(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743

Transformation of a keyword indexed collection into a semantic repository: applicability to the urban domain

Javier Lacasta[1], Javier Nogueras-Iso[1], Jacques Teller[2], and Gilles Falquet[3]

[1]Computer Science and Systems Engineering Dept. Universidad de Zaragoza, Spain
[2]LEMA, Université de Liège, Belgium
[3]Centre universitaire d'informatique, Université de Genève, Switzerland

Abstract. In the information retrieval context, resource collections are frequently classified using thesauri. However, the limited semantics provided by thesauri restricts the collection search and browsing capabilities. This work focuses on improving these capabilities by transforming a set of resources indexed according to a thesaurus into a semantically tagged collection. The core mechanism for building this collection is based on the conversion of the domain specific thesaurus (indexing the collection of resources) into a domain ontology connected to an upper level ontology. The feasibility of this work has been tested in the urban domain by transforming the resources accessible through the European Urban Knowledge Network into a Linked Data repository.

1 Introduction

In the information retrieval context, the resources of a collection are frequently classified and searched using the concepts of thesauri. However, the limited semantics they provide reduce its usability for search and browsing in a collection. Thesauri can be used to expand queries to a collection by including narrower concepts of the selected one (they are about the same theme), but only in a rough way since the lack of semantics in their relations increase the heterogeneity of criteria and interpretations.

With the objective of improving the search and browsing capabilities of a collection, this paper proposes a process to convert a thesaurus indexed collection into a semantically tagged collection stored in a semantic repository. The process to perform this transformation is based on the transformation of the thesaurus used to classify the collection into an ontology. Formal ontologies provide a more detailed structure with improved reasoning capabilities. They describe in detail the meaning of each of the included concepts and the specific types of relations held between them. However, it is a waste of effort to create new models from scratch if a thesaurus already exists in the desired application domain. It is much more suitable to convert them into an ontology by adding the semantics they lack.

The mechanism proposed for the transformation of a thesaurus into an ontology is based on linking it to DOLCE ontology [10] and using this linkage to

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743

refine and extend the original thesaurus model. The method is applied in the urban domain to transform the EUKN resources[1] into a semantically tagged collection. The EUKN thesaurus has been formalized following the described process and the collection has been republished using a Linked Data web service. Linked Data models provide RDF resources and relations between them as valid HTTP URIs, facilitating in that way the access and browsing through the information. This transformation provides, in addition to the thematic entry point to the collection, a more abstract point of view focused on DOLCE categories (activities, events, rational-agents ...) and a set of extended relations between the resources.

The rest of the paper is structured as follows. Section 2 reviews the state of the art. Section 3 introduces the proposed method. Section 4 shows the applicability of the method to the urban domain. Finally, this paper ends with some conclusions and outlook for future work.

2 State of the art in the generation of semantic models

This section reviews the main works dealing with the addition of semantics to collections of resources. It analyzes the works focused on converting resource descriptions into semantic networks, and those related to the formalization of the terminological models used in their classification.

With respect to the conversion of sets of resources into semantic networks, Hearst et al. [5] review the existing approaches. They describe alternatives for searching and browsing collections focusing on the use of faceted search/browsing components based in controlled knowledge models to help to guide the user in the location of the desired resources. In this same field, Hyvöenen [6] describes a content creation process that includes the manual transformation of the thesaurus used for classification of a collection (*Museoalan asiasanasto*) into an ontology and the generation of the collection records in RDF.

