EnviroInfo 2011: Innovations in Sharing Environmental Observations and Information

Integrating Environment into the Linked Data Cloud

Pilar León Araúz[1], Pedro J. Magaña Redondo[2], Pamela Faber[1]

Abstract

Currently, there are many environmental data sets available; however, they lack homogeneity, since they have been created for very different purposes and in very different formats. In this paper we propose a model capable of integrating some of these data sets with EcoLexicon, an e-environment knowledge base. EcoLexicon can contribute to the development of the SEIS and SISE from both a linguistic and knowledge representation perspective. It facilitates learning and communication and is also integrated in an ontological model. This facilitates its integration in the Semantic Web, and, thus supports the development of implementation tools for SEIS, the main objective of the SISE. Our initial proposal within the framework of Linked Data is to integrate EcoLexicon with DBpedia, GeoNames and GEMET.

1. Introduction

One of the main goals of the Shared Environmental Information Space (SEIS) and the Single Information Space in Europe for the Environment (SISE) is to provide some sort of integrated and shared environmental information space. This can be accomplished with a knowledge framework capable of managing and integrating information from different sources. This task has much in common with the Semantic Web initiative, from which the e-environment community can benefit, especially when it comes to the interoperability of previously developed technologies (León Araúz et al., 2009). Currently, there are many environmental data sets available, such as the General Multilingual Environmental Thesaurus (GEMET), Environmental Applications Reference Thesaurus (EARTh), Umwelt Thesaurus (UMTHES), SilvaTerm, and ontologies, such as EnVo, SWEET, and the Earth and Planetary Ontology. However, they lack homogeneity, since they have been created for very different purposes and in very different formats. It is thus our challenge to propose a model capable of integrating some of these data sets with EcoLexicon (http://ecolexicon.ugr.es), an e-environment knowledge base.

2. EcoLexicon: an e-environment resource

EcoLexicon is a multilingual terminological knowledge base (TKB) on the environment. This TKB was initially implemented in Spanish, English, and German. Currently, four more languages are being added: Modern Greek, Russian, French, and Dutch. So far it has a total of 3,271 concepts and 14,644 terms. EcoLexicon can contribute to the development of the SEIS and SISE from both a linguistic and knowledge representation perspective. It is conceived as a knowledge acquisition tool for a wide range of agents involved in environmental communication (i.e. environmental experts, specialized translators, technical writers, lay users, etc.). According to Hrebícek and Pillmann (448: 2009), the aim of the SEIS is to provide environmental information to all citizens. In this sense, EcoLexicon facilitates learning and commu-

[1] Department of Translation and Interpreting, University of Granada, Spain
[2] Andalusian Centre for the Environment (CEAMA), University of Granada, Spain

nication and eliminates terminological confusion. Undoubtedly, one of its main assets is its multilinguality. However, it also includes semantic networks, graphical resources, and contextual information that enhance the representation of conceptual and terminological knowledge. These features also help to raise public awareness of environmental issues and contribute to the standardization of designations in different languages, which also promotes shared knowledge at an international level. EcoLexicon is primarily hosted in a relational database (RDB), but at the same time it is integrated in an ontological model. This facilitates its integration in the Semantic Web, and, thus supports the development of implementation tools for SEIS, the main objective of the SISE (Hrebícek and Pillmann, 2009: 449).

2.1 User interface

Each entry in EcoLexicon provides multiple interrelated modules, such as those shown in Figure 1 for the concept EBB CURRENT. Users are not obliged to view all this information at the same time, but can browse through the interface, depending on their needs. Users can perform both single and combined term searches and obtain different results that can be retrieved from the Search results tab. Alternatively, they can also browse the semantic networks. Furthermore, in the left margin of the screen, users can visualize and expand four modules: Terms, Resources, Definition and Domains.

