(Draft) Conceptual Models for Urban Practitioners. Bologna: Società Editrice Esculapio, 2008, p. 69-84. ISBN 88-7488-208-4.

Generating an urban domain ontology through the merging of cross-domain lexical ontologies

J. Lacasta, J. Nogueras-Iso, F.J. Zarazaga-Soria, P.R. Muro-Medrano
Computer Science and Systems Engineering Department, University of Zaragoza,
María de Luna 1, E-50018 Zaragoza, Spain
{jlacasta, jnog, javy, prmuro}@unizar.es

Abstract: In order to classify resources, digital libraries have traditionally used different types of lexical ontologies, which describe the terminology used in an area of knowledge. This paper analyzes how lexical ontologies covering different areas of knowledge can be merged to generate an enriched urban terminology. This work proposes a method to combine these different perspectives into a single network of urban related concepts. The objective of this network is to facilitate a draft for a more formal (non lexical) urban domain ontology.

Key words: Urbanism, Lexical Ontologies, Ontology Mapping

1. Introduction

Urbanism is usually defined as the study of cities including their economic, political, social and cultural environment. As it can be observed from this definition, this discipline could be considered as an intersection of different domain areas such as economics, politics culture or civil engineering. One way to represent the knowledge behind urbanism is by means of the use of ontologies. The term ontology is used in information systems and in knowledge representation systems to denote a knowledge model, which represents a particular domain of interest. According to (Gruber, 1993) an ontology is "an explicit formal specification of a shared conceptualization". Therefore, given the multidisciplinary character of urbanism, the development of an urban domain ontology requires a revision of all the aforementioned cross-domain areas, capturing the concepts directly involved with the built environment of urbanism.

The purpose of this paper is to reproduce this exercise of revising and merging the knowledge from different domains in order to obtain a better definition of the urban domain. In particular, this work proposes a method for the definition of an urban domain ontology through the merging of thesauri representing the knowledge behind different domains. A thesaurus is a lexical ontology that defines a set of terms describing the vocabulary of a controlled indexing language, formally organized so that the a priori relationships between concepts (e.g., synonymous terms, broader terms, or narrower terms) are made explicit (ISO, 1986). The applicability of thesauri for search and retrieval in digital libraries has promoted the creation and diffusion of well-established thesauri in many different domains. Thus, a thesaurus can facilitate an important source of information when trying to analyze a specific knowledge area.

The proposed merging method takes as input a set of different multilingual lexical ontologies and obtains as a result a more consistent and formalized ontology. The main step of the merging process is the detection of intersections between concepts in the different lexical ontologies. This is performed using lexical similarity techniques that take advantage of the multilingual support given by the input lexical ontologies. Additionally, it is worth noting that the general concepts of the input lexical ontologies are pruned thanks to the use of a thesaurus specialized in urbanism. The output ontology can serve two important objectives. On the one hand, it can be used as a first draft for a more formalized urban domain ontology. On the other hand, the output ontology can be used to analyze to which extent urbanism is represented in the input thesauri.

The rest of the paper is organized as follows. Section 2 summarizes the state of the art in the comparison of lexical ontologies, the basis for the alignment and merging of ontologies. Section 3 describes the merging method for the generation of an urban domain ontology. Then, section 4 tests the feasibility of the method using EUROVOC (EUPO, 2005), GEMET (EEA, 2004), AGROVOC (Lauser et al., 2006) and UNESCO (UNESCO, 1995) as input lexical ontologies. All these resources have been established by well-known organizations and provide a shared conceptualization in the areas of economics, politics, culture and environment. The last section concludes and introduces some ideas on future work.

2. State of the art in comparison of lexical ontologies

In order to extract the urban related concepts and their relations from the analyzed thesauri, it is needed to be able to determine that two concepts of different thesauri are equivalent. This problem of finding relationships (e.g., equivalence or subsumption) between entities of different models is known as ontology alignment. In this area, many different alignment techniques that automatically identify similarities between concepts have been developed (Kalfoglou and Schorlemmer, 2003), but most of the used similarity measures are not adequate for lexical ontologies. The main mapping procedures are based in the following types of analysis:

- Similarity analysis of classes. This includes from simple string comparisons between the class names (Noy and Musen, 1999) to more sophisticated analysis that take into account lexical variants, demorphing or synonymy (Fernández-Breis and Martínez-Béjar, 2002).
- Similarity analysis of properties. This includes comparison in the number, names and types of the properties of the classes (Compatangelo and Meisel, 2002).
- Similarity analysis of relations. Analyze the similarity of names, types and structure of the relations between the classes (Kalfoglou and Schorlemmer, 2002).
- Similarity analysis between instances. Analyze the similarity of the property values of the class instances (Stumme and Maedche, 2001; Doan et al., 2002).

