

How to write and use the Ontology Requirements Specification Document

Mari Carmen Suárez-Figueroa¹, Asunción Gómez-Pérez¹, and Boris Villazón-Terrazas¹

¹ Ontology Engineering Group, Departamento de Inteligencia Artificial,
Facultad de Informática, Universidad Politécnica de Madrid,
Madrid, Spain
{mcsuarez, asun, bvillazon}@fi.upm.es

Abstract. The goal of the ontology requirements specification activity is to state why the ontology is being built, what its intended uses are, who the end-users are, and which requirements the ontology should fulfill. The novelty of this paper lies in the systematization of the ontology requirements specification activity since the paper proposes detailed methodological guidelines for specifying ontology requirements efficiently. These guidelines will help ontology engineers to capture ontology requirements and produce the ontology requirements specification document (ORSD). The ORSD will play a key role during the ontology development process because it facilitates, among other activities, (1) the search and reuse of existing knowledge-aware resources with the aim of re-engineering them into ontologies, (2) the search and reuse of existing ontological resources (ontologies, ontology modules, ontology statements as well as ontology design patterns), and (3) the verification of the ontology along the ontology development. In parallel to the guidelines, we present the ORSD that resulted from the ontology requirements specification activity within the SEEMP project, and how this document facilitated not only the reuse of existing knowledge-aware resources but also the verification of the SEEMP ontologies. Moreover, we present some use cases in which the methodological guidelines proposed here were applied.

Keywords: Ontology Requirements; Competency Questions; Ontology Requirements Specification; and Methodologies.

1 Introduction

One of the key processes in software development is software specification [13], whose aim is to understand and define what functionalities are required from the software product. This process leads to the production of a software requirements document [13], that is, the specification for the software product. It has been proved that a detailed software requirements document provides several benefits [10], such as: (a) the establishment of the basis for agreement between customers and suppliers on what the software product is supposed to do, (b) the reduction of the development

effort, (c) the provision of a basis for estimating costs and schedules, and (d) the offer of a baseline for validation and verification.

When a software application based on ontologies is being developed, ontology requirements should be identified in addition to the application ones. Our experience in building ontology-based applications, in domains as different as satellite data processing¹, finding funding programs², fishery stocks³ and e-employment⁴, has showed that more critical than capturing software requirements was the efficient and precise identification of the knowledge that the ontology should contain. Up to now, application developers already have precise methodologies [13, 10, 18] that help them to define application requirements. However, guidelines included in current methodologies for building ontologies are not enough for defining ontology requirements.

The analysis of the state of the art of Ontological Engineering reveals that most of the methodologies for building ontologies include the ontology requirements specification activity. In this regard, we can mention that METHONTOLOGY [5] identifies the goals of the ontology requirements specification activity; however, this methodology does not propose any method for carrying out the activity. Grüninger and Fox's methodology [8], On-To-Knowledge methodology [14], and the Unified methodology [17] propose the identification of the following aspects for creating the ontology requirements specification document (ORSD): (1) the purpose of the ontology to be developed, (2) the intended uses and users of the ontology to be developed, and (3) the set of ontology requirements that the ontology should satisfy after being formally implemented. Most of the existing methodologies suggest the identification of competency questions (CQs) [8] as the technique for establishing the ontology requirements. CQs are natural language questions that the ontology to be built should be able to answer. CQs and their responses play the role of a type of requirements specification against which the ontology can be evaluated. Although the aforementioned methodologies propose methods for carrying out the ontology requirements specification activity consisting of high level steps, they do not provide detailed guidelines that explain how to carry out each step, what it is needed for obtaining a good ORSD; nor how the ORSD can be used later on in the ontology development (for instance, to search knowledge-aware resources to be reused and to verify the ontology content).

The innovation of this paper lies in the systematization of the ontology requirements specification activity since the paper proposes efficient, prescriptive and detailed methodological guidelines for specifying ontology requirements. Our guidelines are based on the use of the so-called CQs and are inspired by existing methodologies for building ontologies and by available practices and previous experiences in different national and European funded projects. These methodological guidelines help to capture knowledge from users and to produce the ORSD that will be used by ontology engineers in order to develop an ontology that will fulfill the requirements identified. Therefore, the main purpose of an ORSD is to serve as an

¹ <http://www.ontogrid.net>

² <http://esperonto.net/fundfinder>

³ <http://www.neon-project.org>

⁴ <http://www.seemp.org>

agreement among ontology engineers, users and domain experts on what requirements the ontology should cover.

When building knowledge intensive systems, the ORSD will be decisive along the ontology development process because it will facilitate, among other activities, the search and reuse of existing knowledge-aware resources for re-engineering them into ontologies; the search and reuse of existing ontologies, ontology modules, ontology statements (e.g., using Watson⁵ or Swoogle⁶), or ontology design patterns; and the verification of the ontology during the whole ontology development. The methodological guidelines presented in this paper have been created in the context of the *NeOn Methodology* [4] within the NeOn project⁷. They have been validated by users in different ontology-based applications inside and outside the NeOn project.