Under the tag *Terms*, different multilingual choices are shown to the users, who can then click on any of them and obtain terminological information, such as whether a linguistic designation is the main entry term or if it is a synonym, acronym or register and style variant. Information is also provided regarding how terms are used in real contexts. Figure 1 shows that EBB CURRENT has a total of 16 different designations in Spanish, English, German, and Greek since all registers and linguistic varieties are accounted for. For instance, in this case, *ebb tide* is a non-technical variant for the concept EBB CURRENT. Even in large term bases, multilingual variety is rarely represented in an exhaustive way. However, this type of information is invaluable because not only does it provide users with multiple options for specialized text comprehension and production, but it is also useful for conceptual disambiguation (see section 3).

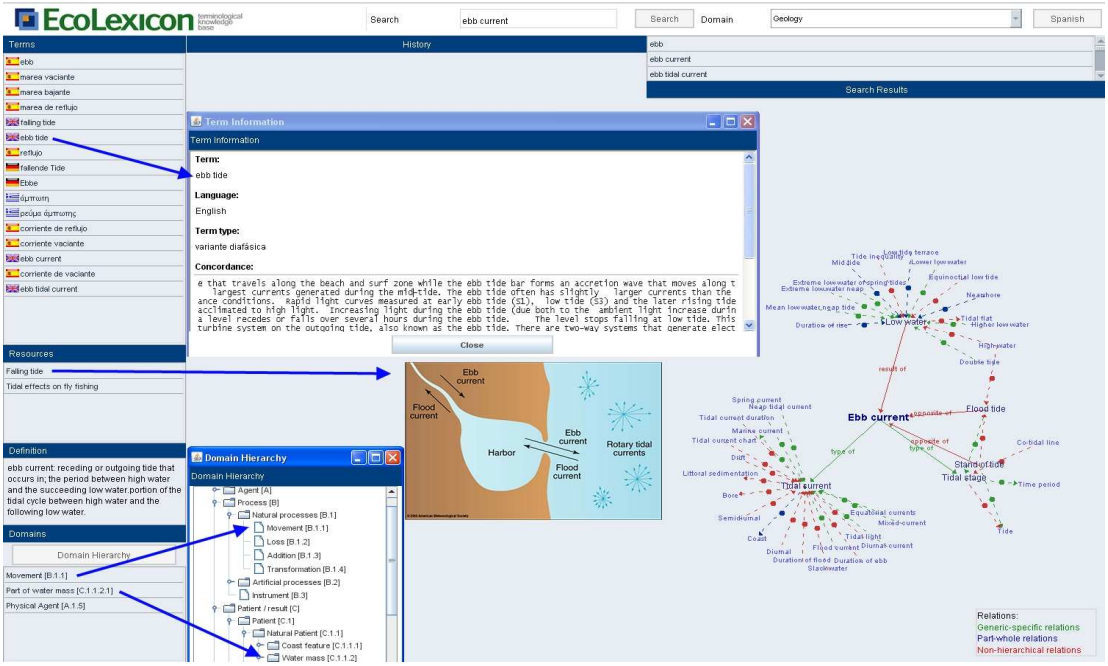


Figure 1. The EcoLexicon user interface: entry for EBB CURRENT

When users click on the tag *Resources*, they are provided with informative URLs or graphical information (images, graphics, charts, etc.) specifically selected, depending on the information contained in *Definitions*, the next tag. Definitions are modelled in terms of the constraints imposed by the domain-specific conceptual categories and relations in EcoLexicon (Faber et al., 2007). This produces a set of templates based on category membership, which is also shown under the tag *Domains*. Figure 1 shows that the definition of EBB CURRENT has the genus TIDE, which, because of its multidimensionality, can potentially belong to three domains: MOVEMENT, PART OF WATER MASS, and PHYSICAL AGENT.

At a more fine-grained level, concepts are displayed in a dynamic network linked to other concepts (right-hand side of the window). As shown in the lower righthand corner, conceptual relations are color-coded and classified into generic-specific, part-whole, and non-hierarchical relations, which contribute to the representation of knowledge natural dynamism (León Araúz and Faber, 2010). Users can click on any of these concepts and thus further expand their knowledge of this domain sector. Nevertheless, problems can arise when it is a question of browsing networks of very general concepts, which are linked to too many other concepts and thus carry an excessive load of information.