These different types of analysis are based in only two different procedures: the analysis of the linguistic similarity between labels and the analysis of the relation between classes. The differences between the available mapping procedures are the techniques used to identify the similarities and the types of analysis considered.

Lexical ontologies have some particularities in their structure with respect to other types of ontologies. They consist of a set of lexical concepts that share a reduced set of property types and relation types. For example, the thesauri structure described in ISO-2788 (ISO, 1986) and ISO-5964 (ISO, 1985) standards is reduced to alternative labels for a lexical term in different languages and a reduced set of possible relation types (*narrower-broader* and a general *related* type that provides little semantic).

As commented in (W3C, 2005), representing each concept of a lexical ontology as a different class produce several problems for its use in resource classification. Therefore, the most usual way to model lexical ontologies is represent each concept as an instance of a general "Concept" class, which define the available types of properties and relations. In this model, instances of a class can be directly used as values of properties in the description of a resource, and therefore, it has been used to create many different lexical ontology formats (Miles et al., 2005; Lauser et al., 2006; Miller, 1990). Its main drawback is that the provided ontological structure is very poor (only one class). The generalized use of this model to represent thesauri makes unnecessary the use of mappings techniques to relate their structure (they are equivalent). Therefore, all the mapping work has been focused in the analysis of similarity between instances.

An additional problem to compare lexical ontologies is the format used to represent them. For decades, the evolution of digital libraries has encouraged the use of lexical ontologies describing the terminology of an area of knowledge in the form of taxonomies, classification schemes or thesauri, promoting in that way the creation and diffusion of well-established collection in different domains. However, the lack of standardization has produced a huge variety of incompatible formats that increase the complexity of the comparison process.

3. Method for the generation of an urban domain ontology

Urbanism can be considered as an intersection of different domain areas such as economics, politics culture or civil engineering. In this context, the process to develop an urban domain ontology, providing explicit and formal specification of the knowledge behind the urbanism discipline, makes necessary to revise all these cross-domain areas and capture all the relevant concepts.

This section describes the process to capture the structure of relations between urban concepts through the analysis and comparison of cross-domain lexical ontologies with a thesaurus structure. The result obtained is a network of related urban concepts that shows the relevance of concept relations. Figure 1 remarks the different steps of the process, showing the inputs and the produced results. Four different tasks can be highlighted and are described in detail in the following subsections: the harmonization of the interchange format used for thesauri, the mapping of concepts, the generation of the network of urban concepts and the visualization of the generated network.

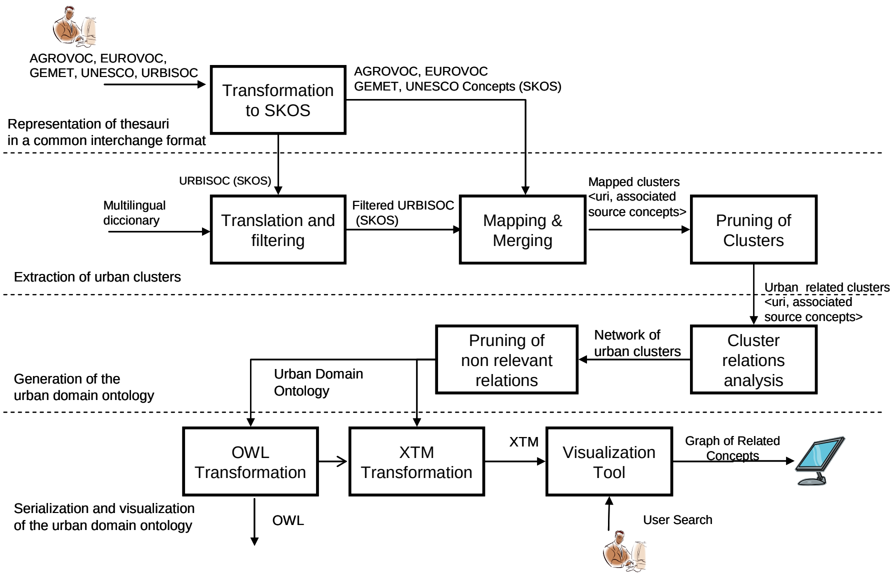

Figure 1: Work flow for the generation of an urban domain ontology

3.1. Representation of thesauri in a common representation format

The lack of a standardized representation format for thesauri has produced the spread of a great variety of incompatible formats. The system described here takes as input a set of thesauri of different knowledge areas. Therefore, to avoid format related issues in the process, thesauri must be provided in a single common format.

