in the Bologna course was suggested, although the subjects in this case worked on the modelling tasks alone and not in pairs. Next there was a second lecture (Lecture 3/4 in Figure 3), on ontology patterns and pattern-based design, thereafter the final task (Task 2 in Figure 3, the same as day 4 in Bologna) was given. In this case the questionnaire was a mix of the questions from days 3 and 4 in Bologna, based on the final questionnaire but where some questions were additionally added and the two final questions were modified (since these subjects only had two tasks to compare and not four). Hence, the set of analysed ontologies, O_2 in Figure 3, is comparable to the ontologies resulting from the second task of both the other sessions, while the set of responses to the questionnaire, Q_3 in Figure 3, is only partly comparable to the sets from the two other sessions. An additional difference was that the time given for each task was only one hour instead of two hours for the other two settings.

2.2.3 User study #3: Master's Course in Jönköping

Finally, the third session was conducted as a part of a Masters course in Information Logistics in Jönköping (SE). The participants were master students at Jönköping University, mainly from the Information Engineering master's program, who had registered for the course in Information Logistics. In this setup the lectures were more or less identical to the ones given in Bologna, except that the parts on Semantic Web were replaced by lectures on Information Logistics. Still, the ontology design, ontology patterns and OWL-parts were the same, and the exercises and questionnaires were identical. The same methodology for problem solving was introduced and the subjects all worked in pairs (except for one group of 3 students). This time, however, the course was given over one month, but the outline was the same as in Bologna. The time of the lectures and exercises was also the same, only spread out over one month of calendar time. The time given for each task was 2 hours. In Figure 4 the 6 days, during one month, containing the lectures and tasks corresponding to the Bologna session, are illustrated.

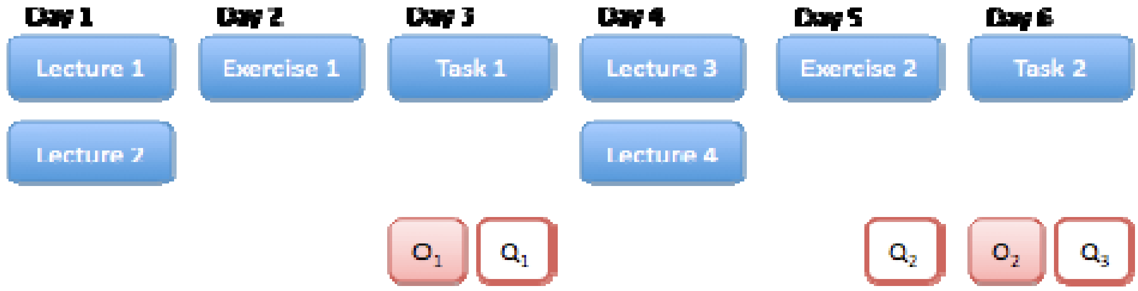

Figure 4. Activities and analysed results during the course duration.

2.2.4 Details on the experiment tasks

In Appendix 1 the tasks given to the participants are shown in detail, exactly as they were given to the participants. In this section we briefly describe the modelling tasks, in order to give an idea of the modelling issues involved.

The first task (Task 1) was, as mentioned previously, set in the music industry domain. The context of the task, i.e. the intent and task of the ontology, was described as follows:

The Warner Bros recording label has decided to manage its own productions by means of an ontology-driven application. They provide the designers with documents describing scenarios that have to be stored in the knowledge base. From these documents one story is extracted and assigned to you as ontology designers.

The requirements of the modelling task was given to the participants in the form of a small ‘story’, representing a typical example of information to be recorded in the knowledge base. The small story given to the participants for the first task was the following:

The “Red Hot Chilli Peppers” are: Anthony Kiedis (vocals), Flea (bass, trumpet, keyboards, and vocals), John Frusciante (guitar), and Chad Smith (drums). During 2005, the band recorded the album “Stadium Arcadium”. The album contains 28 tracks and has been released in May 2006. It includes the track of the song “Hump de Bump”, which was composed in January 2004. The critic Crian Hiatt defines the album as "the most ambitious work of its twenty-three-year career".

The second task (Task 2) was set in the hospital domain. The context was described as follows:

The Italian Ministry of Health wants to monitor the roles taken by employees in hospitals, and is creating a semantic infrastructure for that purpose. The following story is typical of the facts to be represented in its knowledge base.

The subsequent ‘story’ was the following:

Pasquale Di Gennaro is the union representative for male nurses at Ospedale Riunito delle Tre Valli in Nocera Inferiore (IT) from 2002.

2.3 Findings and observations

Below, the three sessions are treated one by one, the backgrounds of the participants are summarised and results of the questionnaires and analysis of the ontologies are given. Some conclusions are drawn specifically for the individual sessions, and the sessions are compared. The data from the questionnaires, the basis of this summary, can be viewed in detail in Appendix 1.

2.3.1 User study #1: PhD Course in Bologna

The results of the experiment in the first setting, the PhD course in Bologna, are discussed in this section. First we summarise the background of the participants, then we present the results of the three questionnaires given, and the analysis of the ontologies (O_1-O_2) from the two experiment tasks.

Participants' background

The total number of participants of the course was 18, although not all of them participated in all exercises and not all answered every questionnaire. The reason for missing results is the fact that not all participants could attend all days of the course. The results of the questionnaires were collected regardless of the participants' participation during the course, but the ontologies were only collected for those pairs that participated in both experiment tasks. Most of the subjects were PhD students or research assistants, with a master or a bachelor as their highest achieved academic level so far, but also some senior researchers participated (4 persons holding a PhD). Subject fields were mainly related to Computer Science, IT and Computer Engineering, but also 3 subjects from Electronic Engineering participated. About half of the participants claimed that their current work or studies were related to ontologies, only 5 stated that it was not related in any way.

Despite this, their experience in using and constructing ontologies proved to be very low. About half of the participants claimed to have little or no experience in working with ontologies, only 6 participants claimed to have more than a few weeks of total experience. 7 participants had never constructed any ontology, and 5 stated that they had only been involved in constructing a single ontology. Of the ones that did construct at least one ontology previously, half had only constructed very small toy examples, and all but one had been simple taxonomies or lightweight ontologies. None had ever used any ontology design patterns for modelling.

With respect to technical experience concerning tools and languages, about half of the participants had tried Protégé 3.x at some point, but only two users had tried the tool used for the experiment previously (TopBraid Composer), and very few had tried any other tools. More than half of the participants were familiar with ER-diagrams and UML modelling, and about half were also somewhat familiar with first-order logic. However, only 5 people had tried OWL (the language used for the experiment), previously, but about half had some experience with RDF as the data model that underlies OWL.

In summary, this user group consisted of mainly inexperienced ontology developers, some that had tried to model example ontologies using tools like Protégé 3.x, but very few were actually experienced and knew the languages and logics behind. Nevertheless, this group consisted of mainly PhD students and researchers, in computer science and related fields, who attended a course to learn about ontologies, so they could be considered highly motivated and probably they would learn quite rapidly.

Experiment tasks

The first few questions with respect to the task and how the participants solved it were identical between the first and the second task (answer sets Q_1 and Q_3 of Figure 2). The following propositions were given and the users were asked to rate those on a five-level Likert scale from strongly disagree to strongly agree:

1. I found the problem description easy to understand.
2. I felt familiar with the domain of the modelling problem.
3. The problem was clearly and unambiguously defined.
4. The modelling problem (the ontology) was small compared to other ontologies I have constructed before.
5. I found the tool used for the experiment easy to use.
6. The modelling problem was easy to solve.
7. I made some mistakes at first and had to redo some parts of the ontology later.
8. There were some problems that I did not manage to solve in a “good” way within the given time limit.

With respect to the first proposition the problem description of the second task was perceived as slightly more difficult by some participants, still 12 out of 14 that answered the question after the second task found the description relatively easy to understand. A majority of the participants also felt familiar with the domain of the task, in both cases. Despite some single different opinions, most participants agreed that both problems were clearly and unambiguously defined.

There was a disagreement whether the ontologies to be constructed were small or not, clearly dependent on the previous experience, and if anything could be concluded at all then the tendency was to perceive the second one as larger than the first task. There was a general agreement that the tool was easy to use, although some found it more difficult in the second task rather than when they first used it. This might indicate that it was in fact not so intuitive to use the content patterns within this tool.

All but two participants found the modelling tasks relatively easy to solve, although some agreed more strongly for the first task than for the second. Just over half the participants stated, in both cases, that they had to rethink some modelling choices after a while and re-model the ontology.

For the final proposition there were fewer people who disagreed for the second task than the first, indicating that the second task was harder and/or that it took more time solving it. This is also supported by the answers after the pattern exercise (Exercise 2, answers recorded in answer set Q_2 of Figure 2) where as many as 60% of the participants felt that they had problems left they did not manage to solve within the given time (compared to 20% and 23% respectively for the first and second task discussed above). One conclusion is that patterns take time to use when first introduced, and it is not certain that they will give any time benefits even after proper training.

After both tasks (Task 1 and 2 in Figure 2) and the pattern exercise (Exercise 2 in Figure 2), the users were asked to list specific problems that they had encountered during the modelling. Their answers are summarized in Table 1. For the first task some of the problems were related to how to model specific things in the task, such as how to model time (Table 1: Time-modelling) or events (Table 1: Event-modelling), how to use unions and other OWL class constructors (Table 1: OWL class constructors), and how to find the correct datatype to use for a literal (Table 1: Correct datatype). The last of those problems is also possibly related to being unfamiliar with the tool used for the experiment. One participant also mentioned the problem of using the theoretical knowledge gained during lectures for solving a practical task (Table 1: Theory to practice).

Three participants mentioned the requirements of the task and the fact that they did not know the exact usage of the ontology, and thereby had a hard time to interpret the requirements (Table 1: Ambiguous requirements/unclear task). This problem is also related to the level of detail that one subject mentioned (Table 1: Level of detail), if the usage of the ontology is not completely clear then it is hard to decide on an appropriate level of detail. Other general problems are related to how to model n-ary relations (Table 1: Modelling n-ary relations) that are not directly expressible in OWL, and how to make ‘good’ modelling choices even with limited experience (Table 1: Modelling choices).

On the second day, after the pattern exercise (Exercise 2 in Figure 2), the main problem, recognised by 4 participants, was the time it takes to get used to, and start understanding, the patterns (Table 1: Learning curve). They recognised that there is quite a long learning curve, when trying to understand something as complex as an ontology pattern. Several participants had problems to match the requirements to the descriptions of the patterns (Table 1: Matching problem to pattern), even though they were given a specific method to do this (by comparing competency questions). To really understand the meaning of the pattern was another issue (Table 1: Semantics of patterns), as well as how to reuse the patterns when imported (Table 1: How to reuse patterns), what properties to specialize etc. Two participants also noted that it was quite time consuming to select and reuse the patterns (Table 1: Time consuming), at least at this stage of learning. Pattern selection was another problem (Table 1: Pattern selection), how to choose the 'best' pattern for the modelling task at hand. Finally, one person still had problems with the OWL language, namely, restrictions (Table 1: OWL restrictions).

Table 1. Problems, as listed by the participants

Did you encounter any specific problems while designing the ontology?

First task	No. of answers	Pattern exercise	No. of answers	Second task	No. of answers
Ambiguous requirements /unclear task	3	Learning curve	4	SPARQL queries	2
Modeling choices	2	Matching problem to patterns	3	Do suitable patterns exist?	1
Correct datatype	2	Semantics of patterns	2	Pattern selection	1
Time-modeling	2	How to reuse patterns	2	Partial reuse of patterns	1
Event-modeling	1	Time consuming	2	Composing patterns	1
Modeling n-ary relations	1	Pattern selection	1	Ambiguous requirements /unclear task	1
Level of detail	1	OWL restrictions	1	Too general patterns	1
OWL class constructors	1			Missing patterns	1
Theory to practice	1			Modeling roles	1

For the second task (Task 2 in Figure 2), the new task where patterns were again used, some participants recognized the difficulty in formulating appropriate SPARQL queries for testing their ontologies (Table 1: SPARQL queries). Some specific issues, such as modelling the roles of persons (Table 1: Modeling roles), were listed together with the general problem of unclear usage of the ontology leading to difficulties interpreting the requirements (Table 1: Ambiguous requirements/unclear task), also noted for the first task. The rest of the issues were related to the patterns, how to know if any suitable patterns existed at all (Table 1: Do suitable patterns exist?), how to select the most suitable pattern (Table 1: Pattern selection), and when reusing it how to know what parts to include, how to discard the rest (Table 1: Partial reuse of patterns), and how to compose several patterns (Table 1: Composing patterns). Some patterns were also perceived as too general (Table 1: Too general patterns), more specific patterns would have been more useful, and some noted that the catalogue was not complete even at this general level (Table 1: Missing patterns), some 'expected' patterns were missing.

Next, a question tried to determine how the patterns were in fact used by the participants, as illustrated in Figure 5. It turned out that almost all participants used the patterns as actual building blocks, not only for inspiration. One user stated that the patterns were used only for inspiration during the first pattern exercise but for the second task none of the users selected this response. The participants could also distinguish between using the complete patterns or only parts of them in their response, but here there was no clear preference, and some participants checked both these alternatives, possibly indicating that it depended on the pattern at hand.

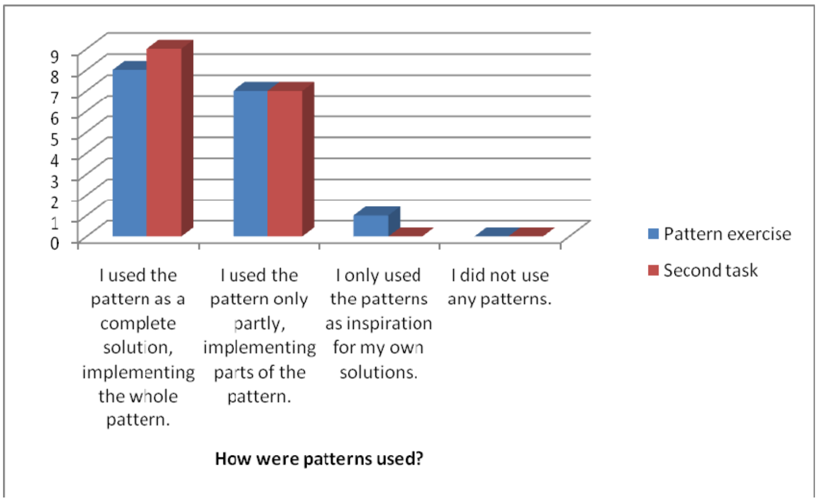

Figure 5. How patterns were used by the participants.

When introducing the patterns, some more specific questions were also asked with respect to the patterns themselves and their usage and usefulness. The following propositions were given, both after the initial pattern exercise and after the second task where patterns were also used, and the participants were again asked to rate them on the same scale as previously:

1. The patterns were clear and easy to understand.
2. The tutorial/course material presented before this exercise was useful for understanding the patterns.
3. The patterns were easy to use.
4. Some of the patterns were “obvious” and trivial.
5. Some of the patterns introduced useful solutions that I did not think of before looking at the pattern.
6. In general, I found the patterns useful.

The responses to the first proposition clearly show that the patterns were not so easy to understand at first, 4 out of 11 participants clearly disagreed and 5 were uncertain after the initial pattern exercise. This improved after the second task where only 2 out of 14 users disagreed, and 11 agreed that the patterns were indeed easy to understand. This clearly indicates that some training and experience is needed in order to understand the patterns properly, although it should also be noted that already one day of training gave this substantial improvement. There was some slight disagreement after the first pattern exercise if the training material was useful for understanding the patterns, but after the second task all but one participant agreed that it was in fact useful.

Next, it was proposed that the patterns were not only easy to understand but also easy to use, where we can again see an improvement from the first pattern exercise to the second task. Although even at the second task two subjects still disagreed with the statement. It should be noted that the tool used had no special pattern support, but the patterns had to be downloaded as OWL-files and imported as any other ontology, whereas it is not so surprising that the process is not found very user friendly. There is no specific support for further pattern operations such as specialization or composition, other than the basic operations available in any ontology editor.

There was a disagreement around the issue if some patterns were obvious and trivial. Many participants noted some patterns that they found trivial, but quite a few also found no such pattern. At least for the next proposition, the other end of the scale that some pattern introduced really new ideas that the participants had not thought of before, there was a slight tendency that more users agreed to this than disagreed. Still, such a question is inherently difficult to answer since nobody can be sure that they would not have come up with a similar idea themselves, given time, so the results should not be interpreted too positively. In the end, it is the final question that is the most interesting one, did the participants find the patterns, in some sense, useful? No participant disagreed after either exercise, and after the second task as many as 64% of the participants strongly agreed that the patterns were in fact found useful.

Additionally, the participants had the opportunity to motivate why they found the patterns useful, and a summary of their responses may be seen in Table 2. Modularization (Table 2: Modularization), i.e., making the ontologies more modular, increased quality (Table 2: Quality) of the constructed ontologies, and the possibility of reusing best practices (Table 2: Reuse) are three of the main benefits that the subjects proposed. Most benefits were listed as single terms without explanation, hence we can here only speculate in what the participants actually meant by their stated benefits, e.g. how they would define ‘increased quality’. Also the fact that an ontology based on patterns is clearer and understandable by other people (Table 2: Understandability) was mentioned by several participants, and one person mentioned the increased re-usability this can lead to (Table 2: Re-usability). The patterns also help to decompose the problem when modelling (Table 2: Problem decomposition), and since parts can be imported directly some ‘routine work’ can be avoided (Table 2: Limit ‘routine work’), compared to just having the pattern as a guideline in a book. Patterns are also viewed as guidelines in themselves (Table 2: Guidelines) and they give new insights into general modelling issues (Table 2: Insights into general problems), and point at common problems. Some subjects found it faster (Table 2: Faster implementation) and easier (Table 2: Easier design) to model using patterns and also noted that patterns made some modelling choices explicit (Table 2: Make choices explicit) and helped to integrate different parts of the ontologies (Table 2: Integration). It is not clear exactly what was meant by integration however.

Table 2. Motivations and proposed rationale for pattern usefulness.

How were the patterns most useful?

Pattern exercise	No. of answers	Second task	No. of answers
Modularization	4	Quality	2
Understandability	4	Reuse	2
Reuse	3	Problem decomposition	2
Quality	3	Modularization	2
Limit ‘routine work’	2	Guidelines	1
Guidelines	1	Insights into general problems	1
Make choices explicit	1	Faster implementation	1
Faster implementation	1	Make choices	1
Problem decomposition	1	Re-usability	1
Insights into general problems	1	Easier design	1
Integration	1		

Even though we have already noted that the first time the participants used the patterns they did not give an optimal effect, and that patterns really need to come with proper training, the users were anyway asked to compare the first task (Task 1) and the initial pattern exercise (Exercise 1). Since this really involved just redoing an exercise they had already done the day before, it was a good opportunity to record their opinions on how they thought that the patterns could be of assistance. Three propositions were given, that the participants had to rate as before:

1. The patterns made it easier to solve the task.
2. The patterns made me solve the task faster.
3. I believe that I solved the task better (constructed a better ontology) now that I had access to the patterns.

For the first and the third proposition there is a clear tendency towards agreement, from a majority of the participants. The strongest seems to be the last one, where only one user disagreed. For the second proposition there is no support, the participants have differing opinions, and if anything there seems to be a slight tendency towards disagreeing with the proposition. A conclusion would therefore be that the subjects perceived that the patterns helped them to make a better ontology, possibly even slightly easier than before, but probably not faster.

This is also supported by their free text comments related to these questions. Several of the participants noted that the final task was performed better because they had now become familiar with both the tool and the patterns, and were able to use them efficiently, but they also noted that using patterns takes time, especially at the beginning when you do not have long experience. Still, a couple of users thought that the first two exercises were anyway slightly easier to solve since they had more help from the problem description and the tasks were easier in themselves.

To compare the subjective opinions of the participants with their actual results, also the ontologies resulting from the first and the second tasks (O_1 and O_2) were analyzed as described previously. The coverage of all the ontologies resulting from the first task was high, between 85% and 93%, on average 89%. Only minor things were usually missing in the ontologies, such as the distinction between a song and its realization as a track on an album. Also for the second task the coverage was high, but slightly lower than for the first task, this time in the range between 70% and 83%, average 79%. This time the general feeling was that in several cases some parts were actually missing because time had run out, rather than that some things missing were just forgotten, since some solutions were really unfinished.

With respect to the usability aspects of the ontologies, as defined previously, the naming of concepts and properties was in most cases (for both tasks) logical and quite reasonable, although usually no general naming convention was applied, whereby some mixed styles were sometimes used and there were cases of misspellings. A summary of the other usability features of the ontologies of the first task can be seen in Figure 6, and the ones after the second task in Figure 7. The fractions have been classified into the following categories: ‘none’, ‘some‘, ‘most’, and ‘all’. Some in this case denotes less than two thirds, and most denotes more than two thirds.

The ontologies resulting from the first task have no comments at all and with respect to labels a few subjects specified labels for all concepts and properties but many did not add any labels. Disjointness axioms were added by a few of the subjects, and in this case they added them for all concepts. The level of axiomatization (formal definitions of concepts, such as subclass or equivalent class axioms, was reasonably high, as well as definitions of inverse properties (which is generally good practice). Still, we can note the improvement in the results of the second tasks, where all ontologies contain at least some of the features involved. Of course there is usually no need for formally defining all concepts in an ontology through axioms, so we are not striving for all of the features to have the value ‘all’. Rather, it is suitable for formal definitions and disjointness axioms to be on this level, while labels, comments and possibly even inverse relation definitions should ideally be defined for all applicable elements. Thereby, in the second task the participants have reached a great deal closer to this goal.

When trying to find common problems in the solutions for the first task three main problems were identified, as illustrated in Figure 8. A problem/mistake related to songs and tracks was displayed by 7 out of 8 ontologies. This problem arises from the task description saying that a certain band recorded a certain song on a certain album. The common mistake is to simply ignore the fact that songs and tracks are different things and model them either as the concept ‘song’ or ‘track’ and add both composition date and recording date as a property of this same concept. Intuitively however we know that it is a song that is composed at some point in time while it is the track that is recorded, and this is only a realization of the song, not the song in itself. Only one ontology observes this issue and models the songs and tracks as different, although related, concepts.

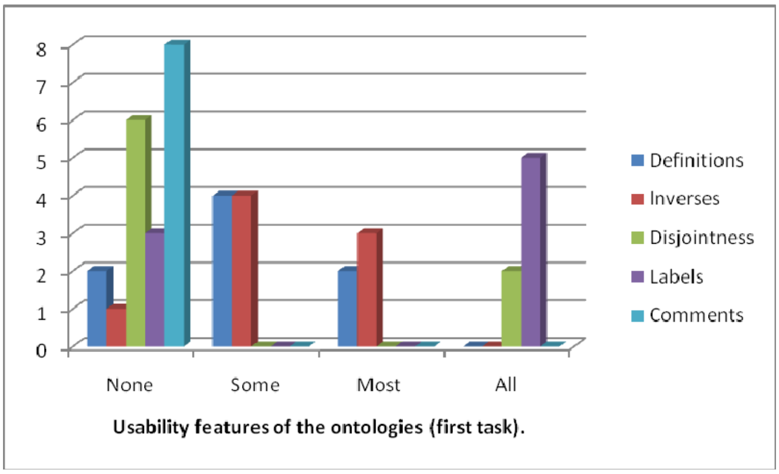

Figure 6. Usability features, results from the first task.

The persons and roles problem/mistake was found in 6 out of 8 ontologies. This issue arises from the fact that a musician is not a specific kind of person, but rather a role that a person can take in some situation at some point in time. Most ontologies do not recognise this distinction, but models musicians and music critics as specializations of the person concept, rather than as roles. The roles that are recognised, e.g., to have a certain role in a band such as a guitar player or a vocalist, are usually modelled statically as the role of a certain person, without taking into account that this role might change over time or that the person may have different roles in different settings, i.e., being member of several different bands with different roles in each.

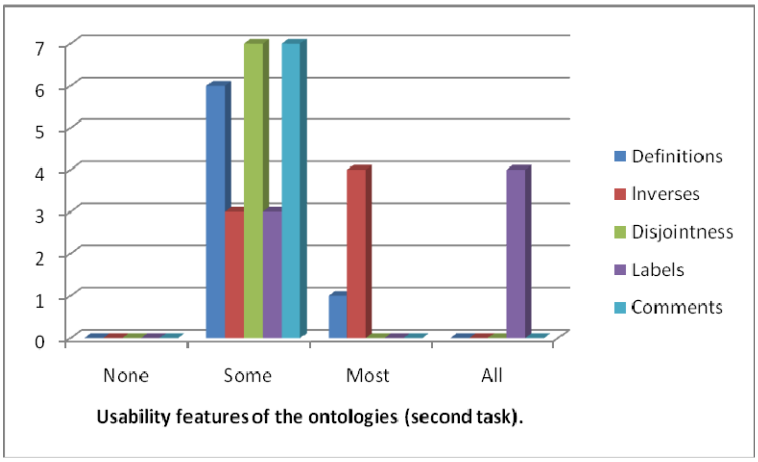

Figure 7. Usability features, results from the second task.

Finally, the third common problem/mistake is related to the notion of album reviews included in the task description. A music critic can write a review of a certain album. This is inherently a ternary relation, between the author of the review, the album being reviewed and the content of the actual review. However, in three of the ontologies this is not recognised as such, and the models only include a binary relation, thereby losing one of the connections, either to the album, the critic or the comment itself.

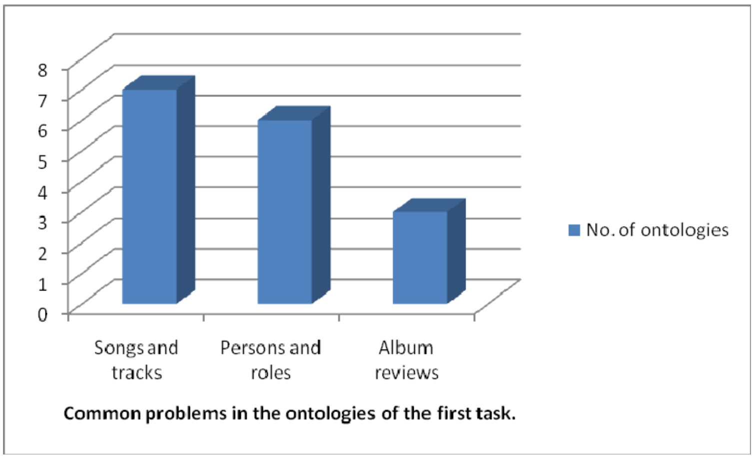

Figure 8. Number of ontologies displaying common problems (first task).

In the second task there were no problems in the task description comparable to the songs and tracks problem, however the other two problems were again manifested in the second description. This time as the professional roles held by people in a hospital, nurses and union representatives, and n-ary relations being present in the situation of representing a certain collective during a certain time period within a certain organisation. In the set of ontologies representing the solutions to the second task all ontologies succeed in observing and correctly representing these issues. Instead only one new common problem could be found in the solutions to the second task, which was the definition of a certain collective. Collectives were usually represented correctly, but in none of the ontologies were the sub-collective of male nurses properly defined. In some ontologies it was included as a named class but in none of the ontologies, even those including concepts such as genders and the group of all nurses, was the notion of male nurses defined. In our opinion this is either a misinterpretation of the task, not feeling the necessity to model this definition, or an issue of leaving the hardest problem to the end and then not having time to solve it, rather than an issue that was incorrectly modelled as the problems within the ontologies of the first task.

The last aspect to be analysed within the set of constructed ontologies was the presence of patterns. First it is interesting to see whether some patterns were used ‘unintentionally’, using the ‘pattern solution’ without knowing about patterns, during the first task, but also if some of the common problems above are such that patterns in the catalogue could have solved them. Above, we discussed three main common problems/mistakes found in several of the ontologies of the first task, the songs and tracks problem, the persons and roles problemand the album review problem. All of these are actually closely related to patterns available in the catalogue, that are intended for exactly these types of modelling issues. The information realization pattern addresses the semiotic aspects of information, meaning the difference between the abstract piece of information and its realization as an object in the world. This pattern exactly conforms to the requirements of the songs and tracks modelling problem, where the song is an abstract piece of information that may be realized on physical media as a recorded track. Next, the persons and roles problem/mistake is related to the agent role pattern, addressing exactly the situation when some agent can take different roles, different from its inherent identity as a type of agent, e.g., a person. Finally, the n-ary relation problems present in the album review issue is addressed by the situation pattern, where situations represent settings where several entities are involved, including possible notions such as time and location although this is not explicitly stated in the general pattern.

It may then be noted that out of the 5 ontologies where the album review problem/mistake was correctly modelled all of them unintentionally used exactly the same type of solution as suggested by the situation pattern. For the two ontologies correctly modelling persons and roles they have both solved the issue in other ways than suggested by the agent role pattern. One solves the issue by defining the roles as concepts by themselves, through formal definitions and restrictions on properties, and the other ontology avoids the problem by simply not including the person concept directly and instead focusing only on the roles, so to model only the roles of the people rather than the people themselves. The last solution may not be a very good one in general, but for this restricted problem description it fulfills the requirements. Finally, the one ontology properly addressing the songs and tracks issue uses an identical solution to the one suggested by the information realization pattern. So we can see that in 6 out of 8 cases where the common problems that were found difficult by the other subjects were actually solved in a correct way, suitable patterns were actually used, although without knowing it.

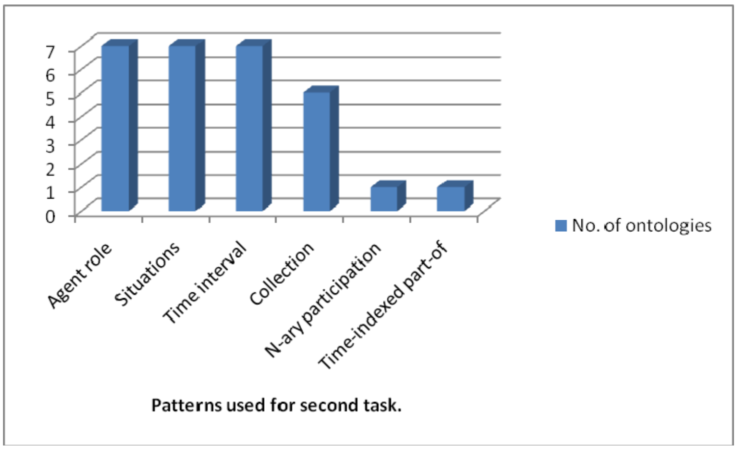

Figure 9. Number of ontologies using specific patterns used in second task.

In the second task where patterns were available, the patterns used are summarised in Figure 9. All ontologies use the agent role, situations and time interval patterns[2], while only some use the collections pattern. One of the ontologies also uses the n-ary participation patterns and another one uses time-indexed part of, these are both used as variations of the same solutions as achieved with the situation pattern. The collection pattern, although very simple, seems not to be entirely intuitive since this is the area where the mistakes are still made in the solutions to the second task (as noted previously). It may actually be the case that the pattern is a bit too simple to assist in solving more complex accounts of different collectives and sub-collectives, whereas some specialisations of the pattern may be called for.

Summary of results for the first session

In summary the participants of the PhD course in Bologna, were inexperienced but highly motivated. Most of them had some experience in different types of modelling, but only a few had any experience in constructing and using ontologies and OWL.

They perceived the patterns as useful, but had some initial problems when understanding and using them. The participants perceived that they constructed ‘better’ ontologies when using the patterns, possibly also slightly easier (although there was some disagreement), but not faster than before. All of the common problems present in the first set of ontologies, solvable using patterns, were solved in a correct way the second time. Only one new common problem was introduced, and this may be due to the formulation of the task rather than the difficulty of the modelling issue. The greatest improvement of the ontologies could be found in the usability aspect, where the ontologies of the second task, for example, contained a lot more labels and comments.

The content patterns were used mostly as complete building blocks, imported and reused within the ontology, rather than just inspiration for own solutions. Nevertheless, the patterns were sometimes hard to understand, it was hard to find and select the right one, to reuse it in a correct

[2] For details on these patterns please visit the ODP portal at http://www.ontologydesignpatterns.org

way and to compose several patterns. This indicates that further tool support for ontology design pattern reuse is definitely needed.

2.3.2 User study #2: Dedicated session in Jönköping

The results of the experiment in the second setting, the dedicated experiment session in Jönköping, are summarised in this section. First we summarise the background of the participants, then we present the results of the two questionnaires given (for this session we remind the reader that the initial pattern training exercise, Exercise 2 in Figure 3, was not performed), and the analysis of the ontologies from the two experiment tasks (O_1 and O_2 in Figure 3). The results were quite similar to the first session and in some cases we therefore simply refer to those results instead of repeating them.

Background

The total number of participants of the session was 8, although not all of them answered all questions in the questionnaires and one participant did not manage to save the resulting ontologies properly, therefore one set of ontologies is missing from this session. The reasons for not answering all questions in the questionnaires are not clear, but could be related to the individual motivation of the participants or the length of the questionnaires. In this session the participants modelled the ontologies individually, unlike the two other sessions where the participants worked in pairs. Most of the participants were PhD students, but additionally one research assistant and two senior researchers participated. Half of the participants in fact already hold a PhD, the rest hold a master or bachelor's degree. Subject fields were mainly related to computer science and IT, but also one participant from the Product Development field. 6 out of 8 of the participants claimed that their current work or studies were related to ontologies, 2 stated that it was not related.

Their experience in using and constructing ontologies proved to be a bit higher than the first group in Bologna. As many as 6 out of 8 of the participants claimed to have more than a few months of experience working with ontologies, and only one claimed to have no experience at all. Nevertheless, only 3 users claimed to have been directly involved in projects designing ontologies, although two of those users had participated in several projects. Of the three people that had been involved in ontology design the focus seemed to be on medium sized ontologies, but quite light weight in nature. None had ever used any ontology design patterns for modelling.

With respect to technical experience with tools and languages, 6 of the participants had tried Protégé 3.x at some point, but only two had tried the tool used for the experiment (TopBraid Composer) previously. A few had also tried other tools, primarily OntoEdit or OntoStudio. Most participants were somewhat familiar with ER-diagrams and UML modelling, and more than half had also tried both first-order logic, RDF and OWL at some point.

In summary, the group consisted of mainly inexperienced ontology developers, although slightly more experienced than the previous group. Some had actually participated in projects constructing lightweight domain ontologies, although most subjects seemed more familiar with using ontologies, or ontologies as a concept, rather than how to construct them. This group consisted of mainly PhD students and researchers, in Computer Science and IT-related fields, but they attended the session because they were asked to participate in an experiment, which could affect their motivation, but being researchers they should still be able to learn modelling skills quite rapidly.

Experiment tasks

The first few questions with respect to the task and how the participants solved it were as in the first session identical between the first and the second tasks. With respect to the understandability of the task the problem description of the second task was also in this session perceived as more difficult by some participants, still 5 out of 8 found the description easy to understand also for the second task. None of the users felt unfamiliar with the domain of the task and most participants agreed that both problems were clearly and unambiguously defined. The ontologies were perceived as relatively small compared to previous experience, although it should be noted that many participants did not respond due to lack of previous experience. Not all participants felt the tool was easy to use for the first task, but this is not very surprising since these people did not have any previous practice using the tool (as opposed to the participants in Bologna doing a practice exercise, Exercise 1, the day before the first experiment task, Task 1). This is supported by the fact that for the second task most participants agreed that the tool was relatively easy to use.

All participants found the first modelling problem relatively easy to solve, while 6 out of 8 found the second task relatively difficult. Again, this is most likely an effect of the lack of training, since for these participants this was their first encounter with the patterns and the task was generally perceived a bit more difficult than the first. Most users agreed that for both tasks they had to re-model some parts after a while. With respect to the time to solve the tasks many participants disagreed for the first task, and thought the time was sufficient, while for the second task all the respondents agreed that there were some problems they had to leave unsolved. Here it should both be taken into consideration that this group had less training than the previous one, and that they had less time to complete both the tasks. The time was obviously not enough for solving the second task even though (or perhaps also because of) they had the pattern catalogue this time.

The participants then listed specific problems that they had encountered during the modelling. Their answers are summarized in Table 3, but similarly as for the previous session we do not always have a clear motivation or explanation for the problems listed. For the first task most of the problems were related to how to model certain constructs in OWL, such as n-ary relations (Table 3: Modeling n-ary relations), data types (Table 3: Data types) and properties in general (Table 3: Creating properties). One participant also had problems to use the tool properly (Table 3: Tool), e.g., accidentally closing windows and not finding the right view again, and another noted that the time for completing the task was too short (Table 3: Lack of time), here we remind the reader that this group only had one hour for each task instead of two. Finally one participant who was previously used to more traditional information modelling formalisms noted that it was hard to remember not to use classes as property values (Table 3: Classes as property values), and the participant had to remodel the ontology in several cases.