In addition to the guidelines, we also present a particular ORSD resulting from the ontology requirements specification activity carried out within the SEEMP project⁸, whose goals were to develop a knowledge intensive and interoperable architecture based on ontologies for public e-Employment services (PES), and to enable a federated market place of employment mediation agencies through a peer-to-peer network based on interoperation.

The rest of the paper is organized as follows: Section 2 includes the methodological guidelines for the ontology requirements specification activity, and the ORSD obtained in the SEEMP project as an example. Section 3 shows how the methodological guidelines and the ORSD were used in different cases. Section 4 explains in detail how the ORSD was used within the SEEMP project to search and reuse knowledge-aware resources and to evaluate the resultant ontology. Finally, Section 5 provides the conclusions.

2 Methodological Guidelines for Ontology Requirements Specification

In this section we explain the guidelines set out to help ontology developers in the ontology requirements specification activity. Such guidelines have been created in the context of the *NeOn Methodology* [4] for building ontology networks. This methodology takes into account the existence of multiple ontologies in ontology networks, the collaborative ontology development, the dynamic dimension, and the reuse and re-engineering of knowledge-aware resources. One of the key elements in this methodology is the set of 9 scenarios identified for building ontologies and ontology networks.

It is worth also mentioning that in the framework of the NeOn Methodology there are prescriptive methodological guidelines⁹ for carrying out different processes and activities involved in the ontology development: ontology requirements specification (presented in this paper), scheduling ontology projects, reuse and re-engineering non-

⁵ <http://watson.kmi.open.ac.uk/WatsonWUI/>

⁶ <http://swoogle.umbc.edu/>

⁷ <http://www.neon-project.org>

⁸ <http://www.seemp.org>

⁹ Deliverables D5.4.1, D5.3.2, and D5.4.2 (<http://www.neon-project.org/>)

ontological resources, reuse ontological resources, reuse ontology design patterns, ontology modularization, ontology localization, ontology evaluation and ontology evolution.

To create the methodological guidelines in the NeOn Methodology, we based on the different studies carried out to revise the state of the art of ontology development, on the experience of developers on building ontologies in different projects, and on the analysis of various project use cases. All the methodological guidelines are described using a filling card and a workflow.

Thus, taking into account the aforementioned methodological work [4], the filling card for the ontology requirements specification activity is provided in Fig. 1. Such filling card explains the information of this activity in a practical and easy way.

Ontology Requirements Specification	
<p><i>Definition</i></p> <p>Ontology Requirements Specification refers to the activity of collecting the requirements that the ontology should fulfill, e.g. reasons to build the ontology, target group, intended uses, possibly reached through a consensus process.</p>	
<p><i>Goal</i></p> <p>The activity states why the ontology is being built, what its intended uses are, who the end-users are, and what the requirements the ontology should fulfill are.</p>	
<p><i>Input</i></p> <p>A set of ontological needs.</p>	<p><i>Output</i></p> <p>Ontology Requirements Specification Document (ORSD).</p>
<p><i>Who</i></p> <p>Software developers and ontology practitioners, who form the ontology development team (ODT), in collaboration with users and domain experts.</p>	
<p><i>When</i></p> <p>This activity must be carried out at the beginning of the ontology project and in parallel with the knowledge acquisition activity.</p>	

Fig. 1. Ontology Requirements Specification Filling Card.

The output of the ontology requirements specification activity is the ORSD. A template (shown in Table 1) for the creation of the ORSD document is also provided in this section. This template contains information about the purpose, scope, implementation language, intended end-users, intended uses, requirements, and pre-glossary of terms of the ontology which is being built.

Table 1. ORSD Template.

Ontology Requirements Specification Document Template	
1 Purpose	
	<i>The main general goal of the ontology. In other words, the main function or role that the ontology should have.</i>
2 Scope	
	<i>The general coverage and the degree of detail that the ontology should have.</i>
3 Implementation Language	
	<i>The formal language that the ontology should have.</i>
4 Intended End-Users	
	<i>The intended end-users expected for the ontology.</i>
5 Intended Uses	
	<i>The intended uses expected for the ontology.</i>
6 Ontology Requirements	
a. Non-Functional Requirements	
	<i>The general requirements or aspects that the ontology should fulfil, including optionally priorities for each requirement.</i>
b. Functional Requirements: Groups of Competency Questions	
	<i>The content specific requirements that the ontology should fulfil, in the form of groups of competency questions and their answers, including optionally priorities for each group and for each competency question.</i>
7 Pre-Glossary of Terms	
a. Terms from Competency Questions	
	<i>The list of terms included in the competency questions and their frequencies.</i>
b. Terms from Answers	
	<i>The list of terms included in the answers and their frequencies.</i>
c. Objects	
	<i>The list of objects included in the competency questions and in their answers.</i>

Next, in Fig. 2 we present the methodological guidelines for carrying out the ontology requirements specification activity in a detailed and prescriptive manner, showing the main tasks involved, their inputs, outputs and actors. The tasks shown in Fig. 2 are explained below.