The British standards BS-5723 (BS, 1987), BS-6723 (BS, 1985) and their international equivalent (ISO-2788 and ISO-5964) propose models to manage lexical ontologies but lack a suitable representation format. Currently, the British Standards Institute IDT/2/2 Working Group is developing the BS-8723 standard that will be promoted to ISO when finished and whose 5th part will describe the exchange formats and protocols for interoperability of lexical ontologies. To establish the interchange format for lexical ontologies, the IDT/2/2 Working Group is involved in the SKOS project (Miles et al., 2005), a W3C initiative for the representation of simple knowledge organization systems such as thesauri, classification schemes and other types of controlled vocabularies (see figure 2). The involvement of the IDT/2/2 Working Group in the SKOS project will probably produce that the approved representation format will be SKOS related. Currently, several important thesauri of different areas are being transformed to this format. Having into account all those reasons, it has been the format selected as input of our analysis system.

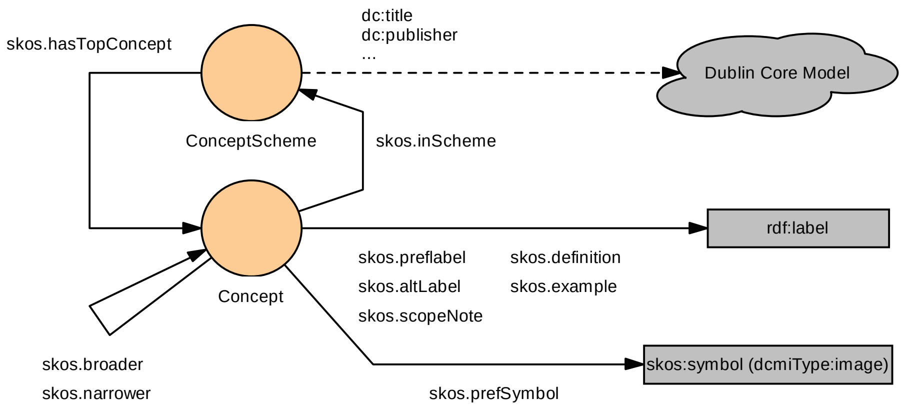

Figure 2: SKOS-Core Model

In order to facilitate the transformation of the different analyzed thesauri into SKOS, we have developed a customizable tool (developed in Java) that converts several file formats, according to a traditional thesaurus model, into the SKOS representation format. Figure 3 shows the mapping established by the tool between a classical thesaurus model and the SKOS representation model. The Unified Modelling Language (UML) notation is used for the representation of the two models. As it can be observed from the figure, the following transformations are applied:

- A concept scheme is created to represent the source thesaurus.
- Each thesaurus preferred term generates a new concept in the SKOS representation (except if it is not used for classification).
- Each translation derives a new preferred label in the language of the label.
- Each term related by a UF/USE relation (synonymy relation) is converted into an alternative label of the related concept.
- The RT relations between terms are converted to *skos:related* relations between the corresponding concepts. The same happens with the BT and NT relations that are converted to *skos:broader* and *skos:narrower*.
- The description of a term is converted into the definition of its associated concept.
- The concepts whose associated term is marked as TT are included in the concept scheme as top terms.
- Another important item is the URI that must be created for each SKOS concept. It has to be generated by using the information provided by the source format. In the example, the term value used as the preferred label of the concept can be converted into an URI by adding an *http://* prefix.

It is worth noting that this tool can be customized to different source formats (text files, relational databases). It provides a common infrastructure for the parsing of source formats and the writing of SKOS output files (in RDF format). Therefore, the support of a new source format is reduced to the development of a new plug-in component. These plug-ins are simple Java classes which conform to a specific interface and extend the functionality of an abstract class to deal with the specific aspects of each new format.

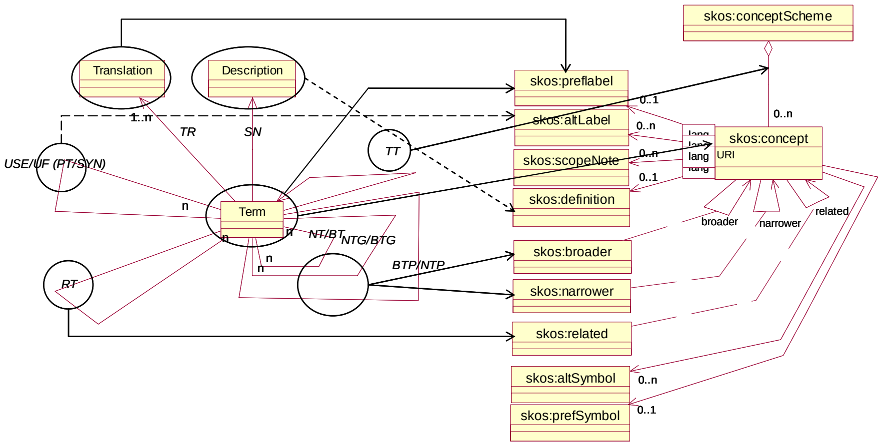

Figure 3: Mapping between a traditional thesaurus model and the SKOS-Core model

Once the thesaurus has been transformed, different statistics about the number of concepts and relations of the source and destination format are computed to verify if the obtained SKOS is correct. SKOS presents many restrictions that have to be validated to assure that the generated file is correct. Among them, the following can be highlighted:

- At least one *ConceptScheme* must exist, and all the contained concepts must refer to a *ConceptScheme*.
- Every concept must have one broader concept except when it is a top term. Then, it must be referenced as *top concept* in the *ConceptScheme* structure and not have a broader term.
- Each concept must have one and only one preferred label for each available language that is unique along the thesaurus.
- All the relations between concepts must reference existent concepts (orphan relations are not allowed).
- The structure of *broader*/*narrower* relationships must not contain cycles.
- The *related* relation is symmetric, so if “A” is *related* with “B”, “B” must be *related* with “A”.
- The *broader* relation is the inverse of the *narrower* one, so if “A” is the *broader* of “B”, “B” must be *narrower* of “A”.

3.2. Extraction of urban clusters

The objective in this step is to extract the concepts related to urbanism from the analyzed thesauri. In order to search urban concepts in the cross-domain thesauri, a set of terms of a thesaurus specialized in urbanism is used as seed for this search. In addition, the relations between the concepts present in the urban thesaurus are used in the next step as a base for the construction of urban domain ontology.

As commented previously, from the available mapping techniques, only the analysis of the values of the properties of the instances is useful. Here, the linguistic similarity between the preferred and alternative labels has been considered for the mapping. The analysis of other properties as definition, scope notes and relations is left for future work.

In the mapping process, every concept of every thesaurus (including the urban one) is compared with every concept of the other thesauri to find equivalences. Two concepts are considered equivalent when at least one of the labels of a concept (preferred and alternatives) is equal to a label in the other concept. Here, the use of multilingual thesauri has the advantage of having labels in different languages to compare (the labels used to describe two concepts may differ in a language but be equivalents in other one). In order to improve the results, plurals, accents and capital letters have been removed. This approach could be enriched with misspellings detection, stemming and word order analysis among others, but given the strict rules used to define the labels used in a thesaurus no much improvement would be expected.

Equation 1 measures the relevance of the mapping obtained between two concepts. The higher the number of labels two concepts share from the total they have, the higher is their equivalence. This equation can be applied to each obtained mapping, but it is used here to analyze the quality of the thesaurus mapping with respect to the urban one showing the relevance that each different knowledge area gives to urbanism (see the experiments in section 4).

$$probabilityOfEquivalence = \frac{2*numberOfMatchedLabels}{totalNumberOfLabelsInTheTwoConcepts} \qquad (1)$$

Each set of mapped concepts is grouped into a cluster (group of equivalent concepts), which is identified with one of the URIs of the original concepts. Figure 4 shows a simplified example of the cluster generated for the “Zonas Urbanas” concept (“Urban areas”) mapping different thesauri and considering only the Spanish labels. In the example, it can be seen that the “Area Urbana” concept of GEMET is included in the cluster thanks to the presence of this label in the concepts of EUROVOC and AGROVOC. In addition, the relevance of the mappings is included to show that some of them are stronger than others.

Not all the clusters obtained in the mapping process are useful; many contain concepts not related to urban terminology. Therefore, only the clusters that contain a concept from the urban thesaurus and those with at least a concept directly related (*broader, narrower* and *related* relations) to another one in a cluster of the first case are stored. The rest are considered as not relevant to urbanism and they are pruned from the system. To maintain the consistency, the relations of the remaining concepts with the deleted ones are also eliminated.

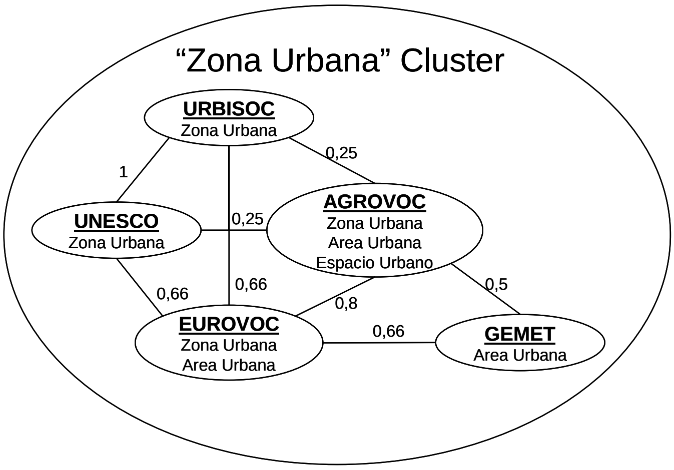

Figure 4: Example of a cluster derived from the mappings

The reason to include clusters that do not directly contain urban terminology is to provide an extension of urban concepts from the point of view of other areas of knowledge.