Without the final objective of transforming a set of resources into a semantic model (but that can also be applied for this task), there are works that analyze other alternatives for the formalization of thesaurus. In this area of work, Tudhope et al. [13] analyze the specialization of the associative thesaurus relations into richer subtypes to find new application possibilities for retrieval. Similarly, Golbeck et al. [3] describe the process used to transform the National Cancer Institute (NCI) thesaurus into an ontology. They show the rules applied to transform each concept, property and relation of the thesaurus into formal equivalents. Other works go a bit further by proposing semiautomatic processes to perform these formalization tasks. Kawtrakul et al. [8] and Soergel et al. [12] present a (semi-)automatic process to refine the relations of the AGROVOC thesaurus based on the analysis of the AGROVOC categories that classify the thesaurus concepts. They are used to establish abstract relations that are applied as general transformation rules for their member relations. In a similar

[1] http://www.eukn.org/E_library

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743

way, Khosravi and Vazifedoost [9] propose a re-engineering process of the ASFA Persian thesaurus using automatic ontology learning methods. It is based on the definition (by experts in the field) of general and specific rules used to transform the thesaurus relations.

The need of formalization is not restricted to thesauri. Collections using other semi-structured or unstructured knowledge models for classification also benefit from the formalization of the used models. For example, Van Damme et al. [2] show how folksonomies and other unstructured vocabularies can be used to construct ontologies. They describe a multiple approach for deriving ontologies from folksonomies, such as the statistical analysis of the folksonomies, the use of online lexical and semantic web resources, ontology matching (and mapping) approaches, and the computer guided human review. Finally, focusing on controlled lists of terms, Aleksovski et al. [1] propose a method to match two unstructured lists of terms through a background ontology using different disambiguation and heuristic techniques.

3 Providing semantics to a keyword indexed collection

The proposed method for adding semantics to a collection of resources is described in figure 1.

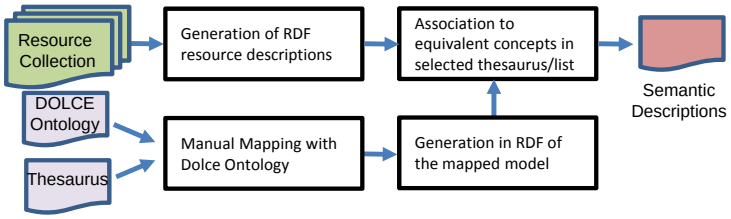

Fig. 1. Generation process of a semantic description of a resource collection

The first step is the generation in RDF of the resource descriptions (or to process them into a structured model if they already exist). This step is dependent on the original source (e.g., HTML, XML, word documents...) and the techniques and processes required to do that can be completely different. The objective is to obtain for each resource a Dublin Core [7] metadata record expressed in RDF containing at least: the resource name, a description, a set of keywords that classify them, and a reference to the original resource. The Dublin Core metadata model was selected for resource description because its extensive use in the digital library field facilitates the access to resources from a wide audience. Next, those fields whose values can be linked to a knowledge model (e.g., keywords, authors, dates, locations) are processed to convert them into independent RDF items linked to Dublin Core RDF resources. Due to lexical heterogeneity of labels (e.g., plurals, misspellings errors...), in order to avoid

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743

the creation of duplicated resources, it is needed to harmonize the labels. Figure 2 exemplifies how Dublin Core metadata containing controlled values can be transformed into independent resources. Additionally, the requirements of the library used later in the construction of a Linked Data based system force all the relations between resources to be bidirectional (i.e., they have an inverse).