After the second task, where patterns were first introduced, there was a larger set of problems reported. The main problem was confusion around how to reuse the patterns (Table 3: How to reuse patterns) properly, what to specialize and how. Also understanding the patterns posed a problem (Table 3: Understanding patterns), at least during the limited amount of time available. How to match the requirements of the task to the specification of the pattern was also a difficult issue (Table 3: Matching problem-patterns), as well as more specific modelling problems such as modelling roles and their connection to persons and their duration in time (Table 3: Person/role and temporal roles). Some participants also noted that there were patterns missing (Table 3: Missing patterns) that they would have liked to use, that some of the patterns were almost too trivial (Table 3: Trivial patterns) and did not help all that much, and in some cases the use of patterns increased the complexity of the model (Table 3: Complexity of the model) so that the participants started to question if it was really motivated to use the pattern at all for this small task. Finally, a participant noted that it was really time consuming (Table 3: Time consuming) to understand and use the patterns, at least at this stage, when not being very familiar with the patterns.

It turned out that all participants, also in this session, used the patterns as reusable building blocks, not only for inspiration, as illustrated in Figure 10. The users could also distinguish between using the complete patterns or only parts of them in their response, but here there was no clear preference. Probably some patterns were usable as a whole while some were only partly reused. Some free text comments with respect to how patterns were used were also collected and primarily the participants felt that the patterns did not cover everything they wanted to model (not very surprising) so they had to extend and specialize the patterns. Some participants found the very general patterns quite useful, they were highly reusable in many situations, while another participant instead viewed this as a drawback stating that some patterns were too general and trivial to provide any benefit when reusing them.

Table 3. Problems, as listed by the participants in Jonköping.

Did you encounter any specific problems while designing the ontology?

First task	No. of answers	Second task	No. of answers
Modeling n-ary relations	2	How to reuse patterns	5
Lack of time	1	Understanding patterns	3
Tool	1	Matching problem-patterns	1
Data types	1	Missing patterns	1
Creating properties	1	Trivial patterns	1
Classes as property values	1	Time consuming	1
		Complexity of model	1
		Person/role and temporal roles	1

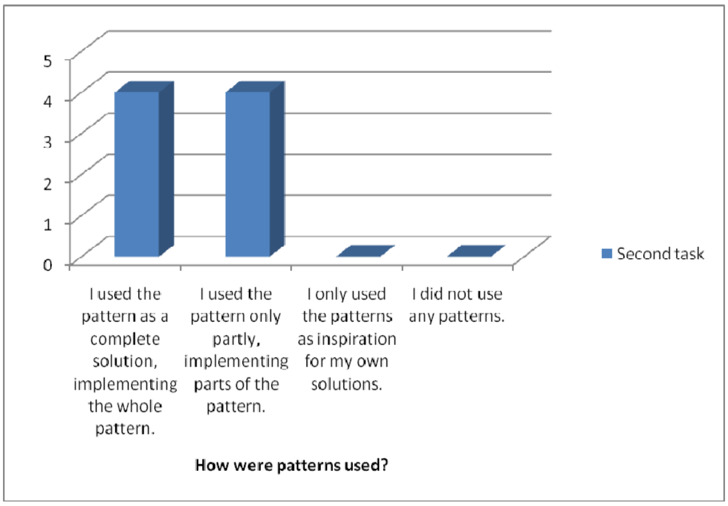

Figure 10. How patterns were used in Jonköping.

When introducing the patterns specific questions were again posed with respect to pattern usage and usefulness. 6 out of 8 participants felt that the patterns were to some extent hard to understand. Whether the tutorial and course material presented helped to understand the patterns or not there was no agreement, many people were unsure. Half of the participants also disagreed with the statement that the patterns were easy to use, only one agreed. Whether or not some patterns were trivial there was again a disagreement but at least there was a slight tendency towards agreement that some patterns actually did provide solutions that the subjects did not think of beforehand. Half of the people (4 out of 8) found the patterns useful, only 2 did not find them useful at all, the rest were unsure.

Comparing these results to the other two sessions we may note that these results are in many ways similar to the ones achieved after the first pattern training exercise in those sessions. The users found the patterns difficult to understand and to use and they sometimes did not see the point in using them. This again supports the conclusion that has been mentioned previously, that patterns are not usable (perhaps no even useful) without proper training and experience. Although the extent of the needed experience and training may vary, the other two sessions show that even with a very limited amount of training the results improve considerably.

The subjects had the opportunity to motivate why they found the patterns useful, and a summary of their responses can be seen in Table 4 below. As in the previous session, all suggested benefits were not motivated, hence it is sometimes not clear exactly what the participants mean with terms such as ‘easier’ or ‘more general’. Nevertheless, increased quality of the constructed ontologies (Table 4: Quality), and easier design (Table 4: Easier design) due to the direct reuse of already implemented components (Table 4: Reuse) were the benefits proposed by several of the subjects. Others found them useful since patterns provide example solutions to a problem (Table 4: Provide example solution) and will probably make the constructed ontology more general (Table 4: More general solution), in terms of generalisation rather than generality of the concepts involved. One subject also found the design process to be faster when using the patterns (Table 4: Faster implementation).

Table 4. Motivations and proposed rationale for pattern usefulness.

How were the patterns most useful?	
Second task	No. of answers
Quality	2
Easier design	2
Faster implementation	1
Reuse	1
More general solution	1
Provide example solution	1

Also in this session the participants were asked to compare between using patterns and not using patterns. In this case it meant comparing between the first and the second task. We used the same propositions as for the other two sessions, although it was of course harder for the participants in this group to compare between different tasks rather than between performing the same task with and without patterns. The only clear tendency was that for the proposition concerning if they were faster when using patterns, a majority of participants disagreed. This is also supported by the free text comments of the participants, regarding these issues. Most participants agreed that it was very hard to use the patterns for the first time, especially on this slightly more difficult problem. It took time to understand both the problem description and the patterns and with lack of pattern experience it was not easy to use them

The ontologies resulting from the first and the second experiment tasks were analyzed as described previously. The coverage of 6 out of 7 of the ontologies resulting from the first task was quite high, between 82% and 88%. Considering that these people constructed the ontologies in one hour less than the group in Bologna, this is very good. Nevertheless, one ontology had a coverage of 39%, giving an average for the total set of 79%, and one additional ontology was lost due to technical reasons when storing it and could therefore not be included in the analysis. In the ontologies with high coverage only very minor things were usually missing, such as forgotten instances or a forgotten restriction on a property. Also for the second task the coverage was high in most cases, but slightly lower than for the first task, this time in the range between 59% and 88% for 6 out of 7 ontologies. Similarly to the first task, one ontology had a coverage as low as 23%, reducing the average to only 64%, and similarly one ontology was lost due to technical reasons. As in the Bologna group the general feeling was that some parts were actually missing because time had ran out.

With respect to the usability aspects, no general naming convention was applied, similarly to the first session. A specific issue noted was that many times plurals were used in concept and property names, which is generally misleading and bad practice if the concept is intended to denote a set of individuals and not a set of classes. A summary of the other usability features for the first task can be seen in Figure 11, and for the second task in Figure 12. An improvement can be noted in the results of the second tasks, where most of the ontologies contain all of the features involved, except for one that is lacking labels and two that are still lacking disjointness axioms. In the second task the subjects have reached a great deal closer to the goal of producing understandable and reusable ontologies.

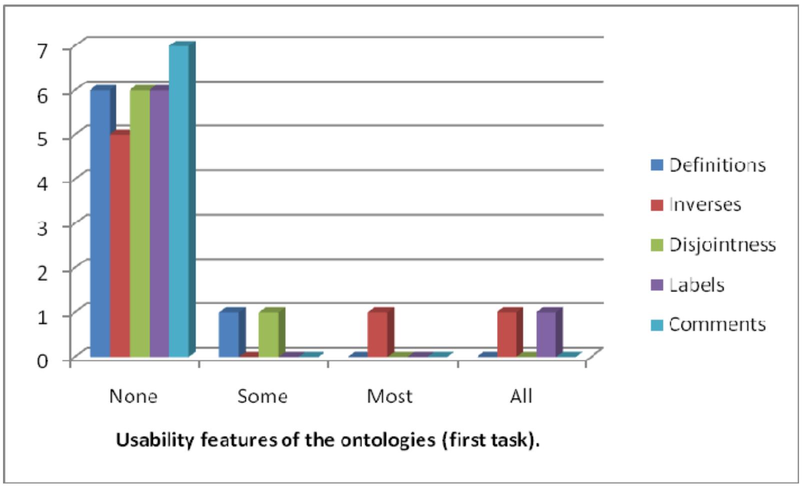

Figure 11. Usability features, results from the first task.

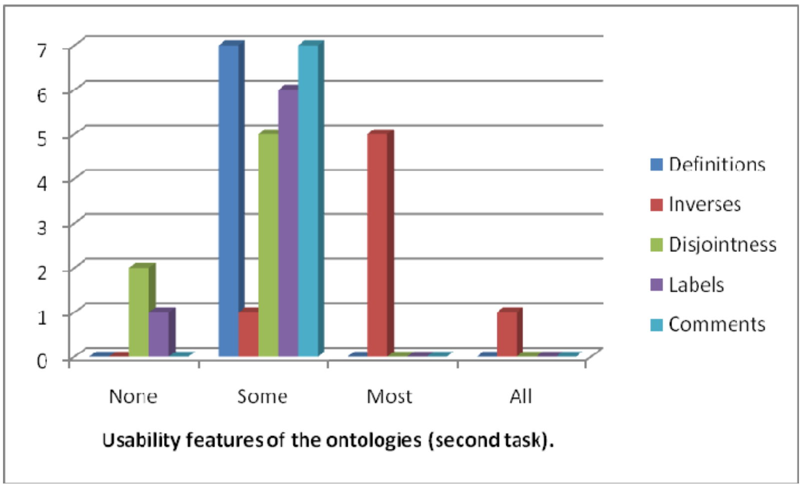

Figure 12. Usability features, results from the second task.

When trying to find common problems in the solutions for the first task three main problems were identified, as illustrated in Figure 13. These are the same three problems as were already identified, and described, for the Bologna session, the songs and tracks problem, the persons and roles problem and the ternary relation connected to the album reviews. All 7 ontologies display the songs and tracks problem/mistake and the person and roles problem/mistake, while only 4 out of 7 ontologies show the problem of defining a ternary relation for album reviews. Additionally one of the ontologies is not within OWL-DL, due to the fact that classes are used as property values. This is not a *common* problem since only one ontology displays this problem/mistake, but it is severe enough to be worth to mention.

In the set of ontologies representing the solutions to the second task all ontologies succeed in correctly representing the roles held by different people, although in some cases the additional n-ary nature of the problem in this specific task was not considered (roles being held for a specific time period or in a specific location or a specific group) by all. It seems that the n-ary relation problem still was not grasped by all participants, because 3 ontologies displayed problems expressing instances of n-ary relations. Collectives were usually represented correctly, but in none of the ontologies was the sub-collective of male nurses properly defined, just as for the ontologies from the Bologna session. All the ontologies for the second task are in OWL-DL, none uses any classes as property values as was the case for the first task.

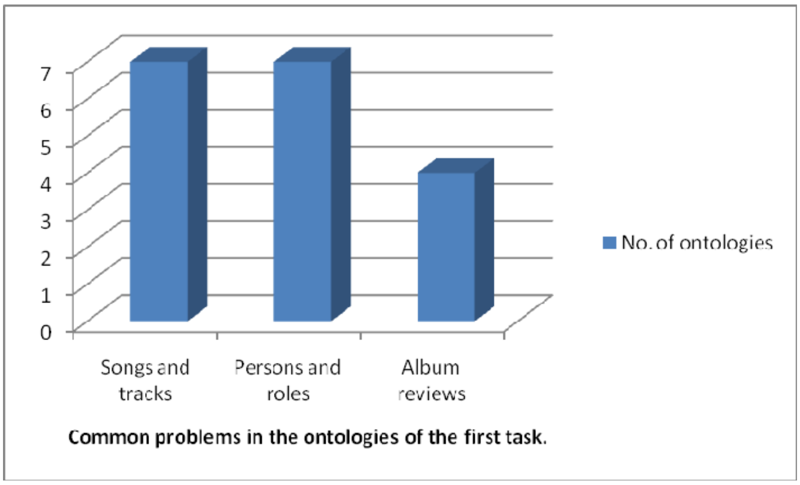

Figure 13. Number of ontologies displaying common problems (first task)

The last aspect to be analysed within the set of constructed ontologies was the patterns. All of the common problems discussed above are closely related to patterns available in the catalogue, as we already saw when discussing the Bologna session. It may be noted that out of the 3 ontologies where the album review problem was correctly modelled, all of them unintentionally used exactly the same type of solution as suggested by the situation pattern. The issue of using classes as property values was related to the notion of roles of the musicians in a band, and in this case, the roles were modelled as concepts rather than instances. It is of course not certain that they would not have been modelled similarly even if the agent role pattern would have been available, but at least this pattern and the examples included in its description might have reduced the risk of making this mistake.

In the second task where patterns were available, the patterns used are listed in Figure 14. All ontologies but one use the situation pattern. The time interval pattern was used by 5 out of 7 ontologies, while only 3 used the collection and agent role patterns. On the other hand the agent role pattern is a specialisation of the object role pattern, which was used in two of the ontologies. One ontology also uses the time-indexed part-of pattern and one uses the n-ary participation patterns, both are used as a specialisation of the situation pattern to address the n-ary relation problem. Finally, constituency is used in one ontology, to solve similar issues as the collection pattern. The choice of pattern is a bit more diverse here than in the other two sessions, but this is no surprise since the subjects in this case had no prior pattern training, no dedicated pattern exercise other than theoretical lectures, and thereby were not equally familiar with the patterns. Nevertheless, the participants did not do much worse than the other two groups. 3 of the ontologies fail to completely solve the n-ary relation problem, although all of them have imported the situation pattern. Sometimes the pattern is misused and sometimes all properties are simply not used in the instantiated situations. Other than that the remaining common issues are similar to the other groups, and mainly related to modelling the collectives and sub-collectives correctly.

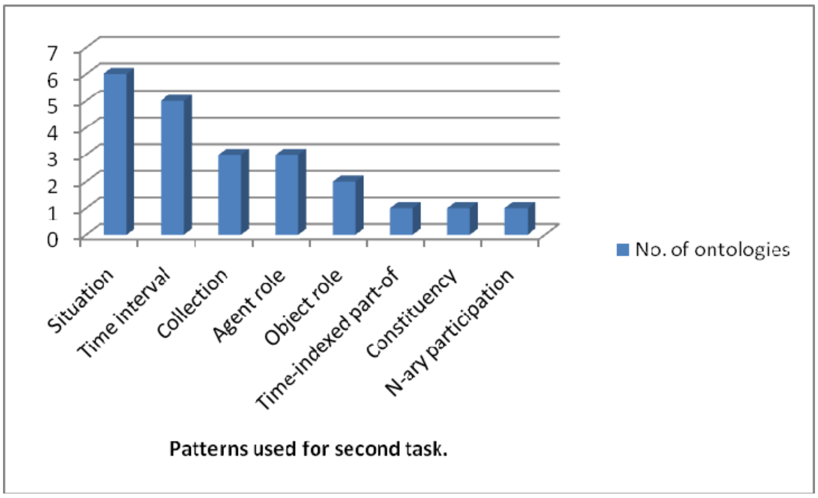

Figure 14. Number of ontologies that uses specific patterns used in the second task.

Summary of results for the second session

In summary, the participants of the dedicated experiment session in Jönköping were slightly more experienced that the participants in Bologna, but perhaps less motivated to really learn the modelling techniques. Most of them had some experience in constructing or using ontologies, but many had only worked with non-DL formalisms and only constructed simple lightweight ontologies.

About half of the group perceived the patterns as useful, but most of the participants also had problems to understand and use them. This was especially visible since this group got less training and had less time to complete the tasks. The users disagreed regarding whether they constructed 'better' ontologies when using the patterns, as well as if the construction was made easier by the patterns. Most participants agreed on the fact that they were at least not faster than without using the patterns. All of the common problems present in the first set of ontologies, solvable using patterns were attempted to be solved by using the patterns the second time, but in some cases the solutions were still not correct.

The content patterns were used mostly as complete building blocks, imported and reused partly or completely within the ontology. Nevertheless, the patterns were sometimes hard to understand, it was hard to find and select the right one, to reuse it in a correct way and to compose several patterns. This again indicates that further tool support for ontology design pattern reuse is definitely needed. From this second session we can also learn that patterns need to be properly introduced and exercised, before any of the benefits can be expected. Nevertheless, the quality of the ontologies still improved, even with this insufficient training, although at the expense of coverage since the patterns took time to understand and use.

2.3.3 User study #3: Master's Course in Jönköping

The results of the experiment in the third setting, the master's course in Jönköping, are summarised in this section. First we summarise the background of the participants, then we present the results of the three questionnaires given (Q_1-Q_3), and the analysis of the ontologies (O_1-O_2) from the two experiment tasks (Task 1 ad Task 2). Results similar to the other two sessions are only mentioned briefly, we focus on the differences rather than repeating the same findings.

Background

The total number of participants in this setting was 19, although not all of them answered all questions in the questionnaires. Just as for the first session in Bologna, all questionnaire responses were recorded and analysed, but the ontologies were only collected and analysed from those pairs participating in both experiment tasks (modelling was again done in pairs of two participants). The reason for missing questionnaire responses is not clear. However, in this group we could note that some participants at times had problems understanding the questions, but the reasons could also be related to individual motivation or the length of the questionnaires, as mentioned regarding the first session. Most of the participants were master students, all but one who probably took the course as an elective one. Subject fields were mainly related to computer science, IT, information engineering and information systems, but also a couple related to management. All participants except one stated that their current work or studies were related to ontologies in some way.

Their experience in using and constructing ontologies ranged from none to a few weeks, which is generally less than for the two other groups. 13 participants claimed to have participated in one or more ontology design project, which in their case referred to a set of labs in a previous course on information modelling. This also shows in their answers regarding the size and complexity of the constructed ontologies, which was stated by a majority to have been a small example ontology, either just a taxonomy or a light weight semantic net-like structure.

With respect to technical experience with tools and languages, 14 of the participants had tried Protégé 3.x, which was the tool used for the previous course and 6 claimed to have tried the tool used for the experiment (TopBraid Composer) previously. Most participants were somewhat familiar with ER-diagrams and UML modelling, also due to previous courses, about half had some knowledge of first-order logic, and a few had also tried RDF and OWL at some point.

In summary, the group consisted of mainly inexperienced ontology developers, the least experienced compared to the other two groups. The part of the group that had taken a previous course on information modelling had actually constructed some small lightweight ontologies previously, and used some ontology editor for this task. The group consisted of master students, primarily in the Information Engineering master program, but all of them were also exchange students, many of them from Asia. This means that some do not speak very good English, and some have shown in previous courses to be very inclined to always agree with the teacher, which is most likely a cultural issue. The consequence of this is that some questions may not have been correctly understood, and some answers are not easy to interpret. Additionally there is a risk that some single answers may not have been given completely honestly, if the student believed that the teacher wanted a certain response.

For the sake of this experiment this means that their results and answers should be interpreted in the light of their background as presented above, but also in the light of being the group where instructions and questions may easily have been misunderstood by some individual. Their learning speed and skills as a group could be considered as lower than for the other two sessions, since these are students and not researchers, which does not in any way hamper this experiment but could be useful for interpreting the results.

Experiment tasks

The problem descriptions were perceived as equally easy to understand, only one participant disagreed for the second task. With respect to the familiarity with the domain after the first task two participants stated to feel unfamiliar with the domain, but after the second task no subjects stated this. Most people agreed that both problems were clearly and unambiguously defined, although one person stated that it was not completely true for the first task, and for the second task more users tended to only agree to some extent or be uncertain.

There was a clear disagreement about the size of the ontologies; some felt that they were larger than what they had experienced previously and some that they were smaller. This is not so surprising since the ontologies they might have constructed before were actually similar in size, since they were similar lab assignments in another course. The tool seemed to create more problems for this group than for the previous ones, but still the clear trend was that more people agreed that it was easy to use for the second task, when they had become more used to the tool.

6 out of 19 (32%) found the first modelling problem a bit difficult, while only 2 out of 14 (14%) felt that the second problem was a bit difficult to solve. The participants felt the task was relatively easy to solve increased from 42% for the first task to 64% for the second task. This group of users obviously had some problems at first, since their background was the weakest this is not so surprising, but then really developed their skills during the exercises and then even felt that the less well specified task was easier than the first one.

Most participants agreed that for both tasks they had to remodel some parts after a while, and re-think modelling decisions, but slightly more so for the first than the second task. For the final proposition, this group is the only one where the ratio of people agreeing that they did not have time to really solve all problems in a ‘good’ way decreased for the second task. This may be due to that the people in this group found the first task more difficult, due to lack of experience, than the other groups, and thereby the second task was in comparison easier and faster to solve.

Just as in the previous sessions the participants were also asked to list specific problems that they had encountered during the modelling. Their answers are summarized in Table 5, below. For the first task most of the problems were related to how to use the tool properly (Table 5: Tool), and in some cases how to find the specific OWL constructs within the tool. The subjects found it hard to make certain modelling choices (Table 5: Modelling choices), e.g., how to know what was a ‘good practice’ when one could model in several different ways, and n-ary relations (Table 5: Modelling n-ary relations) again posed an especially hard problem. One user also felt that it was hard to interpret the requirements (Table 5: Ambiguous requirements/unclear task), especially without having a clear description of the purpose of the ontology.

Table 5. Problems, as listed by the participants in the third study.

Did you encounter any specific problems while designing the ontology?

First task	No. of answers	Pattern exercise	No. of answers	Second task	No. of answers
Tool	3	Pattern selection	2	Modelling choices	2
Modeling choices	2	How to reuse patterns	2	How to reuse patterns	1
Modeling n-ary relations	2	Performance	2	Pattern selection	1
Ambiguous requirements /unclear task	1	Understanding patterns	1		
		Matching problem-patterns	1		
		Tool	1		
		Errors in patterns	1		

After introducing the patterns for the first time there were quite a few problems stated. The most frequent were how to select among patterns (Table 5: Pattern selection) and how to reuse them (Table 5: How to reuse patterns), i.e., how to fit them to the task at hand. Two participants were concerned about the performance issues with respect to using the constructed ontologies (Table 5: Performance), although this was not a concrete problem of the modelling task. Mainly they were concerned that the situation pattern would blow up the number of instances and thereby hamper the performance of the constructed ontology, however they did not propose a better solution for addressing the n-ary relations problem, even for this specific task. Furthermore, one participant had problems to understand the patterns (Table 5: Understanding patterns), another problem was to match the requirements to the patterns (Table 5: Matching problem-pattern) in order to assess their relevance and suitability, and again one subject mentioned that the tool was not really intuitive to use (Table 5: Tool). Finally, one participant claimed that the patterns contained errors (Table 5: Errors in patterns), but from his reasoning we can conclude that he rather meant that the OWL language contains errors. In this case, he was opposing the way of modelling an n-ary relation using the situation pattern, possibly claiming that this should be solvable directly in the language.

For the second task, when again using the patterns there were not many accounts of problems. Nevertheless, it was again noted that it is hard to make the right choices when modelling, especially for inexperienced ontology engineers. Also the problems of how to select and how to use the patterns were again mentioned.

Next, a question tried to determine how the patterns were in fact used by the people, as shown in Figure 15. It turned out that most participants used the patterns as reusable building blocks, not only for inspiration for their own solution, 4 selected this response after the first pattern exercise and only one after the second task. The users could also distinguish between using the complete patterns or only parts of them in their response, and there was a slight preference towards using patterns only partly, especially after the second task when the subjects were more used to the patterns. Probably some patterns were usable as a whole while some were only possible to partly reuse.

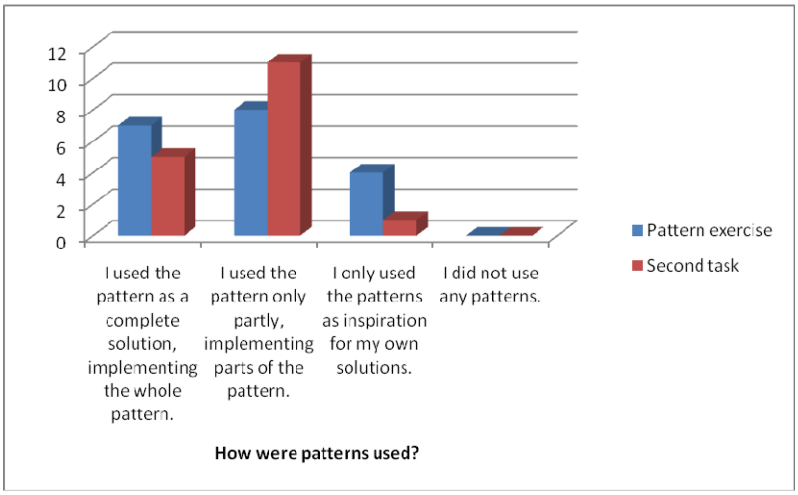

Figure 15. How patterns were used in the third session.

The patterns were not so easy to understand at first, one third of the people disagreed with the statement that patterns are easy to use after the first pattern exercise (Exercise 1), but then only 1 out of 14 still disagreed after the second task (Task 2). Similarly to the results from the session in Bologna, this clearly indicates that some training and experience is needed in order to understand the patterns properly, but that even a short training session will give considerable effect. A majority of the users found the tutorial and material given useful for understanding the patterns, although in this case there were a few more people disagreeing after the second task. No clear reason for this can be found, we can only speculate that the nature of the course planning had something to do with this, since the last session was scheduled quite far in time from the other sessions and thereby some of the continuity of the lectures and labs might have been lost.

Next, it was proposed that the patterns were not only easy to understand but also easy to use, where we can again see an improvement from the first pattern exercise to the second task. After the first pattern exercise not even half of the participants (7 out of 15) agreed to this statement, while after the second task 9 out of 14 subjects agreed. Nevertheless, just as in the Bologna session even for the second task two people still disagreed with the statement. As pointed out previously, the tool used has no special pattern support, neither for finding and reusing patterns, nor for pattern-specific operations, and this may be a contributing reason to the opinion that they are not very easy to use.

Within this group there was a tendency to agree that some patterns were obvious and trivial, more than in the other groups. At a first glance this contradicts the fact that this group was the most inexperienced, but on the other hand this may also be the reason for the opinion. Since an inexperienced person may also have trouble appreciating all the subtle underlying problems of a modelling issue. This may lead to the first impression that they already know the solution and do not need any assistance, when they in reality did not really understand all aspects of the problem. Nevertheless, this is only a hypothetical explanation, which cannot be proven based on this data.

For the next proposition, the other end of the scale that some pattern introduced really new ideas that the people had not thought of before, there was a more solid agreement that this was actually the case for some patterns. However, a couple of participants disagreed to this as well. Finally, did they find the patterns in some sense or another useful? The answer is definitely yes. Only one person disagreed after the second task, and after the second task as many as 64% agreed that the patterns are indeed useful (although not an agreement as strong as in the Bologna group).

The participants had the opportunity to motivate why they found the patterns useful, and a summary of their responses can be seen in Table 6. Increased quality of the constructed ontologies (Table 6: Quality), the possibility to reuse best practices (Table 6: Reuse), and to help decompose the problem (Table 6: Problem decomposition), are three of the main benefits that the participants observed. It was proposed by some that ontology patterns make the design process easier (Table 6: Easier design), and that an ontology based on patterns is more modular in its structure (Table 6: Modularization). The patterns also help to give insights into generic problems (Table 6: Insights into general problems) and to provide a more general solution (Table 6: More general solution), in terms of generalizability and reusability. Finally, one person noted that importing a pattern as a component limits the ‘routine work’ (Table 6: Limit ‘routine work’) of defining all concepts and properties ‘from scratch’.

Table 6. Motivations and user rationale for pattern usefulness.

How were the patterns most useful?

Pattern exercise	No. of answers	Second task	No. of answers
Quality	3	Easier design	2
Easier design	2	Reuse	2
Reuse	1	Problem decomposition	2
Modularization	1	Limit “routine work”	1
More general solution	1	Quality	1
Insights into general problems	1	Insights into general problems	1

Also for this group we asked the participants to compare the first task and the initial pattern exercise. Since this really involved just re-doing an exercise they had already done the day before, it was a good opportunity to record their opinions on how they thought that the patterns could be of assistance. There is a clear tendency towards agreeing that pattern usage made the participants solve the tasks better, 9 out of 14 agree, and only one disagrees to some extent. Also regarding the ease of solving the tasks a majority of the people agree that patterns makes it easier, but there is also quite a large fraction of uncertain individuals. With respect to the time aspect there is even less support and more unsure users, 5 out of 15 are unsure. Although, 6 out of 15 do agree with the proposition.

The coverage of the 9 ontologies resulting from the first task were quite high, between 71% and 92%, average 80%, which was on the same average level as the other group in Jönköping and slightly lower than the group in Bologna. Considering that these participants had two hours to construct the ontologies their coverage level could be expected to be comparable to the one of the Bologna group, but with their more limited experience and background in modelling this result is quite reasonable. For the second task the coverage was also reasonably high, between 64% and 82% for all 9 ontologies, average 72%.

With respect to usability aspects of the ontologies, as defined previously, the naming of concepts was similar to the other two sessions. Unfortunately this did not improve notably to the second task. A summary of the other usability features of the ontologies of the first task can be seen in Figure 16, and the ones after the second task in Figure 17.

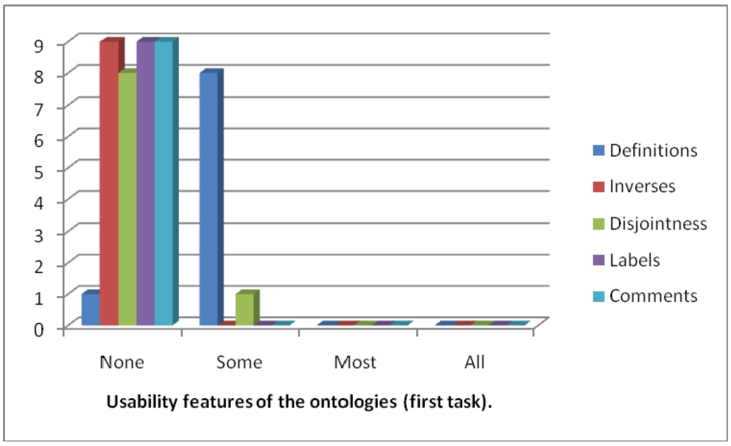

Figure 16. Usability features, results from the first task.

When trying to find common problems in the solutions for the first task four main problems were identified, as illustrated in Figure 18, overleaf. These are the same three problems as were already identified and described for the Bologna session and the first session in Jönköping. Additionally, there were three ontologies displaying issues related to mixing the A-box and T-box levels, thereby these three ontologies were not even in OWL-DL. All 9 ontologies display the songs and tracks mistake, 8 out of 9 the persons and roles mistake, and 7 out of 9 ontologies shows the problem of defining a ternary relation for album reviews.

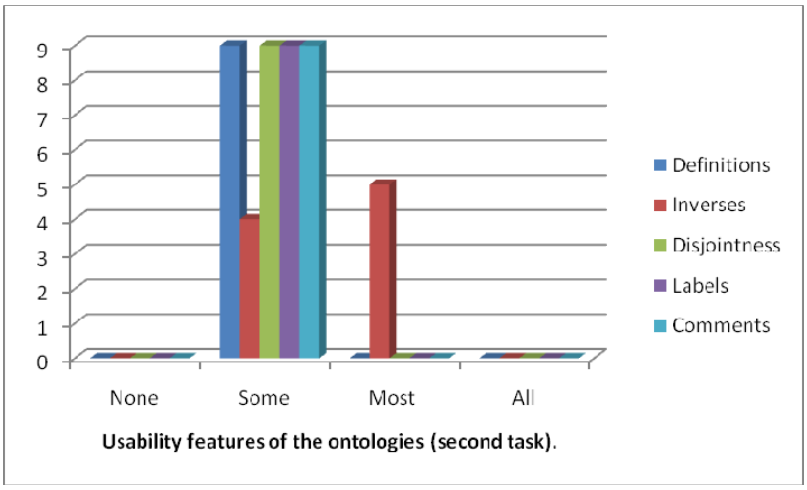

Figure 17. Usability features, results from the second task.

Three of the ontologies are not in OWL-DL, due to the fact that the A-box and T-box levels are mixed, e.g., classes are used as property values or properties are used as property values. The places where this is used is most often places related to some of the other common mistakes, such as trying to solve the ternary relation of album reviews through instantiating a datatype property on a binary relation between reviewer and the album. Finally, one ontology also showed some issues with respect to confusion of *part-of* with the subclass relation, defining a band member as a subclass of the band. This cannot be stated as a *common* problem since only one ontology displayed it, but it is still severe enough to be mentioned.

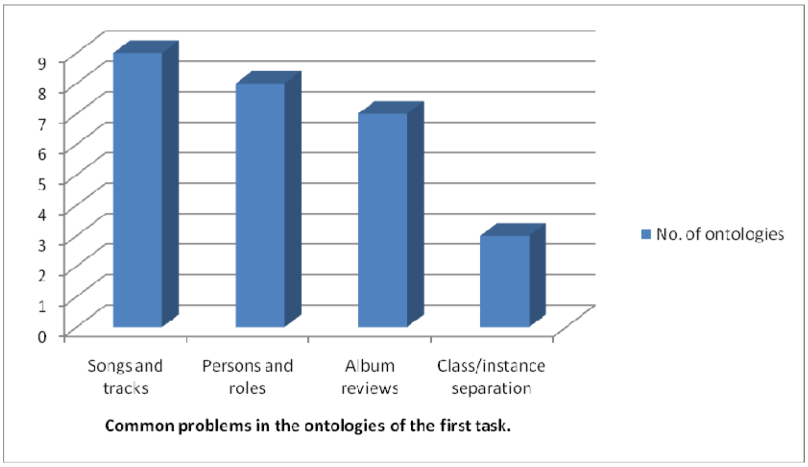

Figure 18. Number of ontologies displaying common problems (first task).

In the set of ontologies representing the solutions to the second task all ontologies succeed in correctly representing the issue of roles held by different people. It seems that the n-ary relation problem still was not grasped by all participants however, because 5 ontologies still displayed problems expressing n-ary relations. Especially to connect a situation to the time period for that situation seems to constitute a problem. Collectives were usually represented correctly, but in none of the ontologies was the sub-collective of male nurses defined, just as for the ontologies from the Bologna session and the previous Jönköping session. All the ontologies for the second task are in OWL-DL, none use, for example, any classes as property values as was the case for the first task. However, there is again one ontology displaying the problem of confusing part-of (or a similar relation) with subclass, stating that ‘city’ is a subclass of ‘country’.

The first three common problems are, as noted previously, closely related to patterns available in the catalogue. The information realization pattern addresses the songs and tracks issue, the agent role pattern addresses the persons and roles issue, and finally the situation pattern addresses the n-ary relation problem. There is no specific pattern to prevent from mixing the A-box and T-box levels, but since these issues occur mainly when trying to solve the n-ary relation problem, possibly the situation pattern could also have prevented these problems. It may be noted that out of the 2 ontologies where the album review issue was correctly modelled both of them unintentionally used exactly the same type of solution as suggested by the situation pattern. In the single ontology correctly solving the persons and roles issue a very similar approach as to the one suggeted by the agent role pattern was applied.

In the second task where patterns were available, the patterns used are listed in Figure 19. All ontologies use the agent role and situation patterns. The time interval and collection patterns are used by 7 out of 9 ontologies, while only 3 used the classification pattern. Two of the ontologies also use the time-indexed part-of and one uses the constituency pattern. Time-indexed part-of is in this case used in combination with the situation pattern and constituency is used instead of collections. Still, 5 out of 9 ontologies did not correctly model all n-ary relations present, even though all of those actually attempted to use the situation pattern. Similarly as for the other two sessions the only other common problem remaining was to correctly define sub-groups of the collectives, i.e., the group of male nurses.