Additionally, in this section we provide, as an example, the ORSD created within the SEEMP project, in which ontology developers have followed the guidelines and used the template proposed in this paper. For a detailed specification please refer to [11].

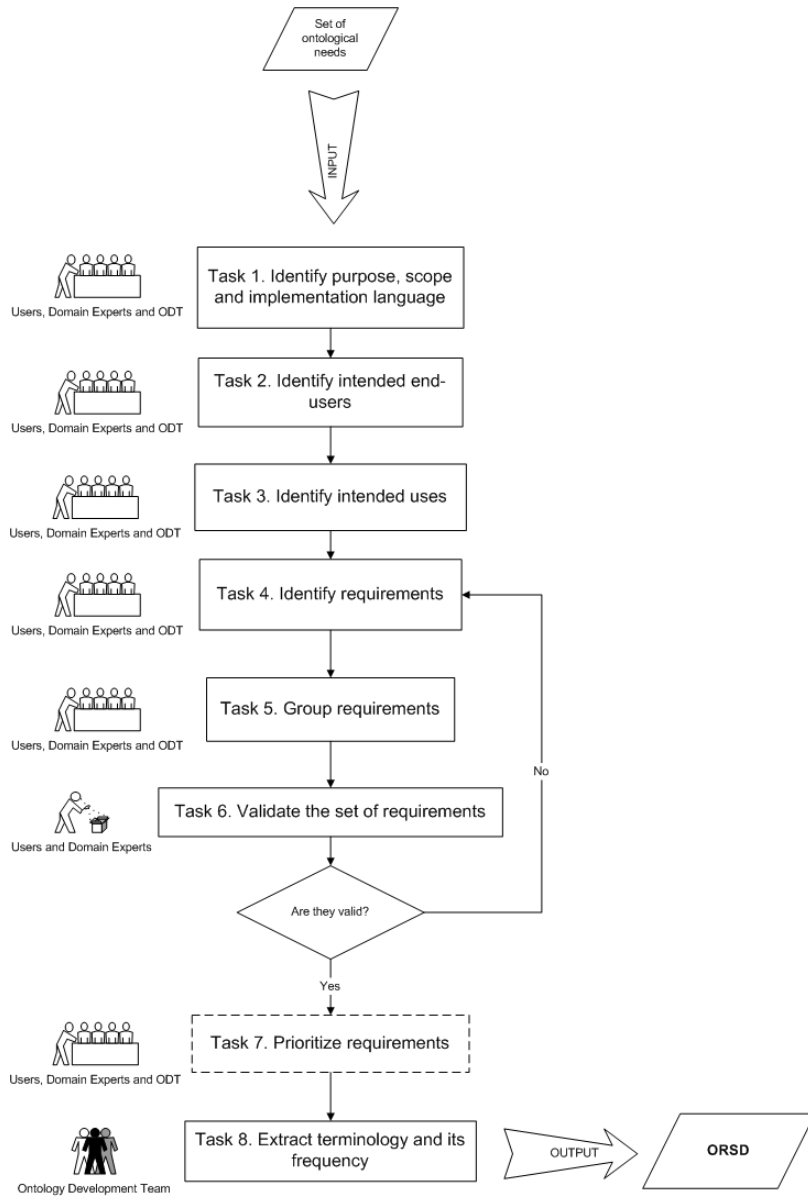


Fig. 2. Tasks for Ontology Requirements Specification.

Task 1. Identify purpose, scope and implementation language. The objective is to determine the main goal of the ontology, its coverage and foreseeable granularity, and its implementation language (e.g., OWL, RDFS, WSMML, etc.). The ontology development team has a set of interviews with users and domain experts to carry out this task taking as input a set of ontological needs, that is, the necessity of having the knowledge represented in the form of an ontology. Users and domain experts are

crucial to identify the purpose and scope of the ontology; the formal language to be used for implementing the ontology should be decided by ontology developers. The task output is included in slots 1-3 of the template shown in Table 1. The output in the SEEMP ORSD is shown in Table 2.

Task 2. Identify intended end-users. The goal of this task is to establish who the main intended end-users of the ontology are. The ontology development team has a set of interviews with users and domain experts to carry out this task taking as input a set of ontological needs. The task output is a list containing the intended users of the ontology to be built, which is included in slot 4 of the template shown in Table 1. The output in the SEEMP ORSD is shown in Table 2.

Table 2. SEEMP ORSD Slots 1 to 5.