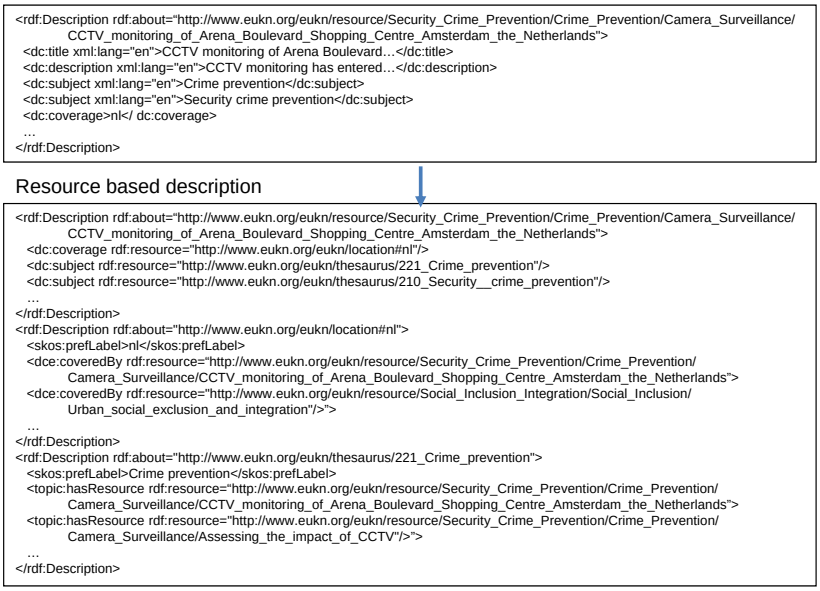

Fig. 2. Example of transformation of properties into resources

Once the descriptions of the resources have been processed, the next step is to replace their (string) subjects with equivalent concepts from well-known knowledge models (e.g., a thesaurus) that provide a better organization and description of concepts. The required matching process is similar to the one performed to identify equivalent properties in different Dublin Core descriptions. The main difference comes from the existence in the knowledge models of different alternative names (synonyms), descriptions, and similarity relations that can be used to help in the matching process. This improves the collection organization; however, the semantics provided by a simple knowledge models is limited. If no description or scope note is provided, a concept meaning can be ambiguous. Additionally, the hierarchical relations contained in thesauri only indicate a general meaning containment and the associative ones only provide an abstract level of relatedness. The formalization of these models helps to identify each concept meaning and the relations between them.

The approach adopted to formalize a simple knowledge model has been to interrelate it with other already existent formal models. As base for the formalization process it has been decided to use a top level ontology such as DOCLE, a

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743

model focused on describing data types and general relations independent of the context [4]. Mika et al [11] describe some reasons why knowledge models should be linked to a top level ontology. On the one hand, it provides additional semantics about the concepts and relations to determine if the model is coherent. On the other hand, it aggregates the concepts and relations into abstract categories that can be used to automate the establishment of domain oriented relations. The objective here is to link the thesaurus with the top level ontology through *subclass-of* relations. Using DOLCE as top level ontology, it is possible to perform a manual matching process between the desired thesaurus concepts and the DOLCE classes contained in the *perdurant*, *endurant*, and *quality* branches. *Perdurants* comprise events, processes, phenomena, activities and states; *endurants* describe entities that maintain their identity along the time, although their properties (e.g., color, size...) may change; finally, *qualities* provide entities that can be perceived or measured (e.g., shape, color).

Once the matching has been performed, the relations between the concepts in the original model (e.g., hierarchical and associative relations) can be automatically refined through inference of the corresponding relations in the mapped ontology. For example, in DOLCE ontology it is described that two *physical-objects* may hold a *part-of* relation. Therefore, *ICT infrastructure* is a narrower of *Technical infrastructure* and those concepts can be classified as DOLCE *physical-objects*, it can be automatically deduced that they hold a *part-of* relation between them.

The resulting model containing the collection resources linked to a formalized thesaurus can be directly stored in a semantic repository such as SESAME or JENA to allow the execution of semantically rich queries and the creation of faceted guided search systems. For example, the DOLCE enriched ontology can be used to search activities, rational-agents, regulations and so on. The publication of the collection as Linked Data is immediate once it is stored in a semantic repository. Tools as PUBBY allow the creation of Linked Data services from a semantic model by transforming the URIs of the resources into valid URLs.

4 Applicability of the method to the EUKN collection

The described process has been applied to the European Urban Knowledge Network (EUKN) collection with the objective of adding semantics to it and providing alternatives for search and browsing. The following subsections describe: the features of the EUKN collection and its thesaurus; the results of applying the transformation method to link the EUKN thesaurus to an upper level ontology; and the publishing process of the EUKN collection as Linked Data together with a discussion of the advantages of the new approach for searching and browsing the collection.