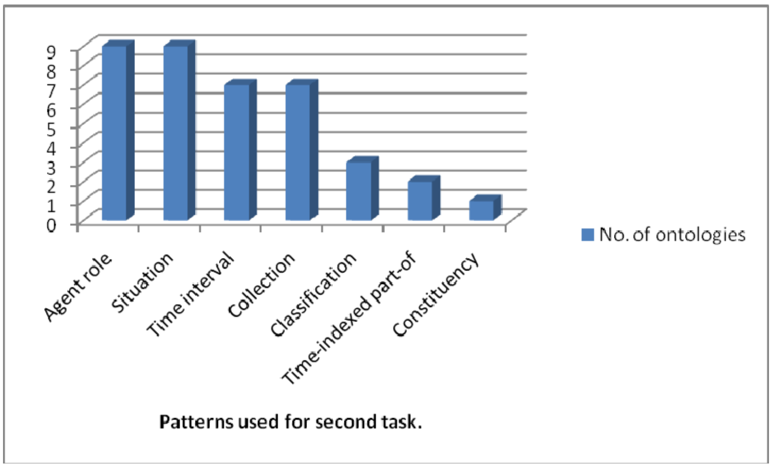

Figure 19. Number of ontologies using specific patterns used in the second task (third session).

Summary of results for the third session

In summary, the participants in the master's course in Jönköping were less experienced than the participants in both other sessions, and also perhaps less motivated to answer questionnaires as well as not being equally used to learning new techniques and using new tools. Most of them had some experience in constructing small example ontologies, due to a previous course in information modelling, but most of them had then only worked with non-DL formalisms.

A majority of the group perceived the patterns as useful, but many of the users also had problems to understand and use them. Some also questioned the relevance of the patterns and the solutions they provided, instead proposing to change the language in itself (for example, regarding n-ary relations). A majority of the participants agreed that they constructed 'better' ontologies when using the patterns, and that possibly the construction was slightly easier when using the patterns. On the other hand, many disagreed with the fact that they were faster when using the patterns.

All of the common problems/mistakes present in the first set of ontologies, solvable using patterns, were attempted to be solved by using the patterns the second time, but in some cases the solutions were still not correct. Also, the problem of mixing A-box and T-box levels could possibly have been solved by using patterns, since it occurred when trying to solve the n-ary relation problem. Only one new common problem was introduced in the set of solutions of the second task, as in the other two sessions, and this may be due to the formulation of the task rather than the difficulty of the modelling issue. The greatest improvement of the ontologies could be found in the usability aspect, similarly to the results from the two other sessions.

The content patterns were used mostly as complete building blocks, imported and then partly reused within the ontology. Nevertheless, the patterns were sometimes hard to understand, it was hard to find and select the right one, to reuse it in a correct way and to compose several patterns. This group also had more problems with using the tool than the previous groups. This again indicates that further tool support for ontology design pattern reuse is definitely needed.

2.4 Further analysis and discussion

Below the results from the different sessions are compared and the study is summarised in terms of answers to the set of questions that were posed at the beginning of this chapter. Some strengths and weaknesses of the study are noted, and possible future extensions are outlined.

2.4.1 Comparative analysis

Although the three groups were slightly different in their composition, and the settings were also in some respects different, the results achieved are remarkably similar. The most significant difference is the problems experienced by the participants of the dedicated experiment session in Jönköping, where the lack of training and lack of time to solve the tasks considerably affected the results and opinions. In all other respects the three sessions agree with respect to the general questions posed as research questions guiding this study.

In order to draw some general conclusions from the study, all the three settings have been analysed together when trying to answer the questions stated at the beginning of the description of this study. We remind the reader of the questions:

1. Are the content patterns perceived as useful by the participants?
2. Are the ontologies constructed using content patterns 'better' in some modelling quality sense, than the ontologies constructed without patterns?
3. Are the tasks given to the participants solved faster when using patterns?
4. How do participants use the patterns provided, and what support for pattern selection and usage would be beneficial?
5. What common problems in modelling solutions can be identified, both when not using patterns and when using the available content patterns?

The first question intends to determine if the patterns are perceived as useful and therefore worthwhile in the first place. To answer this we may note that in all three settings a majority of the participants answered that the patterns were in fact useful, 86%, 50% and 64% of the participants in the respective setting, and that only very few stated that the patterns were not useful, 0%, 25% and 7%, respectively. The least support is found in the second setting where there was no pattern training provided before exposing the participants to the task. From this we draw the conclusion that content ontology design patterns are in fact perceived as useful for modelling ontologies, but only with proper training and given time to get used to and understand the patterns. However, it is not an extensive training that is needed; the improvement can be seen already after one additional training exercise of 2 hours, as applied in two out of three settings of this experiment.

In more detail, the patterns are mainly perceived as helping with the modelling, by guiding the development, and by improving the quality of the solutions. The increased quality of the solutions, ontology quality as defined by the measures in section 2.2, is also supported by the analysis of the produced ontologies, and this provides the answer to the second question. The greatest quality improvement can be seen with respect to the usability of the ontologies, i.e. the understandability and clarity of the ontologies, how well they follow good practices, such as to define concepts where applicable, to add disjointness axioms, and to comment and provide inverses for relations. In this case ontologies from all three settings showed considerable improvements for the second task when patterns were applied. Also with respect to common problems, such as general and well-known modelling problems/mistakes, the quality of the ontologies improved in all three settings. However, some patterns were occasionally not used correctly, which indicates that better support for understanding and using patterns is needed. Even despite those problems the error rate, number of common problems, was much lower in the ontologies constructed using patterns.

The third question proposed that patterns might speed up the ontology development process and let ontology engineers solve the modelling tasks faster. From this study we cannot find any evidence that this is the case. If anything, the people perceived that they were slower when using the patterns, at least in the beginning when they were unfamiliar with the patterns. Nevertheless, there are also comments stating that some individual participants perceived to be faster and that the patterns helped to reduce the amount of routine work to be done by the designer. This could indicate that some tasks are made easier by the patterns, and that by reusing the patterns as imported components, some simple tasks are taken care of by the pattern, e.g., specifying all obviously needed properties, so that the designers could instead focus on the more complex parts of the problem. Still, there was no indication that this gained time balanced out the time spent on understanding, selecting and reusing the patterns, thereby there was still no considerable reduction of overall design time.

With respect to the fourth question, what pattern support is needed, most people state problems with finding possible patterns in the catalogue, matching them to the requirements of the current task, selecting the most suitable pattern, and reusing, e.g., specialising and composing patterns. Current tools such as TopBraid Composer or the NeOn Toolkit in its present release, do not give any support for pattern reuse, but in order to really benefit from the use of patterns such support is definitely needed. Tool support for searching and browsing a pattern catalogue, automatically matching requirements to competency questions of pattern, comparing and evaluating possible choices and their consequences, as well as performing operations on actual patterns when reused, are important requirements for a tool that aims to support pattern-based ontology design. Fortunately, in NeOn these are exactly the use-cases already proposed for the eXtreme Design (XD) plug-in that is being developed for the NeOn toolkit. Some part of this support also needs further research, such as automatic matching between requirements and pattern competency questions, but basic pattern matching has already been proposed in tools like OntoCase (see D3.8.1 [23]) and could quite easily be incorporated in the future versions of the XD plug-in of the NeOn Toolkit.

With respect to common problems/mistakes, as addressed by the final question, some common problems were discovered and they were mostly the same over all three settings. Of course this is not a complete list of common problems, but it shows common problems of the specific modelling tasks given to the participants. On the other hand, the common problems found are quite generic and not domain or task specific in themselves, whereby we can state that these constitute a list of general common problems, although an incomplete list. Since the modelling tasks were engineered in order to be partly coverable by the existing content patterns, in order to be able to compare the tasks and see that the patterns actually were used to solve them, we did not find many additional 'common problems' that could indicate the need for additional patterns. Nevertheless, if given other tasks probably more such problems could be identified, that could also constitute candidates for developing new patterns.

What could be noted was that in some cases the participants requested ready-made compositions of existing patterns, such as combinations of situations with time intervals and roles in order to represent the points in time when a certain person held a certain role in a certain setting. Such specialisations of patterns could be useful to add to the catalogue, if they are not too complex to understand. This is something that needs to be investigated further, how to achieve a good trade-off between specificity and complexity on one hand, and on the other hand reusability and understandability. The only common problem/mistake discovered in the second task was the issue of representing collectives and restrictions on the membership of a collective, such as the male nurses collective. A pattern was present, the collection pattern, but this pattern is very simple and general. Perhaps this pattern needs to be described better, or specialised into several variants in order to be easier to reuse, and provide better results when reused. One issue that is raised here is the question of how to provide support for complex definitions/restrictions by the means of content patterns. It needs to be further investigated how content patterns could support this better, or if this is a task more suited for generic logical patterns.

2.4.2 Identified strengths and weaknesses

Some limitations of this study can be noted, based on both the settings and the results collected. The settings were in general small, in total only 45 people contributed to this study, and in each session there were less than 20 participants. This is a reasonable number, and we are able to draw general conclusions from this study, but in order to be able to do a more quantitative analysis a larger group would be needed. Quantitative studies could answer questions such as "How much time is saved?" or "How much better are the ontologies?", rather than qualitative questions as used for this study, such as "In what respects are patterns perceived as useful?". The focus of this study was to collect the opinions of the users and verify them through analysing the ontologies they produced, while additional quantitative studies may focus even more on measurable results rather than opinions.

The background of the users is also a potential limitation of this study, since most participants were in fact at a similar level of experience. Most had some basic knowledge of ontologies, but were novice users in terms of ontology design. No participant was completely unaware of the concept of ontology; neither were there any experienced ontology developers skilled in OWL and experienced using the tools. The results presented here can therefore only be generalised to hold for this specific type of users, which is on the other hand quite a representative group typical for people designing ontologies, especially on the Semantic Web. Patterns can be seen as a means of teaching 'best practices', hence the group of novice users are in our opinion one of the most important. Nevertheless, additional experiments should be conducted also with experienced ontology developers, as well as when teaching truly novice users the basics of ontologies.

Ideally the experiment should also have studied the results of the exact same task in the exact same setting both with and without patterns. This was not possible due to practical limitations, such as the availability of participants. Courses had to be used as experiment settings in two of the sessions and then it was not possible to divide the group in two, one that did not get any pattern training and the other one getting this possible advantage. Instead the study was conducted through first recording the prior knowledge and level of modelling ability of the users, and then introducing the patterns and studying the new results. In such a setting it can never be ruled out that any improvement actually is caused by the additional experience gained by the subjects during the course of the experiment, rather than the introduction of patterns. Nevertheless, through studying common problems present in the ontologies and issues perceived by the participants, we are able to through logical reasoning to show that the most likely explanation is actually that the patterns helped the users.

In the second session we would have been able to use the original setup, dividing the participants into two groups. However, this was the smallest group and we decided that it was more valuable to replicate the settings used in the other two sessions as closely as possible, rather than changing the setting for this group. Changing the setting would have meant that the results could not be directly compared to the other two sessions.

An additional remark is that not all relevant variables have been varied during this study. The tool was the same during all sessions, whereby it cannot be ruled out that some of the common problems and perceived difficulties are actually due to the tool rather than the modelling problem, the OWL language, or the patterns. Similarly the modelling tasks were the same in all sessions, thereby it cannot be ruled out that some of the problems or opinions depend on the presentation of the tasks rather than the underlying modelling problems. However, these limitations do not affect the analysis and comparison between modelling with and without patterns, but it reduces the external validity, the generalisability, of results such as the lists of common problems and lists of specific perceived problems by the participants.

Strengths of this study are on the other hand that the results from the three settings are very similar, even though the settings were slightly varied. This in contrast to the above limitations increases the generalisability of the results. Also the amount of open ended, free-text, questions posed in the questionnaires is a strength, since the users were thereby free to express additional opinions and were not restricted by only predefined answer alternatives.

2.4.3 Perspectives on further work

In addition to the issues noted above, with respect to the limitations of this study, there are additional aspects that could be studied. In summary we propose to study the following aspects more in detail in future experiments:

- The effects of more varying levels of experience, e.g., expert users, on the use and usefulness of content patterns.
- The effects of different tool support on the use and usefulness of content patterns.
- The effects of different development methods on the use and usefulness of content patterns.
- The effects of pattern presentation on the use and usefulness of content patterns.
- The effects of task presentation, content and size on the use and usefulness of content patterns.
- The use and usefulness of other types of patterns, in addition to content patterns.

When the XD plug-in for pattern-based design within the NeOn toolkit is available this is proposed as the platform for experimenting further with pattern-based design. Many of the issues identified in this study will be supported through the plug-in and it will constitute an interesting study to repeat a similar experiment with this support present.

3. Supporting ontology (requirement) specification

One of the critical activities when developing ontologies is to identify their functional and content requirements. Such an activity is the ontology (requirement) specification, whose goal is to state why the ontology is being built, who its intended uses are, who the end-users are, and which requirements the ontology should fulfil. The precise identification of the knowledge that the ontology should contain and the writing of the ontology requirement specification document (ORSD), containing the identified ontology requirements, should be performed in the ontology (requirement) specification activity.

The main purpose of an ORSD is to serve as an agreement among ontology engineers and users and domain experts on what the ontology should represent. The ORSD will be decisive during the ontology development process because it will facilitate, among other activities, the search and reuse of existing knowledge-aware resources for reengineering them into ontologies, the reuse of existing ontologies, ontology modules or ontology statements (e.g., using Watson[3] or Swoogle[4]) and the ontology verification during the whole ontology development.

3.1 Overview and objectives

Methodological guidelines to carry out the ontology (requirement) specification activity have been included in the NeOn methodology for building ontology networks. In this deliverable, we describe an experiment, which we performed to learn about the understandability and usability of the proposed methodological guidelines for carrying out the ontology (requirement) specification activity.

The main goal of the experiment is to test the benefits of using the methodological guidelines for obtaining the ontology requirement specification document (ORSD) as output of the ontology (requirement) specification activity. Such methodological guidelines were proposed in the context of the NeOn methodology for building ontology networks [6].

Software developers and ontology practitioners involved in developing ontologies will obtain a benefit of this experiment that will serve people involved in the creation of the guidelines to improve, if necessary, the proposed methodological guidelines and to validate them.

3.2 Assumptions and user study setup

In this experiment we propose a questionnaire about the methodological guidelines for the ontology (requirement) specification activity, to be answered by people carrying out the experiment.

People carrying out the experiment have different experience levels and background in databases, software engineering, etc., but no extensive experience in ontology engineering.

3 http://watson.kmi.open.ac.uk

4 http://swoogle.umbc.edu

3.3 Findings and observations

This experiment has been divided into two phases, each of which is reported here as a separate study (mainly for the purpose of clarity and to enable comparison):

1. Experiment carried out with preliminary methodological guidelines on ontology (requirements) specification is described in Section 3.3.1.
2. Experiment carried out with final methodological guidelines on ontology (requirements) specification included in deliverable D5.4.1 [15] is described in Section 3.3.2.

3.3.1 User study #1: Preliminary Guidelines

In this first study, we carried out the experiment 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, but not in ontology engineering.

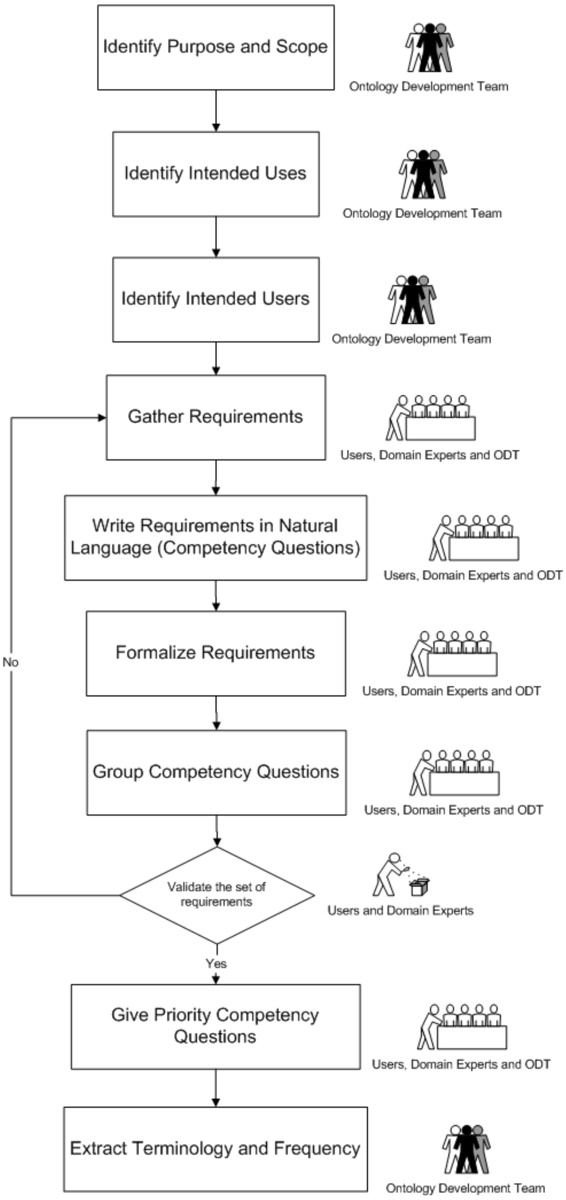

Figure 20. Workflow corresponding to the preliminary methodological guidelines for ontology (requirement) specification

The experiment was carried out during November 2007 by 14 master students working in groups of two attending the “Artificial Intelligence” master course at Facultad de Informática.

As mentioned before, this study was performed with preliminary methodological guidelines for the ontology (requirement) specification activity. Figure 20 shows the workflow corresponding to such guidelines.

The experiment consisted of the following parts:

1. Lectures provided to students the proposed preliminary methodological guidelines for the ontology (requirement) specification activity.
2. Student groups followed the methodological guidelines proposed to carry out the ontology (requirements) specification activity.

   Students were divided into two sets and each set followed a different instantiation of the methodological guidelines (Guidelines-1 and Guidelines-2 included in Appendix 1).

   Students had two weeks for carrying out the experiment using the preliminary methodological guidelines.
3. Students documented in detail each task performed during the ontology (requirement) specification activity, using the preliminary methodological guidelines.
4. Students filled in a questionnaire about the preliminary methodological guidelines. Such a questionnaire is included in Appendix 1.

3.3.2 User study #2: Revised guidelines

In this second study, we carried out the experiment 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, but not in ontology engineering.

The experiment was carried out during November 2008 by 12 master students working in groups of one or two attending the “Artificial Intelligence” master course at Facultad de Informática.

This study was performed with final methodological guidelines for ontology (requirement) specification included in D5.4.1 [15]. Figure 21 shows the workflow corresponding to such methodological guidelines. Such guidelines were produced from the preliminary ones and taking into account the general comments obtained during the user study #1.

The experiment consisted of the following parts:

1. Lectures provided to students the methodological guidelines for the ontology (requirement) specification activity included in D5.4.1.
2. Student groups followed the methodological guidelines to carry out the ontology (requirements) specification activity.

   Students had two weeks for carrying out the experiment using the provided material.
3. Students documented in detail each task proposed in the methodological guidelines and performed during the ontology (requirement) specification activity.
4. Students filled in a questionnaire about the methodological guidelines for the ontology (requirement) specification activity included in D5.4.1. Such a questionnaire is included in Appendix 3.

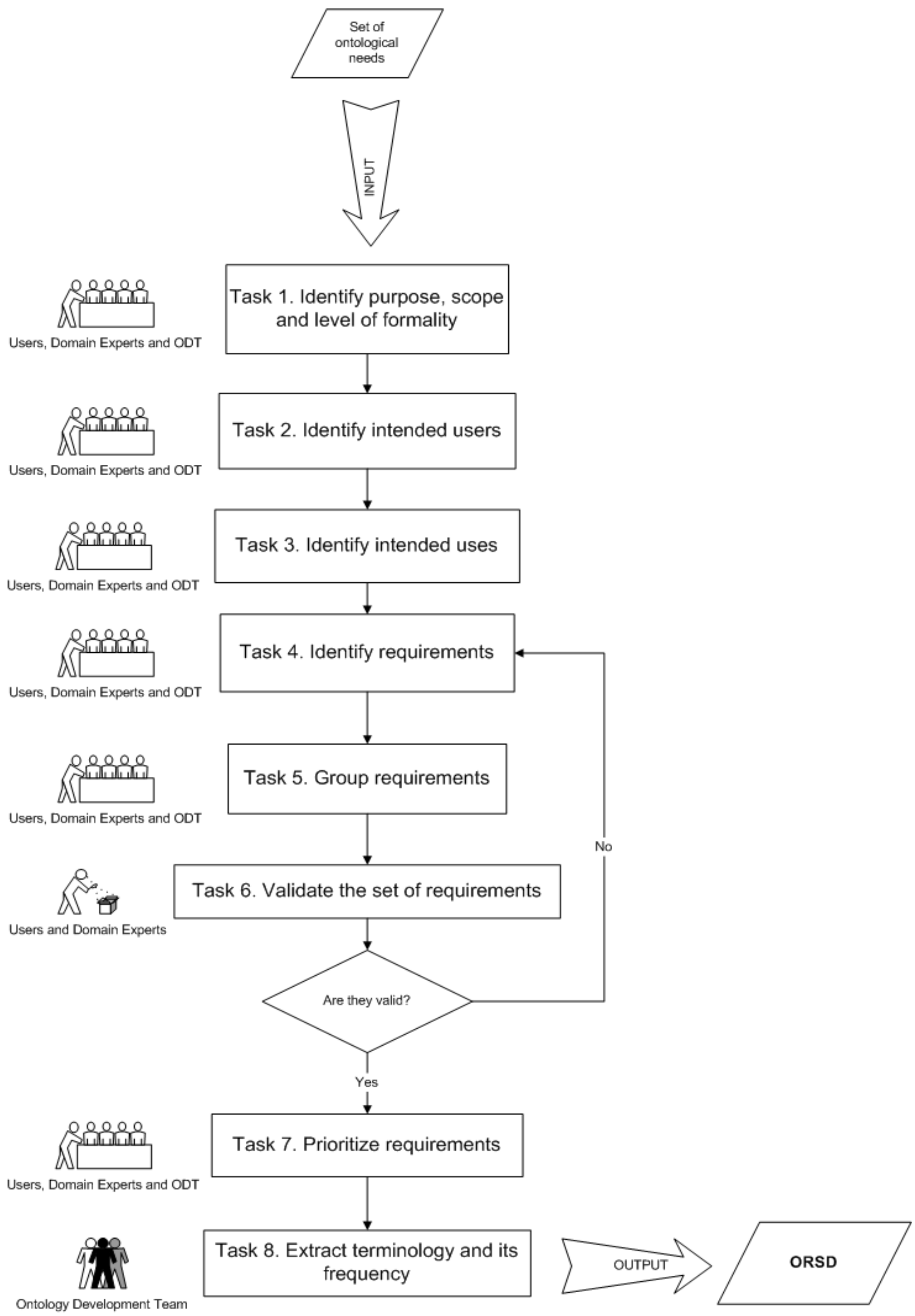

Figure 21. Workflow corresponding to the methodological guidelines for ontology (requirement) specification included in D5.4.1

3.4 Further analysis and discussion

In this section we include the analysis of the two user studies carried out with the methodological guidelines for the ontology (requirement) specification activity.

3.4.1 Analysis of user study #1

As already mentioned, in user study 1 the students used preliminary methodological guidelines for the ontology (requirement) specification activity. In this section, we analyze the questionnaire filled in by the 7 groups of master students (composed of 2 people) that carried out the experiment.

Regarding *general issues* we asked about the guidelines in the questionnaire (Appendix 2), these are the main conclusions*:*

- 86% of the students thought the preliminary methodological guidelines were well explained.

  General comments regarding this aspect were:

  - Some tasks are explained in a very brief way.
  - Some kind of pre-knowledge about the general process of building ontologies is needed in order to obtain benefit of the guidelines.

- 100% of the students commented that more detail was needed in the preliminary guidelines. Concretely, most of the comments were in the following line:

  - More detail on how to apply the techniques should be included in the guidelines, concretely on how to build the mind map diagram and on how to write the competency questions (CQs) (for further details on this see [15]).
  - Detailed explanations should be included in the tasks corresponding to the CQs development and validation.
  - More help about recommended tools should be also included in the guidelines.

- 72% of the students though that more techniques and especially more tools should be provided as recommendation in the methodological guidelines.
- 43% of the students did not miss an integrated tool to carry out the proposed tasks in the ontology (requirement) specification activity; however, 57% missed such an integrated tool.

  In the latter case, the general comment was that in large ontology projects, to have an integrated environment with all the tools (for writing CQs, for grouping CQs, for validating CQs, and for extracting terminology) would be very useful.

The following suggestions were provided by the students to improve the preliminary methodological guidelines:

- More detail on how to carry out some tasks, and more variety of techniques and tools to be used should be included.
- More free tools and examples should be included.
- More detail in the three first tasks of the ontology (requirement) specification activity should be included.
- The task of validating CQs should be divided into subtasks to facilitate the performance of such task, and more examples on how to validate CQs should be included.

Additionally, we obtained the following general comments given by students:

- One of the most difficult parts was to decide the group criteria in the task of grouping CQs and to validate the set of CQs.
- Mind map tools are very useful, especially to sort and classify ideas and the questions themselves.
- To group CQs is useful because it permits clearly to identify the essential parts to be cover by the ontology.
- Modulation approach is a good idea. It lets you find faster the questions, and increases the specification cohesion.
- To extract the terminology and its frequency is useful to know the terms that will form part of the core of the ontology. This provides the necessary input (terminology to be used) to the conceptualization activity and/or to the reuse process. However, tools for extracting automatically the terms were not useful in most of the cases.

3.4.2 Analysis of user study #2

As already mentioned, in user study 1 the students used preliminary methodological guidelines for the ontology (requirement) specification activity. In this section, we analyze the questionnaire filled in by the 7 groups of master students (composed of 1 or 2 people) that carried out the experiment.

The main conclusions obtained form the questionnaire filled by master students (Appendix 2) are:

- 100% of the students commented that the proposed methodological guidelines for the ontology (requirement) specification activity were well explained and very detailed.

  General comments regarding this aspect were:

  - Most of the students mentioned that the provided example clarifies the guidelines.
  - Due to the students having no experience in ontology engineering, they indicated that they would have had more problems during the ontology (requirement) specification activity if they had not had the proposed methodological guidelines.

- 86% of the students considered that the methodological guidelines needed more detail.

  General comments regarding this aspect were:

  - Concretely, they mentioned that more detail is needed in the task of writing CQs, and also more detailed explanation and/or more examples are needed in the task of grouping CQs.
  - Other students considered that the guidelines should explain with more details the task of giving priority to CQs. In this case, if several groups are involved in the ontology (requirement) specification activity a protocol to achieve consensus is likely to be needed and such a protocol should be proposed and explained in the guidelines.
  - Another comment was that the provided example is very useful, and that more examples should be provided in the guidelines.

  Additionally, 14% of the students considered that the level of detail provided in the guidelines is enough and that the complete example provided in the guidelines facilitates the performance of the activity.

- 72 % of the students considered that the methodological guidelines were not complete.

  General comments regarding this aspect were:

  - Some of the students expressed that task 1 in the guidelines was not well explained and that the meaning of “level of formality” was not clear enough.
  - Other point not well explained according to the students’ opinion was the task of obtaining the terminology. Students commented that it was not clear enough if the instances are objects and when a concrete term should be considered as an object.

- 57% of the students answered that no more techniques and/or tools were needed in the guidelines.

  On the contrary, 43% of the students considered that including more tools in the guidelines could be useful. They suggested the following tools: (1) a visual tool that provides the needed information about each task of the guidelines, and (2) some tools to calculate the terminology frequency.

- 57% of the students considered that an integrated tool for carrying out the proposed tasks was not needed.

  However, 43% of the students mentioned that an integrated tool that guides the user through the different tasks to be carried out and provides the information need in each task based on the methodological guidelines should be very useful. Additionally, students commented that such a tool should integrate a mind map editor and a tool to count the terminology.

- 86 % of the students considered that the guidelines for the ontology (requirement) specification activity were useful.
- 100% of the students expressed their intention of using again the methodological guidelines for the ontology (requirement) specification activity. But, some of the students commented they would prefer to have such guidelines integrated in an ontology (requirement) specification tool.
- 100% of the students commented that they found it useful to write the ontology requirement specification document before going into the ontology development.

  Concretely, most of the comments were in the following line:

  - Similar to any software development, a previous specification of what is required in the final product is necessary. Otherwise the final product (in this case the ontology) could not fulfil the expectations.
  - CQs are a good way to think about the problem before going into the development.
  - Ontology requirements allow having an ontology verification element, as in any software development.
  - An ontology (requirement) specification is needed during the ontology development in order to focus in the knowledge to be covered by the ontology.
- 86% of the students considered that they will create ontology (requirement) specifications in the future when the ontology is to be used in a real application.

Additionally to the previous comments, the following suggestions were provided by the students to improve the methodological guidelines:

- It would be useful to include in the guidelines standard CQs applicable to any domain.
- To translate the guidelines to other languages would be also valuable.
- To include more examples and complete ontology requirement specification documents (ORSDs) from different ontology projects would be very useful as guidelines.
- To have a graphic tool that creates the final ORSD would be worthwhile.
- To have a reference card with the summary of the information provided by the guidelines (task input, task output, actors, etc.) would be very interesting and useful.

3.4.3 Identified strengths and weaknesses

After analysing the results obtained in the two user studies performed with the guidelines for the ontology requirement specification activity, we can mention as strengths the following ones:

- The final methodological guidelines for the ontology requirement specification activity used in the user study 2 are well explained according to 100% of the students. This means that the changes we performed in the preliminary guidelines (used in the user study 1) to obtain the final guidelines allowed us to improve the guidelines.
- All the students performed the user study 2 agreed that they would use again the methodological guidelines for the ontology requirement specification activity. However, in some cases, students mentioned that they would prefer to use such methodological guidelines included in a tool for performing the ontology (requirement) specification activity.

With respect to the weaknesses, we can comment the following:

- The three first tasks in the methodological guidelines should be better explained.
- More detail should be included in the tasks about writing and validating CQs and giving priority to CQs.
- The task about extracting terminology should be clarified.

It is worth to mention that the percentage of students that missed an integrated tool for performing the ontology requirement specification activity and those that did not miss such a tool was similar in both user studies.

3.4.4 Perspective further work

Based on the analysis carried out with the data extracted from the questionnaires, we are currently working on:

- Including more detail in the three first tasks of the methodological guidelines.
- Including more detail in the tasks about writing and validating CQs and giving priority to CQs.
- Clarifying the task about extracting terminology.

Additionally, we are studying the possibility of:

- Including more examples and complete ontology requirement specification documents (ORSDs) from different ontology projects in the methodological guidelines.
- Creating a reference card with the summary of the information provided by the guidelines (task input, task output, actors, etc.).
- Implementing an integrated tool for performing the ontology requirement specification activity.

4. Supporting ontology localization

One of the activities identified in the ontology network development process is the Ontology Localization Activity that supports the building of multilingual ontologies. The Neon Glossary of activities [12] defines ontology localization as the activity that consists in adapting an ontology to a particular language and culture. The main difference between ontology localization and a simple translation of ontology labels differs is that translation is only one of the activities in localization; in addition to translation, ontology localization includes many other tasks such as localization management, testing and ontology update. The NeOn deliverable D5.4.2 [16] describes in detail the methodological guidelines for supporting this activity. Our concern here is to describe a set of experiments that can show the effects and benefits of using ontology localization guidelines. There are several different aspects of guidelines that need to be studied and several types of effects of guidelines usage that need to be defined and measured. So far no indisputable evidence has been put forward to support the benefits of using ontology localization to build a multilingual ontology. Only in the software engineering field, where the localization is used to adapt a software product to a specific region or language, can we find some evidence of the benefits of localization.

The localization activity (when discussed generally, and specifically for ontology engineering) is commonly suggested to give two kinds of benefits: to guarantee high productivity and outstanding quality. High productivity is here concerned with reducing the human effort to localize an ontology manually. Outstanding quality is here concerned with the quality of the obtained translations. The experiment described below intends to address all of these issues.

4.1 Overview and objectives

Methodology experts participating in the NeOn project have identified Ontology Localization as one of the crucial activities in the ontology network development process to support the construction of multilingual ontologies. In this deliverable, we propose an experiment to learn about the understandability and usability of the methodological guidelines for carrying out the ontology localization activity.

The main goal of the experiment is to test the benefits of using the proposed methodological guidelines and additional material included in D5.4.2 [16] for obtaining a multilingual ontology as output of the ontology localization activity.

4.2 Assumptions and user study setup

In this experiment we worked with a questionnaire about the methodological guidelines for the ontology localization activity, to be answered by people carrying out the experiment.

People carrying out the experiment have different experience levels and background in databases, software engineering, etc, but no extensive experience in ontology engineering.

4.3 Findings and observations

The experiment was carried out in the “Artificial Intelligence (AI)” master course at the Facultad de Informática (Universidad Politécnica de Madrid) with master students, having background in databases, software engineering, and artificial intelligence, but no extensive practical experience in ontology engineering. We proposed a questionnaire about the use of methodological guidelines for ontology localization activity. Figure 22 shows the workflow corresponding to such guidelines. For interpreting the results, we analyzed the answered questionnaires and extracted some statistics.

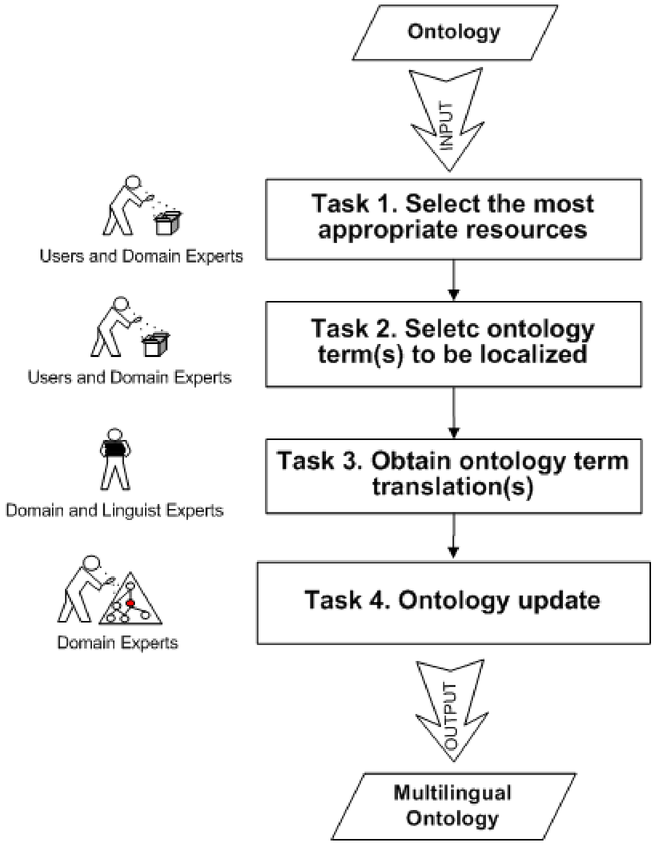

Figure 22. Workflow corresponding to the methodological guidelines for ontology localization included in D5.4.2

The questionnaire includes the following questions:

1. Are the proposed guidelines well explained?
2. Is more detail needed in the guidelines? If so, please explain in detail in which sense and in which tasks
3. Are this guidelines complete? If not, what is missing?
4. Do you think more techniques and tools should be provided?
5. How can we improve the proposed guidelines?
6. Did you find these localization guidelines useful?

The experiment is divided in the following phases:

1. Lecture will provide to students the proposed guidelines included in deliverable D5.4.2.
2. Student groups will analyze the methodological guidelines proposed to carry out the ontology localization activity.
3. Students will fill in a questionnaire about the proposed guidelines.

4.4 Further analysis and discussion

The experiment included 6 questions about localization guidelines solved by 15 students, and as a general conclusion we can say that students did not have problems with the use and understanding of each one of the tasks identified in the methodological guidelines. In the following, we provide some observations extracted from the analysis of the experiment results:

- 95% of the comments provided by the students to question 1 indicated that guidelines were well explained.
- For the comments obtained to question 2: "Is more detail needed in the guidelines?", we can say that 85% of the students consider that more detail is not necessary in the guidelines, however 15% think there is an opportunity to improve the explanations of i) how to select the most appropriate linguistic assets (step 1 in the guidelines), and ii) how to obtain the ontology term translations (step 3 in the guidelines).
- In question 3: "Are the guidelines complete?", 95% of the evaluators believe that the guidelines to perform the localization activity are complete. However 5% consider it necessary to enhance the guidelines to support the evaluation of the obtained translations.
- For the comments obtained to question 4: "Do you think more techniques and tools should be provided?", we can say that all evaluators believe that the techniques and tools to execute each activity of the guidelines are sufficient.
- The generalized comment to question 5 "How can we improve the proposed guidelines?" is to include more examples of how to use the proposed guidelines for the ontology localization activity and what results are expected.
- Finally, with respect to question 6 "Did you find these localization guidelines useful?", all students believed that the guidelines were useful, but also necessary.