SEEMP Reference Ontology Requirements Specification Document	
1	Purpose
	The purpose of building the Reference Ontology is to provide a consensual knowledge model of the employment domain that can be used by public e-Employment services.
2	Scope
	The ontology has to focus just on the ICT (Information and Communication Technology) domain. The level of granularity is directly related to the competency questions and terms identified.
3	Implementation Language
	The ontology has to be implemented in WSML language.
4	Intended End-Users
	<p>User 1. Candidate who is unemployed and searching for a job or searching another occupation for immediate or future purposes</p> <p>User 2. Employer who needs more human resources.</p> <p>User 3. Public or private employment search service which offers services to gather CVs or job postings and to prepare some data and statistics.</p> <p>User 4. National and Local Governments which want to analyze the situation on the employment market in their countries and prepare documents on employment, social and educational policy.</p> <p>User 5. European Commission and the governments of EU countries which want to analyze the statistics and prepare international agreements and documents on the employment, social and educational policy.</p>
5	Intended Uses
	<p>Use 1. Publish CV. Job seeker places his/her CV on the PES Portal.</p> <p>Use 2. Publish Job Offer. An Employer places a Job Offer on the PES Portal.</p> <p>Use 3. Search for Job Offers. The Employer looks for candidates for the Job Offer through PES Portal.</p> <p>Use 4. Search for Employment information. Job Seeker looks for of general information about employment in a given location at the PES Portal.</p> <p>Use 5. Provide Job Statistics. The PES Portal provides employment statistics to the Job Seeker and Employer.</p>

Task 3. Identify intended uses. The development of an ontology is mainly motivated by scenarios related to the application that will use the ontology. The goal of this task is to obtain the intended uses and use scenarios of the ontology. The ontology development team has a set of interviews with users and domain experts to carry out this task taking as input a set of ontological needs; the purpose here is to obtain the uses of the ontology within the application, and to have a general idea of the application requirements. The task output is a list of intended uses in the form of scenarios, which is included in slot 5 of the template shown in Table 1. Such scenarios describe a set of general ontology requirements that the ontology should satisfy after being formally implemented. The scenarios should be described in natural language and can be expressed in UML as use cases. The output in the SEEMP ORSD is shown in Table 2.

Task 4. Identify requirements. The goal of this task is to acquire the set of requirements that the ontology should satisfy. Taking as inspiration the Software Engineering field, in which requirements are divided into functional¹⁰ and non-functional¹¹ requirements [13], we also divide ontology requirements into the following two types, whose definition is different those in Software Engineering:

- *Non-functional ontology requirements* refer to the characteristics, qualities, or general aspects not related to the ontology content that the ontology should satisfy. Examples of non-functional requirements are: (a) whether the terminology to be used in the ontology must be taken from standards, (b) whether the ontology must be multilingual, or (c) whether the ontology should be written following a specific naming convention.
- *Functional ontology requirements*, which can be also seen as content specific requirements, refer to the particular knowledge to be represented by the ontology, for example in the SEEMP case, the knowledge about curriculum vitae with candidate skills, education level, expertise, previous work experience, or about job offers including job location, salary, etc..

The ontology development team should interview users and domain experts taking as input a set of ontological needs, and obtain as result the initial set of ontology requirements (non-functional and functional) of the ontology to be built. To identify functional requirements, the ontology development team uses as main technique the writing of the requirements in natural language in the form of the so-called CQs. Mind map tools [1] and Excel can be used for obtaining the requirements. If people are geographically distributed, wiki tools, such as Cicero¹², can also be employed.

Some strategies for identifying CQs are:

- **Top-Down:** The team starts with complex questions that are decomposed in simpler ones.
- **Bottom-Up:** The team starts with simple questions that are composed to structure/create complex ones.

¹⁰ Functional requirements refer to required behavior of the system, that is, functionalities the software system should have.

¹¹ Non-functional requirements refer to implicit expectations about how well the software system should work. That is, these requirements can be seen as aspects about the system or as ‘non-behavior’ requirements.

¹² <http://cicero.uni-koblenz.de/wiki>

- Middle out: The team starts just writing down important questions that are composed and decomposed later on to form abstract and simple questions respectively.

The output of this task is (1) a list of non-functional ontology requirements written in natural language, which is included in slot 6a of the template shown in Table 1, and (2) a list functional ontology requirements in the form of CQs and their associated answers that is the input of task 5. The list of non-functional requirements in the SEEMP ORSD is shown in Table 3; the list of CQs and their responses are grouped in task 5 and included in slot 6b of SEEMP ORSD shown in Table 4.

Table 3. SEEMP ORSD Slot 6a.

6	Ontology Requirements
	a. Non-Functional Requirements
	<p>NFR1. The ontology must support a multilingual scenario in the following languages: English, Spanish, Italian, and French.</p> <p>NFR2. The ontology must be based on the international, European or de-facto standards in existence or under development.</p>

Task 5. Group requirements. The goal of this task is to group the list of CQs obtained in task 4 into several categories. Users, domain experts and the ontology development team should classify the list of CQs written in natural language using a hybrid approach that not only combines pre-established categories such as time and date, units of measure, currencies, location, languages, etc., but it also creates categories for those terms that appear with the highest frequencies in the list of CQs.

Techniques such as card sorting can be used when the grouping is done manually, whereas techniques for clustering natural language sentences or for information extraction can be used when the grouping is done automatically. In addition, mind map tools can help to display graphically the CQs in groups or Cicero if the grouping is done collaboratively.

