ONTOP
A Process to Support Ontology Conceptualization

Elis Montoro Hernandes, Deysiane Sande and Sandra Fabbri
Computing Department, Federal University of São Carlos, São Carlos, Brazil

Keywords: Ontology Engineering, Collaborative Glossary, Conceptualization Phase, Information Visualization.

Abstract: Although there are tools that support the ontology construction, such tools do not necessarily take heed to the conceptualization phase in its need of execution resources. The objective of this paper is to present the ONTOP Process (ONTOlogy conceptualization Process) as an effective means of enhancing the conceptualization phase of the ontology construction. This process is supported by the ONTOP-Tool which provides an iterative way to defining a collaborative glossary and uses a visual metaphor to facilitate the identification of the ontology components. Once the components are defined, it is possible to generate an OWL file that can be used as an input to other ontology editors. The paper also presents an application of the both process and the tool, which emphasizes the contributions of this proposal.

1 INTRODUCTION

Sharing large volume of data and information through web technologies is, currently, a constant need. Appropriate mechanisms for this are the target of many researches mainly in the context of semantic web (Daconta, Obrst and Smith, 2003). Ontologies have been a common help resource.

According to Gruber (1993), ontology is a formal and explicit specification of the description of concepts in a domain. Ontologies represent the semantic of a domain and can be used by many applications.

Based on the literature (Gruber, 1993; Gómez-Pérez, Fernández-López and Corcho, 2004), it is possible to identify some advantages provided by an ontology: i) improvement of the communication among the involved people since it leads to a particular sense of the vocabulary and meaning of the domain terms; ii) formalization of the knowledge avoiding ambiguities and inconsistencies; iii) representation of the domain knowledge allowing its dissemination and reuse. In addition, ontology allows the knowledge improvement making it possible for different teams develop applications in different moments and with different purposes.

Due to these reasons it is important that domain experts participate in development of the ontology process aiming to avoid mistaken definitions.

In literature there are several approaches to ontology development (Gómez-Pérez et al., 2004). Corcho, Fernández-López and Gómez-Pérez (2003) discuss ontology methods that had been used since the 90's and comment that none of them have reached the maturity. However, Methontology (Fernández-López, Gómez-Pérez, Pazos-Sierra and Pazos-Sierra, 1999) is a method that has been considered one of the most complete (Corcho et al., 2003).

Among the development activities, Methontology is composed by the phases of Figure 1, which are present in the majority of the ontology development processes. Aiming at supporting the execution of these phases some tools and languages were proposed in literature. In this research we particularly use the ontology editor Protégé-2000 (Noy, Fergerson and Musen, 2000; "Protégé-2000", 2010) and the language OWL (Ontology Web Language) ("OWL", 2009) due to the features they provide for our work and to their acceptance in ontology area.

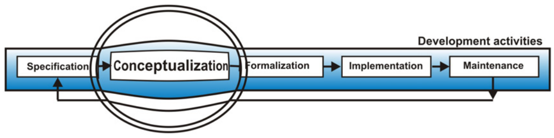

Figure 1: Ontology Development Activities of Methontology. Adapted (Corcho et al., 2003).

Montoro Hernandes E., Sande D. and Fabbri S. (2010).
ONTOP - A Process to Support Ontology Conceptualization.
In *Proceedings of the 12th International Conference on Enterprise Information Systems - Databases and Information Systems Integration*, pages 58-65
DOI: 10.5220/0002908100580065

In spite of such proposals, none of the tools support the conceptualization phase highlighted in Figure 1. The objective of this phase is to organize the non structured knowledge, which was acquired in the previous specification phase. The conceptualization phase converts domain information into a semi-formal specification using a set of intermediate representations. This phase is considered the most important phase for the ontology identification an it initial activity is the construction of a glossary (Corcho et al., 2003). Apart from the glossary, many initial definitions and ontology components, like classes and their relationships, are established in this phase.