One of the objectives of the research on ontology localization should be to improve the detail and completeness of the guidelines, as suggested by the responses to the questions two and three and for this reason some actions are proposed in this deliverable (see Section 4.4.2).

4.4.1 Identified strengths and weaknesses

Based on the comments obtained in experiment #1 we can say that the majority of students found that the methodological guidelines were useful and understandable. The main weaknesses included a more complete description of some tasks of the methodology. Some examples are:

- A more detailed description of the criteria to choose a technique to help in the localization activity.
- The lack of basic guidelines to select a localization tool depending on the type of ontology to be localized, or
- An exhaustive description of the different levels of difficulty that can be found in the translation of ontology labels.

4.4.2 Prospective further work

Based on the analysis carried out with the data extracted from the questionnaires, we are currently working on:

- Including more detail in the two first tasks of the methodological guidelines.
- Adding a new task to support the evaluation of the translations of each ontology term.

Additionally, we are studying the possibility of including more examples of test cases solved by ontology localization guidelines.

5. Supporting the establishment of ontology lifecycles

To build an ontology or ontology network, ontology developers should devise first a concrete plan for the ontology (network) development. That is, they should establish the ontology (network) life cycle. To do this, ontology developers should answer two key questions:

a) Which ontology lifecycle model is the most appropriate for their ontology project?

b) Which particular processes and activities should be carried out in their ontology lifecycle?

5.1 Overview and objectives

Methodological guidelines to carry out partially the scheduling activity have been included in the NeOn Methodology for building ontology networks, in particular in D5.3.2 [13]. In this deliverable we describe an experiment to learn about the understandability and usability of the proposed guidelines for helping software developers and ontology practitioners to decide (1) which ontology network life cycle model is the most appropriate for their ontology network and (2) which concrete processes and activities should be carried out in their ontology network life cycle; and at the end, to establish the concrete ontology network lifecycle.

The main goal of the experiment is to test the benefits of using the proposed methodological guidelines for obtaining the ontology network lifecycle, which is one of the outputs of the scheduling activity.

Software developers and ontology practitioners involved in developing ontologies will obtain a benefit of this experiment that will enable people involved in the creation of the guidelines to improve, if necessary, the proposed methodological guidelines and to validate them.

5.2 Assumptions and user study setup

In this experiment we propose a questionnaire about the methodological guidelines for establishing the ontology (network) lifecycle, answered by people carrying out the experiment.

People carrying out the experiment had different experience levels and a background in databases, software engineering, etc., but no extensive experience in ontology engineering.

5.3 Findings and observations

This experiment was carried out with the methodological guidelines for establishing the ontology (network) lifecycle included in deliverable D5.3.1 [12]. Figure 23 shows schematically how the establishment of the ontology (network) life cycle should be carried out according to the proposed methodological guidelines.

5.3.1 User study #1

In this first phase/study, we carried out the experiment within the “Artificial Intelligence (AI)” master course at Facultad de Informática (Universidad Politécnica de Madrid) with master students, having a background in databases, software engineering, and artificial intelligence, but not in ontology engineering. The experiment about the establishment of a particular ontology network lifecycle was carried out during November 2007 by 16 students, working in groups of two, attending the “Artificial Intelligence” master course at Facultad de Informática.

As mentioned before, this study was performed with the methodological guidelines for establishing the ontology (network) life cycle included in deliverable D5.3.1 [12]. The experiment is divided in the following parts:

1. Lectures provided to students the proposed methodological guidelines for establishing the ontology (network) lifecycle, explanations and additional documents (Collection of ontology network life cycle models, NeOn Glossary of Activities, and Table of Required-If Applicable Activities).
2. Student groups followed the methodological guidelines to establish the ontology network lifecycle for their ontology project.

   Students had two weeks for carrying out the experiment using the provided material (methodological guidelines and additional documentation).
3. Student groups documented in detail each task performed, using the methodological guidelines, during the establishment of the ontology (network) lifecycle.
4. Students filled in the questionnaire about the proposed methodological guidelines. A copy of this questionnaire is included in Appendix 4.

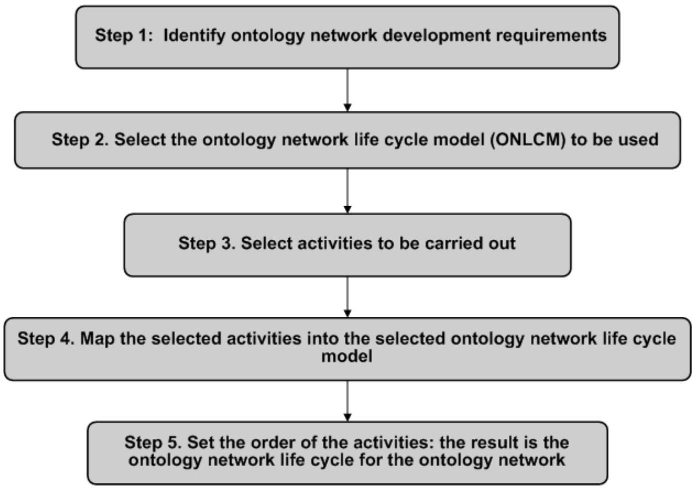

Figure 23. Steps to be carried out for the ontology (network) lifecycle establishment.

5.3.2 User study #2

In this second phase/study, we carried out the experiment 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, but not in ontology engineering. The experiment about the establishment of a particular ontology network lifecycle was carried out once during November 2008 by 8 students, working in groups of one or two, attending the "Artificial Intelligence" master course at Facultad de Informática.

As mentioned before, this phase/study was performed with the methodological guidelines for establishing the ontology (network) lifecycle included in deliverable D5.3.1 [12]. The experiment was divided in the following parts:

1. Lectures provided to students the proposed methodological guidelines for establishing the ontology (network) lifecycle, explanations and additional documents (Collection of ontology network lifecycle models, NeOn Glossary of Activities and Table of Required-If Applicable Activities).

2. Student groups followed the methodological guidelines to establish the ontology network lifecycle for their ontology project.

   Students had two weeks for carrying out the experiment using the provided material (methodological guidelines and additional documentation).

3. Student groups documented in detail each task performed, using the methodological guidelines, during the establishment of the ontology (network) lifecycle.

4. Students filled in the questionnaire about the proposed methodological guidelines. Such a questionnaire is included in Appendix 5.

5.4 Further analysis and discussion

In this section we include the analysis of the two user studies carried out with the methodological guidelines for the ontology (network) lifecycle establishment.

5.4.1 Analysis of user study #1

In this section we analyze the questionnaire filled in by the 8 groups of master students (each group composed of 2 people) that carried out the experiment. Regarding general issues we asked about the guidelines in the questionnaire (Appendix 4), these are the main conclusions:

- 100% of the students considered that the guidelines were well explained. Their general comments regarding this aspect were:
    - Guidelines were clear, structured, well- written, concise, and understandable.
    - Guidelines were in general well explained, but some parts were quite confusing.
- 62% of the students commented that the guidelines needed more detail. In particular, most of the comments were along the following lines:
    - Steps 2 and 3 in the guidelines were very well explained. However, steps 4 and 5 were not detailed enough.
    - More detail is needed in the NeOn Glossary, because some activities were not well explained and it was easy to mistake one for another.
    - Figures in the collection of models (concretely for incremental, iterative and evolutionary models) should be included.
    - In general, students commented that more examples should be included in the guidelines.
- 62% of the students considered that the guidelines were reasonably complete. However, 38% considered that something was missing. In the latter case, general comments were the following:
    - Some steps required more details.
    - It would be very useful to have a mapping between the activities and the phases in each model.
    - More real examples explaining why the different decisions were taken should be included in the guidelines.

With respect to the second step of the guidelines, the following general comments were provided:

- The decision tree to select the ontology network life cycle model (ONLCM) was very useful.

- If ontology developer knows about software engineering models, then students considered that the collection of ontology network lifecycle model was well explained.

In other cases, they mentioned that more details should be included. In particular, waterfall and its variants and spiral models were considered well explained; however the rest of the models in the collection are considered to be not detailed enough. Students suggested that it would be good to include a detailed description of the evolving prototyping model (at least), and graphical representation of all the models.

With respect to the third step, students provided us with the following comments:

- The set of natural language questions was especially useful, as the activities were not familiar. These questions were very intuitive to understand and it was easy to select each activity.
- Students considered that the set of natural language questions was very useful to guide naïve users in the activity selection.
- Students commented that the NeOn Glossary was comprehensive, and they found very useful the references to other activities in the activity definitions, concretely, useful to distinguish those activities that sounded very similar.
- Additionally, students suggested that it would be useful to present in the NeOn Glossary figures for sub-activities (like in the case of evaluation).

In the case of the fourth step, most of the students commented that this step was not very useful because it did not provide real guidelines to distribute the activities along the selected model. The mappings among activities and phases in each model were missing and they are crucial.

Regarding the fifth step, students asked in general for some kind of examples.

Finally, we obtained the following general comments and suggestions given by students:

- To improve the methodological guidelines, (a) step 4 and step 5 should be improved a lot, and (b) more explanations and examples should be included.
- To carry out the establishment of the ontology network life cycle in a faster way, students commented that experience in ontology engineering would be needed.

  Additionally, they suggested (a) having an automatic way to map activities with phases in the models, and (b) including the selection of scenarios in the guidelines.

5.4.2 Analysis of user study #2

In this section we analyze the questionnaire filled in by the 6 groups of master students (each group composed of up to 2 people) that carried out the experiment. Regarding *general comments* we asked about the guidelines in the questionnaire (Appendix 5), these are the main conclusions:

- 83% of the students considered that the guidelines were well explained. General comments regarding this aspect were:
    - Guidelines were explained with an adequate level of detail.
    - Guidelines were concrete, but more details could be included in some steps.
    - The order in the guidelines should be improved.
- 67% of the students commented that the guidelines needed more detail. Concretely, most of the remarks were in the following lines:
    - More details were required in the description of the ontology activities, because some of them are confusing.

  - More details and examples should be included in the ontology (network) lifecycle models.
  - Complete real examples should be included in the guidelines.

Regarding *step 2* of the guidelines, the following general comments were provided by the students:

- The decision tree to select the ontology (network) life cycle model was very useful.
- In general, the collection of models is clear and adequately explained. It was very useful to have the ontology models related to software engineering ones.
- The different phases of ontology (network) life cycle models should be explained in more detail.

  Waterfall models are adequately explained; however, there are no graphics, examples, and detailed explanations for the other models.
- Model descriptions are detailed enough; sometimes too much (common details in similar models should be not repeated)

With respect to *step 3*, students provided us with the following comments:

- The set of natural language questions in combination with the classification of activities in different groups (management, development, support) was useful to select activities.
- The set of natural language questions was a quick reference to select activities; however, sometimes it is confusing because some activities are very similar.
- The NeOn Glossary of Activities was in general well explained, but the number of activities was excessive.

More details are required in the following activities: ontology modularization, ontology forward engineering, ontology environment study, ontology elicitation, ontology population, ontology reverse engineering, ontology formalization, and ontology modification.

In the case of *step 4*, most of the students mentioned the following:

- The classification of activities into groups (management, development, support) should be used in this step to facilitate its execution.
- Mappings among activities and phases in each model were missing, and they should be included.
- Real examples should be included in the guidelines.

Regarding *step 5*, the following general comments were provided by the students:

- To establish priorities or restrictions among activities could be useful to order them.
- Real examples to build the Gantt diagram should be included

Finally, we obtained the following general comments and suggestions given by students:

- To improve the methodological guidelines, it should be useful:
  - To include in which phases we should include each activity.
  - To include complete real examples with explanations about each decision taken in each step.

  - To have a reduced version of the methodological guidelines for a quick overview of them, and to have another version of the guidelines more extended with all the details for consulting and for going in them into depth.
  - To translate the guidelines into other languages (e.g., Spanish).

- To carry out the establishment of the ontology network lifecycle in a faster way, students commented that experience in ontology engineering would be needed.

  Additionally, they suggested having an integrated automatic assistant, like an expert system, to guide them during the establishment of the ontology (network) lifecycle.

5.4.3 Identified strengths and weaknesses

After analysing the results obtained in the two user studies performed with the methodological guidelines for establishing the ontology (network) life cycle, we can mention as strengths the following:

- The methodological guidelines for establishing the ontology (network) lifecycle were clear and understandable according to most of the students.
- The decision tree for selecting the ontology (network) lifecycle model and the set of natural language question to help in the activity selection were considered by the students very useful.
- Regarding the collection of ontology (network) lifecycle models, students stated that it was very valuable to have references to similar models in software engineering.

With respect to the weaknesses, we can comment on the following:

- Steps 4 and 5 should be improved a lot, because they currently do not provide concrete guidelines to help ontology developers.
- More details and clarifications should be included in some activities of the NeOn Glossary.
- More detail and figures should be included in the collection of ontology (network) lifecycle models.
- Mappings among activities and phases in each ontology (network) lifecycle model should be provided because they are crucial in guidelines to help ontology developers to establish the life cycle.
- Real examples on how to use the proposed methodological guidelines should be included.

5.4.4 Prospective further work

Based on the analysis carried out with the data extracted from the questionnaires, we already modified and improved the methodological guidelines included in deliverable D5.3.1 [12], obtaining the methodological guidelines presented in deliverable D5.3.2 [13]. Such methodological guidelines proposed to establish the ontology (network) life cycle following the task shown in Figure 24, overleaf.

The methodological guidelines presented in deliverable D5.3.2 [13] already include the following changes:

- Revision and update of the collection of ontology (network) lifecycle models, including more detail and figures for all the models.
- Inclusion of natural language questions to select the ontology (network) lifecycle model.
- Inclusion of natural language questions, related to the nine identified scenarios in the NeOn methodology, to decide among the different versions of a particular model.

- Inclusion of natural language questions, also related to scenarios, to select which processes and activities should be included in the lifecycle.
- Incorporation of the mappings among processes and activities and phases in each ontology (network) lifecycle model. This part was included in task 3, which was step 4 in the guidelines used during the experiments.

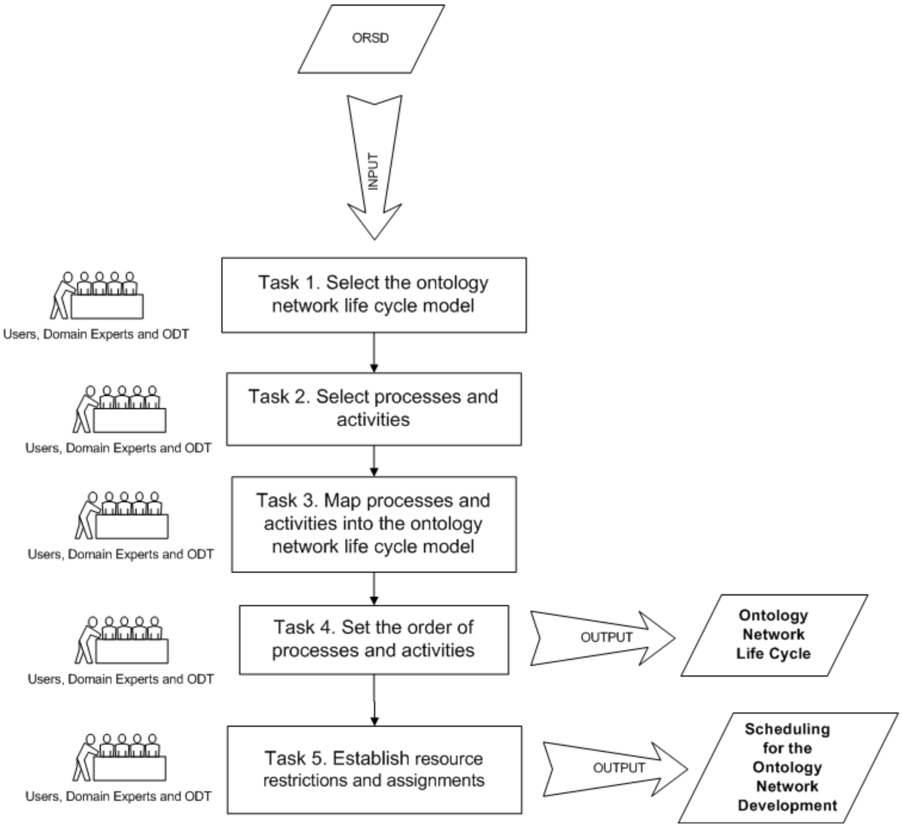

Figure 24. Tasks for (ontology development) scheduling.

Additionally, based on the analysis carried out, we are currently working on:

- Including more details and clarifications in some activities of the NeOn Glossary.
- Including real examples on how to use the proposed methodological guidelines.
- Improving step 5 (that is task 4 in the current guidelines shown in Figure 24).
- Implementing the NeOn plug-in gOntt [13], which guides ontology developers during the establishment of the ontology (network) lifecycle.

Finally, we are also studying the possibility of producing a reduced version of the methodological guidelines for a quick overview, to complement the set of methodological guidelines for establishing the ontology (network) lifecycle.

6. Testing The Collaboration Server

The main motivation of this experiment was to evaluate the models and strategies proposed for the management of ontology changes to support the development and maintenance of ontologies in a collaborative scenario.

6.1 Overview and objectives

In general, we studied on the one hand the conceptual models that provide the foundations to represent the required information and on the other hand the implementation support. The following attributes were studied:

- Conceptual Models
  - Change representation model:
    - The adequacy with respect to the users' requirements
  - Workflow model
    - The adequacy of the model with respect to the users' actions
- System Implementation
- The overall usability and performance of the system. According to the ISO standard 9241-11 [8], usability[5] refers to:
  - Effectiveness
  - Efficiency
  - User satisfaction

6.2 Assumptions and user study setup

The experiment was conducted at FAO headquarters during the last week of October 2008. We needed two days, one for the set-up of the collaborative infrastructure in FAO computers and one for running the experiment. Following the typical behaviour of the ontology editors in FAO and because of time constraints, during the first day, the tester configured the collaborative infrastructure as the follows:

- One server running:
  - NeOn Collaboration Server
  - Oyster (server mode)
- Three clients, each running:
  - NeOn Toolkit extended with:
    - Registry Plug-ins
    - Change Management Plug-in

5 The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use

- Collaboration Plug-ins

The ontology used for the experiment was species-v1.0-model.owl[6]. This ontology is the schema of one of the most important ontologies of FAO for the fishery domain. The tester uploaded this ontology to the server as part of the configuration.

On the second day, before running the actual experiment, the tester gave a brief introduction (30 min) of the system and the goal of the experiment to the FAO team composed of three ontology editors (Subject Expert A, Subject Expert B and Validator A). Each of the editors was in charge of maintaining the ontologies of the fishery domain and they had a different background profile:

- Subject Expert A had a great knowledge about the fishery domain but has never used the NeOn toolkit and in general he has a small knowledge of computer systems or modelling design issues.
- Subject Expert B had a fair knowledge about the fishery domain and the NeOn toolkit and some knowledge about modelling design issues.
- Validator A had also a fair knowledge about the fishery domain, the NeOn toolkit and also about modelling design issues.

All ontology editors were in the same room and each was provided with a detailed and personalized guide of the tasks he had to perform including the initialization of his NeOn Toolkit installation (i.e. each of them had to configure his client as in a real situation). In a nutshell[7], each subject expert (SE) had to perform 6 main tasks while the validator (V) had to perform 4 main tasks, as follows: every ontology editor was requested to configure and start the collaboration support within his NeOn toolkit (T1), then each subject expert was requested to make several changes to the ontology concurrently (SE's-T2), visualize the results of their changes and analyze the information provided by the system (SE's-T3) and submit their changes to be approved (SE's-T4). The chosen changes were 34 (17 changes for each SE) realistic modifications to the ontology including real information according to FAO experts.

Then the validator was requested to analyze the changes performed and to approve/reject them (V-T2). The subject experts were then requested to perform some additional actions according to the workflow to test the possible subject expert actions (e.g. delete a rejected change, modify an approved change, etc.) (SE's-T5 and T6). Finally the validator was requested also to perform some additional actions to test the possible validator actions (e.g. reject to be approved a change, delete an approved change, etc.) (V-T3 and T4).

During the experiment the tester was taking note of the behaviour of the editors, their questions and problems, and at the end of the experiment, each editor fulfilled an online survey consisting of 60 questions (50 of the standard SUMI [17] questionnaire[8] and 10 specific for the collaborative ontology development).

6.3 Findings and observations

As a general conclusion we can say that the results of the evaluation are very positive. The analysis of the results of the experiment that we conducted at FAO shows several good points of the infrastructure as well as some issues that could be improved as part of our future work. In particular, the results showed:

---

6 Available at http://www.fao.org/aims/neon.jsp

7 Complete sets of steps usually carried out by ontology editors involved in the experiment are provided in Appendices 6, 7, and 8.

8 http://sumi.ucc.ie/index.html

- Our models (change representation, workflow model) are adequate with respect to the ontology editors' needs. That is, representative changes and workflow operations from our use case could be captured and represented correctly by our models along with their required information.
- The overall system effectiveness was positive (90% or above) which demonstrates the good capability of our infrastructure to produce the overall goal i.e. collaborative ontology development.
- The efficiency of the system was in general satisfactory. A very positive point is that the time users required to complete their tasks was better than with their previous approach. Regarding the frequency of help use, it is understandable that users asked frequently for assistance, taking into account that they had only a brief introduction to the collaborative infrastructure (and the experiment) (30 min), in addition to the fact that they did not use the NeOn toolkit regularly. Finally, as most of the problems we found during the experiment were related to the NeOn Collaboration server, which is not part of this work, we feel satisfied with the results. Note that the problem related to the server crash was just a lack of memory of the process, which can be easily fixed.

Finally, the results of the survey to measure the user satisfaction showed that users were in general highly satisfied with the infrastructure and they agreed on its usefulness and correctness. For instance, the questions of the survey that evaluated the editors' satisfaction regarding the collaborative ontology development show that editors think our infrastructure is better than the previous approach, i.e., it is faster and they prefer it (see the overall result for the collaborative ontology editing survey in Figure 25). Moreover, ontology editors actually liked the main features of the system (e.g. the integrated view of the workflow, the management of changes in a collaborative environment, etc.) as we can see from the feedback received in the textual answers.

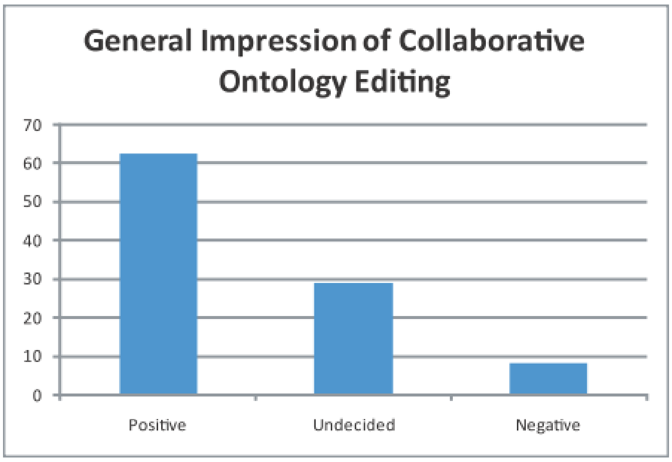

Figure 25. Impressions of subject experts and validators in collaborative editing

6.4 Further analysis and discussion

Nevertheless, the feedback received also shows some aspects that can be improved. In general those aspects are related to improvements in the user interfaces (which are very useful for our future work) such as being able to select multiple changes in one click, being able to sort the changes according to different criteria, or improving the readability of the change information. We also received a few comments regarding the NeOn collaboration server such as speed issues when performing some operations (e.g. adding individuals), or the problem when the server crashed. However, although we used the NeOn collaboration server in this scenario, it is not part of the work we are evaluating.

The overall results for each of the five SUMI dimensions have a similar pattern. In all cases, only around one fourth (25%) of the total answers were negative, another 25% (approximately) were undecided and at least half of the answers (50%) were always positive. So, in general we concluded that the results were fairly satisfying, especially if we take into account that users had only a brief introduction to the system before the experiment and none of them had much experience with the NeOn toolkit. The results show that in general users liked the infrastructure and that they find it self-explanatory. Furthermore, users didn't feel that the software does not help them in their work, or that they are being controlled by the software when carrying out a task or it is difficult to master the system (i.e. learn new features).

6.4.1 Identified strengths and weaknesses

During the experiment and as part of the analysis of the results, we learned important lessons from which we can get some recommendations. For instance, we found out that sometimes users were interested to see only specific changes (e.g., from specific users, from specific type, etc.), in specific order or grouped by some criteria, instead of having the complete history of changes in chronological order (as it is at this moment). Another interesting observation is that users wanted to have a quick view of the changes related to a specific ontology element instead of having again a complete list of changes.

Also, we could observe that users can easily get doubtful (i.e., try to repeat the action) whenever there is a small delay with the communication with the server. From these (and other) observations we got some recommendations to improve our infrastructure, specifically at the GUI level. First we should improve our views with additional features such as: sorting, grouping and filtering. Second, we should also add additional user-friendly features to our interfaces such as the ability to select several changes in one click or refreshing automatically the views when opening them. Third, we should provide a tighter link between the ontology navigator and the information displayed in our views.

Finally, in the case of the NeOn collaboration server, as it is an external component of our infrastructure, we learned that we should provide additional resources (memory) to avoid any crash and to improve the speed of the communications.

7. Testing The LabelTranslator

LabelTranslator is a NeOn plug-in that suggests translations of ontology labels among English, Spanish and German, with the purpose of localizing ontologies [4, 5]. The system 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. A detailed description of the second version of the LabelTranslator NeOn plug-in can be found in report D2.4.2 [10].

In the following sections we describe the experiments designed to measure the quality of translation of the new algorithm used for performing the localization activity and to assess the usability of the second version of LabelTranslator plug-in.

7.1 Overview and objectives

Initially, a preliminary experiment involving PhD students was designed and conducted in March 2008 (section 8 of D5.6.1 [6]). In this experiment we evaluated the aspects related to the translation ranking techniques, where the task was to select the most appropriate translation for each ontology label. In particular, the study looked into two concrete ranking methods: the obtained output using manual and automatic operation, and the quality of translation. Some metrics were used to evaluate the quality of the ontology translations. This was done on the basis of comparing the translations provided by an expert (gold standard) with the translations provided by the ranking algorithm used in the LabelTranslator plug-in. In this sense the experiment was used as a starting point to provide valuable experience and to identify the strengths and weaknesses of the translation-ranking algorithm.

Some changes have been implemented in the new version of the translation-ranking algorithm, so we were interested in re-evaluating the new algorithm and also in evaluating the user satisfaction of the new version of the tool.

The goal of the first experiment carried out this year and described here is to re-evaluate the quality of the new translation-ranking algorithm implemented in the second version of our ontology localization system. To evaluate the quality of the obtained translations we decided to apply a manual evaluation rather than an automatic evaluation metric because current MT evaluation metrics are based on shallow features. Most metrics work only at the lexical level. However, the labels used in ontology terms are rich and ambiguous, allowing for many possible different ways of expressing the same idea.

The goal of the second experiment is to assess the user satisfaction of the LabelTranslator system for carrying out the ontology localization activity.

7.2 Assumptions and user study setup

The first experiment was carried out in the “Artificial Intelligence (AI)” master course at the Facultad de Informática (Universidad Politécnica de Madrid) with master students, having background in databases, software engineering, and artificial intelligence, but no extensive practical experience in ontology engineering.

The second experiment involved 10 participants, most of whom were PhD students at the Facultad de Informática (Universidad Politécnica de Madrid) with a good command on ontology engineering.

7.2.1 User study #1: Quality of Translations

For this experiment we selected two ontologies from the set of Knowledge Web [2] ontologies used in our first attempt to measure the quality of translation. The selected ontologies registered the worst values in the quality of the output of the ranking method. Therefore, our goal is re-evaluate the quality of translation on these ontologies, but using the new algorithm implemented in the second version of our ontology localization system.

The two selected ontologies are in English and our aim is to localize them into Spanish. In this experiment we decided to use a questionnaire that allows collecting the assessments of the students about the capacity of the translation algorithm to provide correct translations according to the context. The questions used to evaluate the quality of the translation deal with the weaknesses found in our first evaluation (see section 8 in D5.6.1 [6]). The questionnaire included the following questions:

1. Are the translations in the target language correct? If not, can you mark the level of correctness? 30%, 50%, 70%, 90%, other
2. If they are not correct, what are the types of errors, in your opinion?
   a. Lack of the correct equivalent
   b. Errors in lexis/terminology
   c. Errors in Syntax
3. Are the compound labels translated correctly? If not, what are the main problems encountered?

This experiment was divided in the following phases:

1. Student groups will perform the ontology localization following the new algorithm used to rank the translations.
2. Student groups will compare the quality of the translations provided by the algorithm in front of the expected translations (according to ontology domain).
3. Students will fill in a questionnaire about the quality of translation obtained.

7.2.2 User study #2: User Satisfaction

For this purpose we conducted an experiment following the Software Usability Measurement Inventory (SUMI) method [9]. The SUMI questionnaire includes 50 items for which the user selects one of three responses (“agree", “don’t know", “disagree"). The following sample shows the kind of questions that were asked:

- This software responds too slowly to inputs.
- I would recommend this software a my colleagues
- The instructions and prompts are helpful
- I sometimes wonder if I am using the right command
- Working with this software is satisfactory
- I think that this software is consistent

The experimenters met with all participants for 10 minutes to explain the purpose of the evaluation session and present the methodology of SUMI evaluation. Then, participants had 20 minutes to test the LabelTranslator system, and 10 minutes to fill the SUMI questionnaire for user-interaction satisfaction. During these two phases of the experiment users were not allowed to ask questions to the experimenters.

The questionnaire was designed to measure the affect, efficiency, learnability, helpfulness and control [3]. SUMI is also mentioned in the ISO 9241 standard as a recognized method of testing user satisfaction [7].

7.3 Findings and observations

In this section, we provide some findings extracted from the analysis of the experiment results.

7.3.1 User study #1: Quality of Translations

From the first part of the experiment, in which students evaluated the quality of translation obtained by the new translation-ranking algorithm, we can mention the following observations:

- As Figure 26 shows, 33% of the students identified the level of correctness of the translations greater than 80%. The rest of students believed that the obtained translations had a level of correctness greater than 90%.

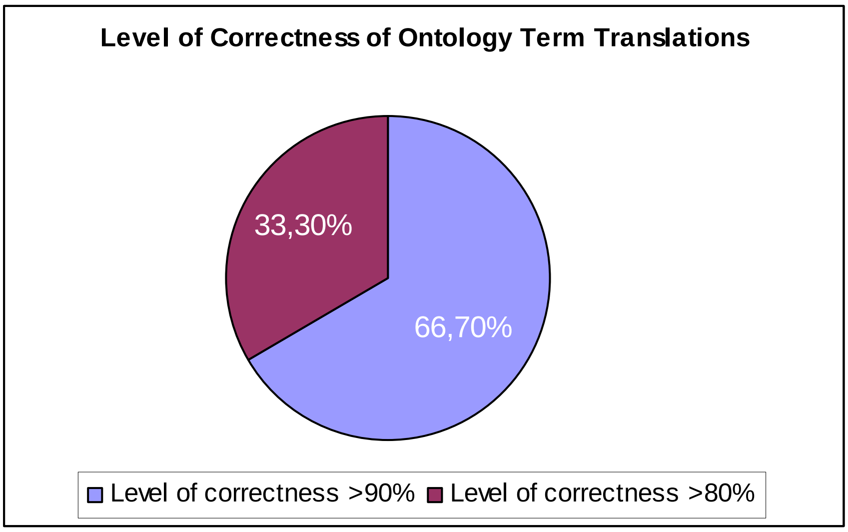

Figure 26. Level of correctness in term translations.

- Figure 27 shows that 83% of errors found in the translation of ontology labels correspond to errors in the terminology used in the translation. Also 17% of the errors correspond to problems in the lack of a correct equivalent in the target language.

1. Finally, to the question “Are the compound labels translated correctly?” the majority of the students believed that the quality of translation of the compound labels was correct. However, they reported errors in those labels that contained tokens with acronyms, for example, “Workshop_URL”, “EPMB_Meeting_Minutes” or “EC_Templates”.

7.3.2 User study #2: User Satisfaction

Figure 28 shows the percentage values for three grades (positive, negative or undecided) of user perception with respect to the goals of each SUMI dimension.

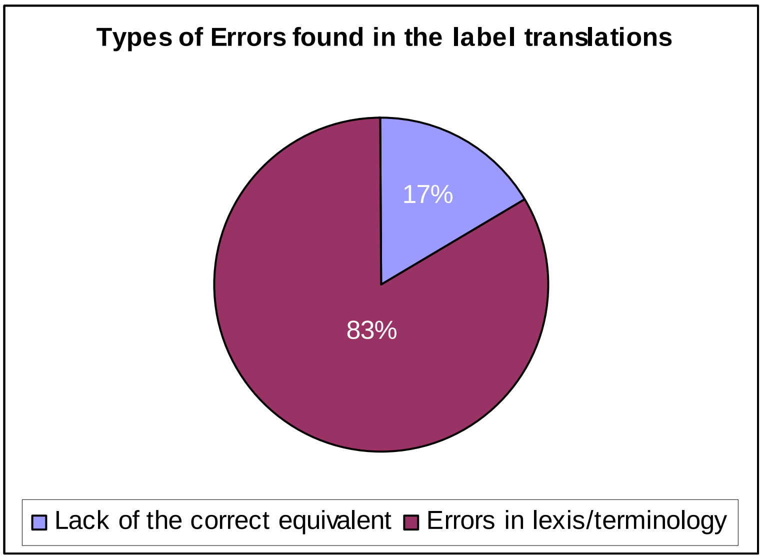

Figure 27. Type of errors found in the translation of ontology labels

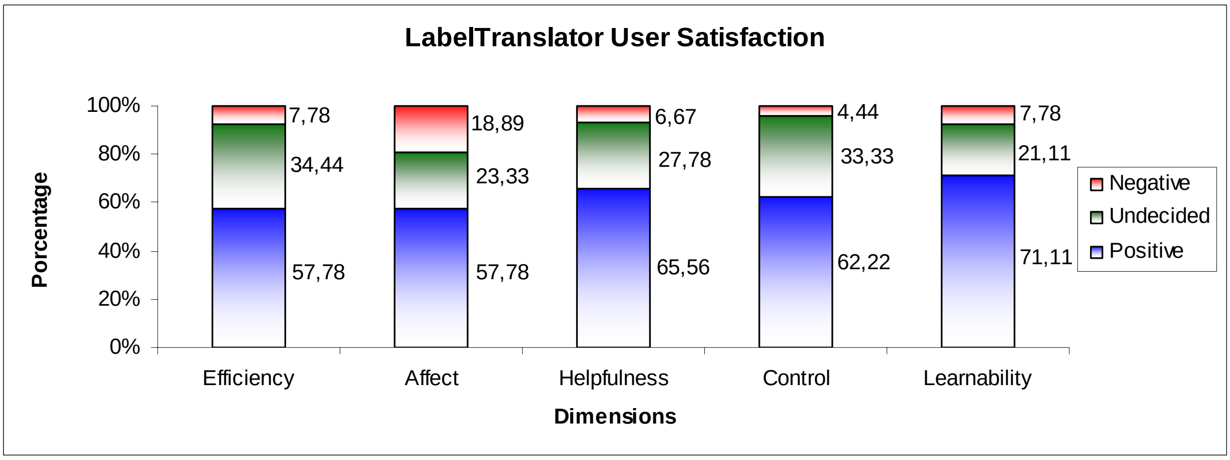

Figure 28. The results of SUMI Questionnaires for LabelTranslator.

7.4 Further analysis and discussion

In this section we include the analysis of the two user studies carried out with the LabelTranslator plug-in.

7.4.1 Analysis of User study #1: Quality of Translations

In this section we analyze the questionnaire filled in by the group of master students that carried out the experiment. Regarding quality of translation issues we asked, these are the main conclusions:

Basically, there was a significant improvement in the translation of the compound labels; this is not surprising, because in the initial translation algorithm we only focused on the definition of lexical templates for compound labels of two or three tokens. The goal of the lexical templates is to produce high quality translations, however, for those compound labels with more than three tokens the algorithm relied directly in the output of the different resources of translation used.