The environmental domain has many concepts that can be represented from very diverse perspectives since it is such a vast knowledge area. This is known as multidimensionality (Rogers, 2004; Kageura, 1997) and is commonly regarded as a way of enriching conventional static knowledge representations. However, this can also lead to an information overload that is a serious obstacle to knowledge acquisition. General versatile concepts, such as WATER, share multiple relations with many other concepts, but they rarely, if ever, activate all those relations at the same time since this would evoke completely different and incompatible scenarios. In this sense, although concepts are entrenched cognitive routines which are interrelated in various ways that facilitate their co-activation, they actually retain sufficient autonomy so that the activation of one does not necessarily entail the activation of all of the rest (Langacker 1987: 162). In line with situated cognition, our claim is that any specialized domain contains sub-domains in which conceptual dimensions become more or less salient, depending on the activation of specific contexts (León Araúz and Faber, 2010; León Araúz et al., in press). Thus, context is a dynamic construct that triggers or restricts knowledge. For instance, the proposition, WATER TREATMENT PLANT *affects* WATER, would only be informative for users whose knowledge search activated a wastewater scenario, whereas WATER *causes* EROSION would only be useful for users that wished to situate the concept in a geological scenario. It is evident that the same query would never lead to the retrieval of both propositions unless users performed their search in a context-free mode.

The area of environmental knowledge was thus divided into a set of contextual domains (e.g. HYDROLOGY, GEOGRAPHY, OCEANOGRAPHY, CIVIL ENGINEERING, ENVIRONMENTAL ENGINEERING, etc.) and the relational power of concepts was constrained accordingly. Contextual domains were allocated in a similar way as in the European General Multilingual Environmental Thesaurus, whose structure is based on themes and descriptors reflecting a systematic, category or discipline-oriented perspective (GEMET 2004).

In EcoLexicon, contextual constraints are neither applied to individual concepts nor to individual relations. Instead, they are applied to conceptual propositions, each of which may be assigned to more than one domain. This constrains versatile concepts, such as WATER, as well as other concepts that are linked to more general ones. For instance, EROSION takes the following shape in a context-free network (Figure 2), which appears overloaded mainly because it is closely linked to WATER, one of the most important agents of this process.

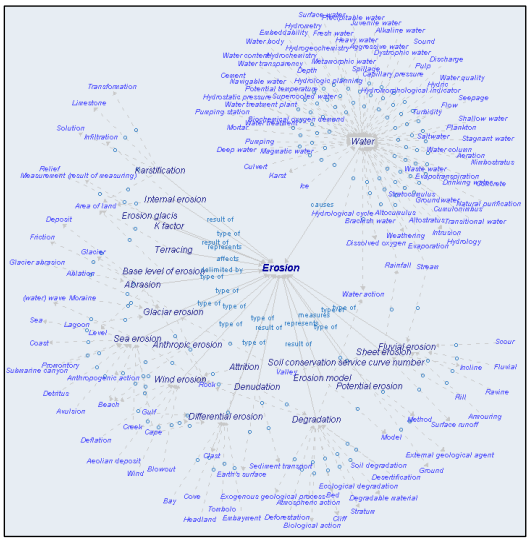
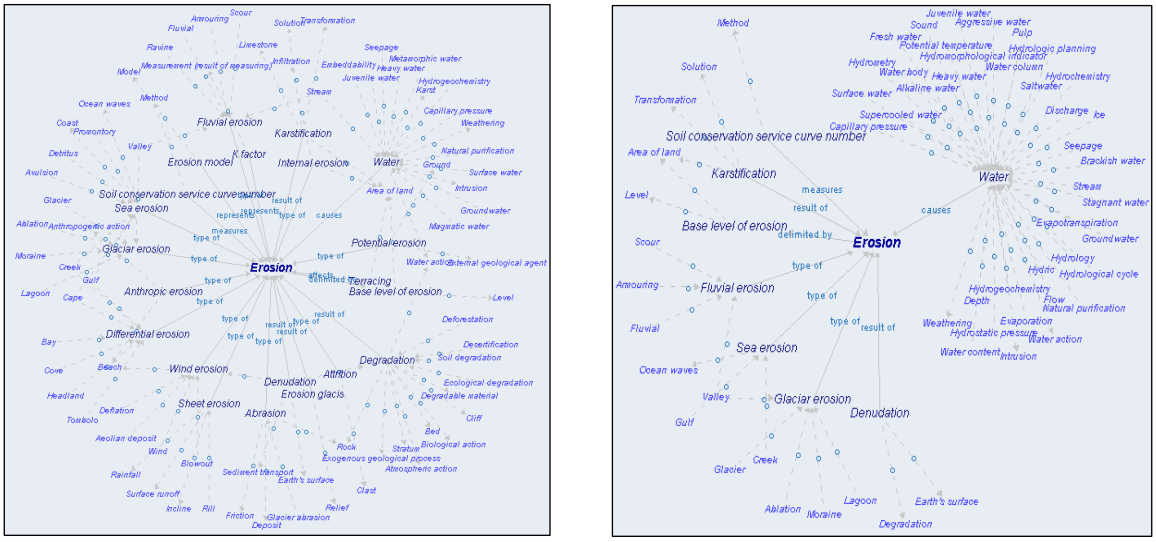