Considering this context, the objective of this paper is to present the process ONTOP (ONTOlogy conceptualization Process) to support the conceptualization phase of the ontology definition and ONTOP-Tool that supports the execution of this process. ONTOP considers the use of a collaborative glossary tool and the use of visualization, which helps a great deal in the identification of classes. In this proposal we are using the glossary available in the free Moodle environment (Modular Object-Oriented Dynamic Learning Environment) (“Moodle”, 2009). The activities proposed in this process are supported by ONTOP-Tool that is responsible for the interaction between the glossary and the visualization, allowing an easier identification of the ontology components. The intention of the example presented in this paper is to explain the process and the functionalities of the tool showing the contribution of the proposal and not exploring the ontology properly.

The remaining of the paper is organized as follows: Section 2 comments the importance of a glossary for ontology definition and mentions the main characteristics a glossary should have; Section 3 provides a brief view on visualization; Section 4 presents the ONTOP process and explains the activities that compose it; Section 5 provides an example of the process application supported by the tool and Section 6 presents the conclusions and future works.

2 THE IMPORTANCE OF A GLOSSSARY FOR ONTOLOGY DEFINITION

Regarding target domain, some authors indicate the glossary as the artifact for knowledge acquisition and documentation (Fernández-López et.al, 1999) (Falbo, Menezes and Rocha, 1998).

A common problem associated with ontology construction is that those who need the ontology are specialists on the domain and often they are geographically distributed. If that is the case, the support of tools becomes essential in order to allow a colaborative development and documentation of the ontology construction. The glossary could be a richer source of information if a greater number of specialists participated in its production. In such case, the glossary would synthetize concepts from different contributors.

Another important reason for using glossaries is that its use is much more friendly to the specialists rather than the use of the notation applied to describe ontology.

Concerning the main characteristics, a glossary should satisfy the auto-reference principle which says that terms used to describe another terms should also be an entry of the glossary. In addition, a glossary should satisfy the principle of minimum vocabulary, i.e., the vocabulary should be as small as possible and does not have ambiguities.

Falbo et al. (1998) emphasize that glossaries should use the concept of hypertext aiming at facilitating the navigation in the document. This characteristic is native in the Moodle environment and was a reason for choosing this environment in this proposal. In addition, Moodle provides other resources like a glossary administrator, different permissions, discussion forums, etc., that facilitate the collaborative work and its management.

3 INFORMATION VISUALIZATION

Visualization is a process that transforms data, information and knowledge in a visual form that explores the natural visual capacity of human beings, providing an interface between two powerful systems of information treatment: the human brain and the computer (Gershon, Eick and Card, 1998). Effective visual interfaces provide a quick interaction with large volume of data, making the identification of characteristics, patterns and tendencies that were masked easier.

In the literature, there are some visualization techniques and tools that implement them (Gershon, Eick and Card, 1998). They present advantages and limitations according to the format and type of the data that will be visualized and to the exploration needs.

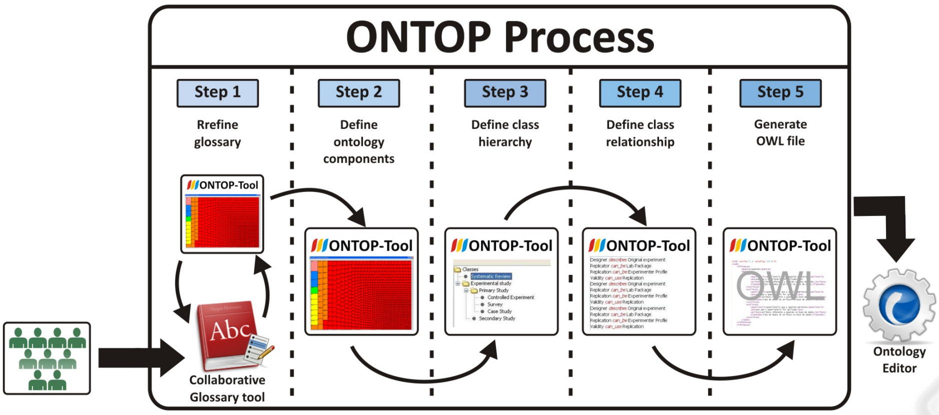

Figure 2: ONTOP process.