3.3. Generation of the urban domain ontology

The clusters generated in the previous step describe the urban terminology used in different knowledge areas, but not how this knowledge is inter-related. This subsection describes the process used to relate these clusters generating a network of urban concepts that can be seen as an urban domain ontology.

The relations of the concepts contained in each cluster are used as a basis for the generation of the relations between clusters. Here, in addition to the basic relations (*broader*, *narrower* and *related*), *sibling* relations (*narrower* of its *broader*) have been also considered. Other relations as the *grandparent* (*broader* of its *broader*) or *grandchildren* (*narrower* of their *narrower*) could be considered as well, but they are less relevant and not taken into account in this paper.

Each relation of a concept in a cluster with other concept from another cluster is marked as a relation between the clusters with the type of relation between the concepts. When two or more concepts of one cluster are related to two or more concepts of another cluster, the relation between these two clusters is marked with the different types of the original relations between the concepts, along with the number of occurrences of each different relation type. The marks indicate us the relevance of the relations between the clusters. That is to say, the more concepts inside the two clusters are related, the more relevant is the relation between these clusters. The result of this inter-cluster relation process is a network of interconnected urban clusters.

Given that all the concepts from the urban thesaurus have been included, the obtained network is based on the structure of relations of this thesaurus. But, in addition to this, it includes the relations and relevance derived from the merging of the source thesauri. See figure 5 in the experiment section as an example of the obtained network.

In many situations, it is not interesting to have a network with all the existent relations but only the most important ones. Therefore, a process to prune the less relevant relations has been created. This process receives as input the complete network of concepts and a weight threshold to determine if a relation is maintained. All the relations with a weight below the threshold are pruned. After the pruning, all the clusters that do not have at least one relation with another one are also eliminated.

3.4. Serialization and visualization of the urban domain ontology

For the serialization of the generated structure, we have proposed the use of the Web Ontology Language (OWL) (Bechhofer et al., 2004) and XTM format (Pepper and Moore, 2001). On the one hand, OWL is a widely accepted language for the definition of formal ontologies based on RDF. On the other hand, XTM is a format for the exchange of topic maps with an emphasis on the find-ability of information. We have selected XTM because of its advantages for the visualization and navigation through the generated network of concepts. It can be easily visualized by a wide range of tools compliant with this format. For instance, we have selected the TMNAV tool created in the TM4J project (TM4J, 2001) but other tools could also have been used.

4. Testing the method in the urban domain

The process described previously has been used to generate a network of urban concepts using GEMET, AGROVOC, EUROVOC and UNESCO as thematic thesauri. These thesauri provide a shared conceptualization in the areas of economics, politics, culture and environment: EUROVOC is a multilingual thesaurus covering the fields in which the European Communities are active (it provides a means of indexing the documents in the documentation systems of the European institutions and of their users); GEMET is a thesaurus for the classification of environmental resources developed by the European Environment Agency and the European Topic Centre on Catalogue of Data Sources; AGROVOC is a specialized thesaurus for the classification of geographic information resources (with special focus on agriculture resources), which has been created by the Food and Agriculture Organization of the United Nations; and UNESCO is a general purpose thesaurus for use in the indexing and retrieval of information in the UNESCO Integrated Documentation Network. The different origins and objectives of these thesauri provide different views of the urban terminology they contain.

From the available thesauri about urbanism, URBISOC (Alvaro-Bermejo, 1988) was selected as a basis for the filtering of urban terminology. URBISOC has been developed by the Spanish National Research Council to facilitate classification at bibliographic databases specialized in scientific and technical journals on Geography, Town Planning, Urbanism and Architecture. This thesaurus contains around 3,600 different concepts labelled in Spanish.

These five thesauri have been published in completely different representation formats. UNESCO and AGROVOC are stored in a database format, but each one with a different database management system and a different table structure; EUROVOC is provided in an XML based format with a specific structure of files and XML tags; GEMET can be obtained in an SKOS based format but with some extensions that make it incompatible with SKOS schema; and URBISOC is provided in the web directly in HTML format. Therefore, the transformation to SKOS has been needed for all of them.

As commented in the process description, the mapping system works better when different labels in different languages are available for each concept. GEMET, EUROVOC, AGROVOC and UNESCO are in Spanish, English, and French between others. However, URBISOC is only provided in Spanish. This produces a limitation in the possibilities of filtering urban concepts (only Spanish labels can be compared). Therefore, before using it in the mapping system, the Spanish labels have been translated into English and French using a multilingual dictionary.

Although the use of a dictionary to translate the labels introduces errors caused by synonymy problems it is believed that the specific character of the urban thesaurus limits the problems of polysemy. The use of a multilingual dictionary has only been needed by the lack of a multilingual thesaurus suitable for filtering urban terminology. If a suitable thesaurus were used, the translation step could be skipped.