Figure 2. EROSION context-free network

When contextual constraints are applied, however, EROSION is only linked to other concepts by propositions applicable in GEOLOGY (Figure 3) or HYDROLOGY (Figure 4).



Figures 3 & 4. *EROSION in the GEOLOGY and HYDROLOGY contextual domains*

2.2 Relational database and ontology

EcoLexicon can undoubtedly benefit from the Semantic Web initiative. Ontologies are a powerful representational model since they add the semantic expressiveness lacking in RDBs. This enriches potential queries because reasoning techniques can be applied to extract implicit information. Nevertheless, Eco-

Lexicon, in the same way as many other resources, was initially not conceived as an ontology. This means that legacy systems (RDB stored information) must be linked to an ontological system (León Araúz and Magaña Redondo, 2010). This is not an easy task since both representational models are very different. In contrast to relational databases, ontologies are highly expressive relational structures that strive to describe concepts in very similar terms to those used by humans. EcoLexicon stores semantic information in the ontology, while leaving the rest in the relational database. In this way, we can continue using the new ontological system, while at the same time feeding the database.

As seen in Figure 5, contextual domains have inspired the design of our ontology classes. The ontology is automatically retrieved from the data stored in our RDB, according to the following assumption: if a concept *c* is part of one or more propositions allocated to a contextual domain *C*, *c* will be an instance of the class *C*. This contextual category structure makes user queries more dynamic since they can perform different searches through the union and/or intersection of our domains. In this way, they can obtain new but still cognitively-sound knowledge networks. For instance, the intersection of HYDROLOGY and GEOLOGY would restrict the conceptual structure to only HYDROGEOLOGICAL propositions, whereas the union of ENVIRONMENTAL, MINING, TRANSPORT, HYDRAULIC and COASTAL ENGINEERING would make up the whole domain of CIVIL ENGINEERING.