In this research we use the Tree-Map technique (Johnson and Shneiderman, 1991) that is illustrated in Figure 2. This technique represents the data as nested rectangles in accordance with their hierarchy. The size of the rectangles is proportional to the number of items that compose the next level of the hierarchy. The size and color variation of Tree-Map makes evident the characteristics of each set of data. This fact enhances the visualization of large sets of data like, for example, a glossary that contains many terms. In addition, Tree-Map uses all the screen space that allows the representation of a great amount of data.

Figure 2 shows an example of Tree-Map. In this case the technique is used in the NewsMap site to group news in accordance with the subject ("NewsMap", 2009). This site uses visualization to show world news allowing that users interact with the site filtering the news by theme, date and country.

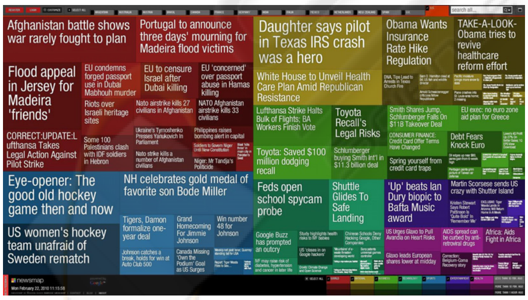

Figure 3: Newsmap site ("NewsMap", 2009).

Visualization allows a broad view of the data as well as the abstraction of new information in a quicker way than if the analysis was done manually. Even when the set of data is small, an appropriated visualization allows an immediate identification of tenuous differences in the data. Many advantages of visualization uses can be viewed through the researches of many authors (Chen, Kuljis and Paul, 2001) (Auvil, Llorà, Searsmith and Searsmith, 2007) (Ichise, Satoh and Numao, 2008).

In this research the Tree-Map technique is used to represent the terms of the glossary such that each box represents a term and the size and the color of the boxes represent the frequency that the term is used in the glossary. In this case, the visualization allows a quick identification of the most cited terms, which are candidate to classes of the ontology.

Although there is, for example, the TreeMap ("TreeMap", 2009) and other tools that implement this technique, they did not have essential resources to help in our problem. Aiming to refine the glossary, we needed that the visualization tool provided two basic operations: string search and edition. This fact leads us to implement ONTOP-Tool that will be presented in Section 5.

4 THE ONTOP PROCESS

ONTOP is a process supported by the ONTOP-Tool which enhances the ontology conceptualization by making use of glossary and visualization (Figure 3). The glossary can be constructed in a collaborative way among the ontology stakeholders, including the domain specialists, through an iterative process of refinement. As the collaborative work is fundamental, we decided to use the glossary of Moodle environment for the reason that it provides some management facilities. Plus the fact that it is possible to export the terms to an XML file so that they can be loaded in the ONTOP-Tool and also visualized through the Tree-Map (Johnson and Shneiderman, 1991), allowing the interaction between the tool and the Moodle. After that, by means of visualization information, it is possible to define the initial ontology components. These components can be exported to an OWL file ("OWL", 2009) and be used by a ontology editor like Protégé-2000 (Noy et al., 2000; "Protégé-2000", 2010) to go ahead with the ontology formalization.

The steps that compose the ONTOP process are the following:

Step 1 – Refine Glossary: the objective of this step is to create, refine and validate the glossary iteratively, counting on the domain specialists' involvement. This is an iterative step where the glossary is imported from and exported to the Moodle environment as well as to and from the ONTOP-Tool, until the glossary is finally able to represent the domain. During this iteration the ontology stakeholders can insert, remove or define the terms.

The Step 1 is composed by the following activities:

1) Create a glossary in the Moodle environment;
2) Share the glossary with the ontology stakeholders so that some specialists can participate in the glossary definition;
3) Refine the glossary with the following actions:
   (i) Export the glossary from the Moodle to an XML file;
   (ii) Import the XML file to the ONTOP-Tool so it can be visualized by means of the Tree-Map;
   (iii) Export the glossary from the ONTOP-Tool to an XML file;
   (iv) Import the XML file to the Moodle
   (v) Go back to (i) until the ontology stakeholders come to an agreement.

Step 2 – Define ontology components: the objective of this step is to identify, among glossary terms, the possible ontology components and, then, classify them in terms of class, class instance, relationship or synonyms. At this point the contribution of the ONTOP-Tool is to make evident the most used terms, pinpointing then as possible ontology components candidates.