4.1 The EUKN collection and its thesaurus

The European Urban Knowledge Network (EUKN) was born in 2004 as a pilot project of different European states to enhance the exchange of knowledge and

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743

expertise on urban development. Nowadays, it is an intergovernmental knowledge network that acts as hub for existing networks of urban practitioners, researchers and policy-makers at all governmental levels. It provides a high-quality knowledge database in the urban field with more than 2,000 resources classified according to a thesaurus that contains 267 concepts to describe the management and control of physical, economic and social forces on urban areas. However, there is a disconnection between the thesaurus creators and their users because only 146 (54%) of those concepts are used in the collection. It may be caused by structural deficits that difficult the location, identification, and usability of the thesaurus concepts. For example, it contains some replicated concepts, the concepts lack a detailed definition and/or scope notes, and the criteria for its hierarchical structure is sometimes unclear (e.g., *Mediation* as narrower of *Community development*).

4.2 Transformation of the EUKN thesaurus into an ontology

The transformation of the EUKN thesaurus into an ontology has been performed linking it to DOLCE top level ontology as described in section 3. The resulting mapping uses 18 DOLCE classes (no more classes have been needed). However, in order to minimize the heterogeneity, the classes being very similar or with few associated concepts (e.g., *planning activities* or *geographical-objects*) have been integrated as part of their corresponding super-classes.

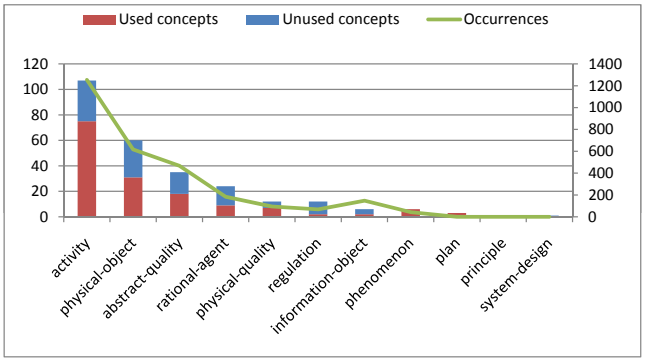

Fig. 3. Analysis of the use of EUKN concepts matched with DOLCE categories for the classification of resources in the collection.

Figure 3 analyzes the results of the thematic association of DOLCE categories with the EUKN concepts. It shows the number of thesaurus concepts associated to each DOLCE category and how the concepts in these clusters are used in the EUKN collection. We have obtained the following classification: 107 *activities*, 61 *physical-objects*, 45 *abstract-quality*, 24 *rational-agents*, 18 *regulations*, 7 *information-objects*, and 6 *phenomenon*. The red section of each bar (Used concepts) identifies the number of concepts classified in that category and used

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743

in the EUKN collection. The blue one (Unused concepts) identifies those that have not been used in the classification. Finally, the graph line (Occurrences) shows the number of EUKN resources classified using concepts of each category (right axis). It can be observed that the thesaurus is mainly focused about the activities, elements and characteristics related to urbanism with other lesser themes such as the persons, groups and regulations involved in it. Only two minor quality issues have been identified. Firstly, the number of resources focusing on *abstract-qualities* and in *information-objects* is too high for the weight that it has been given in the thesaurus. This means a fewer degree of discrimination in searches about these fields. Secondly, some of the identified categories in DOLCE have very few associated concepts and are not even used in the EUKN collection. The first issue is a conceptual problem that would require a restructuring in the next thesaurus version. The second one can be directly fixed by removing the unused leafs to adjust the thesaurus model to the use it has in the collection.

Once the EUKN thesaurus concepts have been linked to a DOLCE class, the next step has been to redefine their broader/narrower relations. This has been done by replacing them with an existing relation between their associated DOLCE classes. However, since DOLCE may provide several possible relations between two classes (e.g., the relation between a geographical-object and a physical-object can be of type *part* or of type *subclass*), a set of rules have been defined to select the most suitable one for each case and apply it to all equivalent cases. For example, if A is the *narrower* of B in the EUKN with A being a *physical-object* and B a *abstract quality*, the *narrower* should be transformed into a *has-quality* relation. Table 1 shows the rules to infer the DOLCE relations that replace the original broader/narrower relations of the thesaurus on the basis of the associated DOLCE classes.