Analyzing the clusters generated by the mapping system we have found that some clusters with URBISOC concepts did not have any specific connection with urbanism. It was caused by the inclusion in URBISOC of very general terminology not specifically related to urbanism. To solve this problem, we decided to get a subset of URBISOC without general concepts by selecting as a start point the “Urban planning” concept (very frequent in urbanism) and adding recursively all the concepts related by *narrower* and *related* relations. The concepts of URBISOC not found through these relations were discarded. This process reduced the size of URBISOC from 3,609 to 3,091 concepts, eliminating most of the general terminology.

Name	Concepts	Mapped Concepts	Mapped Percentage	Liability
GEMET	5244	834	15.9%	0.611
EUROVOC	6649	935	14.06%	0.394
AGROVOC	16896	772	4.57%	0.454
UNESCO	4424	802	18.13%	0.578

Table 1: Relevance of urbanism in cross-domain thesauri

Using this reduced version of URBISOC, the mapping process produced the results shown in table 1. This table shows the number of concepts of each thesaurus, the number of concepts mapped to URBISOC, the percentage of each thesaurus that has been mapped, and the average relevance of the mappings according to equation 1 (section 3.2). It can be seen that around 1 out of 3 urban concepts (3,091) have been mapped to each thesaurus. The percentage of urban terminology in each thesaurus may not seem very high (18% as maximum), but having into account that they are thesauri specialized in other areas of knowledge, these percentages are quite relevant.

Figure 5 displays a screenshot of the TMNAV tool visualizing part of the generated network. It shows the related concepts around the concept “urban population”. The network contains all the relations that have been found, but as commented in the description of the process and shown later, the less relevant ones could have been deleted to obtain a smaller structure.