In the current version of the algorithm we implement two improvements in that respect:

- A recursive function that breaks down the compound labels and then attempts to match the bi-tokens or tri-tokens with the patterns stored in the database.
- For those compound labels whose lexical template is not stored in the database, we have created a method that learns new lexical templates from the translations supplied by the user.

7.4.2 Analysis of User study #2: User Satisfaction

In this section we describe the results obtained for each dimension of SUMI questionnaire:

Efficiency

After analyzing each of the 10 questions for measuring the degree to which users feel that the software assists them in their work, we found out that only one question contributed in particular to the 7.78% of disagreement: "I sometimes don't know what to do next with this software". This means that the great majority of the users did not have problems when using the tool.

Moreover, we found that the two questions that most contributed to 34.44% of indecision were: "If this software stops, it is not easy to restart it" and "I find that the help information given by this software is not very useful". However, this situation can be taken positively because it means that users did not have opportunity to check these events.

Affect

The affect dimension measures the user's general emotional reaction to the software – it may be glossed as Likeability. For this dimension we found that the question that most contributed to 18.89% of disagreement in the user's general reaction to the software was: "I feel safer if I use only a few familiar commands or operations". We believe that we must improve this aspect of the system, so that all functionalities can be perceived with same degree of positiveness by users.

Helpfulness

65.56% of the users believe that the software is self-explanatory (helpful). Moreover, we found that the question that more contributed to 27.78% of indecision was: "This software is awkward when I want to do something which is not standard". This means that the majority of the users did not have the need to find alternative options to perform the available actions in the system.

Control

We consider that the evaluation of the degree to which the user feels that (s)he, and not the product, is setting the pace, is satisfactory, because we only obtained 4.44% disagreement. In the same sense, 33.33% of indecision, correspond to aspects that did not appear in the software such as "Error prevention messages are not adequate", which is positive.

Learnability

The higher value was obtained for this dimension, therefore we believe that the evaluation of the ease with which the users feel they have been able to master the system (learnability) was satisfactory.

7.4.3 Identified strengths and weaknesses

According to the results from the first experiment, the main strength of the current translation algorithm is the improvement in the quality of the translation of compound labels, regardless of the number of elements that make up the compound. In spite of this, some weaknesses have still been detected in the algorithm:

- The module for translating compound labels still has some problems. For example, the misuse or omission of the definite article, wrong prepositions, wrong personal pronouns, and incorrectly translated acronyms.
- The current algorithm has some difficulties in correctly identifying the part-of-speech (PoS) of single and compound labels. Basically, we use PoS tagging as a first mechanism for disambiguation, which help us to easily discard some candidate translations. Any failure in this process produces incorrect translations.

At a higher level, an additional weakness of the system is that it currently accesses a limited or closed set of resources to discover translation candidates. This causes some problems in case of certain localization projects that deal with highly specialized domains of knowledge not covered by the available general resources, or in the case of localization projects in which the ontology has to be compliant with a specific vocabulary or language variant. This is the case of certain organizations that use a specific vocabulary to internally communicate, which is gathered in internal glossaries, such as IBM. Another example of localization projects in which general resources show to be unsatisfactory is in the case of ontologies that will be used in countries where the language has some particularities not shared by the international variant. This would be the case of language variants spoken in certain territories or countries. For example, the most appropriate translation for the English term “computer” in the sense of “a machine for performing calculations automatically” is “ordenador” in Spanish, but for an ontology that will be used in South America, it would be “computadora”.

In this sense, current efforts are focusing on making the translation component more flexible, so that it can easily incorporate additional resources to cover the specificities of a certain vocabulary or language variant.

Based on the comments obtained in experiment #2 we can say that the majority of students found the experience with LabelTranslator satisfactory. However, it is necessary to improve the efficiency and affect aspects, as suggested by the responses to the questions on this dimensions.

7.4.4 Prospective further work

As further work we are applying a follow-up study to investigate precisely both quality and impact of the linguistic resources used in the translation process. In ontology localization, the translation of ontology terms is performed with many resources; therefore we are interested in the estimating of the confidence value of a candidate translation obtained from a specific resource. Concretely, for a given translation produced by any resource, lexical database or multilingual dictionary, we aim to measure the confidence of it being correct, according to some informative features.

8. Refinement of ontology reuse studies

This chapter deals with the ontology reuse experiment. Since it has been decided to develop new technology within NeOn that will likely influence the ontology reuse experience for many potential target groups, we have decided to refine the proposed ontology reuse experiment, and incorporate the new technology that has been developed since the previous designs were made. More specifically, it has been decided to combine Watson, the Alignment Server, Oyster (and thus OMV) and the Topic-Specific Open Rating System (TS-ORS) to produce an online community for ontology users and developers alike. Details of this system, called Cupboard, can be found in [1].

We will in the following discuss the setting for the refined experiment.

8.1 Overview and objectives

The goal of the user study is twofold. On the one hand the proposed reuse methodology is tested, and the other hand it is tested how different technology can facilitate the task of reusing existing ontologies. It will be tested how the integration of different resources directly within the NeOn toolkit affects the behaviour of the users and changes their usage experience.

As a result we hope to gain multiple insights into:

1. How users judge the helpfulness of the NeOn methodology for reuse
2. How tools specifically mentioned in the methodology guidelines can facilitate the individual steps of the process
3. Whether the integration of Watson with the NeOn Toolkit (NTK) through the Watson plug-in reduces user effort or not
4. Whether the integration of Cupboard and its data review and ranking support into the NTK helps the users to further reduce the effort needed to reuse existing axioms

During the experiment, user's interactions will be recorded using a screen-capturing software, so that in the analysis phase it can be seen exactly how the user interacts with the system at all times. We will furthermore log users' actions.

8.2 Assumptions and user study setup

The experiment is dependent on some initial seeding of the new Cupboard system with ontologies, and also with the reviews (and ranking information) for these ontologies. We will ask selected ontology developers and experts who work within the NeOn project to provide such initial reviews, but will also encourage these reviews to come from outside the NeOn consortium.

8.2.1 Proposed tasks

The proposed task is to extend a given fishery ontology with axioms from existing ontologies that contain partial knowledge about the fishery domain following a methodological guideline. So, the task is to reuse existing and available knowledge as found on the Web (i.e., in the web-based ontology sharing systems such as Watson and Cupboard) instead of creating it from scratch. In order to create a more controlled environment, the type of axioms that can be used to extend the ontology (e.g. subclass of relationships or relations) will be defined in the experiment guidelines.

In order to test the effect of our technology on the users' performance, we plan to have one control group and two test groups, each with a different form of technological support.

Group 1 (the control group) will be given access to the NeOn Toolkit (NTK) (which has to be used for plain editing of ontologies) and also access to the selected resources that are available over the Internet. They will not be allowed to install or use the Watson or Cupboard plug-ins for the NTK. This means all the axioms found online that the user decides to add to the initial ontology will have to be entered by hand. The users will also be presented with a version of the NeOn methodology guidelines for reuse, which is adapted to the tools available for them.

Group 2 (the test group) will be given the same resources as the control group, but in addition to those it will be able to use the Watson plug-in for NTK, which will enable them to search for relevant axioms directly from the toolkit and will present them with search results from Watson directly within the NTK. Watson plug-in also enables a one-click integration of found axioms into the initial ontology. The users in this group will not have access to the TS-ORS ranking or the reviews from within the NTK. Therefore, they have to assess the quality of an ontology and the validity of a statement themselves. They will be presented with a version of the methodology that indicates in which steps they can make use of the plug-in.

Group 3 (the test group) will be given the same resources as the previous two groups, but additionally will be able to use the Cupboard plug-in for NTK, which will enable them to retrieve axioms in a personalized ranking that is based on reviews from trusted members of the community. They will be presented with a version of the methodology that indicates in which steps they can make use of the plug-in.

8.2.2 Target groups

The experiment will aim to include both novice users and ontology engineering experts. We will therefore monitor each user's actions and also have each user fill out a questionnaire to determine prior knowledge and experience. After the experiment, the user has to fill out a second questionnaire to gather additional feedback on difficulties, the tools and the proposed methodology. We will try to ensure that the distribution of novice and expert users is even among the control group and test groups.

8.3 Plans and timelines

For better tracking progress, the experiment is divided into several phases.

Phase I (to be completed by early March 2009):

- Have first initial user testing of the Cupboard system
- Decide on a set of ontologies that should be used for the experiment
- Add these ontologies into an ontology space
- Assign ontologies to reviewers

Phase II (completed by end of March 2009)

- Design the guidelines for the experiment
- Develop initial ontology which has to be extended
- Test Cupboard Plugin for NeOn Toolkit internally
- Distribute guidelines and test settings to partners

Phase III (completed by end of April 2009)

- Run the experiments

- o Gather the data at UKARL
- o Start analysing the data

Phase IV (completed by June 2009)

- o Produce detailed analysis
- o Publish results

9. Conclusions and Future Work

In conclusion, this deliverable reports on a range of different studies with real user groups of partial, self-contained sub-sets of methods developed and/or refined in the NeOn project. We have covered several important ontology development activities and processes in this report, and obtained a wealth of useful feedback on how these activities are carried out and supported. Of a particular importance is the formal study of users' perception of ontology design patterns, because many specialist knowledge engineers would agree that following best practices and patterns is in principle useful. However, the success or failure of the pattern-based ontology development depends to a great extent on how the ordinary users internalize the overall idea and what degree of cognitive overhead such users would experience, should they decide to deploy patterns in real ontology development scenarios.

Equally important are the studies focusing on establishing the ontology development lifecycle based on the ontology requirement specification document. As with ontology design patterns, experts usually agree that following a systemic, lifecycle driven approach is recommended; yet, our studies now offer explicit feedback with respect to answering the core question "How do the ordinary users take to using the lifecycle establishment guidelines to get a systemic structure for their project?"

In general, we can say that the user studies reported in this deliverable can be classified in three categories. In the first category we tackled the challenge of obtaining explicit evidence supporting or refuting the prevailing informal view in the community that best practices, design patterns, and formal lifecycle models could be realistically adopted to improve the quality of ontologies. These studies are at the core of almost any ontology development project, and assess our methodology on a kind of "meta" level; i.e., how effective our methodological support is in guiding the overall development, its initialization, setup, etc.

In the second category, we tackled the challenge of establishing how effective our methodological guidance is with respect to concrete ontology development activities. In this category we can include the studies with ontology requirement specification guidelines and with ontology localization. A hallmark of this type of studies is their formative character. In other words, studies with initial, preliminary guidelines and techniques were carried out and followed up by re-design and extension of the support. Subsequently, verification studies were executed to establish "delta" in the iterative development of the aspects of NeOn Methodology.

The third category comprises usability-based testing of the models rather than abstracted methods. One of the reasons for including these studies in the deliverable on methodology is that internal models are often among the primary drivers of a particular technique, and, by implication, are hugely influential on the methodology derived to embrace and support these techniques. Hence, it is fully justifiable to obtain user feedback on effectiveness and usability of a particular setup of models used within such ontology development activities as ontology localization or collaborative ontology development/maintenance/debugging or evaluation.

Let us first summarize and cross-link the positive observations and lessons learnt from our experiments. These generalized linkages will be discussed in section 9.1. Next, we will touch on shortcomings and generalized new challenges that should drive our further work in section 9.2. Finally, we devote some space to the vision of future work in this area in section 9.3.

9.1 Generalization of the adequacy of NeOn Methodology

In general, it is useful to have a range of users being exposed to various aspects of the NeOn Methodology. It is not surprising that our studies were targeted at the novice or inexperienced ontology developers – it is precisely this group of users that prevails in the context of ontologies developed and shared using the Semantic Web frameworks and technologies. Hence, it is this group of users that needs to be influenced in terms of internalizing best practices and systemic guidelines, in order to improve the quality of shared ontologies in the long run. In terms of user type, the reported studies covered mostly doctoral and postgraduate students. However, it is important to note that in addition to the usual cohorts taken from the artificial intelligence courses, we had a rich presence of participants with software engineering, databases and system engineering backgrounds.

Hence, despite the study on ontology design patterns suggesting that most of the users were at the similar level of expertise with respect to designing ontologies, we can argue that the participation of users from the standard software engineering fields is a positive 'outcome' from our studies. In particular, it exposes presumed (and designed rather than harvested) best practices in ontology engineering to the representatives from other fields. With this context in mind, it is encouraging to see a fairly substantial degree of user satisfaction with the way NeOn methods influenced their (often first) experience with ontology development.

For example, between 64% and 86% of participants found the idea of using and following the guidance of ontology design patterns very useful. Far fewer participants expressed their doubts and reservations about pattern usefulness. Although 25% of participants in one cohort not finding design patterns useful may look like a worrying result, this should be interpreted in the context of this group having the fastest turnaround in applying the patterns to the ontology development tasks and having received no formal training or coaching in reading, understanding and applying the patterns. Hence, if anything positive can be emphasized from this study (with similar results coming also from the studies on establishing lifecycles and on supporting requirement specification), it is the importance of *initial training and introduction of the methodological component(s) to the user*. With little background and introduction it is hardly possible to expect much of a voluntary uptake of a fairly comprehensive, and often dense, methodological framework.

It is obviously one question to report on the subjective opinions of the participants, and it is quite another challenge to actually observe some objective change – be it in terms of development speed, outcome quality, or in terms of acceptability of the results of such a principled process. Here, the evidence acquired from the individual studies tends to point to the generalized conclusion that *users are generally slower if they follow a formal methodological guidance, as opposed to carrying out the same (or similar) task ad hoc*. However, this loss in the time scale is often compensated with less iteration in the ontology development process or with the availability of ontologies that are more readily comprehended, understood by the peers if shared.

One can therefore highlight the emerging trend across several studies reported here of seeing *no explicit saving in terms of the total ontology development time*. Usually, the initial losses are due to preliminary activities (e.g., selection of the design pattern, review of the available lifecycle models, or review of the possible translation of a label). Once the user gets through this "preparation", the actual "realization" phase tends to be more straightforward and also faster when compared to unstructured development process.

One interesting finding highlights the fact that following the same guideline (or pattern) is not always equally effective. A very positive outcome from the reported studies is in our improved awareness of the types of tasks where our methods lead to a higher satisfaction. For example, the role of design patterns was more subdued in routine parts of the design process, where people looked at design patterns and guidelines as a source of inspiration rather than something practical. However, in more complex modelling situations (e.g., with n-ary relationships or a mixture of classification and role-taking behaviours), patterns became much more of a guidance element.

Similarly in the study of establishing the lifecycles for a particular ontology development project, it was found that more familiar decision trees in making a "rough cut" decision about the general suitability of a lifecycle model were perceived as more useful compared to less familiar classification of activities into (not necessarily objective) categories. In several studies, the notion of familiarity with different aspects of the pattern/methodological guideline arose and led to *people appreciating the potential value, benefit, yet having some difficulties in translating this value into a concrete situation*. Similarly to our studies highlighting a tight link between initial training or coaching and the readiness to adopt a formal guideline, we can also highlight another generic lesson learnt: *Those parts of methodological guidance that were richly exemplified tended to be perceived as more useful, more effective*. Hence, we definitely emphasize the role of examples, illustrations, real-world instantiations in the descriptions of our methodological components!

A very positive response can be generalized in terms of "predictive power" of the methodological guidelines and patterns in NeOn. Indeed, several studies (including ones with design patterns, ontology requirement specification guidelines, and ontology lifecycle establishment) concluded that most *users are willing to use the guideline/pattern in the future, even in their own ontology development projects*. This is a much more important indicator of effectiveness than the level of user acceptance in the controlled setting – should the user follow *at least a part or sub-set of the method/pattern*, this is likely to accelerate the overall adoption of the entire methodology.

One very useful feedback that is emerging from several studies described in this deliverable is the notion of not only supporting the core, "primitive" activity as a monolithic procedure, but also allow for a decomposition of an activity to a process comprising simple sub-activities. This observation has several implications. *First, the boundary between methodology-supported "activity" and "process" seems to be clear to the users of the NeOn Methodology at the conceptual level*. However, in some cases a normally straightforward methodological guidance applied to a larger, more complex task may become intractable. This happens not because of the methodology being incomplete or insufficient. A more likely source of the troubles is in the possibility that many of the existing methodological guidelines lack the notion of "expanding activity into a process". In other words, what may the user want to consider in addition to the usual issues, if s/he encounters case that needs a sort of "divide and conquer" approach. Some insights into how this link from an activity to a process (or another activity) may be established have been observed in the feedback from the study on lifecycles – *users were keen to see some cross-referencing between different activities* – sort of:

- What other choice do I have except the most recommended one?
- What other activities do I need to be aware of (and to what extent) if I choose this path?
- What activities become irrelevant upon a particular decision?

Although these are obviously hard questions to answer, the positive outcome is that in the NeOn project some of these (ontology development) scheduling and planning choices are planned to be supported by a dedicated NeOn Toolkit plug-in called gOntt, and a possible integration of this scheduling component with a familiar metaphor of providing user support in many professional productive environments by means of wizards and/or cheat sheets. The work on gOntt is in progress at the time of writing this deliverable, yet some visionary sketch of its role and functionality can be found in [18].

Another very interesting and important lesson for the adoption of the methodology is somewhat orthogonal to the content and clarity of the methodological guidelines. In several cases, but most notably in the study of collaboration support, it was observed that *users very quickly resort to doubts about making the right methodological choice* if they do not see a quick response to their action in the tool or with a technique. This points to a great importance of not only the way methods are described and presented to the user, but also of how the methodological guidance *is realized in the respective techniques and toolkit(s)*. In other words, the role of a GUI of a specific plug-in, technique or infrastructural component may easily have a stronger effect on the user adoption of the methodology (and also on the user perception of the methodology effectiveness) than previously thought. As mentioned in the previous paragraph, this is actually a very positive formative feedback to the methodology developers, as it provides further, formal rationale to having tools like gOntt and the presentation of the “instantiated method content” by means of Eclipse cheat sheet widgets that were briefly touched upon above.

9.2 Generalization of the shortcomings of the NeOn Methodology

While most of the quantitative feedback suggests overall satisfaction and positive perception of the components of the NeOn methodology, there are a few areas where we can identify some shortcomings – either to our studies or to the way our methodological guidelines were used and interacted with. The first, and most frequently quoted shortcoming is the discrepancy between the methodological guideline (and the associated best practice and its benefit) and the actual tool support, a technique helping the user to carry out the recommended activity or step.

Here, the example of LabelTranslator provides one of the areas where the alignment between the guidelines and the technique (plug-in) is almost ideal: there is a very extensive functional support in the plug-in for many minor decision points and the realization of those functions is consistent with the respective methodological guideline. On the other hand, many other methodological elements, in particular the design patterns, still lack a well aligned technological support. As was mentioned in chapter 2, there is a clear value in supporting the users in selecting the right pattern, comparing different patterns and their implications, etc. While there is a comprehensive repository of design patterns, the user interaction with it is largely at the level of free text search, which many users did not find to be the most effective support. Nevertheless, we believe that this concrete feedback may act as an additional motivation for the technology developers (a) to provide technique-specific input to designing a particular method and (b) to aim to support even the steps that are less important from the technological viewpoint but make a lot of difference to the user experience.

Another often repeated feedback from the studies described in the previous chapters relates to the degree of abstraction. Due to certain design commitments (mainly, striving to design NeOn methodological guidelines and the methodology so that it is transferable between the tools), there is a degree of abstraction in the description necessary to cover the problem or the activity in general. However, this *generic language seems to be putting some users off, and many users reported back that it was in places hard to instantiate the guidance to a specific scenario*. The fact that not all sections in the description of NeOn methodological guidelines suffered this problem points to the conclusion that this is a matter of minor extension to the existing methods rather than a new design challenge. In fact, users actually pointed to the solution themselves – by providing several concrete examples, possibly with some step-by-step mapping to the generic, they felt more confident in choosing or rejecting a particular design pattern or a particular methodological choice/commitment.

Another aspect that is being tackled in WP5 at the time of writing this report is *the use of consistent terminology in and between the individual activities*. This was less of a problem in the current studies, only a few occasions were noted where people had difficulties with interpreting a particular technical term, but it is an important area to take on board. In particular, *users often expected some cross-referencing to similar sounding activities* – this was the case in the ontology specification and lifecycle establishment studies, and also some cross-references between different (types of) design patterns. Although this shortcoming was not enormous in our current studies it can easily escalate if larger chunks of the NeOn Methodology are tested and the user would be expected to switch from one method to another, from guideline addressing one activity to a series of guidelines addressing preceding or following activities.

Another potential shortcoming relates, to some extent, to the issue of abstraction. Some user feedback suggests that people may be in some specific situations better off if abbreviated guidance was available. For example, people often make some preliminary choices (e.g., with respect to a candidate ontology lifecycle or an applicable design pattern), and to read extensive conceptual background to make such a “rough cut” estimated decision may be off-putting. Hence, some user feedback suggests an existence of shorter, “reference card” guideline overviews in addition to full-scale explained and illustrated methodological guidance.

Generally, *the users were keen to see more figures, more examples and more illustrations from the real world cases included in the actual methodological guidance*. The current balance between specific examples and generic guidance of the methodology components has its origin in aiming at a general methodological support rather than supporting only one concrete tool or domain. One potential solution currently under investigation is the association of a generic method to a somewhat more explicitly grounded guideline. In other words, the methodological support can be seen as a union of more generic statements (“the method”) and a range of different realizations of that method in a particular toolkit and/or problem domain (“the guidelines”). We believe, with the work on gOntt and a pilot of translating some of the methodological content to concrete “cheat sheets” this balance between abstract and concrete may be shifted and made more favourable to the user seeking concrete examples, cases, and scenarios.

One of the re-emerging limitations of this kind of user studies is to what extent we can really compare a control group with a test group. As suggested in the studies on design patterns, on ontology requirement specification or lifecycle establishment, mere terminology needs a certain degree of explanation and introduction. From the educational point of view, it is not possible to train one group of users and not to train a sub-group of students subscribed to a given course – this issue of a ‘trained’ vs. ‘test’ group becomes even more visible in the training setting, where the participants to our studies are actually students or trainees participating in a course with an intent to develop a certain skill.

Ideally a statistically valid experiment should study the behaviour in the same task and in the exactly same setting both with and without the presence of our methodological support. This was largely not possible in many our studies due to practical limitations, such as the availability of participants at a given point in time. Nevertheless, *our studies were designed to the greatest possible extent so that some meaningful implications can be inferred, especially from a range of similar participant groups, if not observed directly*. It is also for these reasons that we included explicit assumptions, prior expectations and limitations in the write-up of the individual user studies, so that the traceability can be preserved as much as possible.

Yet another potential point for criticism arises from the relatively narrow focus of some of the reported studies – e.g., a fairly simple use of design patterns (with no explicit selection, comparison, analysis, combination, etc. of the patterns), or a dedicated label translation (without being a part of a bigger problem, such as ontology re-use or non-ontological resource transformation). The authors of this report are aware of this narrow focus, however, this was a designed feature rather than an accidental bug. We are keen to obtain user feedback and performance data on both aspects – the individual, partial methods, as well as the comprehensive methodology. However, the two have slightly different purposes and thus timing. *Whilst the evaluation of the individual methodological guidelines is more formative, a part of the iterative process of methodology construction, formulation and refinement, the evaluation of the methodology as a whole is much more summative*.

We recognize this duality of needs and therefore work towards addressing both types of evaluation. First the D5.6.* series of deliverables looks into the individual methodological guidelines and their clarity, completeness, user adoption, etc. In a parallel thread, we are presenting in deliverable D5.7.1 a line of proposals to be concatenated into a more comprehensive methodology-centred study. In the latter, the focus is likely to be on the methodology presentation in a given tool, and on the alignment of the methodology with the toolkit(s) and techniques. Nonetheless, both strands are useful, both are needed in order to provide a different picture of the same studied object.

9.3 Future studies with NeOn Methodology

With respect to further work in the area of testing individual methodological guidelines as opposed to testing more comprehensive sequences of methodological guidelines or even methodology as a whole, we can point to several lines of work. First, there are still experiments that need to be done, in particular, with two of the most important ontology development activities in the context of networked ontologies: namely, our refined proposals for evaluating re-use of ontological resources with the help of social ranking and recommendation support. These studies, as well as additional experiments with design patterns will be carried out between M37 and M48, and will be reported in the next deliverable, i.e., D5.6.3.

Some of the possible questions to study with respect to the design patterns only, include the following: the effects of different levels of experience on the use and usefulness of content patterns, the effects of different development strategies on the use of patterns, the effects of pattern presentation on the use and usefulness of content patterns, or the use and usefulness of other types of patterns, in addition to content patterns. To some extent, these tests depend on the availability of the XD plug-in for pattern-based design within the NeOn toolkit, which is proposed as the platform for experimenting further with pattern-based design. Also, many of the raised questions may be incorporated into the experimental proposals for the methodology as a whole in D5.7.1 [18].

In ontology requirement specification studies, one may look into the effect of the discussed reference cards with the summaries of the extensive information provided by the respective methodological guidelines (task input, task output, actors, etc.) One can also study the effect of an integrated tool for performing the ontology requirement specification activity on the adoption and/or perception of the respective methodological guidelines. However, as above, both of these challenges may get incorporated into the comprehensive study of the methodology along the lines of D5.7.1 [18], in particular if plug-ins like gOntt and methodological cheat sheets are pursued.

In terms of the potential extensions to studying the process of establishing lifecycles for an ontology development project, many of the observed issues were already addressed in the revised guidelines and are available in deliverable D5.3.2 [13]. There is a potential value in studying the opportunities to cross-reference and/or to compact the NeOn Glossary of Activities, to make it more wieldy and easier as a referential framework for the users. Obviously, as mentioned in the previous paragraph, there is also a possibility to study the benefits of a reduced version of the methodological guidelines with quick overviews, to complement the “comprehensive” guidance for establishing the ontology (network) lifecycle.

10. References

[1] M. d'Aquin and H. Lewen. Cupboard – A Place to Expose your Ontologies to Applications and Community. In Proc. of the Demo Track of the 6th European Semantic Web Conf., Heraklion (Greece), May 2009.

[2] O. Corcho, A. Lopez-Cima, and A. Gomez-Perez. *The ODESEW 2.0 semantic web application framework*. In Proc. of the 15th International Conference on World Wide Web (Edinburgh, Scotland, May 23 - 26, 2006). WWW '06, ACM Press, New York, NY, 2006

[3] J. Dumas and J. Redish. *A Practical Guide to Usability Testing*. Exeter, UK: Intellect, 1993.

[4] M. Espinoza, A. Gómez-Pérez, and E. Mena. *Enriching an ontology with multilingual information*. In Proc. of the 5th European Semantic Web Conference (ESWC'08), Tenerife, (Spain), June 2008.

[5] M. Espinoza, A. Gómez-Pérez and E. Mena. *LabelTranslator - A Tool to Automatically Localize an Ontology*. In Proc. of the 5th European Semantic Web Conference (ESWC'08), Tenerife (Spain), Springer Verlag, pp. 792-796, June 2008.

[6] R. García-Castro, M. C. Suárez-Figueroa, M. Espinoza, M. Sini, H. Lewen and Eva Blomqvist. *Experimentation with the NeOn methodologies and methods*. NeOn Deliverable D5.6.1, NeOn Project (http://www.neon-project.org). March 2008.

[7] ISO. Guidance on usability specification and measures. ISO, CD 9241-11, 1992.

[8] ISO, editor. ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. 1998

[9] J. Kirakowski, M. Corbett. *SUMI: The Software Usability Measurement Inventory*. British Journal of Educational Technology, 24(3), 1993, pp. 210-212.

[10] E. Montiel-Ponsoda, W. Peters, G. Aguado de Cea, M. Espinoza, A. Gómez-Pérez, M. Sini, *Multilingual and Localization Support for Ontologies*. NeOn Deliverable D2.4.2, NeOn Project (http://www.neon-project.org). August 2008.

[11] M. C. Suárez-Figueroa, S. Brockmans, A. Gangemi, A. Gómez-Pérez, J. Lehmann, and H. Lewen. *NeOn Modelling Components*. NeOn Deliverable D5.1.1, NeOn Project (http://www.neon-project.org). April 2007.

[12] M. C. Suárez-Figueroa, G. Aguado de Cea, C. Buil, C. Caracciolo, M. Dzbor, A. Gómez-Pérez, G. Herrero, H. Lewen, E. Montiel-Ponsoda, V. Presutti. *NeOn Development Process and Ontology Life Cycle*. NeOn Deliverable D5.3.1, NeOn Project (http://www.neon-project.org). August 2007.

[13] M. C. Suárez-Figueroa, M. Fernández-López, A. Gómez-Pérez, K. Dellschaft, H. Lewen, M. Dzbor. *Revision and Extension of the NeOn Development Process and Ontology Life Cycle*. NeOn Deliverable D5.3.2, NeOn Project (http://www.neon-project.org). November 2008.

[14] M. C. Suárez-Figueroa and A. Gómez-Pérez. *First attempt towards a standard glossary of ontology engineering terminology*. In Proc. of 8th International Conference on Terminology and Knowledge Engineering (TKE'08), 2008.

[15] M. C. Suárez-Figueroa, G. Aguado de Cea, C. Buil, K. Dellschaft, M. Fernández-López, A. Garcia, A. Gómez-Pérez, G. Herrero, E. Montiel-Ponsoda, R. M. Sabou, B. Villazon-Terrazas, and Z. Yufei. *NeOn Methodology for Building Contextualized Ontology Networks*. NeOn Deliverable D5.4.1, NeOn Project (http://www.neon-project.org). February 2008.

[16] M. C. Suárez-Figueroa *et al*. *Revision and Extension of the NeOn Methodology for Building Contextualized Ontology Networks*. NeOn Deliverable D5.4.2, NeOn Project (http://www.neon-project.org). February 2009.

[17] Erik P.W.M van Veenendaal. *Questionnaire based usability testing*. In Proc. of the European Software Quality Week, Brussels, 1998.

[18] M. Dzbor, M. C. Suárez-Figueroa and A. Gómez-Pérez. *Plans for evaluating NeOn Methodology*. NeOn deliverable D5.7.1, NeOn Project (http://www.neon-project.org). February 2009.

[19] V. Presutti, A. Gangemi, S. David, G. Aguado de Cea, M. C. Suárez-Figueroa, E. Montiel-Ponsoda, and M. Poveda. *A Library of Ontology Design Patterns: reusable solutions for collaborative design of networked ontologies*. NeOn Deliverable D2.5.1, NeOn Project (http://www.neon-project.org). February 2008.

[20] R. M. Sabou, S. Angeletou, M. d'Aquin, J. Barrasa, K. Dellschaft, A. Gangemi, J. Lehmann, H. Lewen, D. Maynard, D. Mladenic, and M. Nissim. *Methods for Selection and Integration of Reusable Components from Formal or Informal User Specifications*. NeOn Deliverable D2.2.1, NeOn Project (http://www.neon-project.org). May 2007.

[21] A. Gómez-Pérez. *Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases*. Banff Knowledge Acquisition for Knowledge-Based Systems, KAW'99, 1999, 2.

[22] N. Guarino and C. Welty. *Evaluating Ontological Decisions with OntoClean*. Communications of the ACM, 2002, 45, pp.61-65.

[23] J. Völker and E. Blomqvist. *Prototype for Learning Networked Ontologies*. NeOn Deliverable D3.8.1 , NeOn Project (http://www.neon-project.org). February 2008.

Appendix 1 – Design Pattern Experiment Data

In this appendix the data collected from the three experiment sessions is presented, in order to give a detailed background to the analysis and conclusions drawn in chapter 2. The summary of data is divided into three section based on the session where it was collected. Free text answers are usually summarized in tables; other questions are primarily summarized in tables or diagrams. Numbers that are presented without further comments represent the total number of respondents selecting that particular answer. Following are the two task descriptions given to the participants, and an example of the instructions given to the participants (the example is taken from the third session, i.e., the master's course in Jönköping) and belong to the second task:

Task1 – album production

Context

The Warner Bros recording label has decided to manage its own productions by means of an ontology-driven application. They provide the designers with documents describing scenarios that have to be stored in the knowledge base. From these documents one story is extracted and assigned to you as ontology designers.

Story: album production

The "Red Hot Chilli Peppers" are: Anthony Kiedis (vocals), Flea (bass, trumpet, keyboards, and vocals), John Frusciante (guitar), and Chad Smith (drums). During 2005, the band recorded the album "Stadium Arcadium". The album contains 28 tracks and has been released in May 2006. It includes the track of the song "Hump de Bump", which was composed in January 2004. The critic Crian Hiatt defines the album as "the most ambitious work of its twenty-three-year career".

The story above can be simplified by transforming it into 5 sentences:

s1: The "Red Hot Chilli Peppers" are: Anthony Kiedis (vocals), Flea (bass, trumpet, keyboards, and vocals), John Frusciante (guitar), and Chad Smith (drums).

s2: During 2005, the band recorded the album "Stadium Arcadium".

s3: The album contains 28 tracks, and has been released in May 2006.

s4: It includes the track of the song "Hump de Bump", which was composed in January 2004.

s5: The critic Crian Hiatt defines the album as "the most ambitious work of his twenty-three-year career"

Task2 – modelling of nursing

Context

The Italian Ministry of Health wants to monitor the roles taken by employees in hospitals, and is creating a semantic infrastructure for that purpose. The following story is typical of the facts to be represented in its knowledge base.

Story

Pasquale Di Gennaro is the union representative for male nurses at *Ospedale Riunito delle Tre Valli* in Nocera Inferiore (IT) from 2002. At least the following competency questions have to be covered:

Who does represent a certain collective in a certain period?

When does a certain person play a certain role for that hospital?

For which hospital, in which city, a certain person has a certain role?

Where is a certain hospital?

The patterns available were all patterns currently available in the ontology design pattern portal (at http://www.ontologydesignpatterns.org), the catalogue of submitted proposed patterns present was the same for all three sessions but may have changed slightly since.

Example of instructions preceding Task 2

The two PowerPoint slides in Figures A-1 and A-2 were shown in the lectures in the master's course in Jönköping, preceding the second task, as method guidelines for how to solve the tasks.

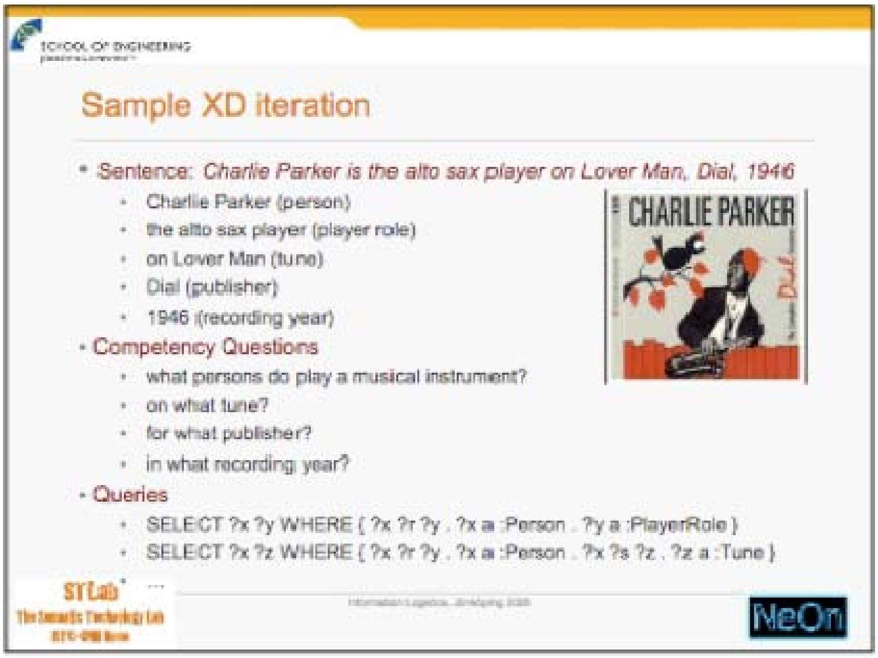

Figure A-1. Slide showing how a sentence can be broken down.