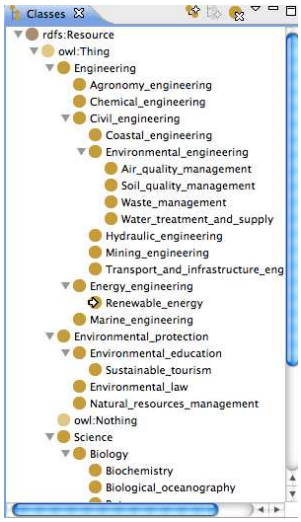


Figure 5. Ontology classes

Furthermore, our conceptual relations can be enhanced by an additional degree of OWL semantic expressiveness provided by property characteristics, such as transitivity. For this purpose, partonymy has been split up into six different relations (*part_of*, *composed_of*, *takes_place_in*, *phase_of*, *located_at*, *delimited_by*) because not all parts interact in the same way with their wholes. For example, if *located_at* were regarded as a *part_of* relation, that would cause fallacious transitivity (Murphy 2003). If a GABION is *part_of* a GROYNE and a GROYNE *part_of* the SEA, an ontology would infer that GABIONS are *part_of* the SEA, which is not a plausible example. Thus, the only relation that can be considered fully transitive is *part_of*, which is only used for physical objects that are sharply bounded in space. For example, from our ontology we can retrieve BERM and BEACH as two wholes of which BERM CREST is a part, despite the fact that the only explicit propositions stored in the ontology are BERM CREST *part_of* BERM and BERM *part_of* BEACH. However, transitivity does not extend to higher levels of the ontology since BEACH is not *part_of* anything but is *located_at* the COAST. The reason for this is that evidently not all COASTS need BEACHES to exist. In this way, the proposition BERM CREST *part_of* COAST is avoided and is thus not inferred.

3. Linked data

Once information can be accessed by using ontological resources, it is also easier to connect it with other environmental resources. Even though reusability is often based on data merging, this has the disadvantage of leading to a heterogeneous blending of very diverse data that were originally stored for very different purposes. Various techniques have been proposed for linking data. Previous attempts use automatic mediation algorithms to map and merge ontology schemas (de Brujin et al., 2006). However, an important drawback is that the schemas do not always remain public. However, many systems provide interfaces to interact with their structured data, namely, APIs (Application Programming Interfaces). These interfaces allow developers to combine information from different data sources and create new services known as mashups (Zang et al., 2008). Nevertheless, given the fact that many APIs are proprietary, it is thusnot possible to set links between data objects.

Linked Data (Berners-Lee, 2006) is an innovative approach to this problem. It uses Semantic Web technologies to publish structured data and, at the same time, set links between data items in one data source and data items in other data sources (Heath et al. 2008). This makes it possible to connect data items in different resources, and at the same time keep the resources independent. In our opinion, this methodology can be successfully applied to EcoLexicon to link its data to other semantically-related data sources. This would also enable other environmental resources to enhance their systems with our information, which would help to build a real environmental community of shared data within the Linked Data framework.

Linked data provides a framework to set RDF links pointing to other data sources on the Web. Technically, an external RDF link is an RDF triple in which the subject of the triple is a URI reference in the namespace of one data set, while the predicate and/or object of the triple are URI references pointing to the namespaces of other data sets (Heath and Bizer, 2011). DBpedia is the linked data version of Wikipedia and is at the core of such an initiative, given that it is one of the central knowledge sources of the web. Connecting the concept WASTEWATER from EcoLexicon to that of DBpedia results in the following statement:

```
<http://ecolexicon.ugr.es/resource/wastewater>
<http://www.w3.org/2002/07/owl#sameAs>
<http://dbpedia.org/resource/Wastewater>
```

3.1 Integrating EcoLexicon into the linked data cloud

Our initial proposal is to integrate EcoLexicon with DBpedia and GeoNames, which are already part of the Linked Data cloud, and GEMET, the GEneral Multilingual Environmental Thesaurus. The GeoNames database integrates data, such as place names in various languages, geographical features, population, and other related information from various sources (i.e. the National Geospatial-Intelligence Agency, the US Board of Geographic Names, the US Geological Survey Geographic Names Information System, etc.). All the information in GeoNames is organized in nine classes called features. In contrast, GEMET was developed as an indexing, retrieval, and control tool for the European Topic Centre on the Catalogue of Data Sources and the European Environment Agency. One of its aims is to define a core general terminology for the environment in 17 European languages and structure it in a set of semantic fields called themes and groups. We decided to link EcoLexicon to these three resources in order to cover a wide scope: (1) DBpedia is the core of Linked Data; (2) GEMET is more connected to environmental institutions and policies; (3) GeoNames opens up the possibility of exploring a geolocalization line in EcoLexicon.

Linking data sources from DBpedia can be quite straightforward since different tools, such as the ontology editor TopBraid Composer, can automatically suggest links. However, due to lexical variation and the lack of univocity in both general and specialized knowledge, automatic mappings are not always viable.

