(Draft) Information Technology and Libraries (ITAL). 2007, vol. 26, nº 3, p. 39-51.

ThManager: An Open Source Tool for creating and visualizing SKOS

J. Lacasta, J. Nogueras-Iso, F.J. López-Pellicer, P.R. Muro-Medrano, F. J. Zarazaga-Soria

Computer Science and Systems Engineering Department, University of Zaragoza,
María de Luna 1, E-50018 Zaragoza, Spain
{jlacasta, jnog, fjlopez, prmuro, javy}@unizar.es
http://iaaa.cps.unizar.es

Abstract. Knowledge Organization Systems denotes formally represented knowledge that is used within the context of Digital Libraries to improve data sharing and information retrieval. To increase their use, and to reuse them when possible, it is vital to manage them adequately and to provide them in a standard interchange format. Simple Knowledge Organization Systems (SKOS) seems to be the most promising representation for the type of knowledge models used in digital libraries, but there is a lack of tools that are able to properly manage it. This work presents a tool that fills this gap, facilitating their use in different environments and using SKOS as an interchange format.

1 Introduction

Unlike the largely unstructured information available on the Web, information in Digital Libraries (DLs) is explicitly organized, described, and managed. In order to facilitate discovery and access, DL systems summarize the content of their data resources into small descriptions, usually called metadata, which can be either introduced manually or automatically generated (index terms automatically extracted from a collection of documents). Most DLs use structured metadata in accordance with recognized standards such as MARC21 (U.S. Library of Congress, 2004) or Dublin Core (ISO, 2003).

In order to provide accurate metadata without terminological dispersion, metadata creators use different forms of controlled vocabularies to fill the content of typical keyword sections. This increase of homogeneity in the descriptions has the objective to improve the results provided by search systems. To facilitate the retrieval process, the same vocabularies used to create the descriptions are usually used to simplify the construction of user queries.

As there are many different schemas for modelling controlled vocabularies, the term knowledge organization systems (KOS) is intended to encompass all those types of schemas for organizing information and promoting knowledge management. As (Hodge, 2000) says, “A KOS serves as a bridge between the users information need and the material in the collection”. Some types of KOS can be highlighted. Examples of simple types are glossaries, which are only a list of terms (usually with definitions), and authority files that control variant versions of key information (such as geographic or personal names). More complex are subject headings, classification schemes, and categorization schemes (also known as taxonomies) that provide a limited hierarchical structure. At a more complex level, KOS includes thesauri and less-traditional schemes such as semantic networks and ontologies that provide richer semantic relations.

There is not a single KOS on which everyone agrees. As (Lesk, 1997) says, while a single KOS would be advantageous, it is unlikely that such a system will ever be developed. Culture constrains the knowledge classification scheme because what is meaningful to one area is not necessarily meaningful to another one. Depending on the situation, the use of one or another KOS, has its advantages and disadvantages, each one having its place.

These schemas, although sharing many characteristics, have been usually treated heterogeneously, leading to a variety of representation formats to store them. Thesauri are an example of the format heterogeneity problem. According to ISO2788 (norm for monolingual thesauri) (ISO, 1986), a thesaurus is a set of terms that describe the vocabulary of a controlled indexing language, formally organized so that the a priori relationships between concepts (e.g. synonyms, broader terms, narrower terms and related terms) are made explicit. This standard is complemented with the ISO-5964 (ISO, 1985), which describes the model for multilingual thesauri, but none of them describe a representation format. The lack of a standard representation model has caused a proliferation of incompatible formats created by different organizations. So, each organization that wants to use several external thesauri has to create specific tools to transform all of them to the same format.

In order to eliminate the heterogeneity of representation formats, the W3C initiative has promoted the development of SKOS (Simple Knowledge Organization Systems) (Miles et al., 2005) for its use in the semantic web environment. SKOS has been created to represent simple KOS such as subject heading lists, taxonomies, classification schemes, thesauri, folksonomies, other types of controlled vocabulary, and also concept schemes embedded in glossaries and terminologies. Although SKOS has been recently proposed, the number and importance of organizations involved in its creation process (and that publish their KOS in this format) indicates that it will probably become a standard for KOS representation.

SKOS provides a rich machine readable language that is very useful to represent KOS, but nobody would expect to have to create it manually or by just using a general purpose RDF editor (SKOS is RDF based). However, in the digital library area, there are not specialized tools that are able to manage it adequately. Therefore, this work tries to fulfil this gap describing an Open Source tool, called ThManager, that facilitates the construction of SKOS based KOS. Although ThManager has been created to manage thesauri, it is also appropriate to create and manage any other models that can be represented using SKOS format.

This paper describes the ThManager tool, highlighting its characteristics. The layered based architecture of the ThManager allows the reuse of the components created for the management of thesauri in other applications where they are also needed. For example, it facilitates the selection of values from a controlled vocabulary in a metadata creation tool, or the construction of user queries in a search client. The tool is distributed as Open Source software accessible through the SourceForge platform (http://thmanager.sourceforge.net/).

The rest of the paper is organized as follows. The next section is a review of the existent thesaurus management tools describing their characteristics, especially the interchange formats provided. Section 2 describes the SKOS model, showing its state of development and numbering their characteristics. Section 3 shows the architecture of the ThManager tool describing their different layers. Section 4 describes the main functionalities provided by ThManager tool to show how it works. Section 5 shows some results of creation, transformation and use of thesauri. The paper finishes with some conclusions and description of future work.

2 State of the art in thesaurus tools and representation models

The problem of creating appropriate content for thesauri is of interest in the digital libraries field and other related disciplines, and an increasing number of software packages have appeared in last years for the construction of thesauri. For instance, the web site of Willpower Information (http://www.willpower.demon.co.uk/ thessoft.htm) offers a detailed revision of more than 40 tools. Some tools are only available as a module of a complete information storage and retrieval system, but others also allow the possibility of working independently of any other software. Among these thesaurus creation tools, one may note the following products:

- BiblioTech (http://www.inmagic.com/). This is a multi-platform tool that forms part of BiblioTech PRO Integrated Library System and can be used to build an ANSI/NISO standard thesaurus (standard Z39.19 (ANSI, 1993)).
- Lexico (http://www.pmei.com/lexico.html). This is a Java-based tool that can be accessed and/or manipulated over the Internet. Thesauri are saved in a text based format. It has been used by the U.S. Library of Congress to manage vocabularies and thesauri such as: the “Thesaurus for Graphic Materials”, the “Global Legal Information Network Thesaurus”, the “Legislative Indexing Vocabulary” and the “Symbols of American Libraries Listing”.
- MultiTes (http://www.multites.com/). This is a windows based tool that provides support for ANSI/NISO relationships plus user defined relationships and comment fields for an unlimited number of thesauri (both monolingual and multilingual).
- TermTree 2000 (http://www.termtree.com.au/). TermTree is a windows based tool that uses Access, SQL Server or Oracle for data storage. It can import/export TRIM thesauri (format used by the Towers Records Information Management system http://www.towersoft.com/) as well as a defined TermTree 2000 tag format.

- WebChoir (http://www.webchoir.com). WebChoir is a family of client-server web applications that provide different utilities for thesaurus management in multiple DBMS platforms. TermChoir is a hierarchical information organizing and searching tool that enables to create and search varieties of hierarchical subject categories, controlled vocabularies, and taxonomies based on either pre-defined standards or a user-defined structure and exported to an XML based format. LinkChoir is another tool that allows indexers to describe information sources using terminology organized in TermChoir. And SeekChoir is a retrieval system that enables users to browse thesaurus descriptors and their references (broader terms, related terms, synonyms...).
- Synaptica (http://www.synaptica.com/). Synaptica is a client-server web application that can be installed locally on a client's intranet or extranet server. Thesaurus data is stored in a SQL Server or Oracle database. The application supports the creation of electronic thesauri in compliance with ANSI/NISO standard. The application allows the exchange of thesauri in CSV (Comma-Separated Values) text format.
- SuperThes (Batschi et al., 2002). SuperThes is a windows based tool that allows the creation of thesauri. It extends the ANSI/NISO relationships allowing many possible data types to enrich the properties of a concept. It can import/export thesauri in XML and in tabular format.
- TemaTRES (http://www.r020.com.ar/tematres/). TemaTres is a web application specially oriented to the creation of thesauri, but it also can be used to develop web navigation structures or to manage the documentary languages in use. The thesauri are stored in a MySQL database. It provides the created thesauri in Zthes (Tylor, 2004) or in SKOS format.
- Finally, it must be mentioned that, given that thesauri can be considered as ontologies specialized in organizing terminology (Gonzalo et al., 1998), ontology editors have sometimes been used for thesaurus construction. In (Denny, 2002) can be found a detailed survey of ontology editors.

All these tools (desktop or web based) present some problems in using them as general thesaurus editors. The main one is the incompatibility in the interchange formats that they support. These tools also present integration problems. Some are deeply integrated in bigger systems and can not easily be reused in other environments because they need specific software components to work (as DBMS to store thesauri). Others are independent tools (can be considered as general purpose thesaurus editors), but their architecture does not facilitate their integration within other information management tools. And most of them are not Open Source tools, so there is no possibility to modify them to improve their functionality.