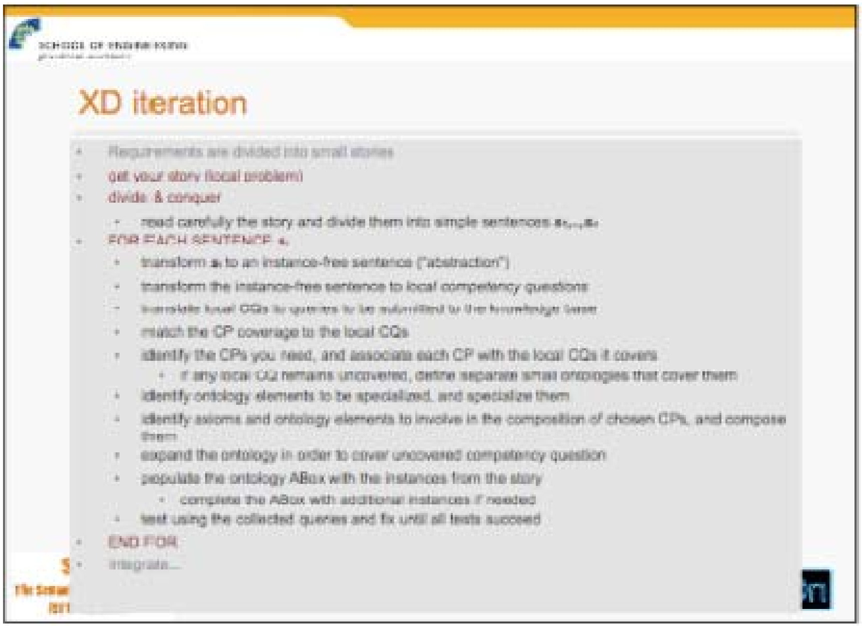

Figure A-2. Slide showing the proposed method.

The following text contains the exact instructions given to the participants of the third session before the second task. PingPong is the Learning Management System used at the university. The instructions were also presented orally and by showing the tool, but no additional examples were introduced at this point.

Lab 4 – Ontology construction using OWL and patterns

Welcome to the fourth lab in the Information Logistics course, fall 2008. When you have completed the exercise be sure to upload the OWL-file in the assignment named "Lab 4 - File upload" at the end of the session. If you forget to do this, the submission will be closed and you have to instead submit the full lab report (see separate instructions).

Lab outline

Before the lab:

1. You should have already filled in the background questionnaire before the first session, but if not do it now. This is a prerequisite to submitting the lab results.

At the lab session (3 hours):

1. The lab starts with a one hour instruction/tutorial where some of the problems you experienced in the other labs will be discussed and some possible solutions will be suggested. Additionally some patterns will be presented and how to use patterns together with the tool will be displayed. Take the opportunity to follow the tutorial and get even more familiar with the tool and especially with the patterns, it will help you in solving the exercise.
2. After one hour you will be instructed to start the exercise. You should solve the exercise in the lab groups. It is recommended that you use only one computer and that one of you models using the tool and the other makes suggestions, corrections and keeps track of the next steps.
3. When there is 10-15 minutes left of the lab session it is time to finish you modelling and to instead answer the lab questionnaire. You will find the questionnaire together with all other lab content in PingPong; it is called "Questionnaire Lab 4".
4. After completing the questionnaire the lab submission assignment will appear in PingPong and you should submit the resulting OWL file. Find the file in your workspace folder and upload it to PingPong, in the assignment Lab 4 - File upload. If you cannot see this assignment you have probably forgot to fill in the questionnaire, do this first and then upload the file.

After the lab:

1. The uploaded results will be marked by the teacher and you will see the result as "status" of your uploaded file. If the status is "completed" then you have passed the lab. If the status is "to be marked" it has not yet been reviewed by the teacher. If the status is "revision required" you have to make some changes/corrections as specified in the teacher's comments and re-submit the file. The resubmission cannot be done in the same assignment, due to technical limitations in PingPong; you have to instead resubmit through the assignment "Lab 4 - File resubmission".
2. The final deadline for completing/passing the lab in this manner (for those that attended the lab session) is January 16. If you still did not pass by then, please refer to the instructions regarding how to write a lab report instead.

Background

This lab is heavily based on the first and second lecture on ontology construction using OWL and patterns, please refer to lecture notes for details. Additional useful information can be found through the following links:

The W3C OWL tutorial page: http://www.w3schools.com/RDF/rdf_owl.asp

An older OWL tutorial (from 2005) with examples: http://owl.man.ac.uk/2005/07/sssw/

All the proposed ontology design patterns available in the pattern portal: http://ontologydesignpatterns.org/index.php/Submissions:Main

A list of suggested patterns for this lab: http://hem.hj.se/~blev/patterns.htm

Also please refer to the help function of TopBraid composer for "how to's" regarding the tool.

Practical information – and how to get started

Important before you start modelling:

1. You can find the tool to use (TopBraid Composer) under the JTH menu item (Start Menu -> All Programs). Note that it takes a few moments before Novell has distributed all applications to the menu, if you do not see the shortcut wait a few minutes.
2. Be sure to set the Workspace path to some folder on your own account, that is, on your G: and NOT on the local machine (otherwise all your results will be lost when you log out). Set the correct folder at startup or change it through the menu item File → Switch Workspace.
3. When you start the editor, make sure that you are displaying the right perspective (the TopBraid perspective). If you are not sure, try to select Open Perspective in the Window menu, select “Other...”, mark TopBraid and click OK.

How to get started:

1. To create a new project: Right click in the Navigator window (or select from the File menu) and select New -> Project... Open the General folder and mark Project, click next. Give the project a name and be sure it will be located in the workspace folder you stated at startup (on you G:). Click Finish.
2. To create a new ontology: Right click on the project folder where you want to create the new ontology. Change the last part of the Base URI (after the last "/") to the name of the ontology (the file name will automatically change accordingly). Make sure the file extension is set to "owl" before you click Finish.
3. There are usually both context menus (right click on any window or item to display its context menu) and icons for performing actions on different objects, such as creating subclasses and subproperties.
4. Since we are working with OWL you can select the option "Start hierarchy with OWLThing" from the classes’ window menu. Otherwise make sure that all classes you define are subclasses of OWL Thing.

While modelling:

1. Are you missing some OWL constructs?
   Use the small downward arrows at the top right of the classes and properties windows to access the menu, select Preferences. Under the heading TopBraid Composer you can find for example settings for the classes and properties views. Here you can select what constructs should be visible.
2. Unable to show the details of a class, property or instance in the central window?
   Be sure you have marked the correct class/property/instance. It is marked if a small white arrow appears on its icon. Mark by double clicking the desired class/property/instance or clicking on its icon.
3. Is there a property missing in the details window when looking at an instance?
   What is displayed is governed by the domain and range of properties. If you want to add another property, remember that you can always drag and drop things from one window to another, drag your property onto the detailed display of the instance!
4. Before running any inferences, configure the inferencing by selecting Inferencing → Configure Inferencing... Make sure that you are using an appropriate reasoner, for example Pellet, with your project (check the box "customize settings for..." in order to change settings).
5. When you run inferences and something goes wrong, be sure to reset inferences before continuing, otherwise all inferred triples will still be present in your ontology. Also note that in case of an inconsistency in the A-box you will have an explanation for the inconsistency in the Inference Explanations tab.
6. To import an ontology (or a pattern) into your ontology, first save the OWL file in your workspace folder. Next, refresh the workspace view to see the file. Drag and drop the file into the imports tab of your current ontology.

Lab 4 – Modelling task

Develop an OWL ontology starting from the below use case, use patterns if you find them helpful.

Context

The Italian Ministry of Health wants to monitor the roles taken by employees in hospitals, and is creating a semantic infrastructure for that purpose. The following story is typical of the facts to be represented in its knowledge base.

Story

Pasquale Di Gennaro is the union representative for male nurses at Ospedale Riunito delle Tre Valli in Nocera Inferiore (IT) from 2002.

At least the following competency questions have to be covered

- Who does represent a certain collective in a certain period?
- When does a certain person play a certain role for that hospital?
- For which hospital, in which city, a certain person has a certain role?
- Where is a certain hospital?

PhD Course – Bologna

The backgrounds of the participants were first recorded through one questionnaire. Below the results from this questionnaire can be seen.

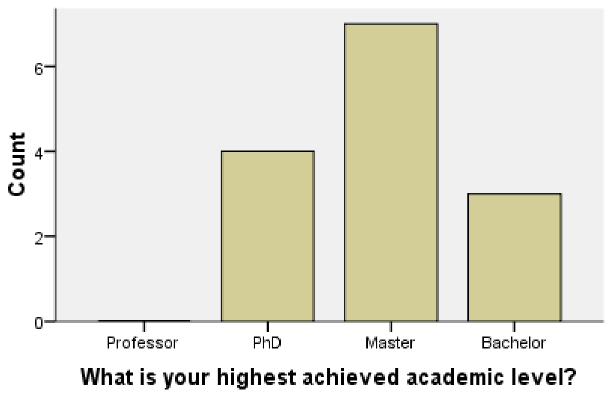

Figure A-3. The academic level of the subjects.

Topics of their studies contained variations of Computer Science (5 subjects), Electronic Engineering (3 subjects), Information Technology (2 subjects), and Computer Engineering (5 subjects).

The current positions of the participants can be seen in Figure A-4, and are mainly PhD students and research assistants, although one senior researcher and one post-doc also answered the question. The ‘other position’ was translated into ‘assignment for research’, which is corresponding to a research assistant. The answers to the question if their current work is in some way related to ontologies are summarized in Figure A-5.

The extent of their experience in working with ontologies is summarized in Figure A-6 and Figure A-7. Subsequently they were asked to characterize the ontologies they had constructed (if any); in terms of size and complexity, and their responses can be viewed in Figure A-8 and A-9.

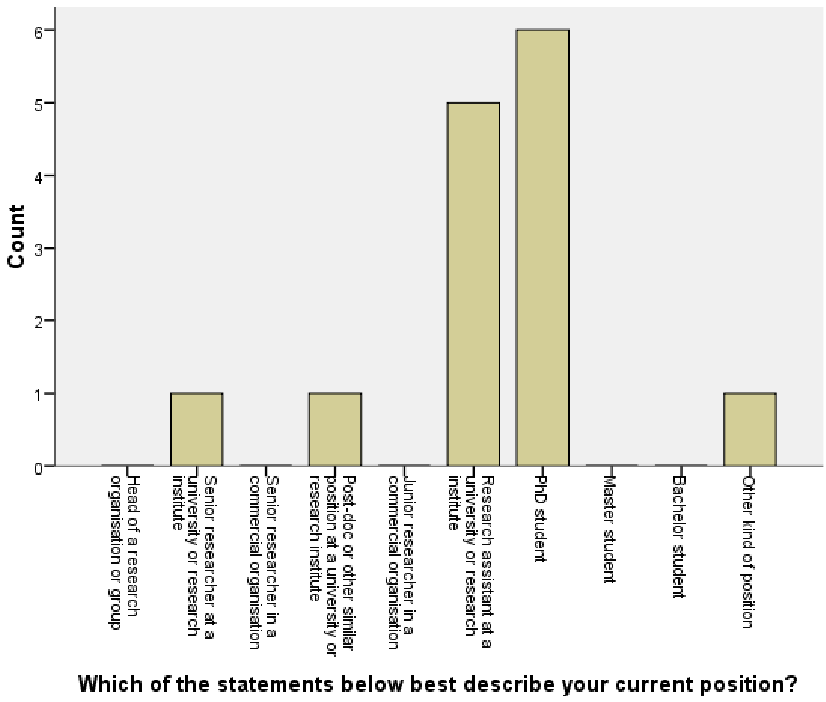

Figure A-4. The current positions of the subjects.

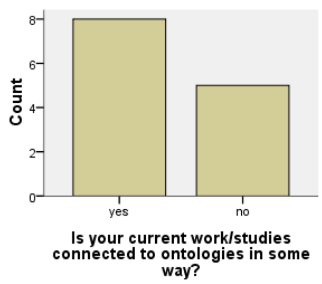

Figure A-5. Connection to ontologies.

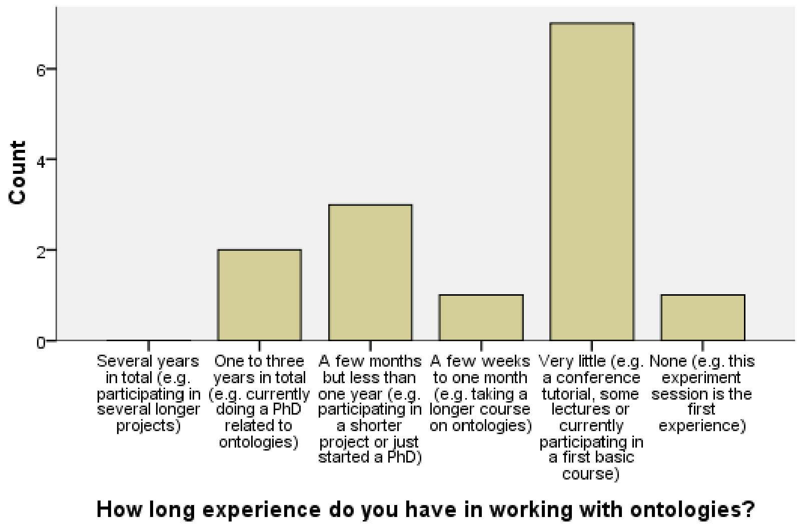

Figure A-6. Experience in terms of amount of time.

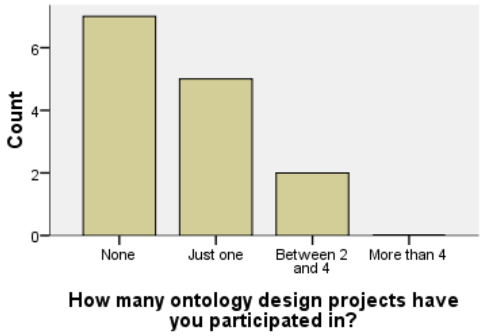

Figure A-7. Experience in terms of number of ontologies.

Figure A-8. Constructed ontologies (if any) in terms of size.

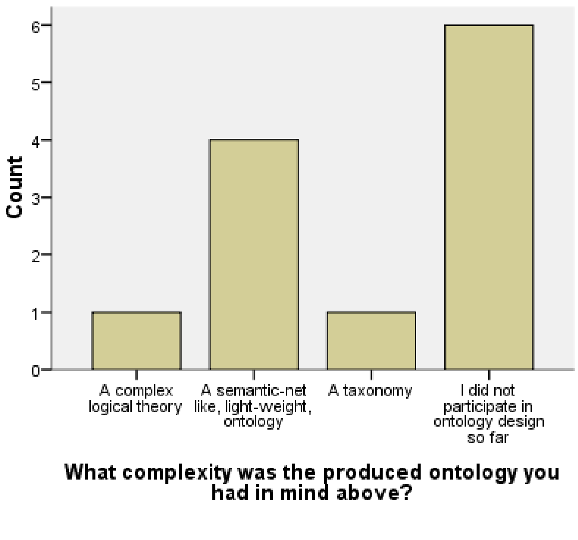

Figure A-9. Constructed ontologies (if any) in terms of complexity.

Next, they were asked to specify their knowledge and experience with respect to a set of tools (see data in Table A-1) and languages (see Table A-2), used for ontology engineering and modelling in general.

Table A-1. Tool experience.

What tools did you use?

	Did not try it	Tried it	Have extensive experience
Protégé 3.x	**4**	**6**	**2**
Protégé 4	**6**	**2**	**0**
TopBraid Composer	**6**	**2**	**0**
OntoEdit / OntoStudio	**8**	**0**	**0**
KAON	**8**	**0**	**0**
NeOn toolkit	**7**	**1**	**0**
SWOOP	**7**	**1**	**0**

Table A-2. Language experience.

What languages did you use?

	Not familiar	Somewhat familiar	Have extensive experience
First-order logic	**5**	**6**	**1**
ER-diagrams	**2**	**9**	**2**
UML	**3**	**7**	**3**
RDF	**4**	**6**	**2**
OWL	**8**	**3**	**2**

Finally, to the final question of the background questionnaire, concerning previous usage of ontology design patterns all of the subjects responded that they had not previously used patterns.

For the following experiment sessions the first set of questions were identical between both tasks, and the detailed resutls can be seen in the referenced diagrams:

1. I found the problem description easy to understand (Figure A-10).
2. I felt familiar with the domain of the modelling problem (Figure A-11).
3. The problem was clearly and unambiguously defined (Figure A-12).
4. The modeling problem (the ontology) was small compared to other ontologies I have constructed before (Figure A-13).
5. I found the tool used for the experiment easy to use (Figure A-14).
6. The modeling problem was easy to solve (Figure A-15).
7. I made some mistakes at first and had to redo some parts of the ontology later (Figure A-16).
8. There were some problems that I did not manage to solve in a “good” way within the given time limit (Figure A-17).

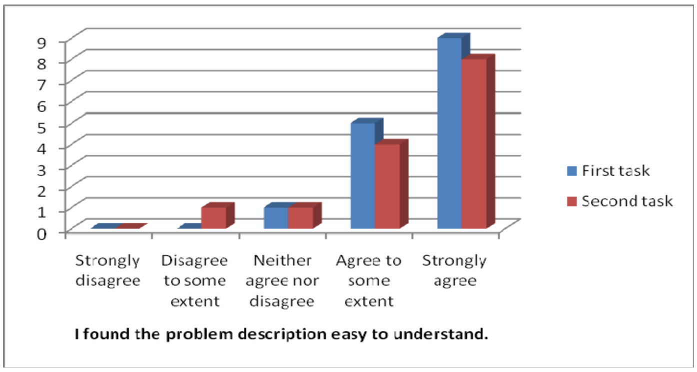

Figure A-10. Problem description.

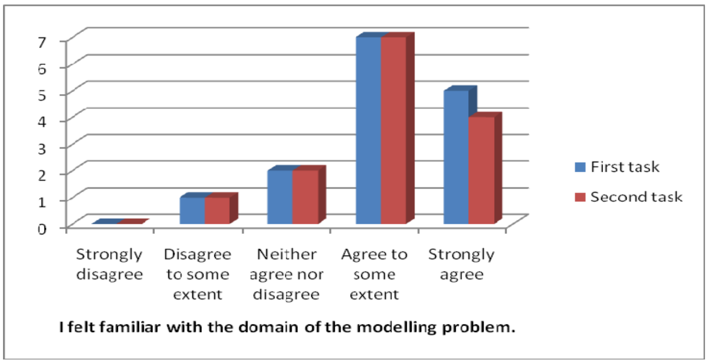

Figure A-11. Problem domain.

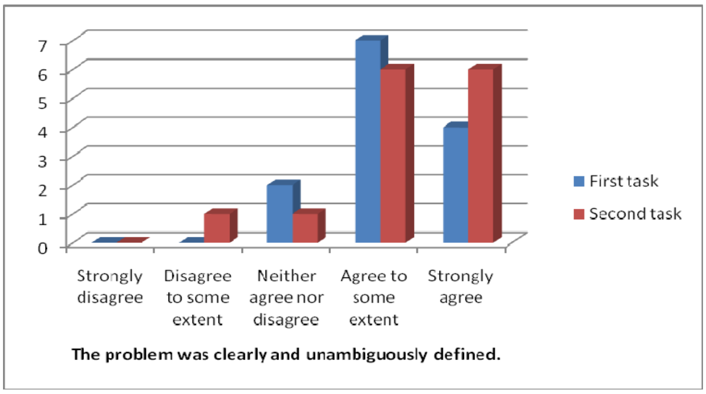

Figure A-12. Clear problem.

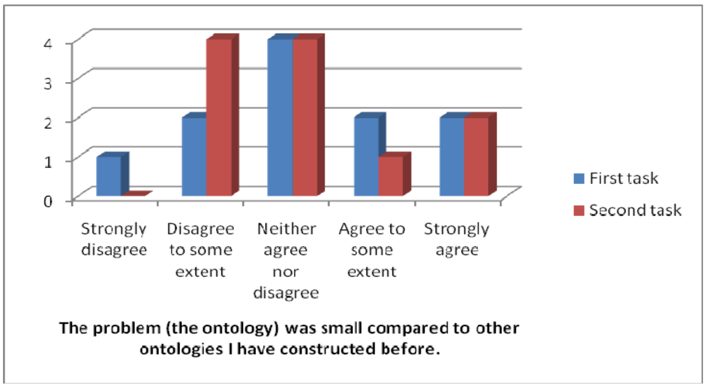

Figure A-13. Size of problem.

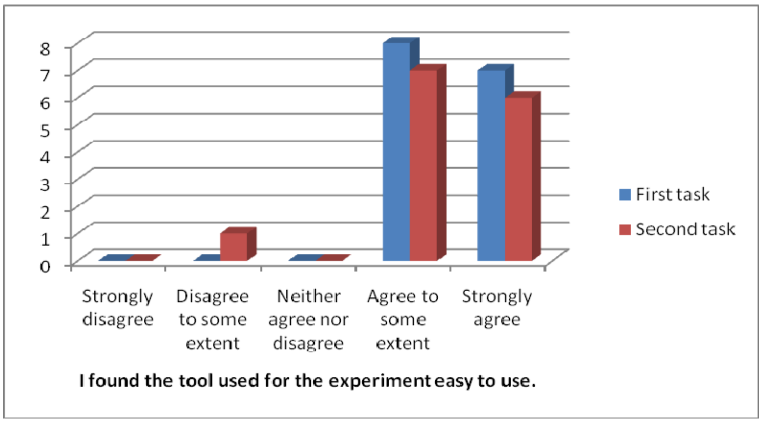

Figure A-14. Tool, ease of use.

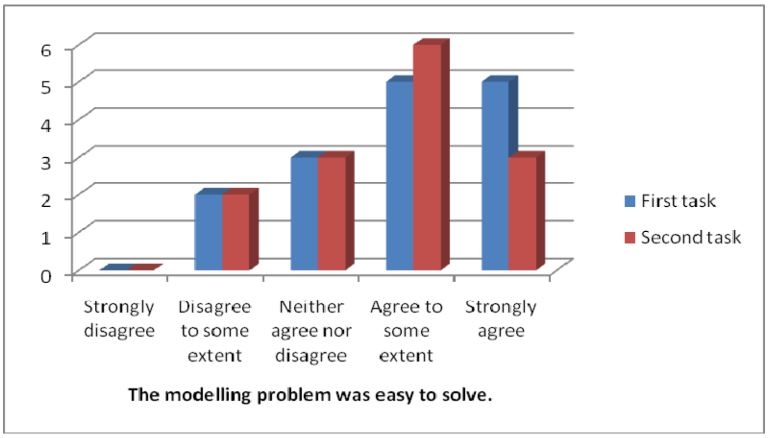

Figure A-15. Easy problem.

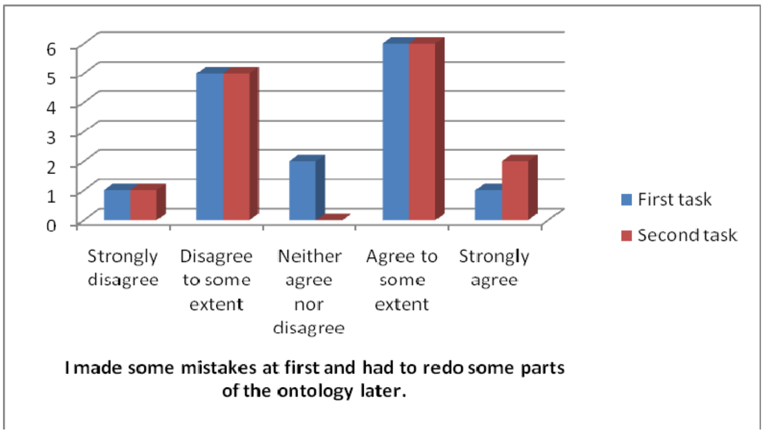

Figure A-16. Remodelling.

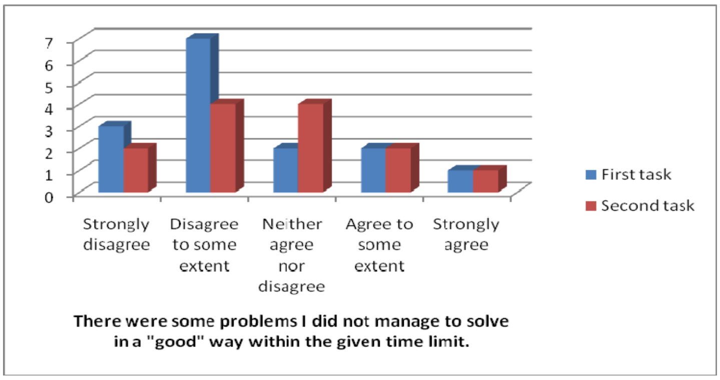

Figure A-17. Remaining problems at time limit.

As a comparison, some data was also collected after the first pattern exercise session, between the first and the second tasks of the experiment. A summary of the same questions as presented above, but for the initial pattern experience (where the first task was redone using patterns) can be seen in Table A-3.

Table A-3. Answers after the first pattern exercise.

	Strongly disagree	Disagree to some extent	Neither agree nor disagree	Agree to some extent	Strongly agree
I felt more familiar with the domain of the modelling problem now than the first time.	**1**	**2**	**3**	**4**	**5**
I found the tool used for the experiment easy to use.	**0**	**3**	**1**	**7**	**4**
The modelling problem was easy to solve.	**0**	**1**	**6**	**5**	**3**
I made some mistakes at first and had to redo some parts of the ontology later.	**0**	**2**	**3**	**6**	**4**
There were some problems I did not manage to solve in a good way within the given time limit.	**0**	**2**	**4**	**1**	**8**

Regarding the actual usage of patterns the subjects were asked how they had used the patterns, the results are shown in Figure A-18.

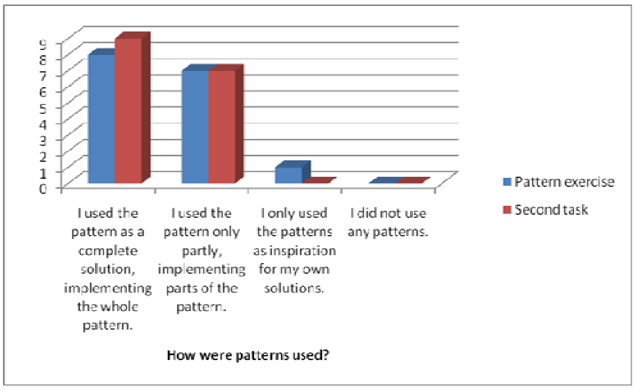

Figure A-18. How patterns were used.

When introducing the patterns, some questions were asked with respect to the patterns themselves and their usage and usefulness. The following propositions were given (both after the initial pattern exercise and after the second task where patterns were also used) and the subjects were again asked to rate them on the same scale as previously (results may be seen in the referenced diagrams):

1. The patterns were clear and easy to understand (see Figure A-19).
2. The tutorial/course material presented before this exercise was useful for understanding the patterns (see Figure A-20).
3. The patterns were easy to use (see Figure A-21).
4. Some of the patterns were "obvious" and trivial (see Figure A-22).
5. Some of the patterns introduced useful solutions that I did not think of before looking at the pattern (see Figure A-23).
6. In general, I found the patterns useful (see Figure A-24).

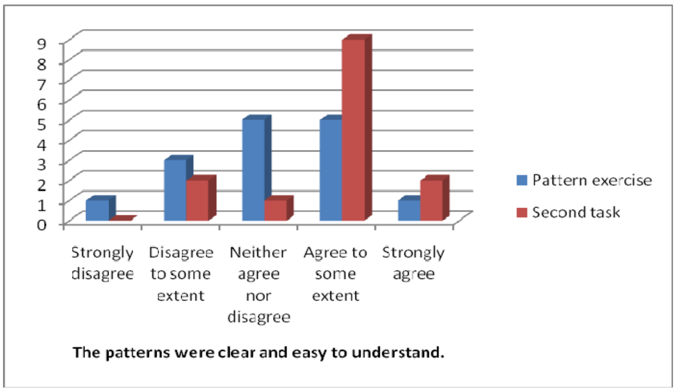

Figure A-19. Understanding the patterns.

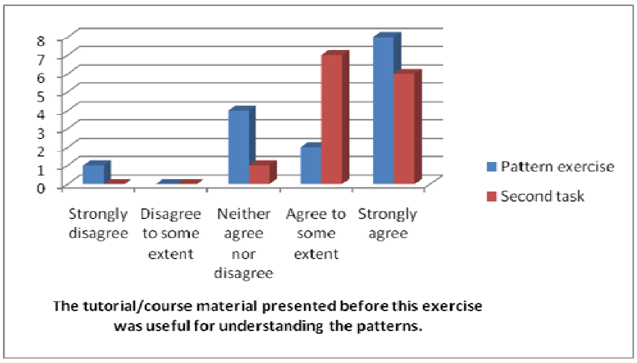

Figure A-20. Usefulness of tutorial.

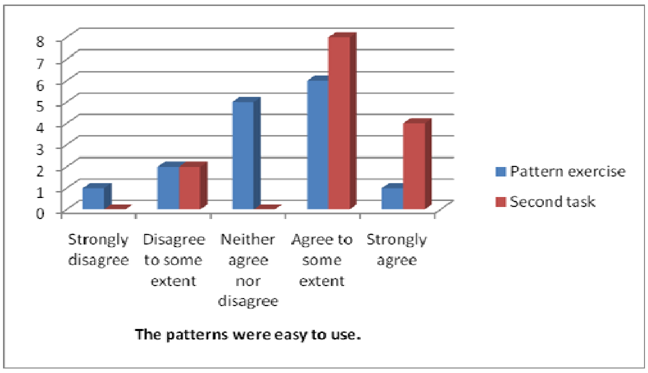

Figure A-21. Ease of use.

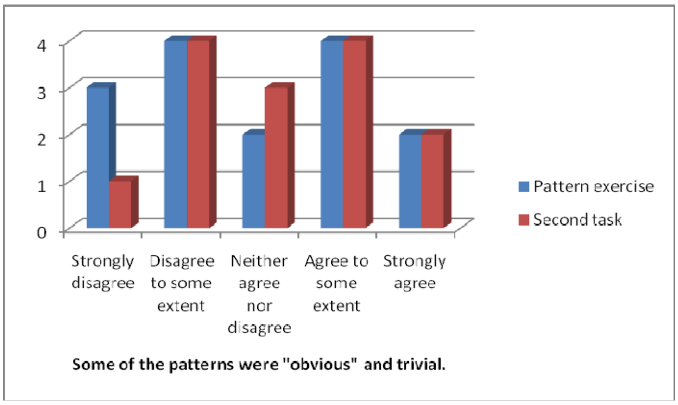

Figure A-22. Trivial patterns.

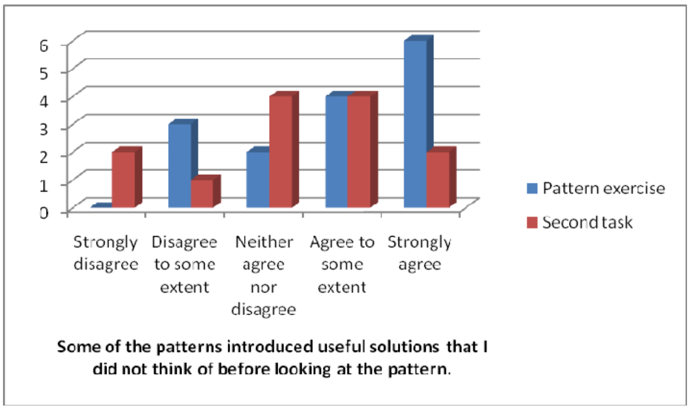

Figure A-23. Novel solutions in patterns.

Below, in Figure A-25, the comparison between the first task and the first pattern exercise, where the same task was remodelled using patterns, id presented. Note that the answers are the subjects' own opinions, i.e. the perceived usefulness of the patterns rather than some objective truth.

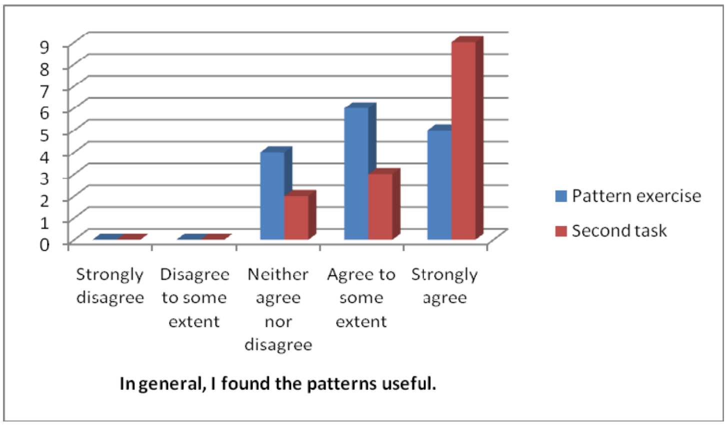

Figure A-24. Are patterns useful?

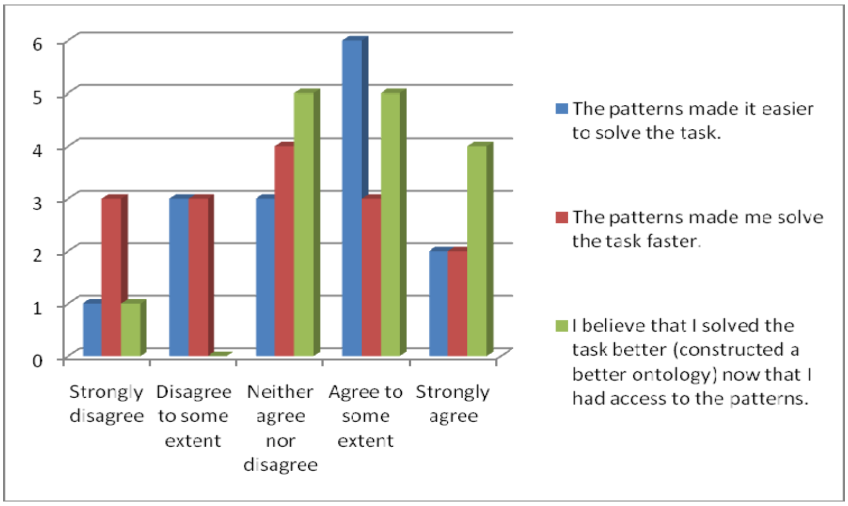

Figure A-25.In what respect are patterns useful?

With respect to the analysis of the actual ontologies, first the coverage of the problem was calculated (as defined in the experiment setup). Below the coverage values of the ontologies for the first and the second task can be viewed in Figures A-26 and A-27, respectively.

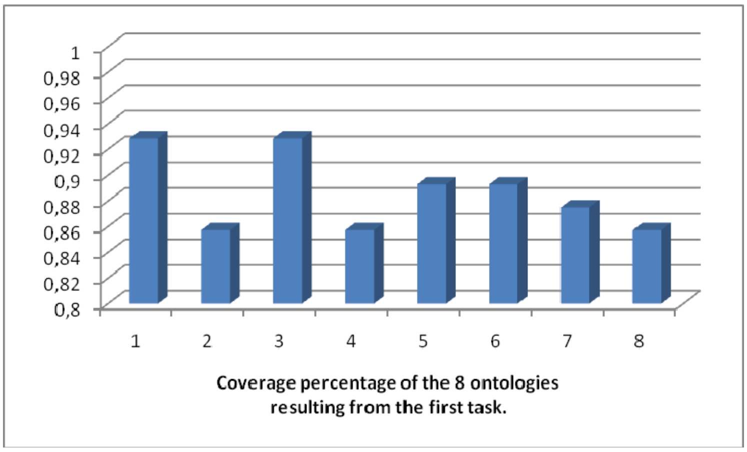

Figure A-26. Coverage of ontologies (first task).

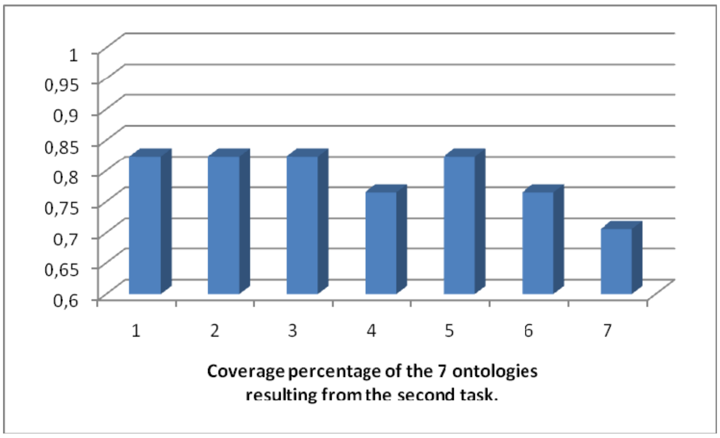

Figure A-27. Coverage of ontologies (second task).