This means that manual work is still necessary and desirable to a certain extent. For example, an automatic mapping of the geological concept CAPE in EcoLexicon to DBpedia resources could be misleading since *cape* is a highly polysemous term. There are fewer problems related to polysemy in GeoNames and GEMET because all concepts are domain-specific, but lexical variation may impair the string matching process.

Nevertheless, instead of mapping one-to-one manual correspondences, we can take advantage of the semantics contained in our resource. Term strings from EcoLexicon can be compared with those from the other three resources, enhanced by EcoLexicon data sets that include multilingual choices and variants, category membership, and semantic relations, such as *is_a* and *part_of*. As previously mentioned, the use of multilingual choices is a powerful method for conceptual disambiguation. Nevertheless, monolingual variants also ensure a systematic matching procedure since not all concepts are designated by their canonical form. This means that a *sameAs* relationship can still be semi-automatically established even between concepts that are designated by different terms in each data set.

The first step in the data linking process is the comparison of the string of all our English variants with the DBpedia, Geonames, and GEMET entries. Since these strings may match various entries in DBpedia and lead to erroneous mappings, conceptual disambiguation is then performed by comparing other language equivalents. In this way, CAPE in EcoLexicon would only be linked to the concept in DBpedia with the Spanish equivalent of *cabo* (geographical landform) and not *capa* (clothing article). Nevertheless, in those cases in which polysemy also occurs at a cross-linguistic level, category membership information must be added to the linking algorithm. If any term belonging to the same contextual class of the search concept appears in any of the RDF properties, then concepts are equivalents.

Contextual classes are also used when mapping EcoLexicon concepts to GEMET and GeoNames. Accordingly, our classes were manually compared to those in both resources in order to restrict the number of concepts to be mapped and ensure context-based correspondences. Since our domains were designed in a similar way to those in GEMET, all themes in the latter coincide with one or more of our classes. For example, GEMET's CHEMISTRY theme matches our classes CHEMISTRY and CHEMICAL ENGINEERING. As for GeoNames, we have ruled out some of the nine features, such as CITY AND VILLAGE since EcoLexicon does not as yet cover instances. Rather our focus is on both natural landforms (PEAK, CAPE, BEACH) and artificial constructions (DAM, BREAKWATER, JETTY) which, based on the data in EcoLexicon, are the most interesting ones to be shown on a map. In our ontology, natural landforms and artificial constructions may only be contained in five of our classes. Thus, the GeoNames features SPOT, BUILDING, and FARM are mapped onto our CIVIL ENGINEERING class, whereas MOUNTAIN, HILL, ROCK and STREAM, LAKE and UNDERSEA are mapped onto GEOGRAPHY, GEOLOGY, OCEANOGRAPHY, and HYDROLOGY.

Finally semantic relations are also used in GeoNames to display more concepts on the map. For instance, once a concept has been matched in the two resources, all of its wholes (if any) and/or its immediate superordinate concept will also be included by transitivity. This means that not only will the concept DOCK be shown on the map, but also its hyperonym COASTAL STRUCTURE and its whole HARBOR. Furthermore, certain categories in GeoNames refer to parts of other concepts, such as *section of stream*, *section of lake*, and *section of harbor*. In these cases, the string after *section of* will be compared with the concepts in EcoLexicon. If there is any match, then all of its parts (if any) would be included.

Semantic relations can also be used in the rare cases of ambiguity in GEMET, given that this is the only resource of the three that is hierarchically organized. Apart from the themes and groups, GEMET is organized in broader, narrower and related terms. In those cases where multilingual choices are not sufficient, all related terms can be mapped onto our networks and be used for disambiguation.

3.2 The case of SPIT, BANK, and WASTEWATER TREATMENT PLANT

To illustrate our data linking proposal, we have chosen three concepts: SPIT, BANK, and WASTEWATER TREATMENT PLANT. Although all of them are included in the three resources, there are certain differences. In EcoLexicon, English and Spanish terms for the concept SPIT are *spit*, *sand spit*, *offshore beach*, *barrier beach*, *playa barrera* and *cordón litoral*. These lexical variations allow us to identify and link the same concept in the three resources, since in Geonames and DBpedia SPIT is designated by *spit*, whereas in GEMET, it is designated by *barrier beach*[3]. However, in DBpedia, *spit* is a polysemic term. It may refer to *rain*, *saliva*, *rapping* or a *coastal landform*, among other things. In this case, contextual categories are not necessary for disambiguation, since comparing the property *label* of all possible forms with our Spanish equivalents leads us to the term *cordón litoral*, which is not ambiguous[4]. Thus, the underlying algorithm of this linking rule uses our data sets related to linguistic variations and multilingual choices.