Pairs of DOLCE classes associated with EUKN concepts	**Relation**
(activity → physical/abstract-quality) (geographical/physical/information-object → abstract-quality) (rational-agent → abstract-quality) (regulation → abstract-quality) (plan → abstract-quality) (physical-quality → abstract-quality) (physical-quality → physical-quality)	has-quality
(activity → rational-agent) (activity → information/physical-object) (activity → regulation) (activity → principle) (phenomenon → geographic-object)	participant
(abstract-quality → abstract-quality) (activity → plan) (phenomenon → activity) (geographic-object → geographic-object) (regulation → plan)	part
(plan → activity) (rational-agent → information-object) (rational-agent → physical-object) (rational-agent → plan) (norm → system-design)	generic-dependent
(geographical-object → physical-object) (rational-agent → rational-agent) (regulation → regulation) (information-object → information-object)	subclass-of
(physical-object → activity) (physical-object → plan)	instrument-of
(activity → activity)	result-of

Table 1. Inferred relation

An example of relations refinement is shown in figure 4. It shows how the 10 narrower relations of the *Environmental sustainability* branch have been processed. In the classification process, the *Environmental sustainability* concept

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743

and three of its children have been tagged as a DOLCE *activity*; the other 7 have been classified as DOLCE *physical-quality*. Following the rules indicated in table 1, the narrower relation between two *activities* must be replaced with the DOLCE *result-of* relation. In the case of narrower relations between an *activity* and a *physical-quality*, the *has-quality* relation has been used. With this improved model, an additional level of knowledge is obtained. Now, it is possible to infer that *Environmental sustainability* is the *result-of* *Waste management and recycling*, *Environmental education*, and *Green public procurement*. Additionally, it can be measured (*has-quality*) through the *Water quality*, the *Air quality* and so on.

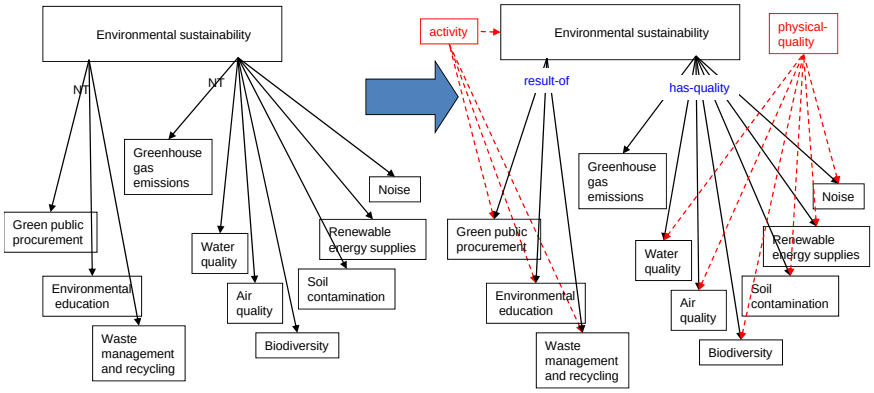

Fig. 4. Transformation of the *Environmental sustainability* concept and its narrower concepts

4.3 Publication of the EUKN collection

The storage of the EUKN collection in a semantic repository and its publication as linked data requires a complete transformation format and the connection with the conceptual model created previously. The first step has been to process the HTML resources and convert them into a Dublin Core RDF model. In this model, information such as the title or description has been stored as property values. However, those fields whose values are shared by several records (e.g., authors, organizations, document types, or coverage) are stored as URIs referencing to independent resources containing the shared information. The thematic information is not replaced with references to new resources, but to the corresponding ones in the knowledge model developed previously. If other fields with replicated values (e.g., the locations, authors or organizations) were also modeled with an ontology, the same association should be done for them (to improve the model quality). Figure 5 shows how a EUKN resource is represented and how it is bidirectionally associated to the *Biodiversity* concept of the transformed EUKN ontology through the Dublin Core *subject* property.