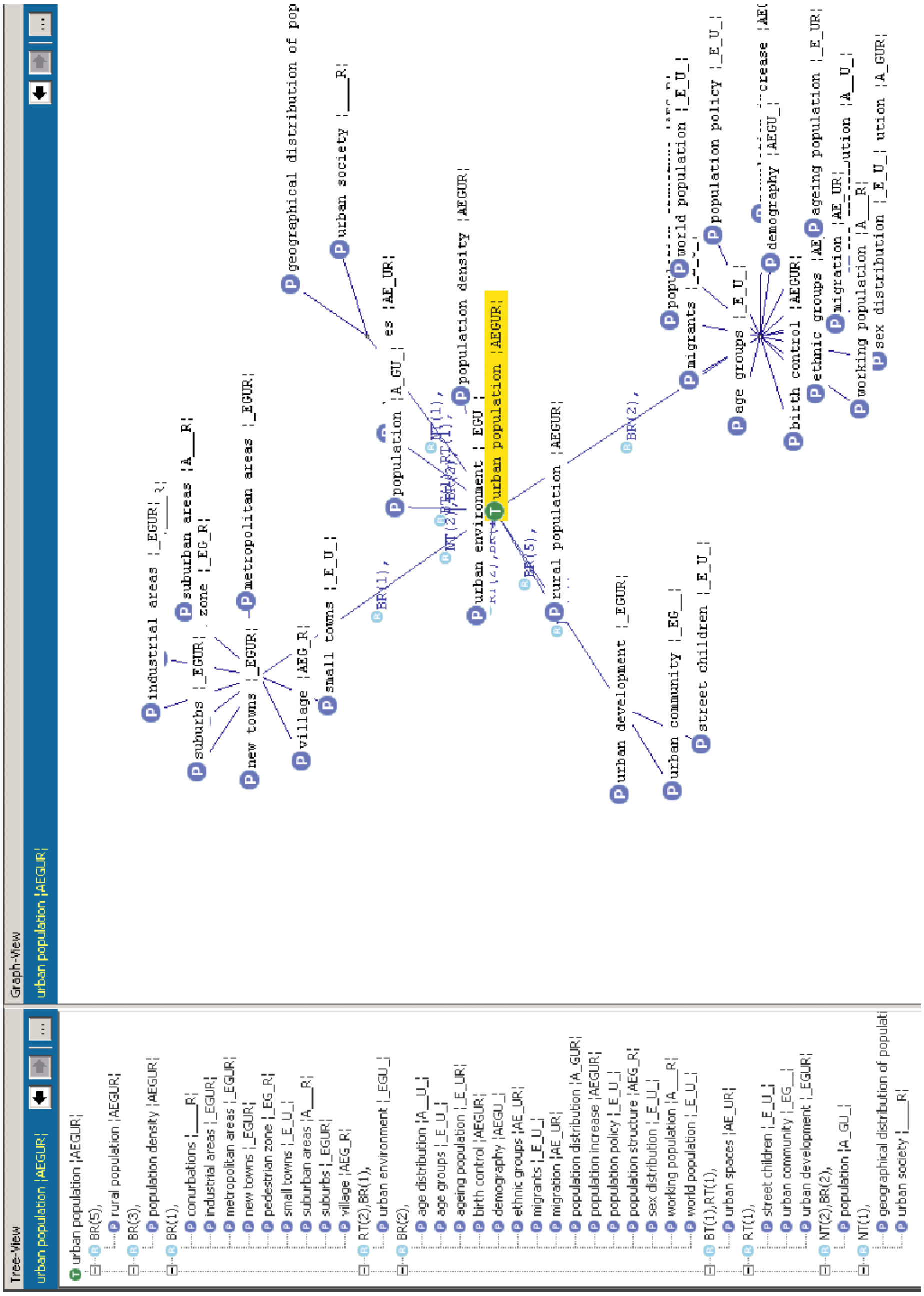

Figure 5: Visualization of a part of the generated urban domain ontology

In the figure, each cluster includes the English label of one of the concepts contained inside (the selection order for the label is UNESCO, GEMET, EUROVOC, AGROVOC and URBISOC), and the initials of the thesauri where the concepts contained in the cluster are from (i.e. AEGUR means that the concepts exist in AGROVOC, EUROVOC, GEMET, UNESCO and URBISOC; _E__R means that the concept does exist only in EUROVOC and URBISOC). The relations are labelled indicating the types and number of the relations found. Here, BT indicates *broader* relation, NT is *narrower*, RT means *related* and BR is used for *siblings*. The total number of relations could also be added but it has been omitted for the sake of clarity. In the example, the "Urban Population" cluster contains concepts of all the thesauri and maintains 2 relations with the "Urban spaces" concept, one of type *related* and another one of type *sibling* (weight 2). This relation is less relevant than the one connecting to the "Rural population" concept where 5 *sibling* relations have been found (weight 5).

Minimum Weight	**Clusters**	**Cluster Size**	**Relations**	**Cluster Relations**
1	6200	2.28	113888	18.36
2	4341	2.66	42823	9.86
3	2878	3.03	17570	6.10
4	2086	3.24	11333	5.43
5	1353	3.54	5622	4.15

Table 2: Size of the network of urban concepts

The output network contains 6,200 concepts with 5,622 relations of weight 5 or greater, 5,711 of weight 4, 6,237 of weight 3, 25,253 of weight 2 and 71,065 of weight 1. Table 2 shows a summary of the size of the generated network when the less relevant relations are pruned. Each row shows the size of the network that includes all the relations of at least "Minimum Weight" weight. Each row shows the number of clusters, the average size of each cluster, the total number of relations between clusters, and the average number of relations of each cluster.

5. Conclusions

This paper has shown a process to generate a network of concepts related to urbanism. The objective is to use this network as a basis for a first draft of an urban domain ontology. The process uses as input a set of multilingual thesauri from different knowledge areas (e.g., GEMET, AGROVOC, UNESCO and EUROVOC) and a thesaurus specialized in urbanism (URBISOC) to be able to select the urban terminology present in the other thesauri.

The main steps of the generation process are the harmonization of the input formats, the mapping between the concepts to generate clusters of equivalent concepts using linguistic similarity measures, and the establishment of relations between the clusters on the basis of the original relations between the concepts contained in different clusters. Finally, in order to facilitate the visualization and reusability of the generated output, it is transformed into XTM and OWL formats.

In the experiment, we found that URBISOC, the specific thesaurus for the filtering of concepts related to urbanism, contains very generic concepts in the top part of the broader-narrower hierarchy. Therefore, these general terms had to be removed before using it as a filtering mechanism.

As regards the mappings established between the source thesauri to obtain the clusters of concepts, we could observe that urban terminology is a relevant part of the analyzed thesauri (up to 18% in UNESCO). Future work will improve the used mapping techniques by having into account the structure provided by the relations between the concepts.

In addition, we have shown, through the experiment, how to reduce the size of the generated network of concepts by pruning the less relevant relations. This pruning is able to reduce the size of the network from 6,200 concepts to only the 1,353 more related. A future improvement as concerns the relations between clusters is to take into consideration the *grandparent* and *grandchildren* relations between thesaurus concepts. The objective is increasing the relevance values of some of the existent relations. For example, two concepts in a thesaurus can be directly related through a narrower relation, however, in other one they may be related through an intermediate concept.

The urban domain ontology obtained as a result of the method proposed has several advantages in comparison with the thesauri used as source and the thesaurus used for filtering in the following areas:

Consensus and focus: The concepts of the resulting network have been selected by consensus thanks to the mappings among the different sources, removing those concepts that are neither common nor focused on urbanism.

Relations: With respect to the relation structure, the total number of available relations is bigger than the existent ones in each of the original sources. Besides each relation has a weight that indicates its relevance. As future work, the semantics of these relations should be enriched. The information provided by definitions, examples, and naming patterns in the properties of the original concepts should help to refine the current relations (e.g., broader relations could be refined as *part of*, *instance of* or *generalization* relations).