The task output is the set of groups of CQs and responses, which is included in slot 6b of the template. The output in the SEEMP ORSD is shown in Table 4.

Usually this task is carried out in parallel with task 4.

Task 6. Validate the set of requirements, including both non-functional and functional ontology requirements. The aim here is to identify possible conflicts between ontology requirements, missing ontology requirements, and contradictions between them. Users and domain experts must carry out this task taking as input the set of requirements identified in task 4 to decide if each element of the set is valid or not. The task output is the confirmation of the validity of the set of non-functional and functional ontology requirements.

The following criteria can be used in this validation task and they are mainly inspired by [9, 2]:

- A set of requirements is correct if each requirement refers to some features of the ontology to be developed.

- Inspired by [19], a set of requirements can be considered complete, if users and domain experts review the requirements and confirm that they are not aware of additional requirements.
- A set of requirements can be considered internally consistent if no conflicts exist between them.
- A set of requirements is verifiable, if there is a finite process with a reasonable cost that tests whether the final ontology satisfies each requirement.
- Each requirement must be understandable to end-users and domain experts.
- An ontology requirement is unambiguous if it has only one meaning; that is, if it does not admit any doubt or misunderstanding.
- A set of requirements is concise, if each and every requirement is relevant, and no duplicated or irrelevant requirements exist.
- A set of requirements is realist, if each and every requirement meaning makes sense in the domain.
- A set of requirements is modifiable if its structure and style allow changing issues in an easy, complete and consistent way.

Task 7. Prioritize requirements. The goal here is to give different levels of priority to the non-functional and functional ontology requirements identified. In the case of functional ontology requirements, priorities should be given to the different groups of CQs, and, within each group, to the different CQs. Users, domain experts and the ontology development team should carry out this task taking as input the requirements identified in task 4 and the groups of CQs written in natural language obtained in task 5. The task output is a set of priorities attached to each requirement and to each group of CQs and to each CQ in a group. The output is included in the slots 6a and 6b of the template.

Priorities in CQs will be used by the ontology development team for planning and scheduling the ontology development and for deciding which parts of the ontology are going to be developed first. This task is optional, but recommended. In fact, if no priorities are given to the groups of CQs, ontology developers will start modeling the ontology without any guidance regarding the requirements that should be implemented first; in this case the waterfall ontology life cycle model should be selected during the scheduling of the ontology project. On the contrary, if different priorities have been assigned to ontology requirements, the iterative-incremental ontology life cycle model [15] should be selected in the scheduling activity.

Table 4. SEEMP ORSD Slot 6b.

6	Ontology Requirements
	b. Functional Requirements: Groups of Competency Questions
	<p><i>CQG1. Job Seeker (14 CQ)</i></p> <p>CQ1. What is the Job Seeker's name? Lewis Hamilton</p> <p>CQ2. What is the Job Seeker's nationality? British; Spanish; Italian; French;</p> <p>CQ3. What is the Job Seeker's birth date? '13/09/1984; 30/03/1970; 15/04/1978</p> <p>CQ4. What is the Job Seeker's contact information? Tel: 34600654231. Email: jsanz@fi2.upm2.es</p> <p>CQ5. What is the Job Seeker's current job? Programmer; Computer Engineer; Computer Assistant</p> <p>CQ6. What is the Job Seeker's desired job? Radio Engineer; Hardware designer; Software Engineer</p> <p>CQ7. What are the Job Seeker's desired working conditions? Autonomous; Seasonal Job; Traineeship; Consultant</p> <p>CQ8. What kind of contract does the Job Seeker want? Full time; Partial time; Autonomous; Seasonal Job</p> <p>CQ9. How much salary does the Job Seeker want to earn? 3000 Euros per month, 40000 Euros per year</p> <p>CQ10. What is the Job Seeker's education level? Basic education; Higher education/University</p> <p>CQ11. What is the Job Seeker's work experience? 6 months, 1 year, 2 years</p> <p>CQ12. What is the Job Seeker's knowledge? Java Programming; C Programming, Database Administration</p> <p>CQ13. What is the Job Seeker's expertise? Software Engineering</p> <p>CQ14. What are the Job Seeker's skills? SQL programming, network administration</p> <p><i>CQG2. Job Offer (11 CQ)</i></p> <p>CQ15. What is the employer's information? CEFRIEL Research Company, Milano, Italy; ATOS, Madrid, Spain</p> <p>CQ16. What kind of job does the employer's offer? Java Programmer; C Programmer, Database administration</p> <p>CQ17. What kind of contract does the employer's offer? Seasonal Job; Autonomous</p> <p>CQ18. How much salary does the employer's offer? 3500 Euros, 3000 USD</p> <p>CQ19. What is the economic activity of the employer? Research; Financial; Education; Industrial</p> <p>CQ20. What is the description of the job offer? Sun Certified Java Programmer</p> <p>CQ21. What are the working conditions of the job offer? Full time; Partial time; Autonomous; Seasonal Job</p> <p>CQ22. What is the required education level for the job offer? Basic education; Higher education/University</p> <p>CQ23. What is the required work experience for the job offer? 1 year, 2 years, 3 years, 4 years, 5 or more years</p> <p>CQ24. What is the required knowledge for the job offer? Java, Haskell, Windows</p> <p>CQ25. What are the required skills for the job offer? ASP Programmer, Data warehouse, Hardware programming</p>