Focusing on the interchange format problem, the ISO-5964 standard (norm for multilingual thesauri) is currently undergoing review by ISO TC46/SC 9, and it is expected that among the new modifications it will include a standard exchange format for thesauri. It is believed that this format will be based on technologies such as RDF/XML. In fact, some initiatives in this direction have already arisen.

- The ADL thesaurus Protocol (Janée et al., 2003) defines an XML and HTTP based protocol for accessing thesauri. As a result of query operations, portions of the thesaurus encoded in XML are returned.
- The Language Independent Metadata Browsing of European Resources (LIMBER) project has published a Thesaurus Interchange Format in RDF (Matthews et al., 2001). This work introduces an RDF representation of thesauri, which is proposed as a candidate thesaurus interchange format.
- The California Environmental Resources Evaluation System (CERES) and the NBII Biological Resources Division are collaborating in a Thesaurus Partnership project (CERES/NBII, 2003) for the development of an Integrated Environmental Thesaurus and a Thesaurus Networking ToolSet for Metadata Development and Keyword Searching. One of the deliverables of this project is an RDF format to represent thesauri.
- The "Semantic Web Advanced Development for Europe" (SWAD-Europe, 2001). One of the activities of this project, SWAD-Europe Thesaurus Activity, has defined the Simple Knowledge Organization System (SKOS), a set of specifications to represent the knowledge organization systems (KOS) on the semantic web (thesauri between them).

The British standards BS-5723 (BS, 1987) and BS-6723 (BS, 1985) (equivalent to the international ISO-2788 and ISO-5964) also lack a representation format. The British Standards Institute IDT/2/2 Working Group is now developing the BS-8723 standard that will replace them and whose fifth part will describe the exchange formats and protocols for interoperability of thesauri. The objective of this working group is to promote the standard to ISO, to replace the ISO-2788 and ISO-5964. Here, it is important to remark that given the direct involvement of IDT/2/2 Working Group with SKOS development; probably the two initiatives will not diverge. The new representation format will be, if not exactly SKOS, at least SKOS based. Taking into account all these circumstances, SKOS seems to be the most adequate representation model to store thesauri.

Given that SKOS is RDF based, it can be created using any tool that is able to manage RDF (usually used to edit ontologies), for example, SWOOP (MINDSWAP Group, 2006), Protégé (Noy et al., 2000) or Triple20 (Wielemaker et al., 2005). The problem with these tools is that they are too complex for editing and visualizing a simple model as SKOS. They are thought to create complex ontologies, so they provide too many options not specifically adapted to the type of relations in SKOS. In addition, they do not allow an integrated management of collection of thesauri and other types of controlled vocabularies as needed in digital library processes (e.g., the creation of metadata of resources, or the construction of queries in a search system).

3 SKOS model

SKOS is a representation model for simple knowledge organization systems, such as subject heading lists, taxonomies, classification schemes, thesaurus, folksonomies, other types of controlled vocabulary, and also concept schemes embedded in glossaries and terminologies. This section describes this model, providing their characteristics, showing the state of development and indicating the problems found to represent some types of KOS.

SKOS was initially developed within the scope of the Semantic Web Advanced Development for Europe (SWAD-Europe, 2001). SWAD-E was created to support W3C's Semantic Web initiative in Europe (part of the IST-7 programme). SKOS is based on a generic RDF schema for thesauri that was initially produced by the DESIRE project (Cross et al., 2001), and further developed in the Limber project (Matthews et al., 2001). It has been developed as a draft of an RDF Schema for thesauri compatible with relevant ISO standards, and later adapted to support other types of KOS. Among the KOS already published using this new format are GEMET (EEA, 2001), AGROVOC (FAO, 2006), ADL Feature Types (Hill and Zheng, 1999) or some parts of WordNet lexical database (Miller, 1990), all of them available in the SKOS project web page.

SKOS is a collection of three different RDF-Schema application profiles. SKOS-Core, to store common properties and relations, SKOS-Mapping, whose purpose is to describe relations between different KOS, and SKOS-Extension, to indicate specific relations and properties only contained in some type of KOS.

For the first step of the development of the ThManager tool, only the most stable part of SKOS has been considered. Figure 1 shows the part of SKOS-Core used. The rest of SKOS-Core is still unstable, so its support has been delayed until it is approved. SKOS-Mapping and SKOS-Extension are still in their first steps of development and are very unstable, so their management in ThManager has been also delayed until the creation of stable versions.

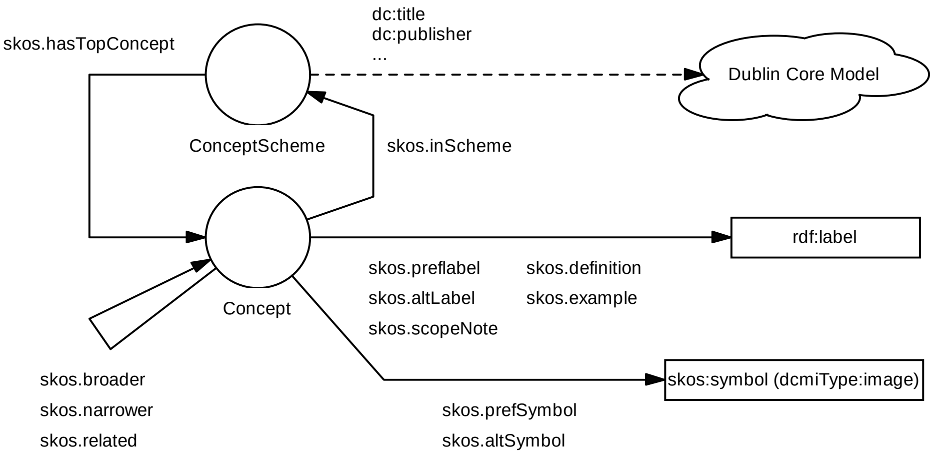


Fig. 1. SKOS Model

In SKOS-Core, a KOS (in our case usually a thesaurus) consists of a set of concepts (labelled as *skos:concept*) that are grouped by a concept scheme (*skos:conceptScheme*). To distinguish between different models provided, the *skos:conceptSchem*e contains a URI that identifies it, but to describe the model content to humans, metadata following the Dublin Core standard can also be added. The relation of the concept scheme with the concepts of the KOS is done through the *skos:hasTopConcep*t relation. This relation points at the most general concepts of the KOS (top concepts), which are used as entry points to the KOS structure.

In SKOS, each concept consists of a URI and a set of properties and relations to other concepts. Among the properties, *skos.preflabe*l and *skos.altLabe*l provide labels for a concept in different languages. The first one is used to show the label that better identifies a concept (for thesauri it must be unique). The second one is an alternative label that contains synonyms or spelling variations of the preferred label (it is used to redirect to the preferred label of the concept). The SKOS concepts also can contain three other properties called *skos.scopeNote, skos.definitio*n and *skos.example*. They contain annotations about the ways to use a concept, a definition or examples of use in different languages. Last, the *skos.prefSymbo*l and *skos.altSymbo*l properties are used to provide a preferred or some alternative symbols that graphically represent the concept. For example, a graphical representation is very useful to identify the meaning of a mathematical formula. Another example is a chemical formula, where a graphical representation of the structure of the substance also provides valuable information to the user.

With respect to the relations, each concept indicates by means of the *skos:inScheme*e relation in which concept scheme it is contained. The *skos.broade*r and the *skos.narrowe*r relations are inverse relations used to model the generalization and specialization characteristics present in many KOS (including thesauri). *Skos.broade*r relates to more general concepts and *skos.narrowe*r to more specific ones. The *skos.relate*d relation describes associative relationships between concepts (also present in many thesauri) indicating that two concepts are related in some way.

With these properties and relations, it is perfectly possible to represent thesauri, taxonomies and other types of controlled vocabularies. However, there is a problem for the representation of classification schemes that provide multiple coding of terms (there is no place to store this information). Under this category one may find classification schemes such as ISO-639 (ISO, 2002) (ISO standard for coding of languages), which proposes different types of alphanumeric codes (e.g. 2 letters and 3 letters). For this special case, the SKOS working group proposes the use of the property *skos.notation*. Although, this property is not in the SKOS vocabulary yet, it is expected to be added in future versions. Given the need to work with these types of schemes, this property has been included in the ThManager tool.

4 ThManager Architecture

This section presents the architecture of ThManager tool. This tool has been created to manage thesauri in SKOS, but it is also a base infrastructure that facilitates the management of thesauri in digital libraries, simplifying their integration in tools that need to use thesauri or other types of controlled vocabularies. In addition, to facilitate its use in different computer platforms, ThManager has been developed using the Java object oriented language.

The architecture of ThManager tool is shown in figure 2. The system consists of three layers: a repository layer where thesauri are stored and identified by means of associated metadata describing them; second, a persistence layer that provides an API for access to thesauri stored in the repository; and third, a GUI layer that offers different graphical components to visualize thesauri, to search by their properties and to edit them in different ways.