Multilingual support: Thanks to the combination of different sources of knowledge with multilingual support, the output network is enriched with alternative terminology in different languages.

Formalism: Since the output network has been generated using a formal language such as OWL, we have increased its usability, facilitating the work with reasoning engines.

Finally, it must be noted that, apart from serving as a first draft for an urban domain ontology, the generated network of urban concepts can be directly applied in information retrieval systems for resource classification, thematic indexing or query expansion.

Acknowledgements

This work has been partially supported by the University of Zaragoza through the project UZ 2006-TEC-04.

References

Alvaro-Bermejo, C., 1988. Elaboración del Tesauro de Urbanismo URBISOC. Una Cooperación Multilateral. In: Encuentro Hispano-Luso de Información Científica y Técnica. II. Salamanca.

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., Stein, L. A., February 2004. OWL Web Ontology Language Reference. W3C, W3C Recommendation 10 February 2004. http://www.w3.org/TR/2004/REC-owl-ref-20040210/

BS, 1985. Guide to establishment and development of multilingual thesauri. BS 6723, British Standards Institute (BSI).

BS, 1987. Guide to establishment and development of monolingual thesauri. BS 5723, British Standards Institute (BSI).

Compatangelo, E., Meisel, H., 2002. Intelligent support to knowledge sharing through the articulation of class schemas. In: Proceedings of the 6th International Conference on Knowledge-Based Intelligent Information & Engineering Systems. Crema, Itally.

Doan, A., Madhavan, J., Domingos, P., Halevy, A., 2002. Learning to map between ontologies on the semantic web. In: The Eleventh International WWW Conference. Hawaii, US.

EEA, 2004. GEneral Multilingual Environmental Thesaurus (GEMET). Version 2004. European Environmental Agency (EEA). http://www.eionet.eu.int/GEMET

EUPO, 2005. European Vocabulary (EUROVOC). European Union Publication Office. http://europa.eu/eurovoc/

Fernández-Breis, J. T., Martínez-Béjar, R., 2002. A cooperative framework for integrating ontologies. Int. J. Hum.-Comput. Stud 56 (6), 665–720.

Gruber, T., 1993. A translation approach to portable ontology specifications. ACM Knowledge Acquisition, Special issue: Current issues in knowledge modeling 5, Issue 2 (KSL 92-71), 199–220.

ISO, 1985. Guidelines for the establishment and development of multilingual thesauri. ISO 5964, International Organization for Standardization (ISO).

ISO, 1986. Guidelines for the establishment and development of monolingual thesauri. ISO 2788, International Organization for Standardization (ISO).

Kalfoglou, Y., Schorlemmer, M., 2002. Information flow based ontology mapping. In: 1st International Conference on Ontologies, Databases and Application of Semantics (ODBASE'02). Irvine, CA, USA.

Kalfoglou, Y., Schorlemmer, M., 2003. Ontology mapping: The state of the art. The Knowledge Engineering Review 18 (1), 1–31.

Lauser, B., Sini, M., Salokhe, G., Keizer, J., Katz, S., 2006. Agrovoc Web Services: Improved, real-time access to an agricultural thesaurus. Quarterly Bulletin of the International Association of Agricultural Information Specialists (IAALD) 1019-9926 (2), 79–81.

Miles, A., Matthews, B., Wilson, M., September 2005. SKOS Core: Simple Knowledge organization for the WEB. In: Proceedings of the International Conference on Dublin Core and Metadata Applications. Madrid, Spain, pp. 5–13.

Miller, G. A., 1990. WordNet: An on-line lexical database. Int. J. Lexicography 3, 235–312.

Noy, N. F., Musen, M. A., 1999. Smart: Automated support for ontology merging and alignment. In: Twelth Workshop on Knowledge Acquisition, Modeling, and Management. Banff, Canada.

Pepper, S., Moore, G. (Eds.), 2001. XML Topic Maps (XTM) 1.0. TopicMaps.org. http://www.topicmaps.org/xtm/1.0/

Stumme, G., Maedche, A., September 2001. Ontology merging for federated ontologies on the semantic web. In: In Proceedings of the International Workshop for Foundations of Models for Information Integration (FMII-2001). Viterbo, Italy.

TM4J, 2001. Homepage of the TM4J proyect. http://tm4j.org/

UNESCO, 1995. UNESCO Thesaurus: A Structured List of Descriptors for Indexing and Retrieving Literature in the Fields of Education, Science, Social and Human Science, Culture, Communication and Information. UNESCO Publishing, Paris. http://www.ulcc.ac.uk/unesco/

W3C, April 2005. Representing classes as property values on the semantic web. http://www.w3.org/TR/swbp-classes-as-values/