Step 3 – Define class hierarchy: the objective of this step is to define the hierarchy of the components identified in step 2. The hierarchy is easily established by the ONTOP-Tool through a drag-and drop action.

Step 4 – Define class relationships: the objective of this step is to attribute the relationships among the classes. Some of these relationships are predefined and obtained from the information generated in Step 2 and some others can be inserted by the user when necessary.

Step 5 – Generate OWL file: the objective of this step is to generate the OWL file which is composed of all the information defined by the user in the previous four steps. This file can be imported by an ontology editor as Protégé-2000, which is used in this research.

5 AN EXAMPLE OF USING ONTOP AND ONTOP-TOOL

In this section we present an example of the process application, detailing the functionalities provided by the ONTOP-Tool. The process is illustrated based on Experimental Software Engineering (ESE) domain. The glossary constructed for this domain counted on the collaboration of domain specialists that composed the program committee of 2006's Experimental Software Engineering Latin American Workshop (Fabbri, Travassos; Maldonado; Mendonça Neto and Oliveira, 2006).

Experimental Software Engineering is a growing area in software engineering and deals with different types of experimental studies, for example, surveys, case study, controlled experiment (Wohlin et al., 2000). Due to limited space we cannot give a deep overview of the domain. In spite this limitation, our main objective is to explain the process steps showing how they work and how they help the ontology conceptualization phase.

The ESE glossary was constructed in the Moodle environment aiming to facilitate the communication among the program committee, which was geographically distributed. Based on this ESE glossary version, ONTOP was applied as illustrated below.

Figure 4 shows the initial screen of the ONTOP-Tool which has buttons for the functionalities needed to execute the process steps.

To execute *Step 1*, for refining the glossary, the user should use the first three buttons. Clicking on the button "Import Moodle Glossary" the tool uploads the XML file that contains the ESE Glossary.

After that, the user should click on the button "Analyse the Glossary" to visualize it like in Figure 5 where:

- each box corresponds to a term;
- each box is colored according to the term frequency;
- clicking on a box it is possible to insert or edit the term definition;

- boxes that have a fading color represent terms that were edited in the current visualization;
- terms can be inserted or excluded as in the Moodle glossary.

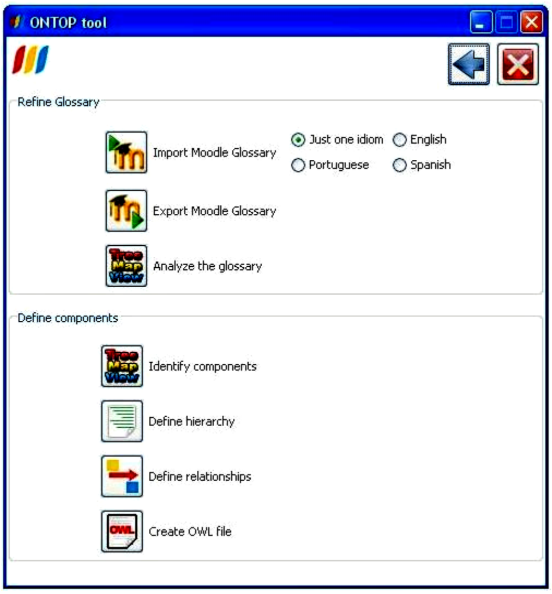

Figure 4: ONTOP-Tool initial screen.

In this example, as the ESE glossary was already constructed by the domain specialists, it was not necessary much iteration to execute the refinement.

However, just to exemplify the contribution of visualization, note that at the left top corner of Figure 5 there are a set of boxes that are grouped because they correspond to terms that do not have a definition. This situation is easily identified in the visual metaphor. To obtain this information in the Moodle environment, the user should verify the terms, one by one. Missing or equal definitions are quickly identified by means of the ONTOP-Tool.

Considering the previous situation, if the user decides to insert a definition to these terms, the color of their boxes is faded (see Figure 6). This is an interesting artifice of the ONTOP-Tool since every time a color is faded in the visual metaphor it means that the corresponding term was edited. The color will persist faded while the user stays in the same functionality.