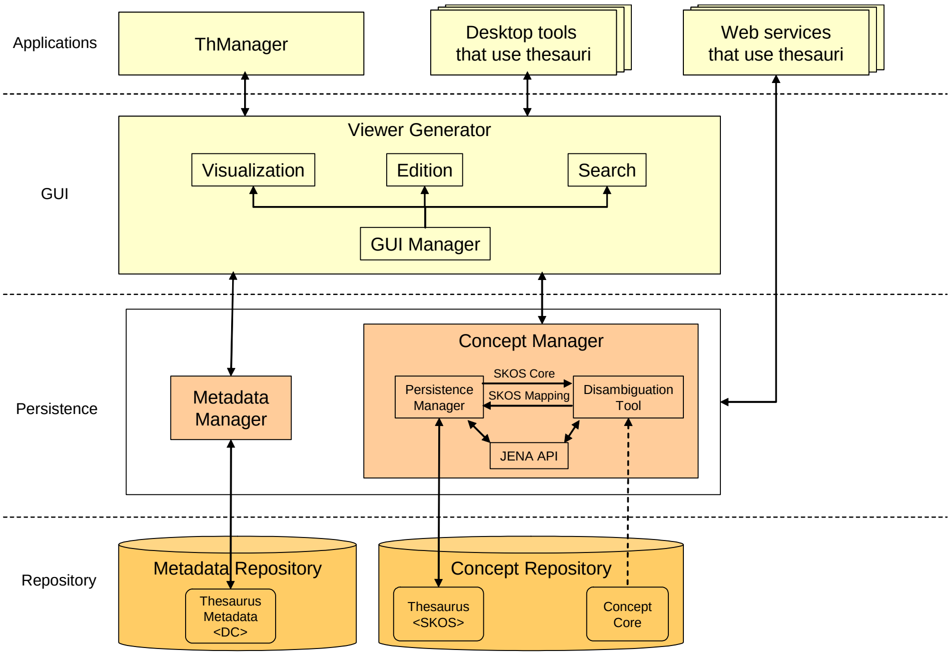


Fig. 2. KOS Manager Architecture

The ThManager tool is an application that uses the different components provided by the GUI layer to allow the user the management of thesauri. In addition, the layered architecture allows other applications to use some of the visualization components or the method provided by the persistence layer to provide access to thesauri.

The main features that have guided the design of these layers have been the following: a metadata-driven design, an efficient management of thesauri, the possibility of interrelating thesauri, and the reusability of ThManager components. The following subsections describe these characteristics in detail.

4.1 Metadata-driven design

A fundamental aspect in the repository layer is the use of metadata to describe thesauri. The ThManager considers metadata of thesauri as basic information in the thesaurus management process, being stored in the metadata repository and managed by the metadata manager. The reason for this metadata-driven design is that thesauri must be described and classified to facilitate the selection of the one that fits better the user needs, allowing the user to search them not only by their name but also by the application domain or the associated geographical area between others. The lack of metadata makes difficult the identification of useful thesauri (provided by other organizations), producing a low reuse of them in other contexts.

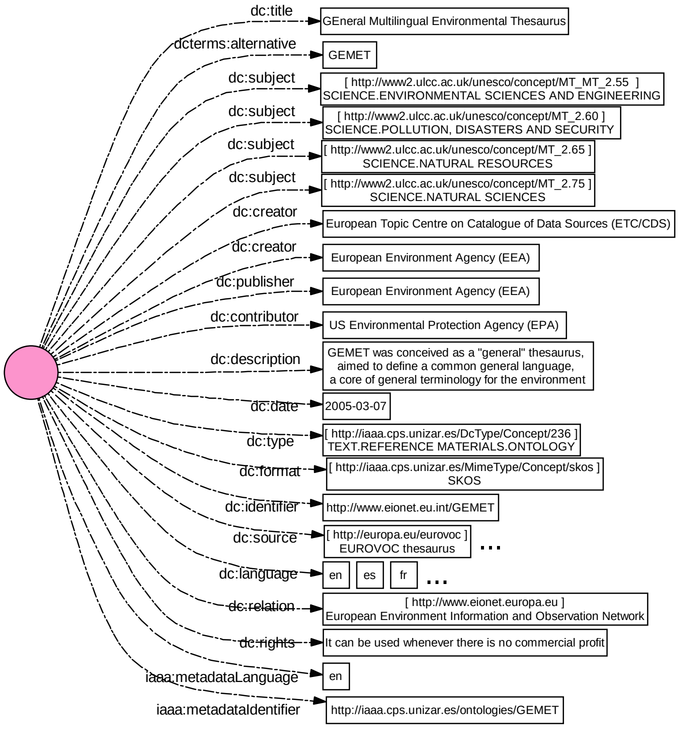

dc:title
GEneral Multilingual Environmental Thesaurus
dcterms:alternative
GEMET
dc:subject
[http://www2.ulcc.ac.uk/unesco/concept/MT_MT_2.55]
SCIENCE.ENVIRONMENTAL SCIENCES AND ENGINEERING
dc:subject
[http://www2.ulcc.ac.uk/unesco/concept/MT_2.60]
SCIENCE.POLLUTION, DISASTERS AND SECURITY
dc:subject
[http://www2.ulcc.ac.uk/unesco/concept/MT_2.65]
SCIENCE.NATURAL RESOURCES
dc:subject
[http://www2.ulcc.ac.uk/unesco/concept/MT_2.75]
SCIENCE.NATURAL SCIENCES
dc:creator
European Topic Centre on Catalogue of Data Sources (ETC/CDS)
dc:creator
European Environment Agency (EEA)
dc:publisher
European Environment Agency (EEA)
dc:contributor
US Environmental Protection Agency (EPA)
dc:description
GEMET was conceived as a "general" thesaurus,
aimed to define a common general language,
a core of general terminology for the environment
dc:date
2005-03-07
dc:type
[http://iaaa.cps.unizar.es/DcType/Concept/236]
TEXT.REFERENCE MATERIALS.ONTOLOGY
dc:format
[http://iaaa.cps.unizar.es/MimeType/Concept/skos]
SKOS
dc:identifier
http://www.eionet.eu.int/GEMET
dc:source
[http://europa.eu/eurovoc]
EUROVOC thesaurus
...
dc:language
en
es
fr
...
dc:relation
[http://www.eionet.europa.eu]
European Environment Information and Observation Network
dc:rights
It can be used whenever there is no commercial profit
iaaa:metadataLanguage
en
iaaa:metadataIdentifier
http://iaaa.cps.unizar.es/ontologies/GEMET

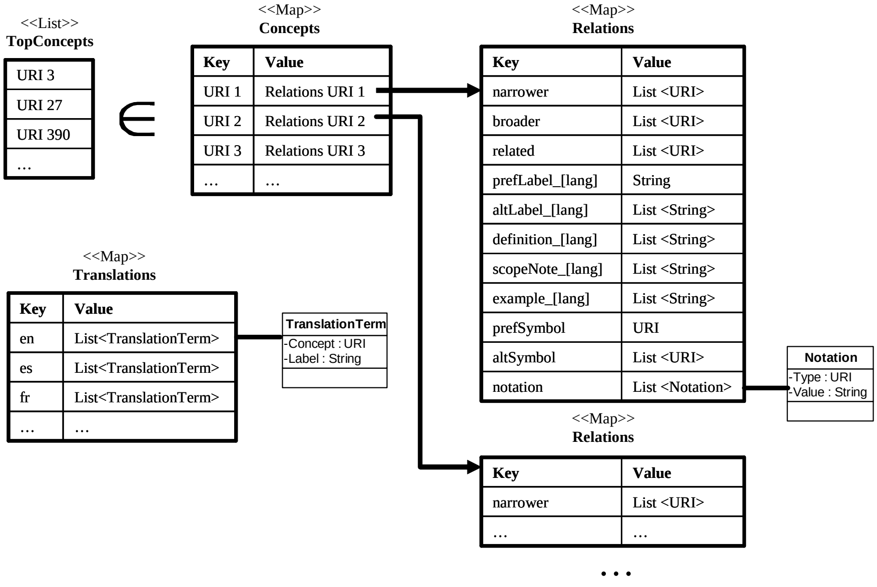
Fig. 3. Metadata of GEMET thesaurus

To describe thesauri in our service, a metadata profile based on Dublin Core has been created. The reason to use Dublin Core as basis of this profile has been its extensive use in the metadata community. It provides a simple way to describe a resource using very general metadata elements, which can be easily matched with complex domain-specific metadata standards. Additionally, Dublin Core can be also extended to define application profiles for specific types of resources. Following the metadata profile hierarchy described in (Tolosana-Calasanz et al., 2006), the thesaurus metadata profile refines the definition and domain of Dublin Core elements, as well as including two new elements (metadata language and metadata identifier) to identify appropriately the metadata records describing a thesaurus. The profile for thesauri has been described using the IEMSR format (Heery et al., 2005) and is distributed with the tool. IEMSR is an RDF based format created by the JISC IE Metadata Schema Registry project to describe metadata application profiles. Figure 3 shows the metadata created for GEMET thesaurus (the resource), expressed as a hedgehog graph (reinterpretation of RDF triplets: resources, named properties and values). The purpose of these metadata is not only to simplify the thesaurus location to a user, but also to facilitate the identification of thesauri useful for a specific task in a machine-to-machine communication. For instance, one may be interested only in thesauri that cover a restricted geographical area or have a specific thematic.