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743


```
<rdf:Description
rdf:about="http://www.eukn.org/eukn/resource/Urban_Environment/Environmental_Sustainability/
        Biodiversity/Urbanisation_can_be_an_opportunity_or_a_threat_for_biodiversity">
    <dc:title xml:lang="en">Urbanisation can be an opportunity or a threat ...</dc:title>
    <dc:subject rdf:resource="http://www.eukn.org/eukn/thesaurus/11_Biodiversity"/>
    <dc:coverage rdf:resource="http://www.eukn.org/eukn/location#eu"/>
    <dc:description xml:lang="en">The report '10 messages for 2010 - Urban Ecosystems',
       published by the European Environment Agency (EEA), provides an overview of the
       relation between urban ecosystems and biodiversity </dc:description> ...
</rdf:Description>

<rdf:Description rdf:about="http://www.eukn.org/eukn/thesaurus/11_Biodiversity">
    <rdfs:subClassOf rdf:resource=
        "http://www.eukn.org/eukn/thesaurus/dolceEq#physical-quality"/>
    <dolce:inherent-in rdf:resource=
        "http://www.eukn.org/eukn/thesaurus/9_Environmental_sustainability"/>
    <topic:hasResource rdf:resource="http://www.eukn.org/eukn/resource/Urban_Environment/
        Environmental_Sustainability/Biodiversity/
        Urbanisation_can_be_an_opportunity_or_a_threat_for_biodiversity"/>
    <skos:prefLabel xml:lang="en">Biodiversity</skos:prefLabel> ...
</rdf:Description>
```

Fig. 5. Example of RDF generated for a resource

The integrated model has been then stored in a JENA semantic repository and accessed through an SPARQL end point provided by the JOSEKI library. The SPARQL endpoint provides the dual functionality to facilitate an open query system for advanced users and as base for the construction of simpler and specific query and browsing components. A thematic graphical query component that transforms the user interaction in corresponding SPARQL queries to the endpoint has been developed. With respect to the browsing through the collection, the Linked Data service PUBBY has been used. PUBBY performs the transformation between the URIs used to link the resources in the collection and valid URLs that provide the desired resource concepts in the web. This provides a simple way to browse an RDF collection but at the cost of using for browsing a set of URIs that are different from the contained in the repository and accessed through the SPARQL endpoint. This issue will have to be dealt to provide homogeneous URIs from both services.

Figure 6 shows three examples of SPARQL queries. The first one demonstrates that thanks to the development of this repository, it is now possible to search resources based on their metadata descriptions (the original collection of HTML records describing the resources did not allow field based search functionality). The second one shows how complex queries can be easily expressed in SPARQL. In particular, it shows how to retrieve all the resources annotated with concepts being part of an urban technical infrastructure. Using a traditional digital library system, we should have first expanded our query to include all the narrower terms of urban technical infrastructure, and later search the metadata database. Finally, the third example shows the potential of inference reasoners. It returns all resources annotated with EUKN concepts directly classified as DOLCE activities (concepts described as subclass of activity), or subclasses of these EUKN concepts.

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743


```
Select ?resUri where { ?resource dc:source ?resUri.
       ?resource dc:title ?title. FILTER regex(?title, "town", "i") }
Select ?title where { ?concept topic:hasResource ?resource. ?resource dc:title ?title.
       ?concept dolce:part-of <http://www.eukn.org/eukn/thesaurus/90_Technical_infrastructure>}
Select ?title where { ?resource dc:title ?title.
       <http://www.eukn.org/eukn/thesaurus/93_Electricity> topic:hasResource ?resource.
       ?concept rdfs:subClassOf dolce:activity. ?concept topic:hasResource ?resource}
```

Fig. 6. SPARQL query examples

With respect to the browsing, the thematic search components provide a first entry point to the desired resources in the collection. The original EUKN thesaurus is provided as a tree in which the selection of a term returns the resources classified according to that term, but also (if requested) it can return those classified according to all their narrowers. Additionally, it shows the enriched EUKN thesaurus model with the DOLCE categories used to organize the concepts and allows browsing through the updated relations. Additionally, the linked visualization facilitates the access to the information providing a rich structure of related resources. Figure 7 shows snapshots of the browsing system[2] and the links that hold between them. It can be observed how in addition to the EUKN thesaurus based navigation, browsing based on DOLCE categories and in other controlled fields is also possible.