The case of BANK is similar to that of SPIT. Nevertheless, it is necessary to add other parameters to the linking rule since *bank* is also polysemic in Spanish. In the same way as in English, *banco* can refer to a geographic landform or a financial institution, and there are not many other common multilingual equivalents in DBpedia for disambiguation. In DBpedia, this domain-specific entry is named, and differentiated from others, such as BANK (GEOGRAPHY)[5]. In order to match this entry and not any of the others it is necessary to add a context-based rule. Therefore, this match will occur in the following contexts: (1) when the word in brackets matches the string of any of our classes; (2) when any term associated with any concept belonging to the same contextual category as the search concept appears in one or more of the following properties: *dbpedia-owl:abstract*, *dcterms:subject*, *rdfs:comment*, or *dbpedia-owl:wikiPageRedirects*. In this case BANK in EcoLexicon belongs to the classes, GEOGRAPHY, GEOLOGY and OCEANOGRAPHY, as do many other concepts, such as SHORELINE, ESTUARY, RESERVOIR, SLOPE, RIVER, MARSH, etc, all of which are contained in the properties *dbpedia-owl:abstract*, *dcterms:subject* and *rdfs:comment*. Furthermore, since the disambiguating word in brackets coincides with the EcoLexicon class GEOGRAPHY, the second step is not required. In GeoNames, there is only one BANK concept. As for GEMET, the concept is also distinguished from others, such as BANK (LAND)[6]. Nevertheless, these steps are not required, since the Spanish equivalent *ribera* matches one of the variations in EcoLexicon.

Finally, WASTEWATER TREATMENT PLANT does not show ambiguity problems because it is a very specialized concept. However, it is a clear example that reinforces the need for term bases to store not only multilingual choices but also monolingual variations. In EcoLexicon there are 17 terms in English, Spanish, and Modern Greek associated with this concept. If lexical variations were not so exhaustively collected in EcoLexicon, we would not be able to link it to the three resources, since each resource gives a different term to designate the concept. In DBpedia, WASTEWATER TREATMENT PLANT is named *wastewater treatment plant*[7], the same as in EcoLexicon. However, the term in GEMET is an orthographic variation (*waste water treatment plant*[8]), whereas the GeoNames term (*sewage treatment plant*) clearly reflects register variation.

This is a good example of how linking data does not always ensure knowledge acquisition by the user, since conceptual modeling does not necessarily follow a concrete pattern in all resources. Consequently, there is no assurance that the content is well structured. The definition of *wastewater treatment plant* in DBpedia does not describe the concept at all. In fact, it is wrongly ascribed to a disambiguation category and it redirects users to different types of WASTEWATER TREATMENT. Nevertheless, it does not offer a

[3] http://www.eionet.europa.eu/gemet/concept?cp=10149&langcode=en&ns=1
[4] http://dbpedia.org/page/Spit_%28landform%29
[5] http://dbpedia.org/page/Bank_%28geography%29
[6] http://www.eionet.europa.eu/gemet/concept?cp=13552&langcode=en&ns=1
[7] http://dbpedia.org/page/Wastewater_treatment_plant
[8] http://www.eionet.europa.eu/gemet/concept?cp=9144&langcode=en&ns=1

proper definition of the PLANT itself The Spanish version of Wikipedia has a good entry for its equivalent (*estación depuradora de aguas residuales*), but there is no link between them.