If the user wishes to share the editions among the ontology stakeholders, he should export this version of the glossary clicking on the button "Export Moodle Glossary" and import it again to the Moodle environment, by means of Moodle functionality.

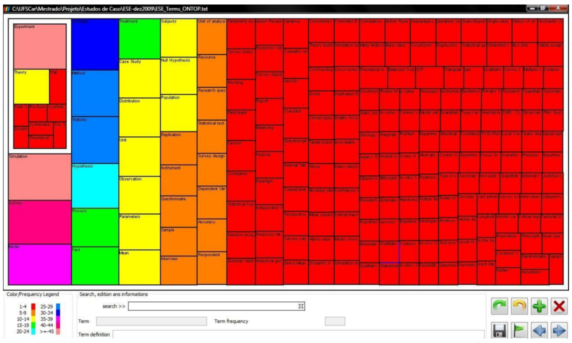

Figure 5: Initial glossary visualization.

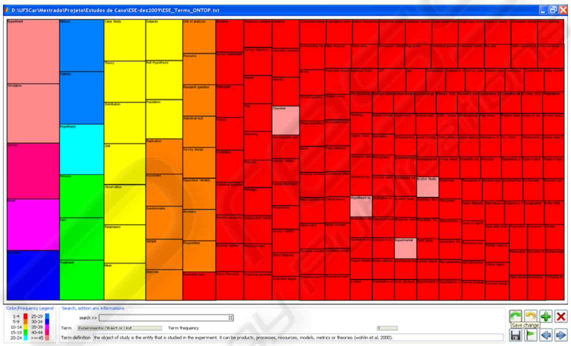

Figure 6: Fade color represents edited terms.

All these activities should be repeated until the ontology stakeholders reach an agreement. Once the glossary is finished, the user can execute the *Step 2*, clicking on the button "Identify components", for classifying the glossary terms as ontology components. Figure 7 shows the screen of this functionality, where the region to insert the definitions is highlighted. The visualization is the same of the previous functionality and, at this moment, one of the contributions of visualization is related to the size or the color of the boxes, since they represent the frequency associated to each term. Terms that have high frequency are candidates to become classes of the ontology.

For example, in Figure 7, the terms Experiment, Simulation and Survey that are highlighted in the figure, correspond to the most referenced in the ESE glossary; they have the largest boxes and colors that correspond to high levels of frequency.

In fact, for the ESE domain, the term "Experiment" is used for defining or expressing many other terms like Controlled Experiment, Experiment Design, Replicated Experiment, etc. The same happens with the term "Simulation" that is used to compose Continuous Simulation and Dynamic Simulation, in addition to define other terms.

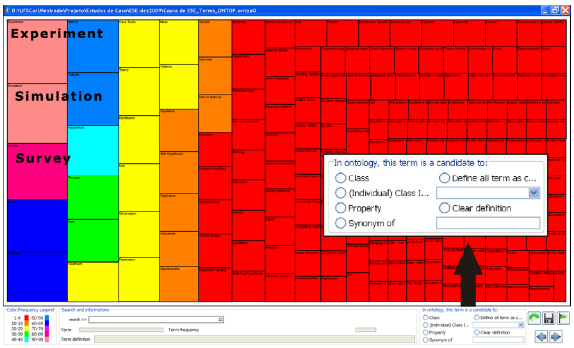

Figure 7: Screen for defining components.

Another contribution of visualization for defining the ontology components is related to the search resource. In this case, when the user classifies a term, the ONTOP-Tool uses this term as a key word for searching other terms that use the defined term in some way. As it is showed in Figure 8 the terms that satisfy the searching are highlighted in the screen. This fact allows that all these terms are classified at the same time, making easier the classification activity.

In Figure 8 all the terms that use "Validity" were highlighted when the user classified that term.