Next, the usability criteria were analyzed. The results were already presented in the result summary previously in this report, but are repeated here in Figures A-28 and A-29 for completeness. The variables have been assessed on a scale from 'none', via 'some' and 'most' to 'all', where none mean that no such features were included in the ontology, some mean that it was included in a few places (less than two thirds of the possible places), most mean that it was included in most places where applicable (more than two thirds of the possible places), and all mean that it was included in every place where applicable.

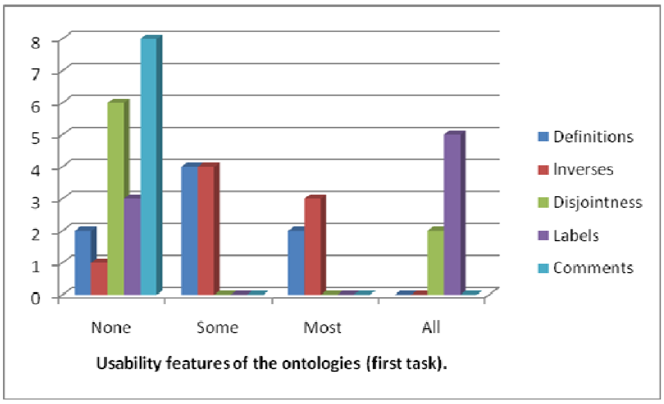

Figure A-28. Usability features, results from first task.

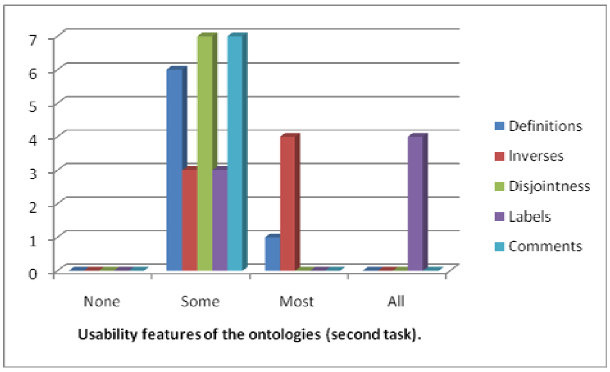

Figure A-29. Usability features, results from second task.

Dedicated experiment session – Jönköping

The background of the subjects was first recorded through one questionnaire. Below the results from this questionnaire may be seen.

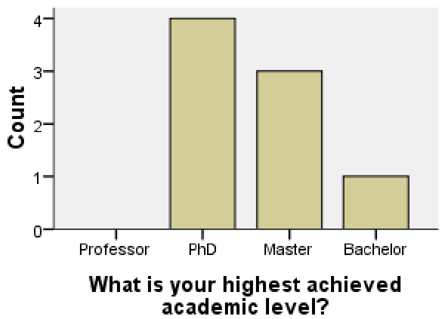

Figure A-30. The academic level of the subjects.

Topics of their studies contained variations of Computer Science and technology (3 subjects), Information Technology (3 subjects), Ontology Matching (1 subject) and Product Development (5 subjects).

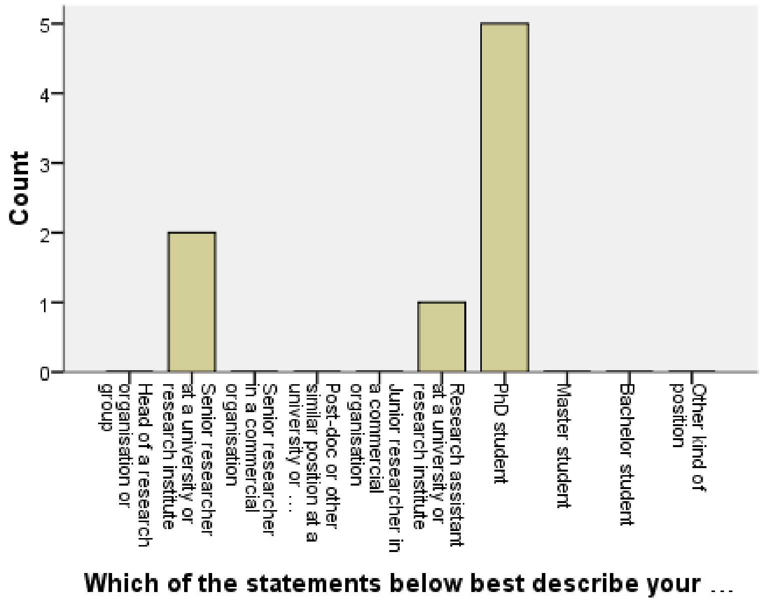

Figure A-31. The current positions of the subjects.

The current positions of the subjects can be seen in Figure A-31, and are mainly PhD students and research assistants, although two senior researchers also answered the question. The answers to the question if their current work is in some way related to ontologies are summarized in Figure A-32.

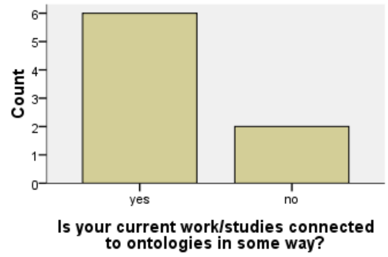

Figure A-32. Connection to ontologies.

The extent of their experience in working with ontologies is summarized in Figure A-33 and Figure A-34. Subsequently they were asked to characterize the ontologies they had constructed (if any), in terms of size and complexity, and their responses can be viewed in Figure A-35 and A-36.

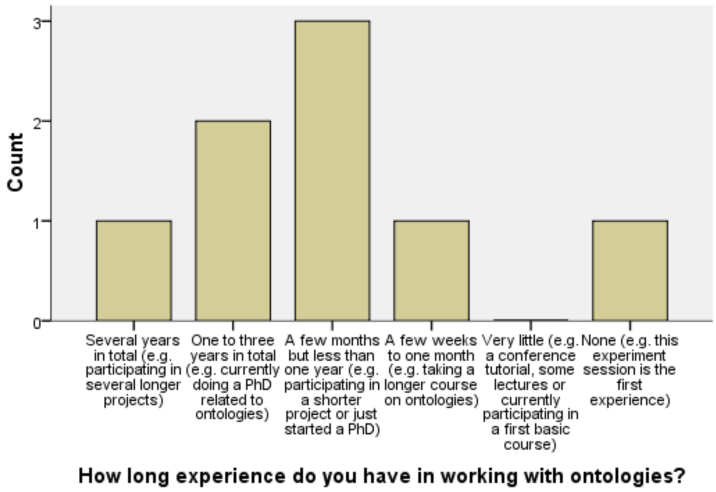

Figure A-33. Experience in terms of amount of time.

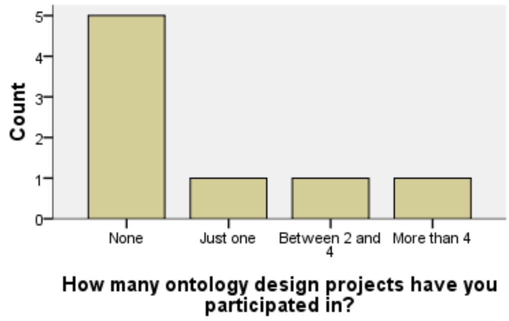

Figure A-34. Experience in terms of number of ontologies.

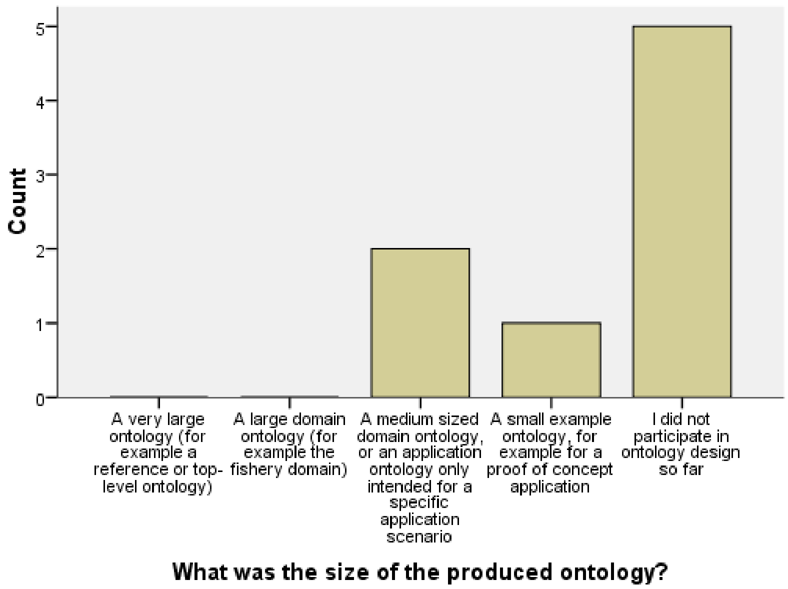

Figure A-35. Constructed ontologies (if any) in terms of size.

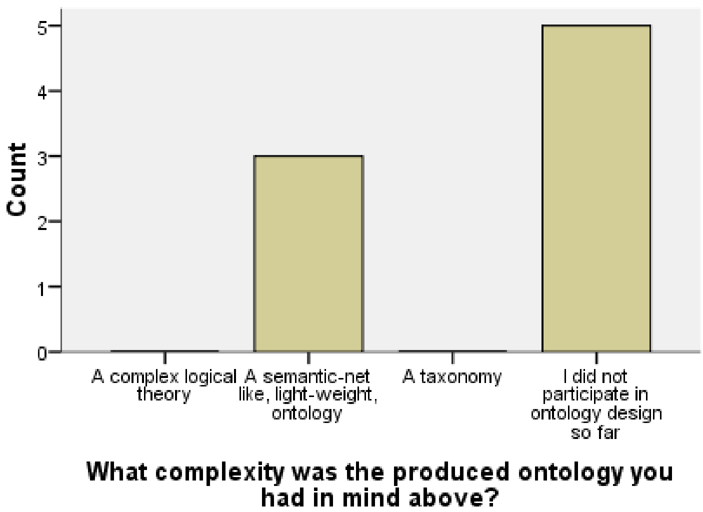

Figure A-36. Constructed ontologies (if any) in terms of complexity.

Next, they were asked to specify their knowledge and experience with respect to a set of tools (see data in Table A-4) and languages (see Table A-5), used for ontology engineering and modelling in general.

Table A-4. Tool experience.

What tools did you use?

	Did not try it	Tried it	Have extensive experience
Protégé 3.x	**2**	**4**	**2**
Protégé 4	**6**	**1**	**1**
TopBraid Composer	**5**	**2**	**0**
OntoEdit or OntoStudio	**3**	**4**	**0**
KAON	**5**	**2**	**0**
NeOn toolkit	**7**	**0**	**0**
SWOOP	**7**	**0**	**0**
Other	**0**	**2**	**0**

Table A-5. Language experience.

What languages did you use?

	Not familiar	Somewhat familiar	Have extensive experience
First-order logic	**0**	**3**	**2**
ER-diagrams	**0**	**3**	**3**
UML	**0**	**5**	**2**
RDF	**2**	**5**	**0**
OWL	**1**	**5**	**0**

Finally, for the last question of the background questionnaire, the question of previous usage of ontology design patterns, all of the subjects responded that they had not previously used patterns.

For the following experiment sessions the first set of questions were identical between both tasks, and the detailed resutls can be seen in the referenced diagrams:

1. I found the problem description easy to understand (Figure A-37).
2. I felt familiar with the domain of the modelling problem (Figure A-38).
3. The problem was clearly and unambiguously defined (Figure A-39).
4. The modelling problem (the ontology) was small compared to other ontologies I have constructed before (Figure A-40).
5. I found the tool used for the experiment easy to use (Figure A-41).
6. The modelling problem was easy to solve (Figure A-42).
7. I made some mistakes at first and had to redo some parts of the ontology later (Figure A-43).
8. There were some problems that I did not manage to solve in a “good” way within the given time limit (Figure A-44).

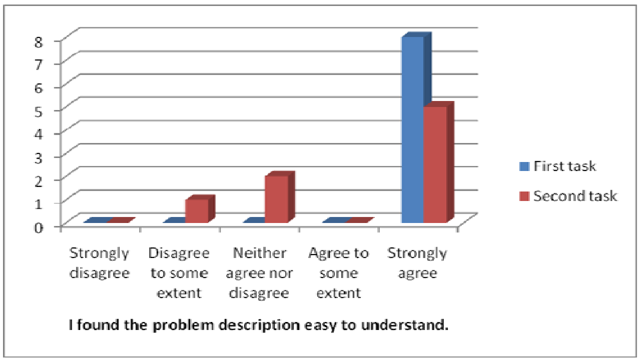

Figure A-37. Problem description.

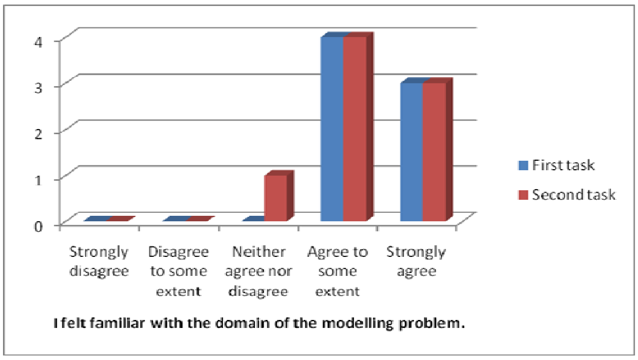

Figure A-38. Problem domain.

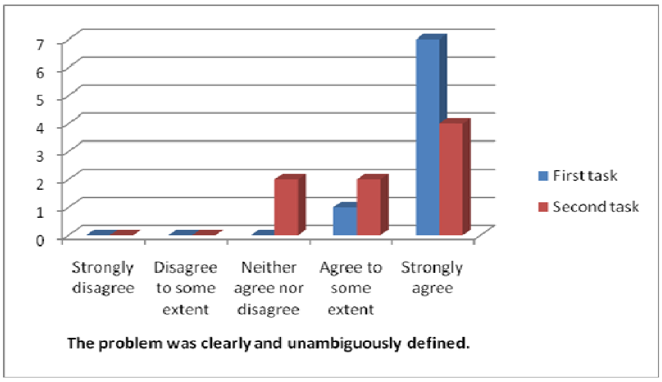

Figure A-39. Clear problem.

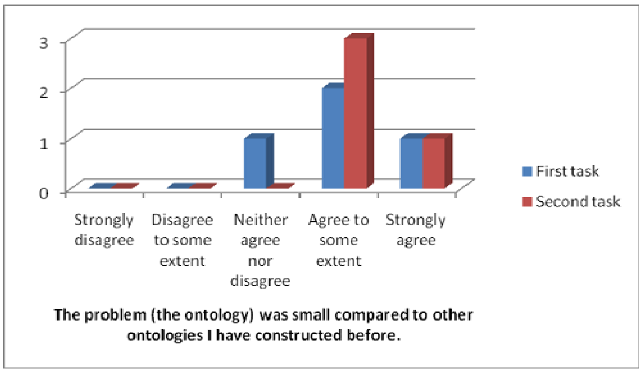

Figure A-40. Size of problem.

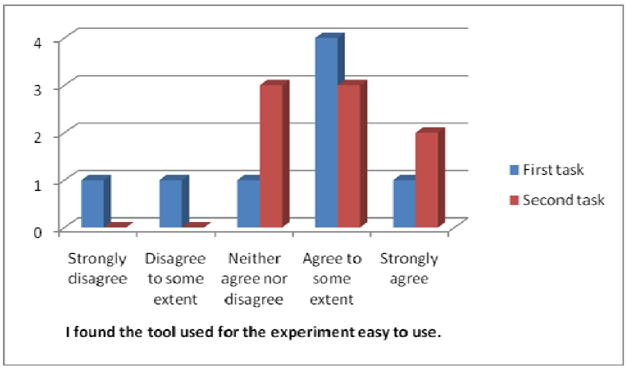

Figure A-41. Tool, ease of use.

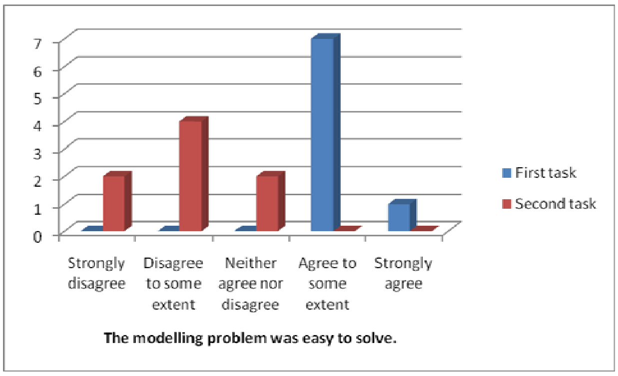

Figure A-42. Easy problem.

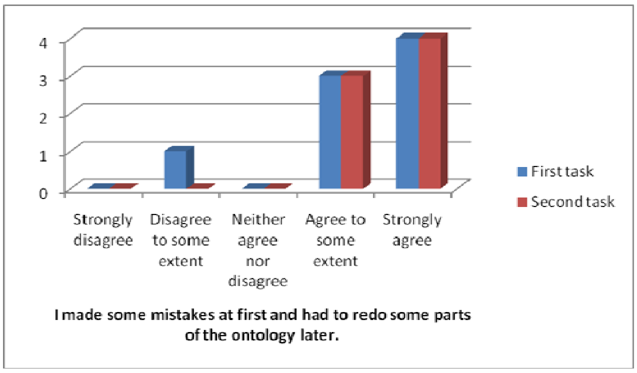

Figure A-43. Remodelling.

Figure A-44. Remaining problems at time limit.

Regarding the actual usage of patterns the subjects were asked how they had used the patterns, the results are shown in Figure A-45.

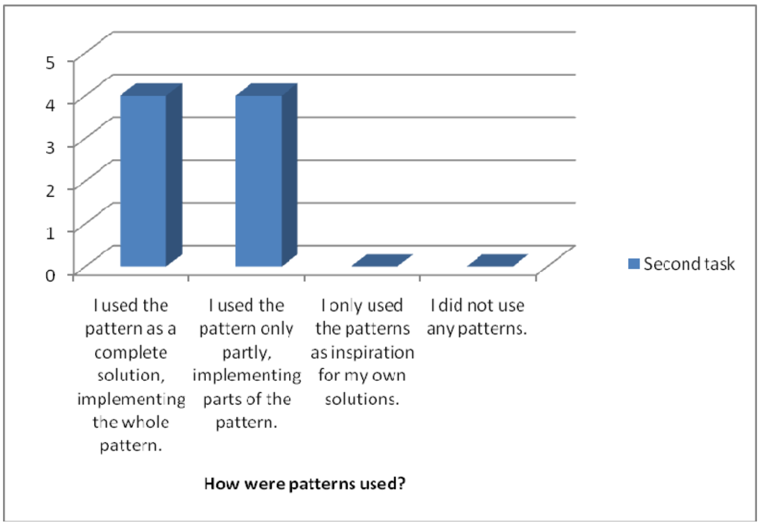

Figure A-45. How patterns were used.

When introducing the patterns some questions were asked with respect to the patterns themselves and their usage and usefulness. The following propositions were given after the second task and the subjects were again asked to rate them on the same scale as previously (their responses may be seen in Table A-6):

1. The patterns were clear and easy to understand.
2. The tutorial/course material presented before this exercise was useful for understanding the patterns.
3. The patterns were easy to use.
4. Some of the patterns were “obvious” and trivial.
5. Some of the patterns introduced useful solutions that I did not think of before looking at the pattern.
6. In general, I found the patterns useful.

Table A-6. Answers after the second task.

	Strongly disagree	Disagree to some extent	Neither agree nor disagree	Agree to some extent	Strongly agree
The patterns were clear and easy to understand.	**3**	**3**	**1**	**1**	**0**
The tutorial/course material presented before this exercise was useful for understanding the patterns.	**0**	**2**	**3**	**3**	**0**
The patterns were easy to use.	**3**	**1**	**3**	**1**	**0**
Some of the patterns were "obvious" and trivial.	**2**	**1**	**1**	**2**	**2**
Some of the patterns introduced useful solutions that I did not think of before looking at the patterns.	**1**	**1**	**2**	**4**	**0**
In general, I found the patterns useful.	**2**	**0**	**2**	**4**	**0**

Below, in Figure A-46, the comparison between the first and the second task is presented. Note that the answers are the subjects' own opinions, i.e. the perceived usefulness of the patterns rather than some objective truth.

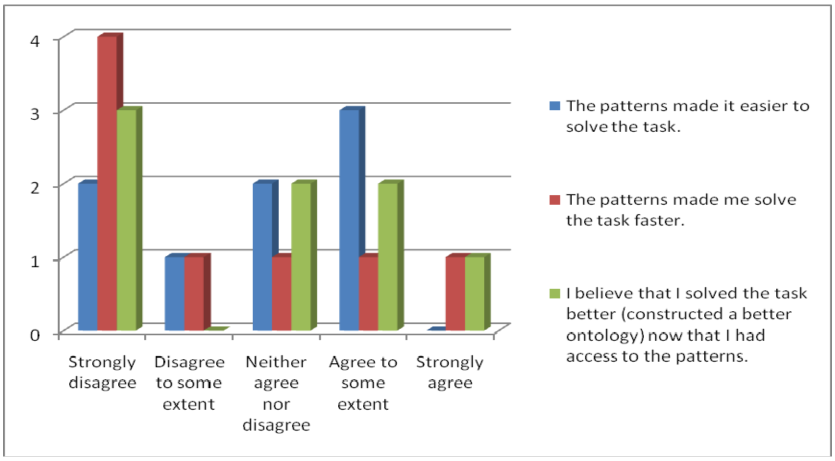

Figure A-46. In what respect are patterns useful?

With respect to the analysis of the actual ontologies, first the coverage of the problem was calculated (as defined in the experiment setup). Below the coverage values of the ontologies for the first and the second task can be viewed in Figures A-47 and A-48, respectively.

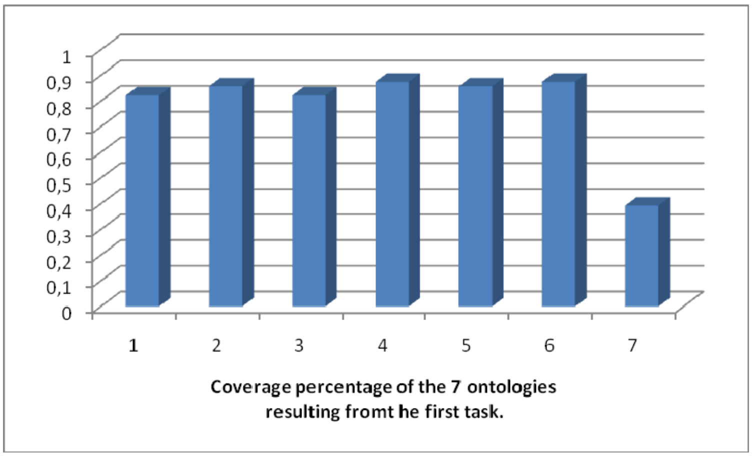

Figure A-47. Coverage of ontologies (first task).

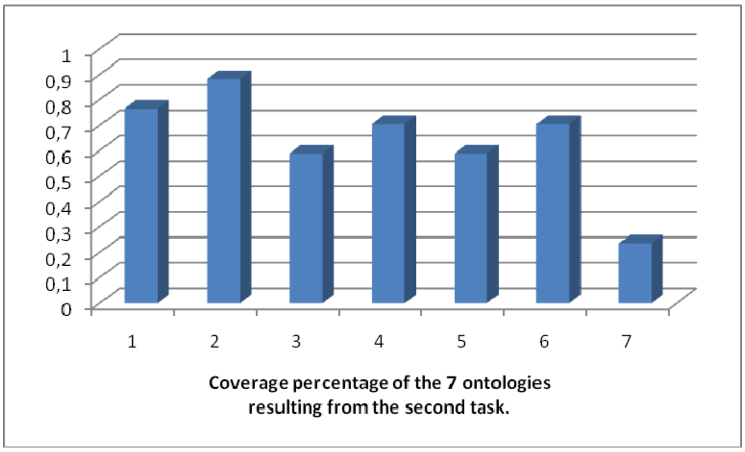

Figure A-48. Coverage of ontologies (second task).

Next, the usability criteria were analyzed. The results were already presented in the result summary previously in this report, but are repeated here in Figures A-49 and A-50 for completeness. The variables have been assessed on a scale from 'none', via 'some' and 'most' to 'all', where none mean that no such features were included in the ontology, some mean that it was included in a few places (less than two thirds of the possible places), most mean that it was included in most places where applicable (more than two thirds of the possible places), and all mean that it was included in every place where applicable.

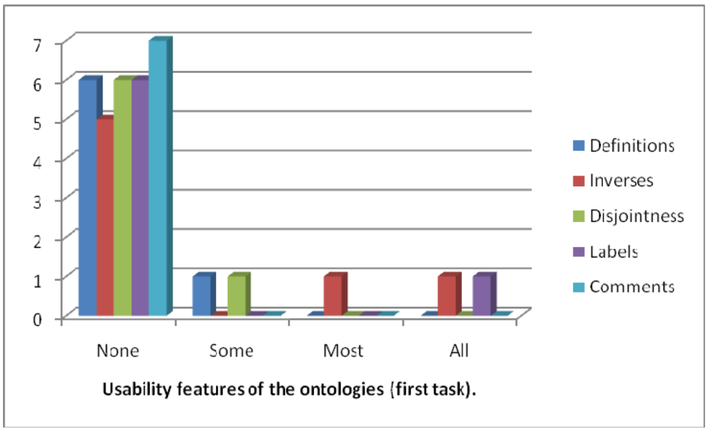

Figure A-49. Usability features, results from the first task.

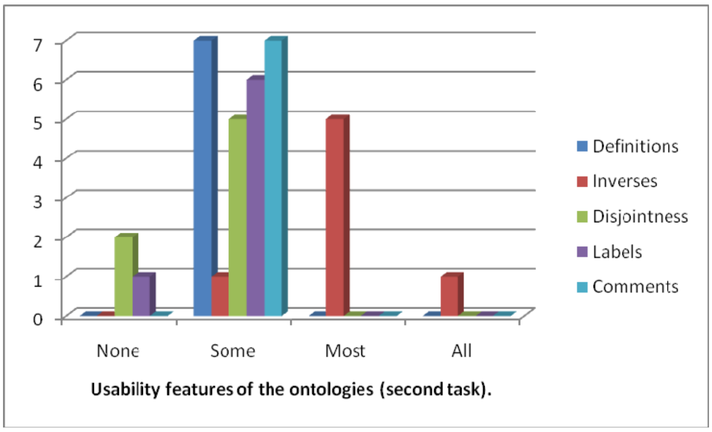

Figure A-50. Usability features, results from the second task.

Master's course – Jönköping

The background of the participants was first recorded through one questionnaire. Below the results from this questionnaire may be seen.

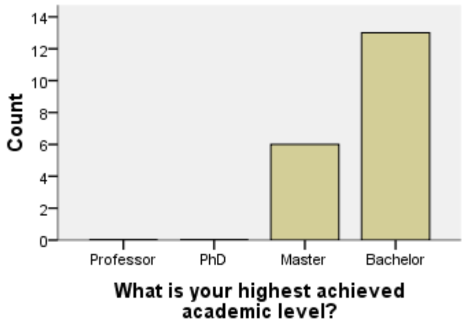

Figure A-51. The academic level of the participants.

Topics of their studies contained variations of computer science and engineering (12 subjects), information engineering (4 subjects), IT and IT management (2 subjects). All were master students except one bachelor student, and all except one claimed to have work or studies in some way related to ontologies. The extent of their experience in working with ontologies is summarized in Figure A-52 and Figure A-52. Subsequently they were asked to characterize the ontologies they had constructed (if any), in terms of size and complexity, and their responses can be viewed in Figure A-54 and A-55.

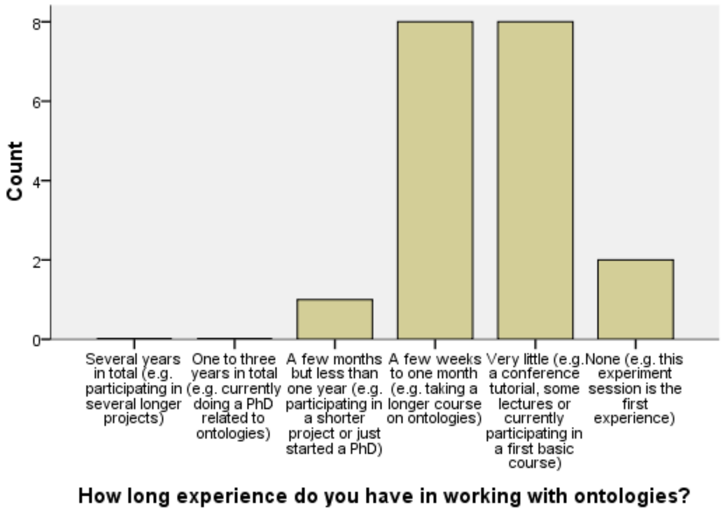

Figure A-52. Experience in terms of amount of time.

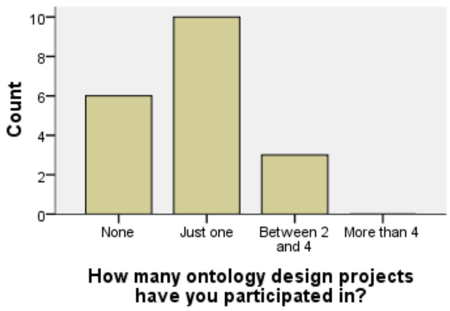

Figure A-53. Experience in terms of number of ontologies.

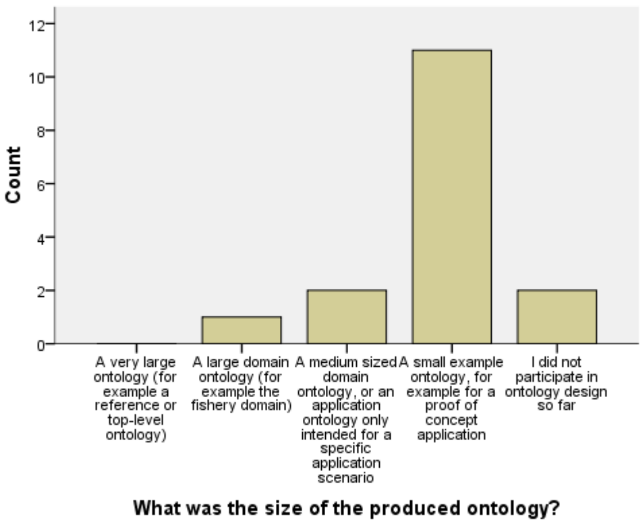

Figure A-54. Constructed ontologies (if any) in terms of size.

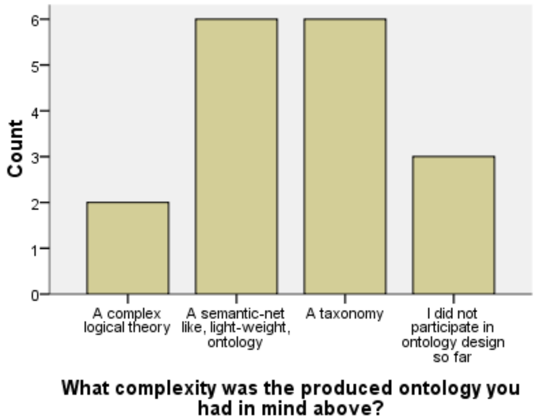

Figure A-55. Constructed ontologies (if any) in terms of complexity.

Next, they were asked to specify their knowledge and experience with respect to a set of tools (see data in Table A-7) and languages (see Table A-8), used for ontology engineering and modelling in general.

Table A-7. Tool experience.

What tools did you use?

	Did not try it	Tried it	Have extensive experience
Protégé 3.x	**4**	**14**	**0**
Protégé 4	**10**	**1**	**0**
TopBraid Composer	**6**	**6**	**0**
OntoEdit or OntoStudio	**9**	**1**	**0**
KAON	**10**	**0**	**0**
NeOn toolkit	**10**	**0**	**0**
SWOOP	**10**	**0**	**0**
Other	**10**	**0**	**0**

Table A-8. Language experience.

What languages did you use?

	Not familiar	Somewhat familiar	Have extensive experience
First-order logic	**2**	**9**	**0**
ER-diagrams	**2**	**12**	**3**
UML	**0**	**14**	**4**
RDF	**4**	**5**	**1**
OWL	**4**	**8**	**1**
Other	**9**	**0**	**0**

Finally, to the question of previous usage of ontology design patterns (the last question in the background questionnaire) all of the subjects responded that they had not previously used patterns. For the following experiment sessions the first set of questions were identical between both tasks, and the detailed resutls can be seen in the referenced diagrams:

1. I found the problem description easy to understand (Figure A-56).
2. I felt familiar with the domain of the modelling problem (Figure A-57).
3. The problem was clearly and unambiguously defined (Figure A-58).
4. The modelling problem (the ontology) was small compared to other ontologies I have constructed before (Figure A-59).
5. I found the tool used for the experiment easy to use (Figure A-60).
6. The modelling problem was easy to solve (Figure A-61).
7. I made some mistakes at first and had to redo some parts of the ontology later (Figure A-62).
8. There were some problems that I did not manage to solve in a "good" way within the given time limit (Figure A-63).

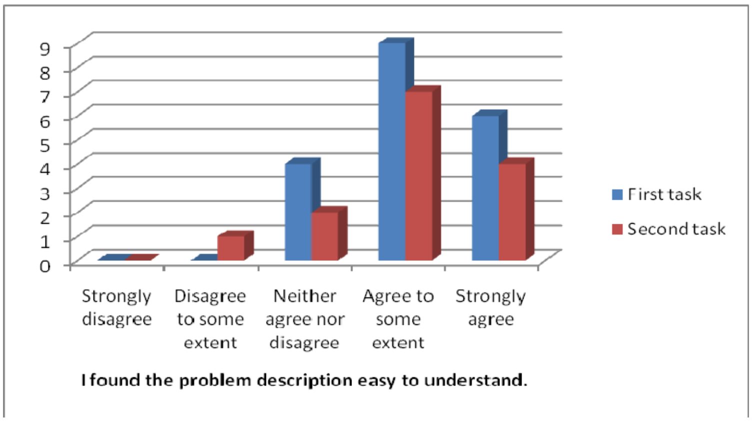

Figure A-56. Problem description.

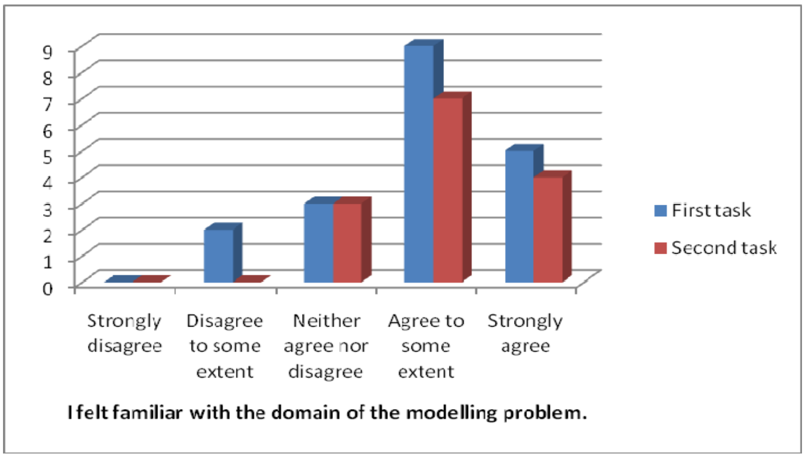

Figure A-57. Problem domain.

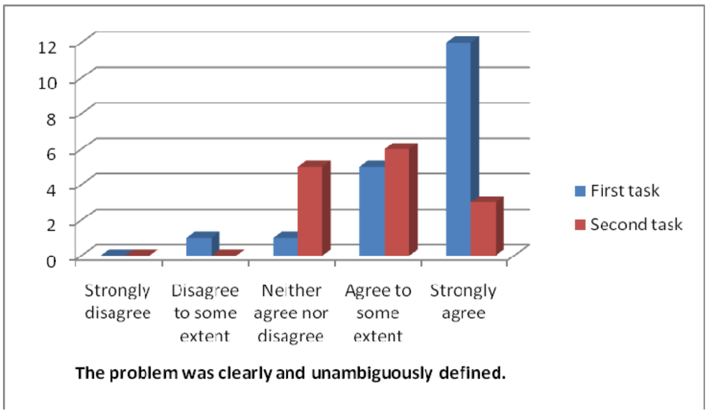

Figure A-58. Clear problem.