The pseudocode of the different matching algorithms is shown in the next table:

```
for each c:context in ecolexicon
for each t:theme_list in gemet
  if context_matching_gemet(c, t) > context_threshold
    for each w:word in c
    for each cp:concept in t
      all_words = language_variants(w)
      for each v:variant in all_words
        v' = stem(v); cp' = stem(cp)
        if str_compare(v', cp') > word_threshold
          multi_e = multilingual_variants(v)
          multi_g = multilingual_variants(cp)
          if multingual_compare(multi_e, multi_g) > multilingual_threshold
             result.add(pair(v, cp))
          n_terms = narrower_terms(cp); t_terms = type_of_terms(v)
          for each n:n_term in n_terms
          for each t:t_term in t_terms
            if narrower_compare(n, t) > narrower_threshold
              result.add(pair(n, t))
```

```
for each c:context in ecolexicon
for each f:feature in geonames
  if context_matching_geonames(c, f) > context_threshold
    for each w:word in c
    for each cp:concept in f
      w' = stem(w); cp' = stem(cp)
      if str_compare(v', cp') > word_threshold
        result.add(pair(v, cp))
        result.add(make_pairs(super(v), cp))
        result.add(make_pairs(part_of(v), cp))
        result.add(make_pairs(v, part_of(section_of(cp)))
```

```
for each w:word in ecolexicon
for each cp:concept in dbpedia
    w' = stem(w); cp' = stem(cp)
    if str_compare(v', cp') > word_threshold
      multi_e = multilingual_variants(v)
      multi_g = multilingual_variants(cp)
      if multingual_compare(multi_e, multi_g) > multilingual_threshold
        result.add(pair(v, cp))
        related_instances = instances_of(context(v))
        for each i:instance in related_instances
          if look_for_text(comment_properties(cp), i) > text_threshold
            result.add(pair(v, i))
```

4. Conclusions and future work

Even though Linked Data is conceived as a simple and efficient way for the integration of heterogeneous information in the Web, making links is not a trivial task when it comes to set correspondences between general and specialized knowledge concepts or concepts designated by different linguistic variations. Different EcoLexicon's features, such as, multilinguality, category membership and semantic relations are useful data sets to ensure concept disambiguation and an accurate string matching process.

Bibliography

Berners-Lee, T. 2006. Linked Data. *W3C Design Issues*.

De Bruijn, J., Ehrig, M., Feier, C., Martín Recuerda, F., Scharffe, F., and Weiten, M. (2006). "Ontology Mediation, Merging, and Aligning". *Semantic Web Technologies*.

Faber, P., P. León Araúz ,J. A. Prieto Velasco and A. Reimerink. (2007). "Linking Images and Words: the description of specialized concepts (extended version)". *International Journal of Lexicography* 20 (1), 39-65.

Langacker, R. W. 1987. *Foundations of cognitive grammar: theoretical prerequisites*, Vol 1. Stanford: Stanford University Press.

León Araúz, P., Reimerink, A., Aragón, A.G. (in press). "Dynamism and context in specialized knowledge". *Terminology*

León Araúz, P. and P. Faber. 2010. "Natural and contextual constraints for domain-specific relations". In *Proceedings of Semantic relations. Theory and Applications*. Malta.

León Araúz, P. and Magaña Redondo, P.J. (2010). EcoLexicon: contextualizing an environmental ontology. In *Proceedings of the Terminology and Knowledge Engineering (TKE) Conference 2010*. Dublin: Dublin City University.

León Araúz, P., Magaña Redondo, P. and Faber, P. (2009). Building the SISE: an environmental ontology. In *TOWARDS eENVIRONMENT Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe*. Brno: Masaryk University.

Murphy, M. L. 2003. *Semantic Relations and the Lexicon.* Cambridge: Cambridge University Press.

GEMET. 2004. "About GEMET. General Multilingual Environmental Thesaurus". Available at: http://www.eionet.europa.eu/gemet/about

Heath, T. and Bizer, C. 2011. *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

Hrebícek, J., Pillmann, W. 2009. Shared Environmental Information System and Single Information Space in Europe for the environment: Antipodes or Associates? In *TOWARDS eENVIRONMENT Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe*. Brno: Masaryk University.

Zang, N., Rosson, M. B. and Nasser, V. 2008. Mashups: who? what? why? In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, p. 3171-3176.