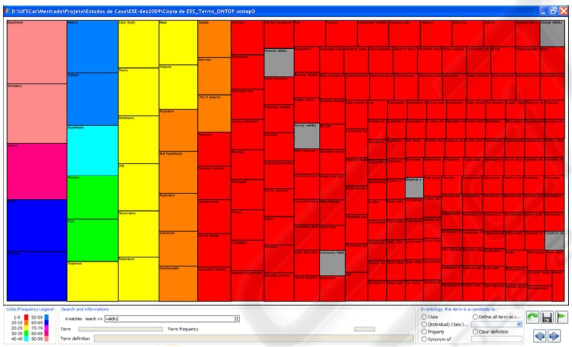

Figure 8: Terms highlighted after a searching.

As it happens during the refinement activity, the color of the boxes becomes fade as the terms are classified. In Figure 9 all the boxes have a fade color. This visual effect allows that the user identifies, quickly, the terms that were defined the ones that were not.

After all the terms were classified, the next step provided by ONTOP-Tool – *Step 3* – corresponds to the button "Define hierarchy" that should be used to organize, in an hierarchical way, the ontology classes defined in the previous step. This functionality uses a drag-and-drop interface which facilitates this operation.

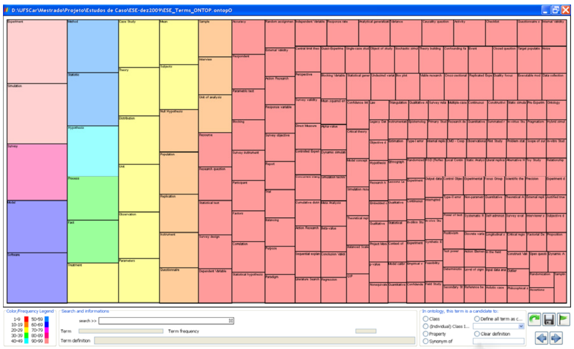

Figure 9: Visualization after the definition of all the terms.

Again, considering the ESE domain, the initial organization of the classes is presented in Figure 10. By means of the drag-and-drop resource the user can reach the organization showed in Figure 11 in a friendly way.

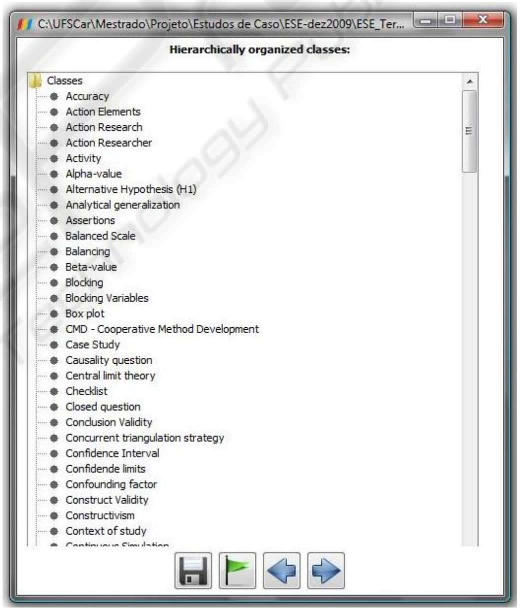

Figure 10: Initial hierarchy of the ontology classes.

Another resource provided by ONTOP-Tool is available through the *Step 4* that corresponds to the button "Define relationships" of Figure 4. The screen related to this functionality is presented in Figure 12. In this interface, the classes defined in Step 2 are presented on the left and on the right side of the screen. Between them it is presented a list of properties. These properties can be provided by the ONTOP-Tool or can be defined by the user in this occasion. The properties that are provided by the tool correspond to the ones that are frequently used by ontologies or the ones that were defined in Step 2.

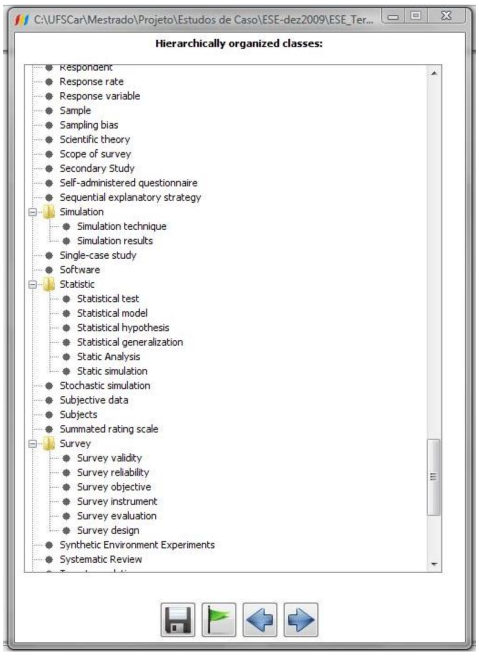

Figure 11: Final hierarchy of the ontology classes.