Task 8. Extract terminology and its frequency. The goal is to extract from the list of CQs and their answers a pre-glossary of terms. This pre-glossary of terms is divided in three different parts: terms from the CQs, terms from the CQs' answers, and terms identified as named entities, which are objects. From the requirements in the form of CQs, we extract terminology (names, adjectives and verbs) that will be formally represented in the ontology by means of concepts, attributes, relations or instances (in the case of named entities). From the answers to the CQs, we extract terminology that could be represented in the ontology as concepts or as instances. From both CQs and corresponding answers, we extract named entities such as countries or currencies. The output is included in the slots 7a, 7b and 7c of the template, respectively. The set of terms with higher appearance frequencies will be used later on for searching knowledge-aware resources that could be potentially reused in the ontology development.

The ontology development team should carry out this task taking as input the list of identified CQs and their answers by using terminology extraction techniques and tools supporting such techniques. The task output is included in slot 7 of the template. The output in the SEEMP ORSD is shown in Table 5.

Table 5. SEEMP ORSD Slot 7.

7	Pre-Glossary of Terms			
	a. Terms from Competency Questions + Frequency			
	Job Seeker	27	Address	1
	CV	2	Nationality	1
	Personal Information	3	Contact (phone, fax, mail)	3
	Name	4	Objective	3
	Gender	1	Job Category	3
	Birth date	1	...	
	b. Terms from Answers + Frequency			
	SW engineer, programmer	5	Research, Financial, Education	4
	British, Spanish, Italian, French	1	1 year, 2 years, 3 years	1
	Autonomous, Seasonal Job,	2	3000 Euros per month	1
	Basic education, Higher education	1	CEFRIEL Research Company	1
	c. Objects			
	Andorra, Angola, Argentina, Australia, Bolivia, France, Italy, Malta, Spain, etc.			
	Euro, Zloty, Great British Pound, US Dollar, Peso, etc.			
	CEFRIEL, ATOS, etc.			

3 Experimentation

Both the methodological guidelines and the ORSD template presented in this paper have been used in the NeOn project and in the development of ontologies in other research and educational projects with positive feedback and interesting comments from the ontology developers involved in each case. In this section we briefly summarize a set of such cases, presenting qualitative results¹³.

The invoice use case, whose main aim is to solve the lack of interoperability between invoice emitters and receivers; and **the nomenclature use case**, whose main objectives are to help in the systematization of the creation, maintenance and keeping up-to-date drug-related information, and to allow an easy integration of new drug resources, within the NeOn project [16, 7]. Ontology developers in both use cases provided positive feedback about the usefulness of the guidelines, the goodness of writing an ORSD prior to the ontology development, and the benefits of using the ORSD as input to other activities (e.g., reuse of knowledge-aware resources and ontology verification). Moreover, ontology developers realized the importance of having guidelines for the ontology requirements specification. However, they commented that examples on how to use the guidelines would be very useful for understanding and applying such guidelines in a better way.

The development of **ontologies** representing the knowledge related with the “*Camino de Santiago*” (Saint James’s Way) in **the Geobuddies**¹⁴ project. Ontology developers included, besides the previous advantages, the usefulness of the pre-glossary of terms during the ontology reuse process and the conceptualization activity.

The hands-on session during the Sixth Summer School on Ontologies and Semantic Web¹⁵, in which we tested the combination of the ontology requirements specification activity and the ontology reuse process using Watson with 50 students. In this case, we obtained a very positive feedback from most of the students who carried out the hands-on. They stated that the proposed guidelines were useful, and that they would like to use anew such guidelines in future ontology developments. They also expressed that the writing of the ORSD before going into the ontology development was by no means a waste of time, and that the ORSD was useful in the ontology development.

Two experiments to test the benefits of using the ontology requirements specification guidelines proposed in this paper were carried out with students at **UPM during the academic years 2007-2008 and 2008-2009**. Such experiments were performed within the “Artificial Intelligence (AI)” master course at *Facultad de Informática (Universidad Politécnica de Madrid)* with master students whose backgrounds included databases, software engineering, and artificial intelligence, but not ontology engineering. One of the experiments was carried out during November 2007 with 14 master students working in groups of two people; the other was performed during November 2008 with 12 master students working in groups of one

¹³ Quantitative results are not provided because to test the same real case using the proposed guidelines and without the guidelines is not feasible due to the effort needed.