4.2 Efficient storage of thesauri

Thesauri vary enormously in size, ranging from hundreds of concepts and properties up to millions. So, the time spent on load, navigation and search processes are a functional restriction for a tool that has to manage them. SKOS is RDF based, and since reading RDF to extract the content is a slow process, the format is not appropriate for inner storage. To provide better access time, ThManager transforms SKOS into a binary format when a new SKOS is imported.

The persistence layer provides a unified access to the thesaurus repository. This layer is used by the GUI layer to access the thesauri but it can also be used by other tools that need to use thesauri outside a desktop environment (e.g., a thematic search system accessible through the web, which needs browsing a thesaurus to facilitate construction of user queries). This layer performs the transformation of SKOS to the binary format when a thesaurus is imported. The transformation is provided using the Jena library (a popular library to manipulate RDF documents that allows storing them in different kinds of repositories, http://jena.sourceforge.net/). Jena provides an open model that can be extended with specialized modules to use other ways of storage, making it possible to easily change the storage format system for another that is more efficient if needed.



<<List>>
TopConcepts

URI 3
URI 27
URI 390
…

∈

<<Map>>
Concepts

Key	Value
URI 1	Relations URI 1
URI 2	Relations URI 2
URI 3	Relations URI 3
…	…

<<Map>>
Relations

Key	Value
narrower	List <URI>
broader	List <URI>
related	List <URI>
prefLabel_[lang]	String
altLabel_[lang]	List <String>
definition_[lang]	List <String>
scopeNote_[lang]	List <String>
example_[lang]	List <String>
prefSymbol	URI
altSymbol	List <URI>
notation	List <Notation>

Notation
-Type : URI
-Value : String

<<Map>>
Translations

Key	Value
en	List<TranslationTerm>
es	List<TranslationTerm>
fr	List<TranslationTerm>
…	…

TranslationTerm
-Concept : URI
-Label : String

<<Map>>
Relations

Key	Value
narrower	List <URI>
…	…

…

Fig. 4. Persistence Model

The data structure used is shown in figure 4. The model is an optimized representation of the information given by the RDF triplets. The *Concept*s map contains the *concepts* and their associated *relations* in the form of key-value pairs: the key is a URI identifying a concept; and the value is a *Relations* object containing the properties of the concept.

A *Relation*s object is a map that stores the properties of one concept in the form of *<property type, property values>* pairs. The keys used for this map are the names of the typical property types in the SKOS model (e.g., *narrower,* or *broader*). The only special cases for encoding these property types in the proposed data structure occur when they have a language attribute (e.g., *prefLabel*, *definition*, or *scopeNote*). In those cases we propose the use of a *[lang]* suffix to distinguish the property type for a particular language. For instance, *prefLabel_en* indicates a *prefLabel* property type in English. Additionally, it must be noted that the data type of the property values assigned to each key in the *Relations* map varies upon the semantics given to each property type. The data types fall into the following categories: a string for a *prefLabel* property type; a list of strings for *altLabel*, *definition*, *scop*e *note,* and *example* property types; a URI for a *prefSymbol* property type; a list of URIs for *narrower*, *broader*, *related* and *altSymbol* property types; and a list of *Notation* objects for a *notation* property type. The data type used for *notation* values is a complex object because there may be different notation types. A *Notation* object consists of *type* and *value* attributes. The *type* attribute is a URI which identifies a particular notation type and qualifies the associated notation value.

Additionally and with the objective of increasing the speed of some operations (e.g. navigation or search), some optimizations have been added. Firstly, the URIs of the top concepts are stored in the *TopConcept*s list. This list contains redundant information, given that those concepts are also stored in the *Concept*s map, but it makes immediate their location. Secondly, to speed up the search of concepts and the drawing of the alphabetic viewer the *Translation*s map has been added. This map contains, for each language in which the thesaurus is, a list of pairs URI, *prefLabe*l in that language -ordered by *prefLabel*. It also contains redundant information, which allows the immediate creation of the alphabetic viewer for a language, making easier the search process and, as it can be seen later, this does not provides a big overhead in load time. In addition, if no alphabetic viewer and search are needed, this structure can be removed without affecting the hierarchical viewer.

This solution has proven to be useful to manage the kind of thesauri we use (they do not surpass the size of 50,000 concepts and about 330,000 properties), loading them to memory in an average computer in a reasonable time, and allowing the immediate navigation and search (see section 6).

4.3 Interrelation of thesauri

The vast choice of thesauri that are available nowadays implies an undesired effect of content heterogeneity. Although a thesaurus is usually created for a specific application domain, some of the concepts defined in thesauri from different applications domains may be equivalent. In order to facilitate cross-domain classification of resources, users would benefit from the possibility of knowing the connections of a thesaurus in their application domain to thesauri used in other domains. However, it is difficult to detect manually the implicit links between those different thesauri.

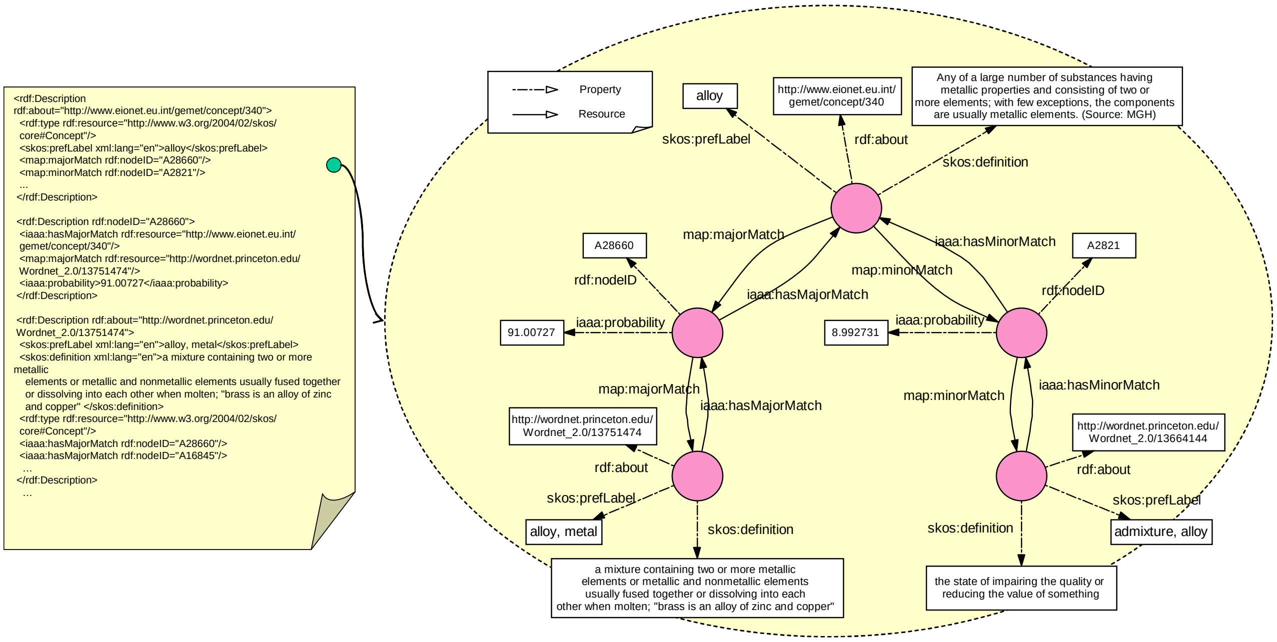
Therefore, in order to facilitate automatically these inter thesaurus connections, the persistence layer of ThManager tool provides an interrelation functionality that relates a thesaurus with respect to an upper level lexical database (the concept core displayed in Figure 2).

The interrelation mechanism is based on the method presented in (Nogueras-Iso, Zarazaga-Soria et al., 2005). It is an unsupervised disambiguation method that uses the relations between concepts as disambiguation context. It applies a heuristic voting algorithm to select for each thesaurus concept the most adequate sense of the used concept core. At the moment, the concept core is the WordNet lexical database. WordNet is a large English lexical database that groups nouns, verbs, adjectives and adverbs into sets of cognitive synonyms (synsets), each expressing a distinct concept. Those synsets are interlinked by means of conceptual-semantic and lexical relations.

The interrelation component has been conceived as an independent module that receives as input a thesaurus in SKOS and returns the relation respect to concept core using an extended version of the SKOS Mapping model (Miles and Brickley, 2004). This model, as commented before, is a part of SKOS that allows describing exact, major and minor mappings between concepts of two different KOS (in this case between a thesaurus and the common core). SKOS Mapping is still in an early stage of development and to provide the needed functionality it has been extended.

The base SKOS Mapping provides the *map:exactMatch*, *map:majorMatc*h and *map:minorMatc*h relations to indicate the degree of relation between two concepts. Given that the interrelation algorithm can not assure that a mapping is 100% exact, only the major and minor match properties are used. The algorithm returns, for each concept, a list of possible mappings with the lexical database: the one with the highest probability is assigned as *major* match, and the rest are assigned as *minor* matches.