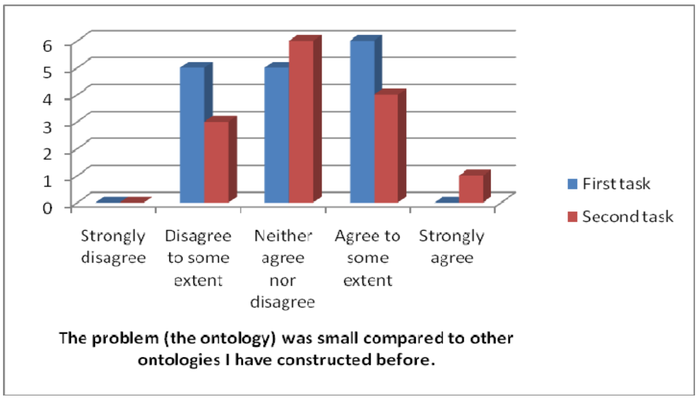

Figure A-59. Size of problem.

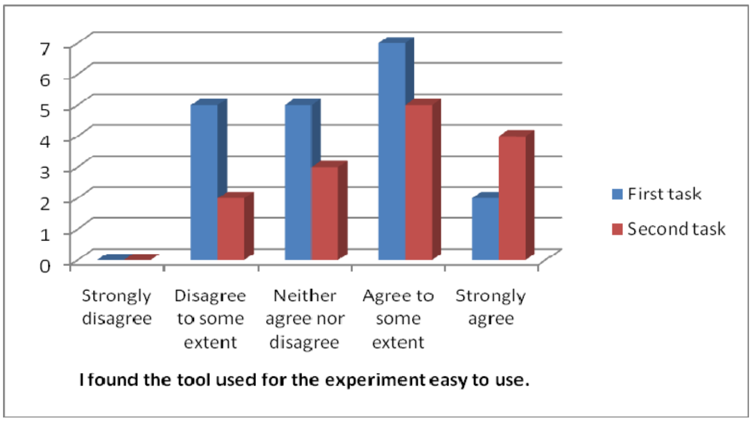

Figure A-60 Tool, ease of use.

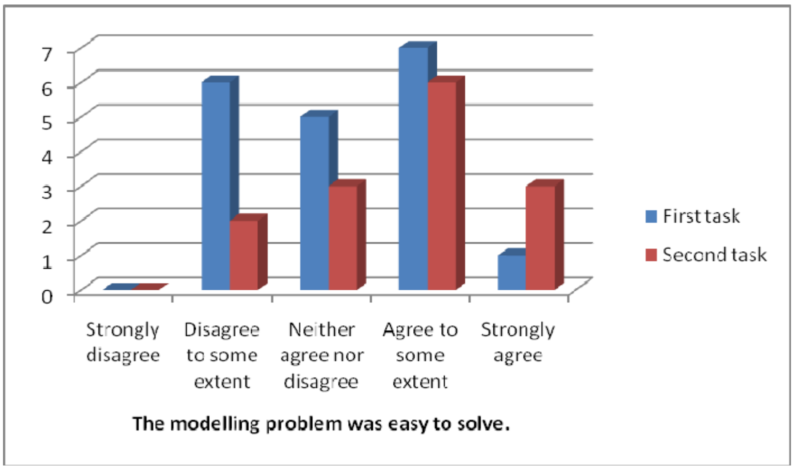

Figure A-61. Easy problem.

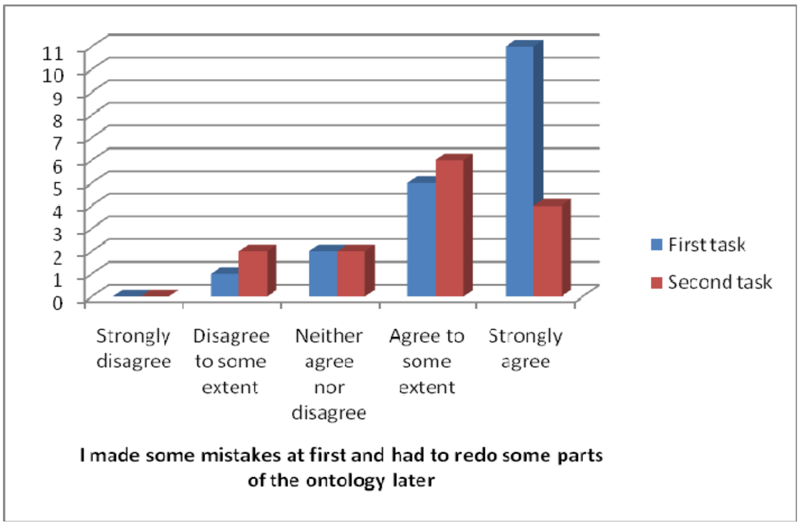

Figure A-62. Remodelling.

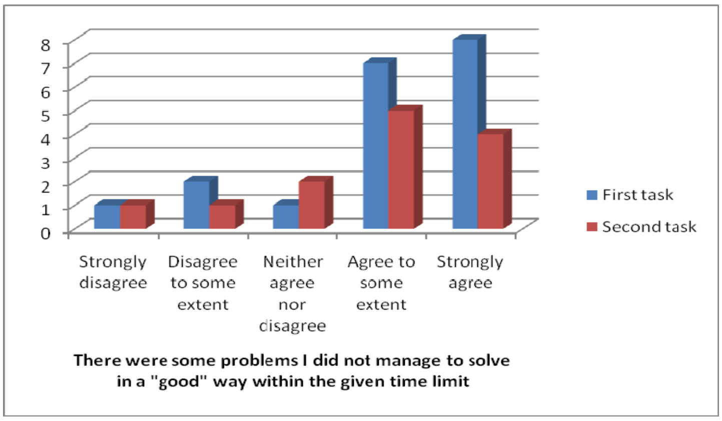

Figure A-63. Remaining problems at time limit.

As a comparison some data was also collected after the first pattern exercise session, between the first and the second tasks of the experiment. A summary of the same questions as presented above, but for the initial pattern experience (where the first task was redone using patterns) can be seen in Table A-9.

Table A-9. Answers after pattern exercise.

	Strongly disagree	Disagree to some extent	Neither agree nor disagree	Agree to some extent	Strongly agree
I felt more familiar with the domain of the modelling problem now than the first time.	**0**	**0**	**3**	**7**	**5**
I found the tool used for the experiment easy to use.	**0**	**0**	**3**	**9**	**3**
The modelling problem was easy to solve.	**1**	**4**	**6**	**1**	**3**
I made some mistakes at first and had to redo some parts of the ontology later.	**0**	**0**	**5**	**7**	**3**
There were some problems I did not manage to solve in a good way within the given time limit.	**1**	**1**	**4**	**4**	**5**

Regarding the actual usage of patterns the subjects were asked how they had used the patterns, the results are shown in Figure A-64.

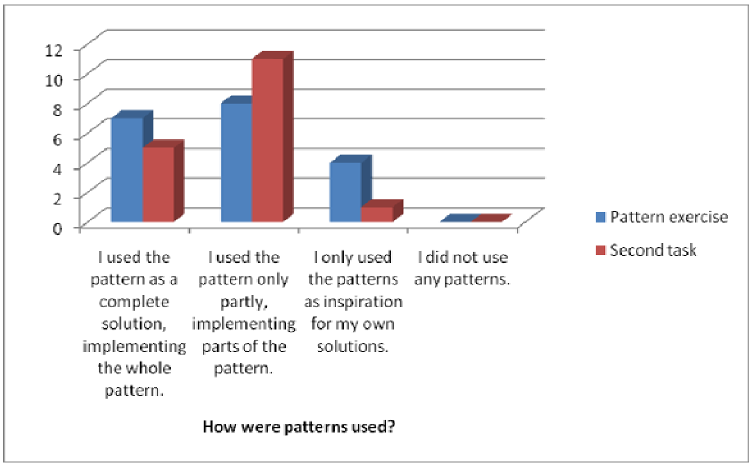

Figure A-64. How patterns were used.

When introducing the patterns some questions were asked with respect to the patterns themselves and their usage and usefulness. The following propositions were given (both after the initial pattern exercise and after the second task where patterns were also used) and the subjects were again asked to rate them on the same scale as previously (results may be seen in the referenced diagrams):

1. The patterns were clear and easy to understand (see Figure A-65).
2. The tutorial/course material presented before this exercise was useful for understanding the patterns (see Figure A-66).
3. The patterns were easy to use (see Figure A-67).
4. Some of the patterns were “obvious” and trivial (see Figure A-68).
5. Some of the patterns introduced useful solutions that I did not think of before looking at the pattern (see Figure A-69).
6. In general, I found the patterns useful (see Figure A-70).

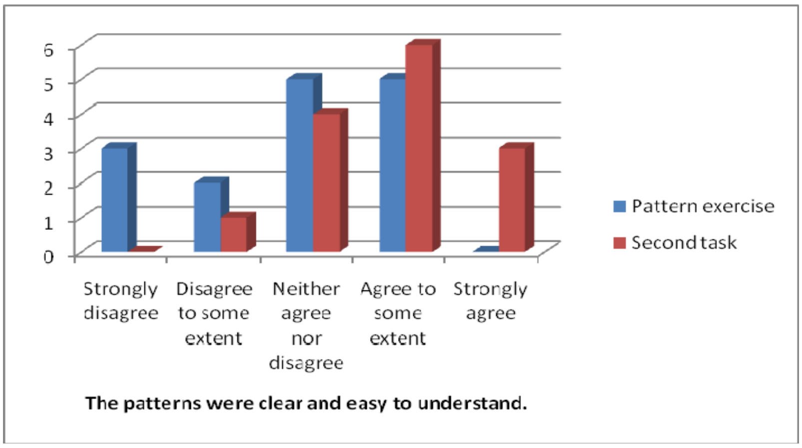

Figure A-65. Understanding the patterns.

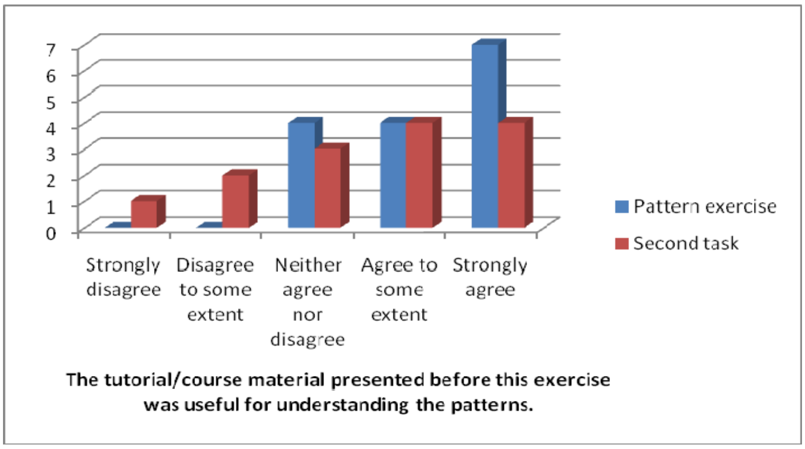

Figure A-66. Usefulness of tutorial.

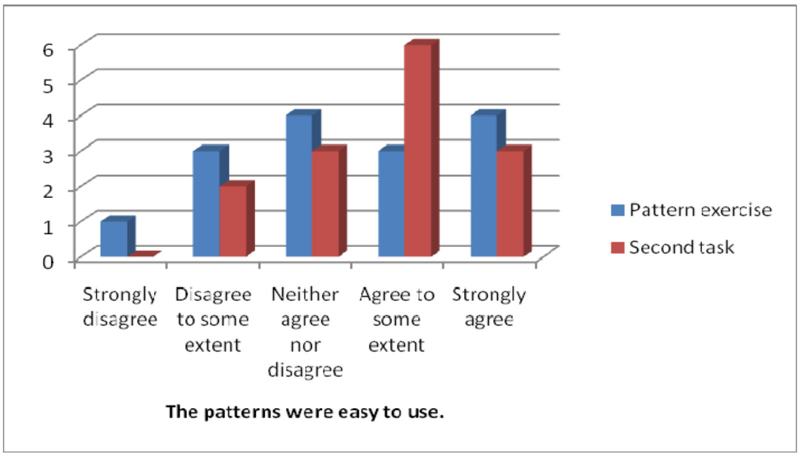

Figure A-67. Ease of use.

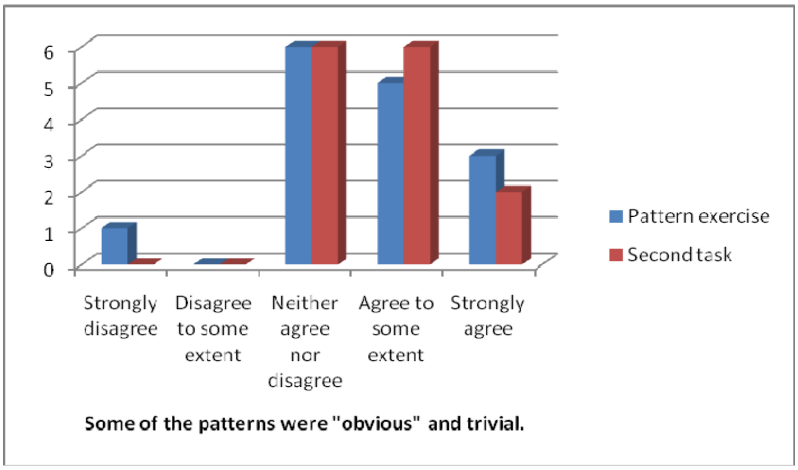

Figure A-68. Trivial patterns.

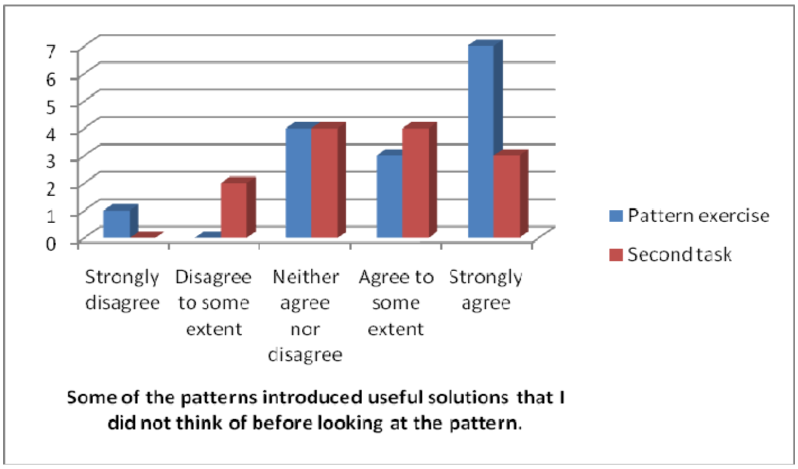

Figure A-69. Novel solutions in patterns.

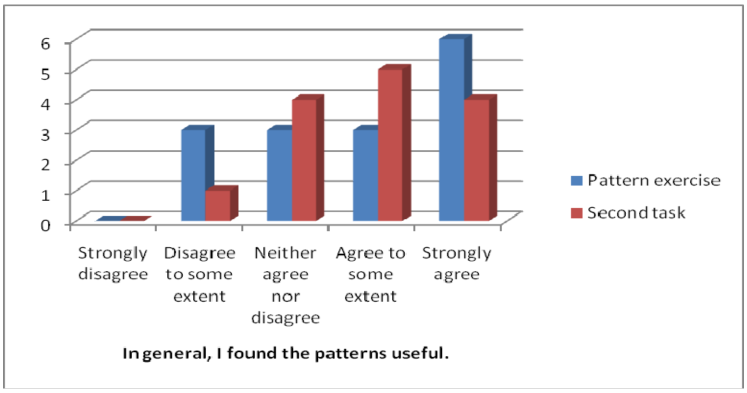

Figure A-70. Are patterns useful?

Below, in Figure A-71, the comparison between the first task and the first pattern exercise, where the same task was remodelled using patterns, is presented. Note that the answers are the subjects' own opinions, i.e., the perceived usefulness of the patterns rather than some objective truth.

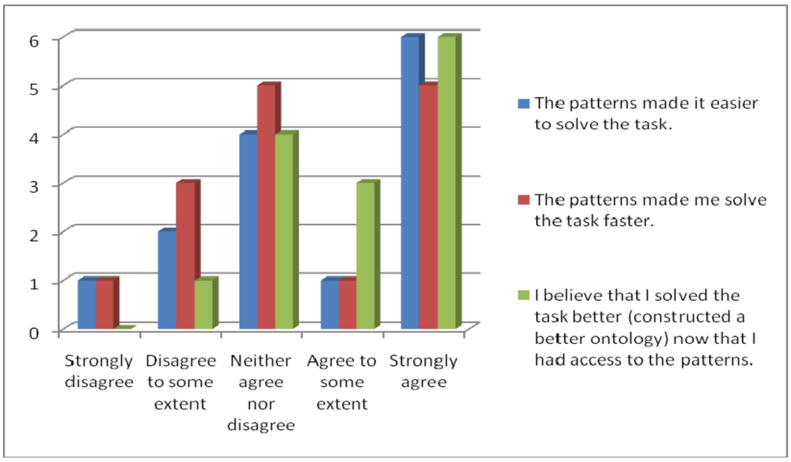

Figure A-71. In what respect are patterns useful?

With respect to the analysis of the actual ontologies, first the coverage of the problem was calculated (as defined in the experiment setup). Below the coverage values of the ontologies for the first and the second task can be viewed in Figures A-72 and A-73, respectively.

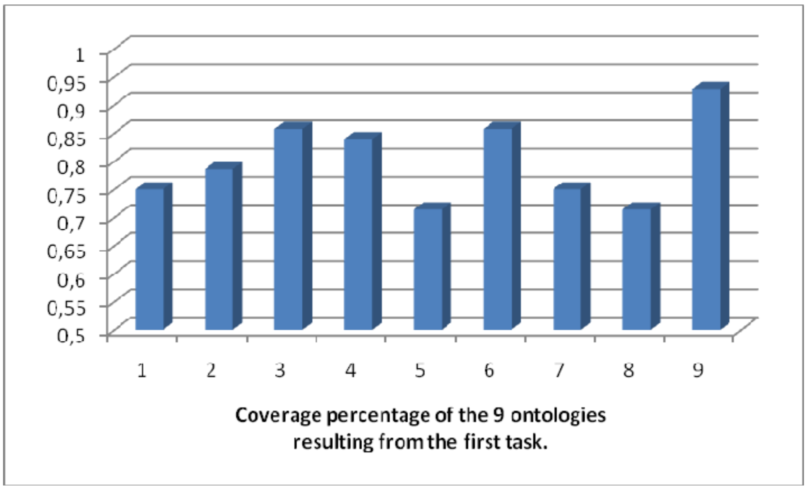

Figure A-72. Coverage of ontologies (first task).

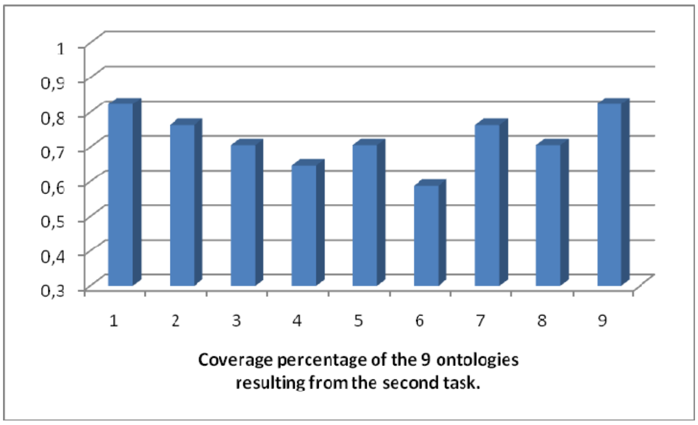

Figure A-73. Coverage of ontologies (second task).

Next, the usability criteria were analyzed. The results were already presented in the result summary previously in this report, but are repeated here in Figures A-74 and A-75 for completeness. The variables have been assessed on a scale from 'none', via 'some' and 'most' to 'all', where none mean that no such features were included in the ontology, some mean that it was included in a few places (less than two thirds), most mean that it was included in most places where applicable (more than two thirds), and all mean that it was included in every place where applicable.

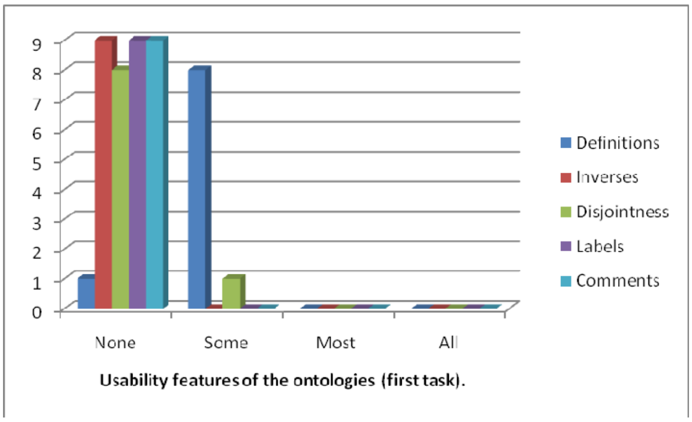

Figure A-74. Usability features, results from the first task.

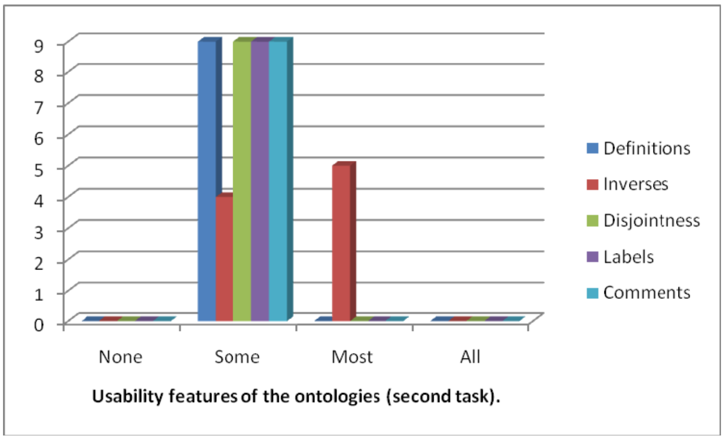

Figure A-75. Usability features, results from the second task.

Appendix 2 – Material Used in Ontology (Requirement) Specification Experiment, User Study #1

Table A2-1. Instantiation 1 of the Preliminary Methodological Guidelines for Ontology (Requirement) Specification

Guideline-1: Sequence of tasks for the ontology (requirement) specification activity, including techniques and tools.

1. Identify Purpose and Scope for the ontology.
2. Identify Intended Uses.
3. Identify Intended Users.
4. Gather requirements for the ontology.
    a. Technique: Exploit scenarios and use cases (using templates, which you should create).
    b. Tool: You can use a wiki (if you want) for writing the informal requirements.
5. Write Competency Questions (CQs).
    a. Technique/Approach: Top-Down.
    b. Tool: MindMap Tool to write the CQs.
6. Group CQs.
    a. Technique: To choose between:
        i. Card Sorting.
        ii. Clustering NL sentences.
    b. Tool: MindMap Tool to represent the groups of CQs.
7. Validate CQs.
8. Extract Terminology and Frequency.
    a. Technique: To choose between:
        iii. Manual extract terms and count their appearance number in CQs.
        iv. Automatic: using terminology extraction and frequency techniques and tools.

Table A2-2. Instantiation 2 of the Preliminary Methodological Guidelines for Ontology (Requirement) Specification

Guideline-2: Sequence of tasks for the ontology (requirement) specification activity, including techniques and tools.

1. Identify Purpose and Scope for the ontology.
2. Identify Intended Uses.
3. Identify Intended Users.
4. Gather requirements for the ontology.
    b. Technique: Brainstorming.
    c. Tool: You can use a wiki (if you want) for writing the informal requirements.
5. Write Competency Questions (CQs).
    d. Technique/Approach: Bottom-Up.
    e. Tool: Excel for writing the CQs.
6. Group CQs.
    f. Technique: To choose between:
        i. Card Sorting.
        ii. Clustering NL sentences.
    g. Tool: Excel for represent the groups of CQs.
7. Validate CQs.
8. Give Priority to CQs.
9. Extract Terminology and Frequency.
    h. Technique: To choose between:
        iii. Manual extract terms and count their appearance number in CQs.
        iv. Automatic: using terminology extraction and frequency techniques and tools.

Table A2-3. Questionnaire about the Preliminary Methodological Guidelines for Ontology (Requirement) Specification

<u>Questionnaire</u>

Questionnaire about the proposed methodological guidelines to carry out the ontology (requirements) specification activity.

Based on the guidelines you followed (1 or 2), please answer each question in 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? If so, in which sense?, and in which tasks?
3. In your opinion, are this guidelines complete? If not, what is missing?
4. Do you think more techniques and tools should be provided?
5. Did you miss an integrated tool for carrying out the proposed tasks?
6. How we can improve the proposed guidelines?

Guidelines 1 and 2. Tasks 1-3:

1. How difficult was to carry out these tasks?
2. Did you expect more guidelines for carrying out these tasks? If so, what kind of guidelines did you miss? And in which tasks?

Guidelines 1 and 2. Task 4:

1. How difficult was to carry out this task?
2. Was the proposed technique useful for you? If not, please explain why.
3. Was the proposed tool useful for you? If not, please explain why.

Guidelines 1 and 2. Task 5:

1. How difficult was to carry out this task?
2. Was the proposed technique useful for you? If not, please explain why.
3. Was the proposed tool useful for you? If not, please explain why.

Guidelines 1 and 2. Task 6:

1. How difficult was to carry out this task?
2. Was the proposed technique useful for you? If not, please explain why.
3. Was the proposed tool useful for you? If not, please explain why.
4. If you chose the clustering NL sentences technique, how difficult was to apply such technique? Did you find tools for applying the technique?
   Please explain in detail how you carried out this task.
5. Do you think to group CQs is useful? If so, please explain cases in which you think the classification of CQs would be useful.

Guidelines 1 and 2. Task 7:

1. How difficult was to carry out this task?
2. Please explain in detail how you carried out the task, which criteria you used, and why you chose those criteria.

Guidelines-1. Task 8:

1. How difficult was to carry out this task?
2. Was the proposed technique useful for you? If not, please explain why.
3. If you chose the proposed automatic technique, how difficult was to apply such technique? Did you find tools for applying the technique?
   Please explain in detail how you carried out this task.
4. Do you think to extract the terminology and its frequency is useful? Please explain cases in which you think so.

Guidelines-2. Task 8:

1. How difficult was to carry out this task?
2. Please explain in detail how you carried out the task, which criteria you used, and why you chose those criteria.

Guidelines-2. Task 9:

1. How difficult was to carry out this task?
2. Was the proposed technique useful for you? If not, please explain why.
3. If you chose the proposed automatic technique, how difficult was to apply such technique? Did

you find tools for applying the technique?
Please explain in detail how you carried out this task.

4. Do you think to extract the terminology and its frequency is useful? Please explain cases in which you think so.

Appendix 3 – Material Used in Ontology (Requirement) Specification Experiment, User Study #2

Table A3-1. Questionnaire about the Methodological Guidelines for Ontology (Requirement) Specification

Questionnaire
Questionnaire about the proposed methodological guidelines to carry out the ontology (requirements) specification activity. Please answer each question in detail and be honest! Thank you very much. *General comments:* 1. Are the proposed guidelines well explained? 2. Is more detail needed in the guidelines? If so, please explain in detail in which sense and in which tasks. 3. In your opinion, are these guidelines complete? If not, what is missing? 4. Do you think more techniques and tools should be provided? 5. Did you miss an integrated tool for carrying out the proposed tasks? 6. How we can improve the proposed guidelines? 7. Did you find useful the ontology requirement guidelines? 8. Do you think you will use again the proposed guidelines for the ontology requirement specification? 9. Did you find useful to write the ontology specification before going into the ontology development? 10. Do you think you will create ontology (requirement) specifications?

Appendix 4 – Material Used in Ontology Lifecycle Establishment Experiment, User Study #1

Table A4-1. Questionnaire about the Methodological Guidelines for Establishing the Ontology (Network) Life Cycle

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

Appendix 5 – Material Used in Ontology Life Cycle Establishment Experiment, User Study #2

Table A5-1. Questionnaire about the Methodological Guidelines for Establishing the Ontology (Network) Life Cycle

Questionnaire

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.

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 optional 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?

General Comments:

1. Are the proposed guidelines well explained?
2. Is more detail needed in the guidelines?
3. How we can improve the proposed guidelines?
4. How the activity of establishing the life cycle for a concrete ontology network could be carried out faster?

Appendix 6 – Experiment Guide Subject Expert1

I. Setting-up the environment

- Start the NTK.
- Configure the Registry Properties: Open Preferences->Oyster Storage Preference and configure it with the following properties:
    - Super Node IP: ServerIP (provided at the experiment)
    - Push Node IP: blank (default)
    - Read Ontologies Locally: checked
- Start the Registry: Either from the Registry menu or by clicking the Registry icon on the toolbar. Note that the icon shape and the message on the Registry menu changes when the registry is running
- Identify to the system: Open Preferences->Collaborative Development Preference, go to the Register section and provide:
    - Your first name
    - Your last name
    - Choose your role
    - Click Register
- Create a new Ontology Development Project with the following properties:
    - Ontology Language: OWL
    - Datamodel Type: CollaborationServer
        - i. Host: ServerIP (provided at the experiment)
        - ii. Port: 8267 (default)
- Open existing ontology species_v1.0_model.owl into the project by selecting “Add Ontology To Project” in the project context menu.

II. Editorial Workflow

1. Start logging the ontology: Right-click the ontology and select “Log Changes”
2. Add Individual 31005_10000 (Species)
3. Add Individual 31005_10001 (Species)
4. Add Individual 31005_10000 DataProperty hasCodeAlpha3 value: DCR. Type: string
5. Add Individual 31005_10000 DataProperty hasID value: 10000. Type: string
6. Add Individual 31005_10000 DataProperty hasMeta value: 31005. Type: string
7. Add Individual 31005_10000 DataProperty hasNameEN value: Yellow-nosed albat. Type: string
8. Add Individual 31005_10000 DataProperty hasNameScientific value: Diomedea chlororhynchos. Type: string
9. Add Individual 31005_10001 DataProperty hasCodeAlpha3 value: PDM. Type: string
10. Add Individual 31005_10001 DataProperty hasID value: 10001. Type: string
11. Add Individual 31005_10000 DataProperty hasMeta value: 31005. Type: string

12. Add Individual 31005_10001 DataProperty hasNameEN value: Great-winged petre. Type: string
13. Add Individual 31005_10001 DataProperty hasNameScientific value: Pterodroma wrong macroptera. Type: string
14. Add Root Class Speciation
15. Add Individual Allopatric
16. Add Individual Peripatric
17. Add Individual ParapatricWrong
18. Add DataProperty hasSpeciation (Species)

III. Analysis of the changes/actions

To see the changes information, open the view “Change Log View”. Read and comment

To see the workflow actions, open the view “Draft View”. Read and comment.

IV. Submit your changes to be approved (Select them and click submit to be approved)

Wait for Validator to finish task II

V. Delete the rejected changes by selecting them and clicking delete button

VI. Open “Approved View” and submit to be deleted the following change:

- DataProperty hasSpeciation (Species)
    - Change Type: AddDataProperty
    - Related Entity: hasSpeciation
    - Author: SE1

Appendix 7 - Experiment Guide Subject Expert2

I. Setting-up the environment

- Start the NTK.
- Configure the Registry Properties: Open Preferences->Oyster Storage Preference and configure it with the following properties:
    - Super Node IP: ServerIP (provided at the experiment)
    - Push Node IP: blank (default)
    - Read Ontologies Locally: checked
- Start the Registry: Either from the Registry menu or by clicking the Registry icon on the toolbar. Note that the icon shape and the message on the Registry menu changes when the registry is running
- Identify to the system: Open Preferences->Collaborative Development Preference, go to the Register section and provide:
    - Your first name
    - Your last name
    - Choose your role
    - Click Register
- Create a new Ontology Development Project with the following properties:
    - Ontology Language: OWL
    - Datamodel Type: CollaborationServer
        i. Host: ServerIP (provided at the experiment)
        ii. Port: 8267 (default)
- Open existing ontology species_v1.0_model.owl into the project by selecting “Add Ontology To Project” in the project context menu.

II. Editorial Workflow

1. Start logging the ontology: Right-click the ontology and select “Log Changes”
2. Add SubClass genus of biological_entity. Right Click biological_entity and add Class genus.
3. Add Root Class Person
4. Add DataProperty name (Person)
5. Add ObjectProperty hasScientificNameAuthor
6. Add ObjectProperty domain (hasScientificNameAuthor,Species)
7. Add ObjectProperty range (hasScientificNameAuthor,Person)
8. Add Root Class Category
9. Add DataProperty description (Category)
10. Add Individual Extinct (Category)
11. Add Individual Endangered (Category)

12. Add Individual Extint DataPropertyValue (description, the last remaining member of the species has died, string,--)
13. Add ObjectProperty hasCategory.
14. Add ObjectPropertyRange (hasCategory,Category)
15. Add ObjectPropertyDomain (hasCategory,Species)

Wait for SE1 to finish task II

16. Add Species Class Super Restriction (AT_LEAST/MIN, 1, hasScientificNameAuthor, ---)
17. Add Species Class Super Restriction (EXACTLY/CARD, 1, hasCategory, Thing)
18. Add Species Class Super Restriction (HAS_VALUE, hasMeta, “31005”)

III. Analysis of the changes/actions

To see the changes information, open the view “Change Log View”. Read and comment

To see the workflow actions, open the view “Draft View”. Read and comment.

IV. Submit your changes to be approved (Select them and click submit to be approved)

Wait for Validator to finish task II

V. Delete the rejected changes by selecting them and clicking delete button

VI. Open “Approved View” and submit to be deleted the following change:

- Individual Extint DataPropertyValue (description, the last remaining member of the species has died, string,--)
    - Change Type: AddIndividualDataProperty
    - Related Entity: Extint
    - Author: SE2

Appendix 8 - Experiment Guide Validator1

I. Setting-up the environment

- Start the NTK.
- Configure the Registry Properties: Open Preferences->Oyster Storage Preference and configure it with the following properties:
    - Super Node IP: ServerIP (provided at the experiment)
    - Push Node IP: blank (default)
    - Read Ontologies Locally: checked
- Start the Registry: Either from the Registry menu or by clicking the Registry icon on the toolbar. Note that the icon shape and the message on the Registry menu changes when the registry is running
- Identify to the system: Open Preferences->Collaborative Development Preference, go to the Register section and provide:
    - Your first name
    - Your last name
    - Choose your role
    - Click Register
- Create a new Ontology Development Project with the following properties:
    - Ontology Language: OWL
    - Datamodel Type: CollaborationServer
        i. Host: ServerIP (provided at the experiment)
        ii. Port: 8267 (default)
- Open existing ontology species_v1.0_model.owl into the project by selecting “Add Ontology To Project” in the project context menu.

Wait for SE’s finish task IV

II. Editorial Workflow

- Start logging the ontology: Right-click the ontology and select “Log Changes”.

Open “To Be Approved view” and analyze changes from SubjectExperts.

Approve all except the following (reject them):

- Individual 31005_10001 DataProperty hasNameScientific value: Pterodroma wrong macroptera.
    - Change Type: AddIndividualDataProperty.
    - Related Entity: 31005_10001
    - Author: SE1
- Individual ParapatricWrong
    - Change Type: AddIndividual

    - Related Entity: Speciation
    - Author: SE1
- Species Class Super Restriction (EXACTLY/CARD, 1, hasCategory, ---)
    - Change Type: AddSubClassOf
    - Related Entity: Species
    - Author: SE2
- SubClass genus of biological_entity.
    - Change Type: AddSubClassOf
    - Related Entity: genus
    - Author: SE2

Wait for SE's finish task VI

III. **Open "To Be Deleted View"** and

- Reject back to Approved change: DataProperty hasSpeciation (Species):
    - Change Type: AddDataProperty
    - Related Entity: hasSpeciation
    - Author: SE1
- Delete permanently change: Individual Extint DataPropertyValue (description, the last remaining member of the species has died, string,--)
    - Change Type: AddIndividualDataProperty
    - Related Entity: Extint
    - Author: SE2

Imagine there is another validator, then do:

IV. **Open "Approved View"** and (i.e. Validator1)

- Reject back to be approved change: Species Class Super Restriction (HAS_VALUE, hasMeta, "31005").
    - Change Type: AddSubClassOf
    - Related Entity: Species.
    - Author: SE2

V. **Open "ToBeApproved View"** and (i.e. Validator2)

Approve the pending change.