¹⁴ <http://www.geobuddies.net>

¹⁵ <http://kmi.open.ac.uk/events/sssw08/>

or two people. Both experiments were carried out with methodological guidelines for the ontology requirements specification activity. Most of the students considered that the guidelines for the ontology requirements specification activity were useful. All of them expressed their intention of using again the methodological guidelines for the ontology requirements specification activity. However, some of the students commented they would prefer to have such guidelines integrated in an ontology requirements specification tool. Students also stated that they found useful to write the ontology requirements specification document before going into the ontology development. Additionally, students commented that (1) the three first tasks in the methodological guidelines should be explained in more detail; (2) the tasks of writing and validating CQs, as well as giving priority to CQ, could be further described; and (3) the task about extracting terminology should be clarified.

As already mentioned, the guidelines for the ontology requirements specification activity proposed here were used in the **development of the SEEMP Reference Ontology**¹⁶. In Section 4 we present how the final ORSD obtained in such a project was used during the ontology development. Such a final ORSD has been presented in Section 2 as an example while the methodological guidelines were explained.

Additionally, it is worth mentioning that comments provided by ontology developers involved in the aforementioned cases are allowing us to improve and update the methodological guidelines for the ontology requirements specification activity.

4 Ontology Requirements Specification and Other Activities

The guidelines presented here for the ontology requirements specification activity were used in the development of ontologies in the public e-Employment services domain within the SEEMP project. In this section we present how the ontology requirements specification document allows (a) a more direct search for available knowledge resources to be reused in the ontology development, and (b) the evaluation of the ontology content.

Terms with the highest frequency included in the pre-glossary of terms of the ORSD were used to direct the search of existing knowledge-aware resources that cover the terminology in question in internet and in standardization bodies (e.g. ISO and proprietary resources in the intranet of companies). We also looked for existing ontologies using Watson and other semantic search engines. Thus, the ORSD was the key input in the search process in the following way:

- We searched for time ontologies describing the terms related with time and date management (e.g., year, month, etc.). We used the comparative study of nine time ontologies presented in [3] and then we selected the DAML Time Ontology¹⁷.
- We searched for standards of occupations and economic activities in high reliable Web sites, domain-related sites and resources within organizations, instead of

¹⁶ The resultant ontology is available at <http://droz.dia.fi.upm.es/hrmontology/>

¹⁷ <http://cs.yale.edu/homes/dvm/daml/time-page.html>

starting from scratch the identification of an exhaustive list of occupations on the IT domain and a list of economic activities. As a result of the search we found four potential non-ontological resources that included terms related to occupation, and three potential non-ontological resources related to economic activities.

- Using NeOn guidelines for reusing knowledge-aware resources [16], we assessed the set of candidate non-ontological resources based on coverage and precision criteria. First, we extracted the lexical entries of the non-ontological resources. Then we calculated the coverage and precision of the non-ontological resource lexical entries against the terms that appeared in the pre-glossary in the ORSD. The non-ontological resources with best coverage and precision, and thus the selected ones, were ISCO-88 (COM)¹⁸ and ONET¹⁹ for occupations, and NACE¹⁸ standard for economic activities. Then, these non-ontological resources were re-engineered into ontologies.
- Regarding geographical location classifications, we did the search in similar sites and we found four non-ontological resources including named entities related with geographical location (e.g. Spain, France). Following the NeOn guidelines, we selected the ISO 3166²⁰ standard and we transformed it into an ontology.
- Finally, we evaluated the resultant SEEMP Reference Ontology by means of the CQs from the ORSD. We formalized the CQs into SPARQL queries, and transformed the Reference Ontology WSMML implementation into RDF(S). In this sense, we checked whether the query results of every formalized CQ were empty.

Thus, the ORSD helped us to find existing consensual knowledge-aware resources to be used in the development of the SEEMP Reference Ontology [12]. This ontology is composed of thirteen modular ontologies, from which ten were built by reusing and re-engineering non-ontological resources, and three were built from scratch. The Reference Ontology has 1985 concepts, 315 attributes, 1449 instances and 1037 axioms. The ontology was built in 6 months. It must be said that at least 2 months were dedicated to the ontology requirements specification activity. Finally, the evaluation of the ontology lasted approximately one month.

Our experience in SEEMP has served us to demonstrate that the approach to build the ORSD following detailed methodological guidelines saves time and effort during the ontology development. We compared our previous experiences developing ontologies with the development of the SEEMP Reference Ontology and we realized that a decrease in the development time occurred. For example, we developed a set of Legal Ontologies [6], without the guidelines described in this paper, spending 10 MM (man months) and, as already mentioned, we spent 6 MM for the development of the SEEMP Reference Ontology. It is worth mentioning that these two ontologies (Legal and SEEMP ones) have similar level of complexity. The ORSD, also, facilitated the reuse of existing ontologies, the reuse and later re-engineering of existing and already consensual non-ontological resources, and the evaluation of the ontology content. In this paper we have demonstrated that with a good ORSD, better ontologies can be built faster.