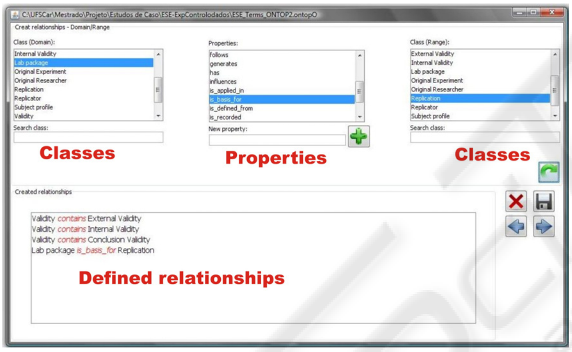

Figure 12: Screen for defining relationships.

The establishment of the relationships requires the following actions:

(i) select a class of the right list, for example, the Lab Package class;
(ii) select a property, for example, is_basis_for;
(iii) select a class of the left list, for example,the Replication class;
(iv) confirm the relationship;
(v) repeat the actions (i) to (iv) until all the relationships are established.

After these actions, the relationship "Lab Package *is_basis_for* Replication" was created.

We observe that ONTOP-Tool creates relationships of Domain-Range type. This kind of relationship indicates that the property links the individuals of the Domain class to the individuals of the Range class. The other kinds of relationship that are used in the context of ontology should be created by the tools that support ontology development, like Protégé-2000.

Finally, the last functionality provided by ONTOP-Tool corresponds to the *Step 5* and to the button "Create OWL file" of Figure 4. This functionality allows the creation of this file that contains all the information defined till now. The OWL file can be imported to several ontology editors. In our research we use Protégé-2000, versions 3.4 and 4.0.

6 CONCLUSIONS AND FURTHER WORKS

To sum up, this paper presented the ONTOP process which supports the conceptualization phase of an ontology development. The tools for ontology development identified in the field literature do not deal with the conceptualization of the domain since they focus on the implementation phase. ONTOP deals with this phase and it is supported by the ONTOP-Tool which facilitates the construction of a collaborative glossary as well as the identification of the ontology components.

Concerning the glossary construction, itself the target domain should be as representative as possible. To reach this objective it is essential that different views and suggestions are considered. This fact implies the involvement of different stakeholders, especially the domain specialists that are often geographically distributed. For the reason we decided to use the Moodle glossary for the fact that the Moodle environment is a free software that provides a good set of glossary management functionalities. By means of the Moodle environment the glossary is easily shared and validated by many stakeholders. Also, as it is possible to export the glossary as an XML file, the ONTOP-Tool provides an iteration activity that enhances its refinement.

Another aiding support provided by the ONTOP and ONTOP-Tool for the conceptualization phase (which is essential for every method that supports the ontology development, including Methontology) is visualization. This resource was adopted in light of two different purposes: to facilitate the glossary refinement (for example, making easier the identification of the definition of terms) and to facilitate the preliminary identification of the ontology components (for instance, using the size of the boxes to select possible components of the ontology).

An additional advantage of our proposal is the fact that, once ONTOP-Tool can automatically generate an OWL file at the end of the process, the next phases of the ontology construction may be carried out from the point where many definitions have already been done. To continue the ontology construction, the OWL file can be imported to a ontology editor like Protégé-2000, among others.

All things considered, it is important to finally point out that although we used the Experimental Software Engineering domain to exemplify the process and the tool, it was not our intention to present a deeper analysis of the ontology itself, but rather explore the ONTOP process and the ONTOP-Tool.

In our further studies, we intend to improve the ONTOP-Tool by adding linguistic processing so that semantic tagging can be used to enhance the identification phase of the ontology components. Another functionality that we intend to add to the ONTOP-Tool is the generation of an XMI file (XML Metadata Interchange). This file would allow classes, properties and relationships to be used by UML tools.