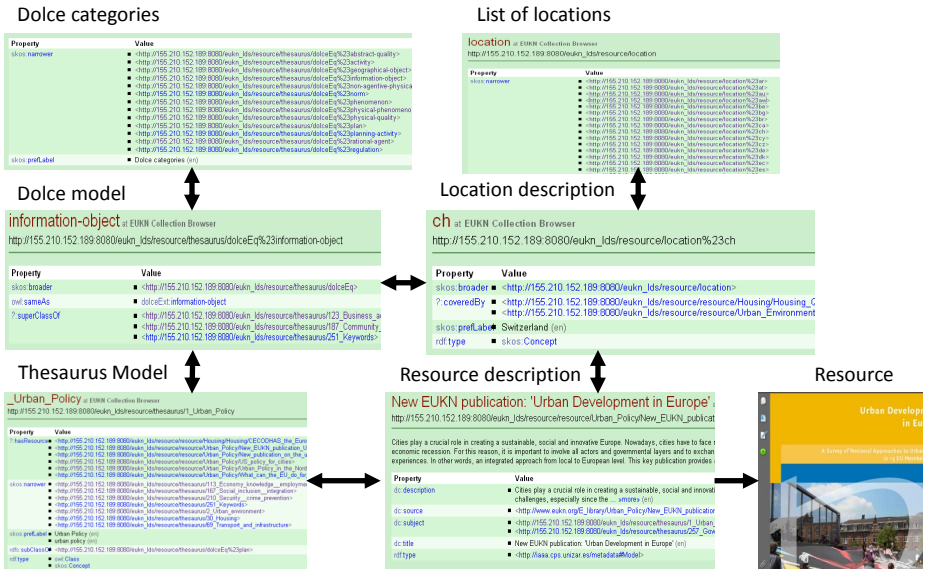

Fig. 7. Overview of the browsing system

[2] http://mularroya04.cps.unizar.es:8080/eukn_sparql/

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743

This system provides an improvement with respect to the original one in the sense that it facilitates guided multi criteria search and browsing through the collection through a conceptual view of the collection instead of (or combined with) a thematic one. Experienced users maintain the classical thematic access to the collection but with an improved model and more precise relations between the concepts. Additionally, the collection can also be accessed through DOLCE categories providing a more generic conceptual access.

5 Conclusions

This paper has described a process for the transformation of a collection of resources indexed with a thematic thesaurus (and described mostly as free text) into a semantically tagged collection that can be accessed and browsed as Linked Data.

The process described is focused on two tasks: enriching the thesaurus used for the collection classification using a top level ontology such as DOLCE, and transforming the textual resource collection into a set of interrelated RDF resources. The enrichment is based on a semiautomatic matching process between the thesaurus and DOLCE where the relations are automatically inferred. The transformation of the collection resources into RDF is left open due to the dependence of the specific collection to process.

The process has been applied to the EUKN collection. The resulting classification ontology and enriched collection has been stored in a JENA semantic repository and accessed through a facet-based search system allowing the browsing through the collection using a PUBBY Linked Data service. The search system provides access to the collection through the original and the enriched thesaurus. It maintains the original access for experience users, but it also offers a conceptual entry-point and collection browsing for inexperience ones.

Future work will firstly focus on validating the generated ontology to measure its quality and correct possible imprecisions in the established relations. When the model is verified, the system is expected to be published and improved through the obtained feedback. Additionally, we want to explore how the mapping with DOLCE can help to measure the thesaurus quality, and improve it, if needed. For example, very heterogeneous categories with few members may indicate a poor concept selection or a dispersion in the thesaurus classification objectives. Moreover, relations between concepts whose DOLCE equivalents do not hold a suitable relation may indicate an organization error in the thesaurus concept structure.

