Software services exploiting the eENVplus framework of interlinked thesauri for metadata management

Andreas Abecker
Riccardo Albertoni
Carlo Cipolloni
Monica De Martino
Yashar Moradiafkan
Roman Wössner
Overview eENVplus Thesaurus Framework

- Not only one thesaurus … but …
- integration of different available thesauri
- cross-walking from one thesaurus to another

Thesaurus Framework

**Modularity**
To add new KOS as a new module plugged in the set of thesauri in the TF

**Openness**
To easily extend each KOS keeping separated the original one

**Interlinking**
Linking among the terms referring to the same concepts in more than one thesaurus in order to harmonize their usage

**Exploitability**
To encode in a standard and flexible format in order to encourage the adoption and its enrichment from third-party systems

**Design Principle**

Simple Knowledge Organization System (SKOS) to encode the thesaurus content

Linked Data best practices to publish the thesaurus in machine-understandable format
Usage Scenarios for LusTRE Exploitation Services

Basic Idea

**THESAURUS EXPLOITATION SERVICE MODULES**

- KeywordCompletion
- KeywordExplanation
- KeywordTranslation
- QueryReformulation
- Keyword Validation
- ThesaurusVisualization

**USE CASES**

- Metadata Creation / Compilation
- Data & Service Querying / Discovery
- Explorative Search

**REST API**

- GetCapabilities
- GetSuggestions
- GetSynonyms
- GetRelatives
- DescribeConcept
- SPARQL

**Thesaurus Framework Server**

- Thesauri Datasources
- Interlinking
- Thesaurus Framework
- Exploitation Server Components

**Exploitation Client Interface**

**WWW**
Usage Scenarios for LusTRE Exploitation Services
Explanations + Remarks

1 Introduction
2 SKOS Essentials
   2.1 Concepts
   2.2 Labels
      2.2.1 Preferred Lexical Labels
      2.2.2 Alternative Lexical Labels
      2.2.3 Hidden Lexical Labels
   2.3 Semantic Relationships
      2.3.1 Broader/Narrower Relationships
      2.3.2 Associative Relationships
   2.4 Documentary Notes
   2.5 Concept Schemes
3 Networking Knowledge Organization Systems on the Semantic Web
   3.1 Mapping Concept Schemes
   3.2 Re-using and Extending Concept Schemes
   3.3 Subject Indexing and SKOS
4 Advanced SKOS: When KOSs are not Simple Anymore
   4.1 Collections of Concepts
   4.2 Advanced Documentation Features
   4.3 Relationships between Labels
   4.4 Coordinating Concepts
   4.5 Transitive Hierarchies
   4.6 Notations
   4.7 On Specializing the SKOS Model

Thesaurus Exploitation Service Modules
- Keyword Completion
- Keyword Explanation
- Keyword Translation
- Query Reformulation
- Keyword Validation
- Thesaurus Visualization

REST API
- GetCapabilities
- GetSuggestions
- GetSynonyms
- GetRelatives
- DescribeConcept
You should extend your metadata management software – we offer you the tools. Concrete prototypes to be developed in eENVplus: (1) eENVplus MD editor (2) JRC Geoportal (3) Disy Preludio

The best way to exploit the thesaurus framework depends on:
- Your users
- Your software environment
- Your content
- ....

We support finding the best individual fit by a very flexible and configurable software architecture
KeywordCompletion may save your time

KeywordExplanation, KeywordTranslation, ... may help you better understand the offered indexing terms and their correct use

- Also if you are not working in your mother language
- Also if you are not working in your preferred domain topic
- Browsing through broader, narrower, related terms and through related domain-specific thesauri may provide further help

KeywordValidation may automatically or semi-automatically provide a „normal form“ for your metadata

Metadata transformation services might be applied to legacy metadata
Simple Usage of KeywordCompletion, KeywordExplanations and QueryReformulation for Improving Metadata Data Search

- Suggested keyword completions when typing
- Search terms automatically extended by synonyms, translations, identical concepts from other thesauri, ...
- Search results from different multilingual, cross-domain sources
Semantic explorative search:

- during metadata compilation or during manual data discovery
- browse through a visualization-based of interlinked thesaurus structures
- keyword translations by cross-walking

**Translations:**
- Arabic: تربة
- Basque: lurzoru
- Bulgarian: Почва
- Catalan: sòl
- Chinese: 土壤
- Croatian: tlo
- Czech: půda
- Danish: jord

**Definitions:**
The top layer of the land surface of the earth that is composed of disintegrated rock particles, humus, water and air. (Source: CED)
Semantic explorative search:
- during metadata compilation or during manual data discovery
- browse through a visualization-based interlinked thesaurus structures
- keyword translations by cross-walking

Mock-Up: LusTRE Exploitation Services for Browsing Conceptual Spaces
HTTP web server implementing specified interface provides access to the exploitation service from outside

JSON-based and RESTful architecture makes it instantly usable for developers working on client-side environment
Operation responses typically deliver keyword objects:

- Enhances simple keyword value
- Contains lexical value, language and concept URI

<table>
<thead>
<tr>
<th>KeywordObject</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ title: String</td>
</tr>
<tr>
<td>+ language: String</td>
</tr>
<tr>
<td>+ conceptUri: String</td>
</tr>
</tbody>
</table>

- Keeping the reference to a concept helps to get hierarchical information, relatives and synonyms
### Request parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Mandatory</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>concept</td>
<td>String</td>
<td>True</td>
<td>A thesaurus concept URI as search criteria.</td>
</tr>
<tr>
<td>maxCount</td>
<td>Integer</td>
<td>False</td>
<td>Maximum number of results to be returned (default = 100).</td>
</tr>
<tr>
<td>thesauri</td>
<td>String</td>
<td>False</td>
<td>The thesauri data source to be searched.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>This parameter enables the limitation the query target to only specific thesauri (All thesauri would be targeted, if not specified).</td>
</tr>
<tr>
<td>service</td>
<td>String</td>
<td>False</td>
<td>The specific service modules to be used by this request.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The parameter value can be a comma-separated list of service modules.</td>
</tr>
<tr>
<td>languages</td>
<td>String</td>
<td>False</td>
<td>The targeting natural language of the query.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>This parameter enables the limitation of the query to target only the thesauri concepts in the enlisted languages. The parameter value can be a list of comma-separated language codes based on the ISO-639-1 standard (All available languages would be targeted, if not specified).</td>
</tr>
<tr>
<td>extraParams</td>
<td>String</td>
<td>False</td>
<td>Placeholder foreseen for the additional extra parameters, in case defined by a service module. The value should be a list of comma separated key/value pairs. The key/values must be delimited using a “:” character. Syntax: extraParams=param1:value1,param2:value2</td>
</tr>
</tbody>
</table>
Request:


Response:

```
{
    "relatives": [
        {
            "relation": {
                "label": "broader",
                "relationUri": "http://www.w3.org/2004/02/skos/core#broader"
            },
            "keyword": {
                "title": "Pedosphere",
                "language": "en",
                "conceptUri": "http://www.eionet.europa.eu/gemet/concept/6094"
            }
        },
        {
            "relation": {
                "label": "narrower",
                "relationUri": "http://www.w3.org/2004/02/skos/core#narrower"
            },
            "keyword": {
                "title": "Soil type",
                "language": "en",
                "conceptUri": "http://www.eionet.europa.eu/gemet/concept/7899"
            }
        }
    ]
}
```
The eENVplus Thesaurus Framework exploitation services allow for a maximum flexible and powerful access to the eENVplus interlinked thesauri knowledge from any MD management tool.

Enhancing your MD management tool by eENVplus thesaurus exploitation services supports multilinguality, broader and deeper domain coverage and cross-domain services.

We invite you:

☐ To take a look at the Thesaurus Framework at: http://linkeddata.ge.imati.cnr.it:2020
☐ To interlink your vocabularies/thesauri with the LusTRE thesauri
☐ To test our Thesaurus Framework exploitation services from autumn 2014 on

... Thank you!