To store the interrelation probability, SKOS mapping has been extended by adding a blank node with the liability of the mapping. Also, to be able to know which concepts of which thesauri are equivalents to one of the common core, the inverse relations of *map:majorMatch* and *map:minorMatch* have been created. An example of SKOS mapping can be seen in figure 5. There, the concept 340 of GEMET thesaurus (alloy) is correctly mapped to the WordNet concept number 13751474 (alloy, metal) with a probability of 91.007%, an unrelated minor mapping is also found, but it is given a low probability (8.992%).



```
<rdf:Description
rdf:about="http://www.eionet.eu.int/gemet/concept/340">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/
  core#Concept"/>
  <skos:prefLabel xml:lang="en">alloy</skos:prefLabel>
  <map:majorMatch rdf:nodeID="A28660"/>
  <map:minorMatch rdf:nodeID="A2821"/>
  ...
 </rdf:Description>

 <rdf:Description rdf:nodeID="A28660">
  <iaaa:hasMajorMatch rdf:resource="http://www.eionet.eu.int/
  gemet/concept/340"/>
  <map:majorMatch rdf:resource="http://wordnet.princeton.edu/
  Wordnet_2.0/13751474"/>
  <iaaa:probability>91.00727</iaaa:probability>
 </rdf:Description>

 <rdf:Description rdf:about="http://wordnet.princeton.edu/
 Wordnet_2.0/13751474">
  <skos:prefLabel xml:lang="en">alloy, metal</skos:prefLabel>
  <skos:definition xml:lang="en">a mixture containing two or more
metallic
    elements or metallic and nonmetallic elements usually fused together
    or dissolving into each other when molten; "brass is an alloy of zinc
    and copper" </skos:definition>
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/
  core#Concept"/>
  <iaaa:hasMajorMatch rdf:nodeID="A28660"/>
  <iaaa:hasMajorMatch rdf:nodeID="A16845"/>
   ...
 </rdf:Description>
   ...
```

Property
Resource
alloy
http://www.eionet.eu.int/gemet/concept/340
Any of a large number of substances having metallic properties and consisting of two or more elements; with few exceptions, the components are usually metallic elements. (Source: MGH)
skos:prefLabel
rdf:about
skos:definition
map:majorMatch
iaaa:hasMinorMatch
A28660
A2821
map:minorMatch
rdf:nodeID
iaaa:hasMajorMatch
rdf:nodeID
91.00727
iaaa:probability
8.992731
iaaa:probability
map:majorMatch
iaaa:hasMajorMatch
map:minorMatch
iaaa:hasMinorMatch
http://wordnet.princeton.edu/Wordnet_2.0/13751474
http://wordnet.princeton.edu/Wordnet_2.0/13664144
rdf:about
rdf:about
skos:prefLabel
skos:prefLabel
alloy, metal
skos:definition
skos:definition
admixture, alloy
a mixture containing two or more metallic elements or metallic and nonmetallic elements usually fused together or dissolving into each other when molten; "brass is an alloy of zinc and copper"
the state of impairing the quality or reducing the value of something

Fig. 5. SKOS Mapping extension

4.4 Reusability of ThManager components

On top of the API layer, the GUI layer has been constructed. This layer contains several graphical interfaces to provide different types of viewers, searchers and editors for thesauri. This layer is used as base for the construction of the ThManager tool. The tool groups a subset of the provided components, relating them to obtain a final user application that allows the management of the stored thesauri, their visualization (navigation by the concept relations), their edition and their importation/exportation using SKOS format.

The ThManager tool has not only been created as an independent tool to facilitate the management of thesauri, but also to allow an easy integration in tools that need to use thesauri. It has been done by combining the information management with specific graphical interfaces in different black box components. Between the provided components, there is a hierarchical viewer, an alphabetic viewer, a list viewer, a searcher and an editor, but more components can be constructed if needed. The use of the GUI layer as a library of reusable graphical components makes it possible to create different tools that are able to manage thesauri with different user requirements with a minimum effort, allowing also the integration of this technology in other applications that need controlled vocabularies to improve their functionality. For example, in a metadata creation tool, it can be used to provide the graphical component to select controlled values from thesauri and automatically insert them in the metadata. Also, it can be used to provide the list of possible values to use in a web search system, or to provide a thesaurus based navigation of a collection of resources in an exploratory search system.

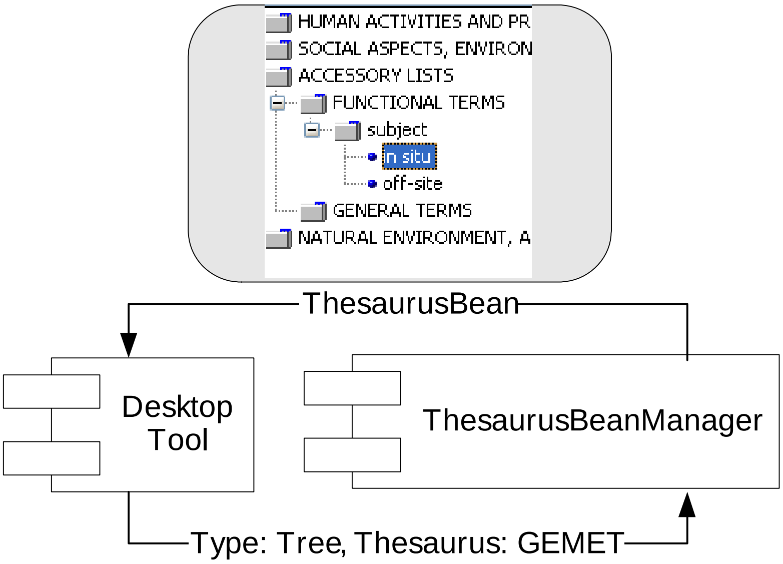


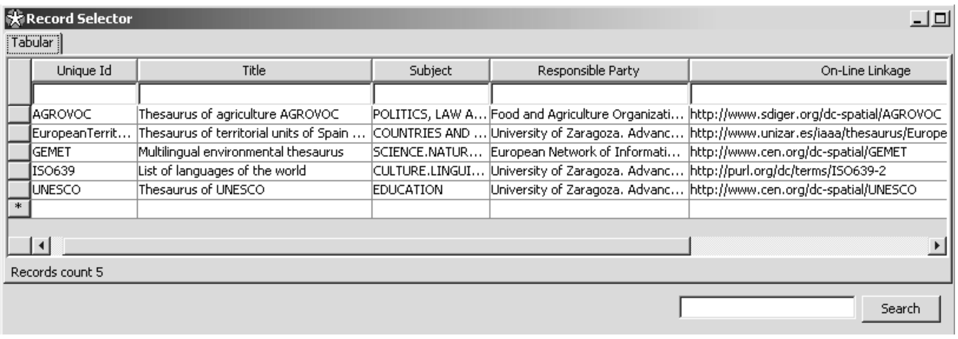
Fig. 6. GUI component integration

Figure 6 shows the integration process of a thesaurus visualization component in an external tool. The provided thesaurus components have been constructed following the Java Beans philosophy (reusable software components that can be manipulated visually in a builder tool), where a component is a black box with methods to read and change its state that can be reused when needed. Here, each thesaurus component is a *ThesaurusBean* that can be directly inserted in a graphical application to use its functionality (visualize or edit thesauri) in a very simple way. The *ThesaurusBeans* are provided by the *ThesaurusBeanManager* that, given the parameters of the thesaurus to visualize and the type of visualization, returns the most adequate component to use.

5 Description of ThManager functionality

ThManager tool is a desktop application that is able to manage thesauri stored in SKOS. As regards to the installation requirements, the application requires 100 MBs of free space on the hard disk. With respect to RAM and CPU requirements, they depend greatly on the size and the number of thesauri loaded in the tool. Considering the number and size of thesauri used as testbed in section 6, RAM consumption ranges from 256 to 512 MBs, and with a 3Ghz CPU (e.g., Pentium IV) the load times for the bigger thesauri are acceptable. However, if the size of thesauri is smaller, RAM and CPU requirements decrease, being able to operate on a computer with just a 1 Ghz CPU (e.g., Pentium III) and 128 MBs of RAM.

Given that the management of ThManager is metadata oriented, the first window in the application shows a table including the metadata records describing all the thesauri stored in the system (figure 7). The selection of a record in this table indicates, to the rest of the components, the selected thesaurus. The creation or deletion of thesauri is also provided here.