With respect to the browsing system, future work will focus on the improvement of the faceted system by providing additional access characteristics such as locations, authors, organizations or dates. We will need to model how each of these elements are organized and create a suitable ontology for their domain values. This will expand the resulting linked model enhancing their capabilities. In this context, the user interface must be improved to show human friendly labels (currently it shows URIs) sorted in an appropriate way (e.g., alphabetically).

(Draft) International Conference on Theory and Practice of Digital Libraries 2011 - TPDL 2011.
Lecture Notes in Computer Science (LNCS). 2011, vol. 6966, p. 372-383. ISSN 0302-9743

6 Acknowledgements

This work has been partially supported by Spanish Government by the projects TIN2009-10971 and "España Virtual" ref CENIT 2008-1030 (through contracts with the National Center of Geographic Information and GeoSpatiumLab), GeoSpatiumLab S.L. and Zeta Amaltea S.L.

References

1. Aleksovski, Z., Klein, M., ten Kate, W., van Harmelen, F.: Matching unstructured vocabularies using a background ontology. Lecture Notes in Computer Science 4248, 182–197 (2006)
2. Damme, C.V., Hepp, M., Siorpaes, K.: Folksontology: An integrated approach for turning folksonomies into ontologies. In: Bridging the Gap between Semantic Web and Web 2.0 Workshop. ESWC 2007. pp. 1–14. Innsbruck, Austria (2007)
3. Golbeck, J., Fragoso, G., Hartel, F., Hendler, J., Parsia, B., Oberthaler, J.: The national cancer institute's thesaurus and ontology. Journal of Web Semantics 1(1), 1–5 (2003)
4. Guarino, N.: Formal Ontologies and Information Systems. In: Amsterdam, I.P. (ed.) Proceedings of FOIS'98. pp. 3–15. Trento, Italy (June 1998)
5. Hearst, M.: Design Recommendations for Hierarchical Faceted Search Interfaces. In: ACM SIGIR Workshop on Faceted Search (August 2006)
6. Hyvönen, E., Salminen, M., Junnila, M., Kettula, S.: A content creation process for the semantic web. In: Proceeding of OntoLex 2004: Ontologies and Lexical Resources in Distributed Environments (May 2004)
7. International Organization for Standardization: Information and documentation - The Dublin Core metadata element set. ISO 15836:2003, International Organization for Standardization (ISO) (November 2003)
8. Kawtrakul, A., Imsombut, A., Thunkijjanukit, A., Soergel, D., Liang, A., Sini, M., Johannsen, G., Keizer, J.: Automatic Term Relationship Cleaning and Refinement for AGROVOC. In: Workshop on The Sixth Agricultural Ontology Service. Vila Real, Portugal (July 2005)
9. Khosravi, F., Vazifedoost, A.: Creating a persian ontology through thesaurus reengineering for organizing the digital library of the national library of iran. In: International Conference on Libraries, Information and Society, ICoLIS 2007. pp. 19–36. Malaysia (June 2007)
10. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., Schneider, L.: Wonderweb deliverable d17: The wonderweb library of foundational ontologies. Tech. rep., ISTC-CNR (2003)
11. Mika, P., Oberle, D., Sabou, M., Gangemi, A.: Foundations for service ontologies: Aligning owl-s to dolce alignment to foundational ontologies. In: Proceedings of the Thirteenth International World Wide Web Conference. New York, USA (May 2004)
12. Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., Katz, S.: Reengineering Thesauri for New Applications: the AGROVOC Example. Journal of Digital Information 4(4), 1–19 (2004)
13. Tudhope, D., Alani, H., , Jones, C.: Augmenting Thesaurus Relationships: Possibilities for Retrieval. Journal of Digital Information 1, Issue 8(41, 2001-02-05) (2001), 22 pages