ACKNOWLEDGEMENTS

The authors would like to thank the Brazilian funding agencies FAPESP, CNPq, CAPES, the institute INEP and the Project *Observatório da Educação* for their support.

REFERENCES

Auvil, L.; Llorà, X.; Searsmith, D. & Searsmith, K. (2007). VAST to Knowledge: Combining tools for exploration and mining. In *IEEE Symposium on Visual Analytics Science and Technology, November 2007* (pp. 197-198). Philadelphia: IEEE Computer Society Press.

Chen, C.; Kuljis, J.& Paul, R. J. (2001). Visualizing latent domain knowledge. *IEEE Transactions on Systems, Man. and Cybernetics, Part C: Applications and Reviews*, 31(4), 518-529.

Corcho, O.; Fernández-López, M. & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point?. *Data & Knowledge Engineering*, 46(1), 41-64.

Daconta, M. C., Obrst, L. J. & Smith, K. T. (2003) *The Semantic Web: A guide to future of XML, Web Services and Knowledge Management*. Indianapolis: Wiley Publishing.

Fabbri, S.; Travassos, G. H.; Maldonado, J. C.; Mendonça Neto, M. G. M. & Oliveira, M. C. F. (Eds.). (2006). ESE Glossary: *Proceedings of ESELAW'06: Experimental Software Engineering Latin American Workshop*. Retrieved January 18, 2010, from Federal University of Rio de Janeiro, COPPE site: http://lens.cos.ufrj.br:8080/eselaw/proceedings/2006/proceedings2006.pdf

Falbo, R. A.; Menezes, C. S. & Rocha, A. R. (1998). A systematic approach for building ontologies. In *6th Ibero-American Conference on Artificial Inteligence, October 1998* (pp. 349 - 360). Lisbon, Portugal: Springer-Verlag.

Fernández-López, M.; Gómez-Pérez, A.; Pazos-Sierra, A. & Pazos-Sierra, J.(1999) Building a chemical ontology using Methontology and the ontology design environment. *IEEE Intelligent Systems & their applications,* 14(1), 37-46.

Gershon, N.; Eick, S. G. & Card. S. (1998). Information Visualization. *ACM Information Visualization Interactions,* 5 (2), 9-15.

Gómez-Pérez, A.; Fernández-López, M. & Corcho, O. (2004). *Ontological Engineering*. London: Springer Verlag.

Gruber. T. R. (1993). Toward principles for the design of ontologies used for knowledge sharing?. *Knowledge Acquisition*, 5(2), 199-220.

Ichise, R.; Satoh, K. & Numao, M.(2008) Elucidating Relationships among Research Subjects from Grant Application Data. In *12th International Conference Information Visualisation, July 2008* (pp. 427-432) New York: IEEE Computer Society Press.

Johnson, B. & Shneiderman, B. (1991). Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In *2nd Conference on Visualization, October 1991* (pp. 284-291). San Diego, California: IEEE Computer Society Press.

Moodle – Modular Object-Oriented Dynamic Learning Environment. (2009). Retrieved March 10, 2009, from http://moodle.org/

NewsMap - Application for Google News. (2009). Retrieved December 20, 2009, from http://newsmap.jp

Noy, N. F.; Fergerson, R. W. & Musen, M. A. (2000). The Knowledge model of Protégé-2000: combining interoperability and flexibility. In: *12th International Conference in Knowledge Engineering and Knowledge Management, January 2000* (pp. 17-32). Berlin, Germany: Springer-Verlag.

OWL - Ontology Web Language. (2009). Retrieved March 10, 2009, from http://www.w3.org/2004/OWL

Protégé-2000– Ontology Editor and Knowledge Acquisition System. (2010). Retrieved January 06, 2010, from http://protege.stanford.edu

TreeMap Tool. (2009). Retrieved February 7, 2009, from http://www.cs.umd.edu/hcil/ treemap

Wohlin, C.; Runeson, P.; Höst, M.; Ohlsson, M. C.; Regnell, B. & Wesslén, A. (2000). *Experimentation in software engineering - an introduction*.Sweden: Springer-Verlag.