Unique Id	Title	Subject	Responsible Party	On-Line Linkage
AGROVOC	Thesaurus of agriculture AGROVOC	POLITICS, LAW A...	Food and Agriculture Organizati...	http://www.sdiger.org/dc-spatial/AGROVOC
EuropeanTerrit...	Thesaurus of territorial units of Spain ...	COUNTRIES AND ...	University of Zaragoza. Advanc...	http://www.unizar.es/iaaa/thesaurus/Europe
GEMET	Multilingual environmental thesaurus	SCIENCE.NATUR...	European Network of Informati...	http://www.cen.org/dc-spatial/GEMET
ISO639	List of languages of the world	CULTURE.LINGUI...	University of Zaragoza. Advanc...	http://purl.org/dc/terms/ISO639-2
UNESCO	Thesaurus of UNESCO	EDUCATION	University of Zaragoza. Advanc...	http://www.cen.org/dc-spatial/UNESCO

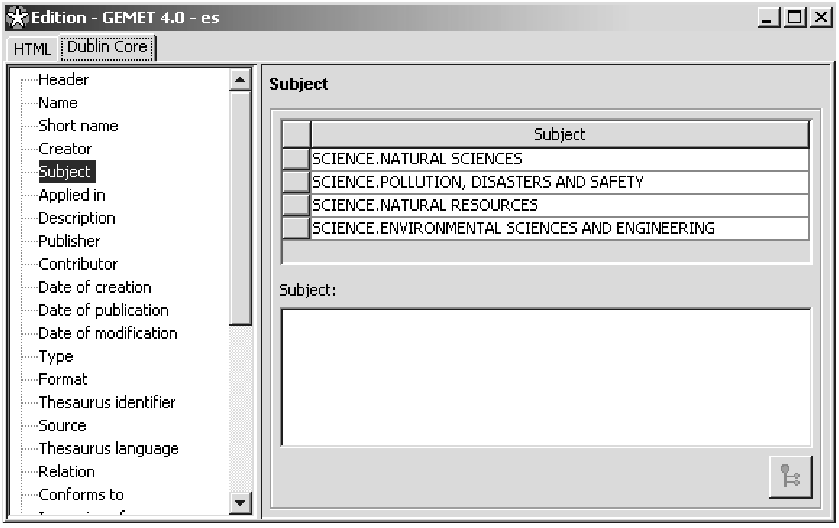
Records count 5

Fig. 7. Thesaurus Selector

The only operation that can be performed when no record is selected is to import a new thesaurus stored in SKOS. To import it, the name of the SKOS file has to be provided. The import tool also contains the option to interrelate the imported thesaurus to the concept core. The metadata of the thesaurus are extracted from inside of the SKOS if they are available, or they can be provided in an associated XML metadata file. If no metadata record is provided, the application generates a new one with minimum information, using as base the name of the SKOS file.

Once the user has selected a thesaurus, it can visualize and modify its metadata, visualize or modify its content, export the thesaurus to SKOS or, as commented before, delete it.

With respect to the metadata describing a thesaurus, a metadata viewer visualizes the metadata in HTML and a metadata editor allows the editing of metadata following the thesaurus metadata profile described in section 4.1 (figure 8). Different HTML views can be provided by adding more CSS files to the application. The metadata editor is customizable. To add or delete metadata elements to the metadata edition window, it is only needed to modify the description of the IEMSR profile for thesauri included in the application.

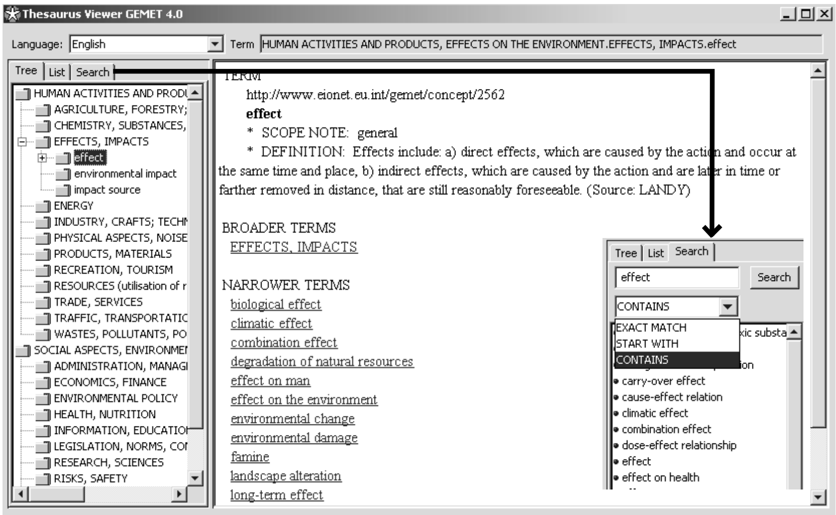


	Subject
	SCIENCE.NATURAL SCIENCES
	SCIENCE.POLLUTION, DISASTERS AND SAFETY
	SCIENCE.NATURAL RESOURCES
	SCIENCE.ENVIRONMENTAL SCIENCES AND ENGINEERING

Fig. 8. Thesaurus Metadata Editor

The main functionality of the tool is to visualize the thesaurus structure, showing all properties of concepts and allowing the navigation by relations (figure 9). Here, different read-only viewers are provided. There is an alphabetic viewer, which shows all the concepts ordered by the preferred label in one language. A hierarchical viewer provides navigation by broader and narrower relations. Additionally, a hypertext viewer shows all properties of a concept and provides navigation by all its relations (broader, narrower and related) by means of hyperlinks. Finally, there is also a search system, which allows the typical searches needed for thesauri (equals / starts with / contains). Currently, search is limited to preferred labels in the selected language, but it could be extended to allow searches by other properties such as synonyms, definitions or scope notes.

All these viewers are synchronized, so the selection of a concept in one of them produces the selection of the same concept in the others. The layered architecture described before allows these viewers to be reused in many situations including other parts of the ThManager tool. For example, in the thesaurus metadata editor described before, the thesaurus viewer is used to facilitate the selection of values for the subject section of metadata. Also, in the thesaurus editor shown later, the thesaurus viewer simplifies the selection of a concept related (by some kind of relation) to the selected, and provides a preview of the hierarchical viewer (to help to detect wrong relations).



Language: English

Term HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE ENVIRONMENT.EFFECTS, IMPACTS.effect

TERM
http://www.eionet.eu.int/gemet/concept/2562
effect
* SCOPE NOTE: general
* DEFINITION: Effects include: a) direct effects, which are caused by the action and occur at the same time and place, b) indirect effects, which are caused by the action and are later in time or farther removed in distance, that are still reasonably foreseeable. (Source: LANDY)

BROADER TERMS
EFFECTS, IMPACTS

NARROWER TERMS
biological effect
climatic effect
combination effect
degradation of natural resources
effect on man
effect on the environment
environmental change
environmental damage
famine
landscape alteration
long-term effect

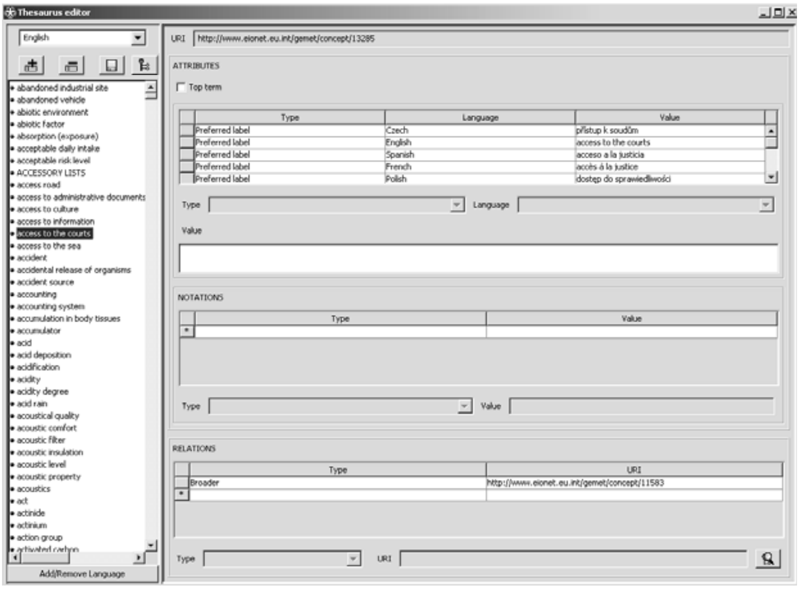
effect
CONTAINS

Fig. 9. Thesaurus Concept Selector

The third available operation is to edit the thesaurus structure. Here, to create a thesaurus following the SKOS model, an edition component is provided (figure 10). The graphical interface shows a list with all the concepts created in the selected thesaurus, allowing the creation of new ones (providing their URIs) or deletion of selected ones. Once a concept has been selected, its properties and relations to other concepts are shown, allowing the creation of new ones and the deletion of others. To facilitate the creation of relations between concepts, a selector of concepts (based in the thesaurus viewer) is provided, allowing the user to add related concepts without the need of manually typing the URI of the associated concept. Also, to see if the created thesaurus is correct, a preview of the hierarchical viewer can be shown, allowing the user to easily detect problems in the broader/narrower relations.

With respect to the interrelation functionality, at the moment, the mapping obtained is shown in the thesaurus viewers but the navigation between equivalent concepts of two thesauri has to be manually done by the user. However, a navigation component still under development will allow the user to jump from a concept in a thesaurus to concepts in others that are mapped to the same concept in the common core.

As mentioned before, for efficiency, the format used to store the thesauri in the repository is binary, but the interchange format used is SKOS. So, a module for thesauri importation/exportation is provided. This module is able to import from SKOS and export to SKOS but also, if the thesaurus has been interrelated with respect to the concept core, it is able to export its mapping to the concept core using the extended version of SKOS mapping above.