¹⁸ <http://ec.europa.eu/eurostat/ramon>

¹⁹ <http://online.onetcenter.org/>

²⁰ <http://www.iso.org/iso/en/prods-services/iso3166ma/index.htm>

5 Conclusions

One of the critical activities when developing ontologies is to identify their requirements. In this paper we have systematized the ontology requirements specification activity by proposing detailed and prescriptive methodological guidelines for specifying ontology requirements, and a template for writing the ORSD. When a particular ORSD is created, it can be used for speeding up the ontology development process. Terms and their frequencies from the pre-glossary in the ORSD can be used for searching and selecting already knowledge-aware and consensual resources that, after a re-engineering process if necessary, allow building ontologies cheaper, faster, and with higher quality. CQs in the ORSD can be used for verifying the correctness of the ontology with respect to the ontology requirements identified.

The NeOn methodological guidelines for the ontology requirements specification activity and the ORSD template presented here have been used in the development of the NeOn ontologies and the development of ontologies in other research and educational projects with interesting feedback from the ontology developers involved in each case.

Finally, it is worth mentioning that the ORSD facilitates the ontology development in different ways: (1) allowing the identification of which particular knowledge should be represented in the ontology, (2) facilitating the reuse of existing knowledge resources by means of focusing the resource search towards the particular knowledge to be represented in the ontology, and (3) permitting the verification of the ontology with respect to the requirements that the ontology should fulfill. As in other disciplines, a good specification (a) establishes the basis for agreement between the users and ontology developers, (b) reduces the development effort, (c) provides a basis for estimating costs and schedules, and (d) offers a baseline for verification.

Acknowledgments. This work has been partially supported by the European Commission projects NeOn (FP6-027595) and SEEMP (FP6-027347), as well as by an R+D grant from the UPM. We are very grateful to Elena Montiel-Ponsoda and Mariano Fernández-López for their revisions and comments.

References

1. Buzan, T. "Use your head". BBC Books. 1974.
2. Davis, A. "Software Requirements: Objects, Functions and States". Upper Saddle River, New Jersey: Prentice Hall, 1993.
3. Fernández-López, M., Gómez-Pérez, A. "Searching for a Time Ontology for Semantic Web Applications". Formal Ontology in Information Systems, Turín, Italy. 2004.
4. Gómez-Pérez A., Suárez-Figueroa, M.C. "NeOn Methodology: Scenarios for Building Networks of Ontologies". 16th International Conference on Knowledge Engineering and Knowledge Management Knowledge Patterns (EKAW 2008). Conference Poster. Italy. 2008.

5. Gómez-Pérez, A., Fernández-López, M., Corcho, O. "Ontological Engineering". Springer Verlag. Advanced Information and Knowledge Processing series. ISBN 1-85233-551-3. November 2003.
6. Gómez-Pérez, A., Ortiz-Rodriguez, F., and Villazón-Terrazas, B. Legal Ontologies for the Spanish e-Government. *Current Topics in Artificial Intelligence*. Springer Berlin / Heidelberg. 301-310, 2006.
7. Gómez-Pérez, J.M., Pariente, T., Buil-Aranda, C., Herrero, G., Baena, A.: NeOn Deliverable D8.3.1. Ontologies for pharmaceutical case studies. 2007.
8. Grüninger, M., Fox, M. "Methodology for the design and evaluation of ontologies". In Skuce D (ed) *IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing*, pp 6.1–6.10. (1995).
9. IEEE Recommended Practice for Software Requirements Specifications. IEEE Std 830-1993.
10. IEEE Recommended Practice for Software Requirements Specifications. IEEE Std. 830-1998.
11. SEEMP Consortium. SEEMP D3.1a. Supporting the State of the Art. July 2006.
12. SEEMP Consortium. SEEMP D4.1. Components Design. March 2007.
13. Sommerville, I. "Software Engineering". 8th edition. International Computer Science Series. 2007.
14. Staab, S., Hans, P., Studer, R., Sure, Y. "Knowledge Processes and Ontologies". *IEEE Intelligent Systems* 16(1):26–34. (2001).
15. Suárez-Figueroa, M. C., Fernández-López, M., Gómez-Pérez, A., Dellschaft, K., Lewen, H., Dzbor, M. NeOn D5.3.2. Revision and Extension of the NeOn Development Process and Ontology Life Cycle. NeOn project. <http://www.neon-project.org>. November 2008
16. Suárez-Figueroa, M.C., Aguado de Cea, G., Buil, C., Dellschaft, K., Fernández-López, M., García, A., Gómez-Pérez, A., Herrero, G., Montiel Ponsoda, E., Sabou, M., Villazón-Terrazas, B. and Yufei, Z.. "NeOn D.5.4.1. NeOn Methodology for Building Contextualized Ontology Networks". February 2008.
17. Uschold, M. "Building Ontologies: Towards A Unified Methodology". In: Watson I (ed.) *16th Annual Conference of the British Computer Society Specialist Group on Expert Systems*. Cambridge, United Kingdom. (1996).
18. Wieggers, E. "Software Requirements 2: Practical techniques for gathering and managing requirements throughout the product development cycle". 2nd ed., Redmond: Microsoft Press. ISBN 0-7356-1879-8. 2003.
19. Wiering, R. "Requirements Engineering: Frameworks for Understanding". New York: John Wiley & Sons, 1996.