Thesaurus editor

English

- abandoned industrial site
- abandoned vehicle
- abiotic environment
- abiotic factor
- absorption (exposure)
- acceptable daily intake
- acceptable risk level
- ACCESSORY LISTS
- access road
- access to administrative documents
- access to culture
- access to information
- access to the courts
- access to the sea
- accident
- accidental release of organisms
- accident source
- accounting
- accounting system
- accumulation in body tissues
- accumulator
- acid
- acid deposition
- acidification
- acidity
- acidity degree
- acid rain
- acoustical quality
- acoustic comfort
- acoustic filter
- acoustic insulation
- acoustic level
- acoustic property
- acoustics
- act
- actinide
- actinium
- action group
- activated carbon

Add/Remove Language

URI http://www.eionet.eu.int/gemet/concept/13285

ATTRIBUTES

☐ Top term

Type	Language	Value
Preferred label	Czech	přístup k soudům
Preferred label	English	access to the courts
Preferred label	Spanish	acceso a la justicia
Preferred label	French	accès à la justice
Preferred label	Polish	dostęp do sprawiedliwości

Type Language

Value

NOTATIONS

Type	Value

Type Value

RELATIONS

Type	URI
Broader	http://www.eionet.eu.int/gemet/concept/11583

Type URI

Fig. 10. Thesaurus Concept Editor

6 Results of the work

This section shows some experiments performed with the ThManager tool for the storage and management of a selected set of thesauri. In particular, this set of thesauri is relevant in the context of the geographic information community. The increasing relevance of geographic information for decision-making and resource management in different areas of government has promoted the creation of geo-libraries and spatial data infrastructures to facilitate distribution and access of geographic information (Nogueras-Iso, Zarazaga-Soria et al., 2005). In this context, complex metadata schemes, such as ISO-19115 have been proposed for a full-detailed description of resources. Many of the metadata elements in these schemes are either constrained to a selected vocabulary (ISO-639 for language encoding, ISO-3166 for country codes…) or the user is told to pick a term from the most suitable thesaurus. The problems for this second case are usually that the choice for thesauri is quite open, the thesauri are frequently large, and the exchange format of available thesauri is quite heterogeneous.

In such a context, the ThManager tool has proven to be very useful to simplify the management of the used thesauri. At the moment, 80 KOS between thesauri and other types of controlled vocabulary have been created or transformed to SKOS and managed through this tool. Table 1 shows some of them, indicating their names (Name column), the number of concepts (NC column), their total number of properties and relations (NP and NR columns), and the number of languages in which concept properties are provided (NL column). To give an idea of the cost of loading these structures, the sizes of SKOS and binary files (SS and SB columns) are provided in Kilobytes (KB).

Table 1. Sizes of some thesauri and other types of vocabularies

Name	NC	NP	NR	NL	LT	BT	JT	SS	SB
ADL FTT	210	210	408	1	0.4	0.047	0.062	103	41
ISOC-G	5136	5136	1026	1	2.4	1.063	1.797	2796	1332
ISO-639	7599	16247	0	6	5.1	1.969	2.89	3870	3017
UNESCO	8600	13281	21681	3	2.1	1.406	2.984	4034	2135
EPSG	4772	9544	0	1	1.8	0.969	1.796	2935	1682
AGROVOC	16896	103484	30361	3	7.5	4.953	14.75	15859	5089
EUROVOC	6649	196391	20861	15	11.1	9.266	15.828	18442	11483
ETU	44991	89980	89976	2	13.3	10.625	17.844	23828	10412
GEMET	5244	326602	12750	21	13.7	11.828	25.61	28010	15048

Additionally, table 1 compares the performance time of ThManager with respect to other tools that load the thesauri directly from an RDF file using the Jena library (time performance has been obtained using a 3Ghz Pentium IV processor). For this purpose three different load times (in seconds) have been computed. The BT column contains the load time of binary files without the cost of creating the GUI for the thesauri viewers. The LT column contains the total load time of binary files (including the time of GUI creation and drawing). The JT column contains the time spent by an hypothetical RDF-based editor tool to invoke Jena and load in its memory model the RDF SKOS files (it does not include GUI creation) containing the thesauri. The difference between the BT and LT column shows the time used to draw the GUI once the thesauri have been loaded in memory. The difference between BT and JT columns shows the gain in terms of time of using a binary storage instead of a RDF based one.

The thesauri shown in the table are the ADL Feature Types Thesaurus (ADL FTT), the ISOC thesaurus of Geography (ISOC-G), the ISO-639, the UNESCO Thesaurus (UNESCO, 1995), the OGP Surveying and Positioning Committee Code Lists (EPSG) (OGP, 2006), the Multilingual Agricultural Thesaurus (AGROVOC), the European Vocabulary Thesaurus (EUROVOC) (EUPO, 2005), the European Territorial Units (Spain and France) (ETU), and the GEneral Multilingual Environmental Thesaurus (GEMET). They have been selected because they are very different in sizes and can be used to show how the load time evolves with the thesaurus size.

Among them, GEMET and AGROVOC can be highlighted. Although they are provided as SKOS, they include non-standard extensions that we have transformed to standard SKOS relations and properties. EUROVOC and UNESCO are examples of thesauri provided in other formats different to SKOS that we have completely transformed into SKOS. The first one was in a XML based format and the other used a plain text format. Another thesaurus transformed to SKOS is the European Territorial Units that contains the administrative political units in Spain and France. Here, the original source was a collection of heterogeneous documents that contained parts of the needed information and that have been processed to generate a SKOS file.

Some classification schemes have been also transformed to SKOS, such as the ISO-639 and the different EPSG codes for coordinate reference systems (including datums, ellipsoids and projections). With respect to controlled vocabularies created (by us) in SKOS using the ThManager tool, there is an extended version of the ADL Feature Types that includes a more detailed classification of features types and different glossaries used for resource classification.

Figure 11 depicts the comparison of the different load times shown in table 1 with respect to the size of the RDF SKOS files. The order of the thesauri in the figure is the same as in the table 1. It can be seen that the time to construct the model using a binary format is almost half the time spent to create the model using a RDF file. In addition, once the binary model is loaded, the time to generate the GUI is not very dependent on thesaurus size. This is possible thanks to the redundant information added to facilitate the access to top concepts and to make faster the load of the alphabetic viewer. This redundant information produces an overhead in the load of the model but without it the drawing time would be much worse (having to generate it on the fly).

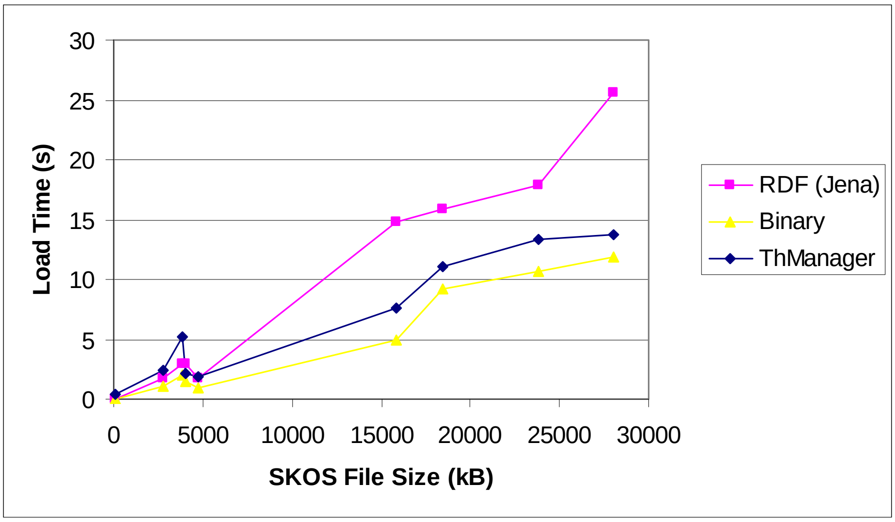


Fig. 11. Thesaurus Load Times

However, in spite of the improvements, for the bigger thesauri considered, the load time starts to be high, given that it includes the load time of all the structure of the thesaurus in memory and the creation the objects used to manage it quickly when loaded. But, once it is loaded, future accesses are immediate (less than 0.5 seconds). These accesses include, opening it again, navigating by thesaurus relations, changing the visualization language, and searching concepts by their preferred labels. To minimize the load time, thesauri can be loaded in background when the application is launched, reducing in that way the user perception of the load time.

Another interesting aspect in figure 11 is the peak of the third element. It corresponds with the ISO-639 classification scheme. It has the special characteristic of not having hierarchy and having many notations. These two characteristics produce a little increase in the model load time, given that the top concepts list contains all the concepts and the notations are more complex than other relations. But most of the time is used to generate the GUI of the tree viewer. The tree viewer gets all the concepts that are top terms, and for each one it asks for their preferred labels in the selected language and sorts them alphabetically to show the first level of the tree. This is fast for a few hundred of concepts but not for the 7599 of the ISO-639. However, this problem could be easily solved if the metadata contained a description of the type of KOS to visualize. If the tool knew that the KOS does not have broader/narrower relations, it could use the structures used to visualize the alphabetic list, which are optimized to show all the concepts of the KOS very fast, instead of trying to load it as a tree.

The persistence approach used has the advantage of not having to use external persistence systems such as a DBMS, and that it provides a very fast access after the load, but it has the drawback of loading all thesauri in memory (in time and space). So, for much bigger thesauri, the use of some kind of DBMS would be necessary. If this change were necessary, minimum modifications would be needed (one class). However, if not all the concepts are loaded, the alphabetic viewer (shows all the concepts) would have to be updated (e.g., showing the concepts by pages) or it would become too slow to work with it.

7 Conclusions

This paper has presented a tool that is able to manage thesauri needed in a digital library, for metadata creation and search processes, using SKOS as the interchange format.

This work makes a revision of the available tools that are able to edit thesauri, highlighting the lack of a formalized way to exchange thesauri and the difficulty of integrating those tools in other environments. This work selects SKOS, from the available interchange formats for thesauri, as the most promising format to become a standard for SKOS representation and highlights the lack of tools that are able to manage it properly.

The ThManager tool is offered as the solution to these problems. It is an Open Source tool that is able to manage thesauri stored in SKOS, allowing their visualization and edition. Thanks to the layered architecture, its components can be easily integrated in other applications that need to use thesauri or other types of controlled vocabularies. Additionally, they can be used to control the possible values used in a web search service to facilitate traditional or exploratory search based on a controlled vocabulary.

The performance of the tool is proved through a series of experiments on the management of a selected set of thesauri. This work analyzes the features of this selected set of thesauri and compares the efficiency of this tool with respect to other tools that load the thesauri directly from a RDF file. In particular, it is shown that the internal representation used by ThManager helps to decrease the time spent for the graphical loading of thesauri, easing the navigation of the thesaurus contents, as well as other typical operations like sorting or change of visualization language.

Additionally, it is worth noting that the tool can be used as a library of components to simplify the integration of thesauri in other applications that require the use of controlled vocabularies. On the one hand, ThManager has been integrated within the Open Source CatMDEdit tool (Zarazaga-Soria et al., 2003), a metadata editor tool for the documentation of geographic information resources (metadata compliant with ISO19115 geographic information metadata standard). The *ThesaurusBeans* provided in ThManager library have been used to facilitate the selection of keywords for some metadata elements. The ThManager component library has also contributed to the development of catalog search systems guided by controlled vocabularies. For instance, it has been used to build a thematic catalog in the SDIGER project (Nogueras-Iso, Latre et al., 2005). SDIGER is a pilot project on the implementation of the Infrastructure for Spatial Information in Europe (INSPIRE) for the development of a Spatial Data Infrastructure to support access to geographic information resources concerned with the European Water Framework Directive. Thanks to the ThManager components, the thematic catalog allows browsing of resources by means of several multilingual thesauri such as GEMET, UNESCO, AGROVOC and EUROVOC.

Future work will enhance the functionalities provided by ThManager. Firstly, it is planned to improve the ergonomics to show the connections between different thesauri. Currently, these connections can be computed and annotated but the GUI does not allow the user to navigate them. As the base technology has been already developed, only a graphical interface is needed. Secondly, the tool will be enhanced to support data types different from texts (e.g. images, documents, or other multimedia sources) for the encoding of property values of concepts. Thirdly, it has been noted that the thesauri concepts can evolve with time. Thus, a mechanism for the managing the different versions of thesauri will be necessary in the future. Finally, improvements in usability are also expected. Thanks to the component-based design of ThManager widgets (Thesaurus Beans), new viewers or editors can be created with little effort to meet the needs of specific users.

Acknowledgements

This work has been partially supported by the Spanish Ministry of Education and Science through the projects TIN2006-00779 and TIC2003-09365-C02-01 from the National Plan for Scientific Research, Development and Technology Innovation. The authors would like to express their gratitude to Juan José Floristán for its support in the technical development of the tool.

Bibliography

American National Standards Institute (ANSI). 1993. Guidelines for the Construction, Format, and Management of Monolingual Thesauri. ANSI/NISO Z39.19-1993. Revision of Z39.19.

Batschi, W.D., B. Felluga, R. Legat, P. Plini, H. Stallbaumer, and K. L. Zirm. 2002. SuperThes: A New Software for Construction, Maintenance and Visualisation of Multilingual Thesauri. In Enviro Info: 1-8. Vienna (Austria). http://www.t-reks.cnr.it/docs/ST_enviroinfo_2002.pdf.

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

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

CERES/NBII. 2003. The CERES/NBII Thesaurus Partnership Project. http://ceres.ca.gov/thesaurus.

Cross, P., D. Brickley, and T. Koch. 2001. RDF Thesaurus Specification. Technical Report 1011, Institute for Learning and Research Technology. http://ilrt.org/discovery/2001/01/rdf-thes.

Denny, Michael. 2002. Ontology building: a survey of editing tools. XML.com (November 2002): 1-6. http://xml.com/pub/a/2002/11/06/ontologies.html.

European Environment Agency (EEA). 2004. GEneral Multilingual Environmental Thesaurus (GEMET). Version 2.0. European Environment Information and Observation Network. http://www.eionet.europa.eu/gemet/rdf.

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

Food and Agriculture Organization of the United Nations (FAO). 2006. Agriculture vocabulary (AGROVOC). Agricultural Information Management Standards. http://www.fao.org/aims/ag%20alpha.htm.

Gonzalo, J., F. Verdejo, C. Peters, and N. Calzolari. 1998. Applying EuroWordNet to Cross-Language Text Retrieval. Computers and the Humanities Special Issue on EuroWord-Net: 113-135.

Heery, R., P. Johnston, D. Beckett, and N. Rogers. 2005. JISC metadata schema registry. In 5th ACM/IEEE-CS joint conference on Digital libraries: 381-381.

Hill, L. and Q. Zheng. 1999. Indirect Geospatial Referencing through Place Names in the Digital Library: Alexandria Digital Library Experience with Developing and Implementing Gazetteers. In Knowledge: Creation, Organization and Use. Proceedings of the 62nd Annual Meeting of the American Society for Information Science: 57–69. Washington, D.C. http://www.alexandria.ucsb.edu/~lhill/paper_drafts/ASIS99_confpaper2_final.pdf

Hodge, Gail. 2000. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. The Digital Library Federation.

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

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

International Organization for Standardization (ISO). 2002. Codes for the representation of names of languages. ISO 639. ISO/TC 37/SC 2.

International Organization for Standardization (ISO). 2003. Information and documentation -The Dublin Core metadata element set. ISO 15836:2003.

Janée, G., S. Ikeda, and L. L. Hill. 2003. The ADL Thesaurus Protocol. http://www.alexandria.ucsb.edu/~gjanee/thesaurus.

Lesk, Michael. 1997. Practical Digital Libraries. San Francisco: Books, Bytes, and Bucks.

Matthews, B. M., M. D. Wilson, K. Miller, and J. Ryssevik. 2001. Internationalising data access through LIMBER. In Third international workshop on internationalisation of products and systems: 1-14. Milton Keynes (UK). http://epubs.cclrc.ac.uk/bitstream/401/Limber_IWIPS.pdf.

Miles, Alistair and Dan Brickley, ed. 2004. SKOS Mapping Vocabulary Specification. W3C. http://www.w3.org/2004/02/skos/mapping/spec/2004-11-11.html.

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

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

MINDSWAP Group. 2006. SWOOP A Hypermedia-based Featherweight OWL Ontology Editor. Maryland Information and Network Dynamics Lab. Semantic Web Agents Project. http://www.mindswap.org/2004/SWOOP.

Nogueras-Iso, J., M. Latre, R. Béjar, P. Muro-Medrano, and F. Zarazaga-Soria. 2005. SDIGER: Experiences and identification of problems on the creation of a transnational SDI. In Proceedings of Jornadas Técnicas de la Infraestructura de Datos Espaciales de España: 1-10. Madrid (Spain).

Nogueras-Iso, J., F. J. Zarazaga-Soria, and P. R. Muro-Medrano. 2005. Geographic Information Metadata for Spatial Data Infrastructures - Resources, Interoperability and Information Retrieval. Springer Verlag.

Noy, N. F., R. W. Fergerson, and M. A. Musen. 2000. The knowledge model of Protégé2000: Combining interoperability and flexibility. In 2th International Conference on Knowledge Engineering and Knowledge Management: 1-20. Juan-les-Pins (France).

OGP Surveying & Positioning Committee. 2006. Surveying & Positioning. OGP Surveying & Positioning Committee. http://www.epsg.org.

Semantic Web Advanced Development for Europe (SWAD-Europe). 2001. Semantic Web Advanced Development for Europe Thesaurus Activity. http://www.w3.org/2001/sw/Europe/reports/thes.

Tolosana-Calasanz, R., J. Nogueras-Iso, R. Béjar, P. R. Muro-Medrano, and F. J. Zarazaga-Soria. 2006. Semantic interoperability based on Dublin Core hierarchical one-to-one mappings. International Journal of Metadata, Semantics and Ontologies (IJMS&O). In press.

Tylor, Mike. 2004. The ZTHES specifications for thesaurus representation, access and navigation. http://zthes.z3950.org.

United Nations Educational, Scientific and Cultural Organization (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. Paris: UNESCO Publishing. http://www.ulcc.ac.uk/unesco.

U.S. Library of Congress. 2004. MARC standards. Network Development and MARC Standards office. http://www.loc.gov/marc.

Wielemaker, J., G. Schreiber, and B. Wielinga1. 2005. Triple20 - an RDF/RDFS/OWL visualisation and editing tool. Lecture Notes in Computer Science (LNCS) 3729: 773–785.

Zarazaga-Soria, F. J., J. Lacasta, J. Nogueras-Iso, M. P. Torres, and P. R. Muro-Medrano. 2003. A Java Tool for Creating ISO/FGDC Geographic Metadata. In Proc. of the GI-days 2003 conference: 17–30. Münster (Germany).