KNOWLEDGE REPRESENTATION IN ECOLEXICON[1]

Pamela Faber
Pilar León Araúz
Arianne Reimerink
Universidad de Granada (Spain)

ABSTRACT: EcoLexicon, a multilingual terminological knowledge base (TKB) on the environment, provides an internally coherent information system covering a wide range of specialized linguistic and conceptual needs. Our research has mainly focused on conceptual modelling in order to offer a user-friendly multimodal interface. The dynamic interface combines conceptual, linguistic, and graphical information and is primarily hosted in a relational database that has been recently linked to an ontology. One of the main challenges we have faced in the development of our TKB is the information overload generated by the domain. This is not only due to its wide scope, but especially to the fact that multiple dimensions are not always compatible but context-dependent. As a result, overloaded concepts have been reconceptualised according to two contextual factors: domain membership and semantic role.

Keywords: TKB, specialized knowledge representation, dynamism.

INTRODUCTION

EcoLexicon[2] is a multilingual knowledge resource on the environment. So far it has 3,115 concepts and 11,678 terms in Spanish, English and German. Currently, two more languages are being added: Modern Greek and Russian. It is aimed at users such as translators, technical writers, environmental experts, etc., which can access it through a friendly visual interface with different modules devoted to both conceptual, linguistic, and graphical information.

---

[1] This research has been supported by project FFI2008-06080-C03-01/FILO, from the Spanish Ministry of Science and Innovation.

[2] http://manila.ugr.es/visual

Each entry of EcoLexicon provides a wide range of interrelated information. In Figure 1, the GROYNE entry is shown. Users do not have to see all this information at the same time, but can browse through the different windows and resources according to their needs.

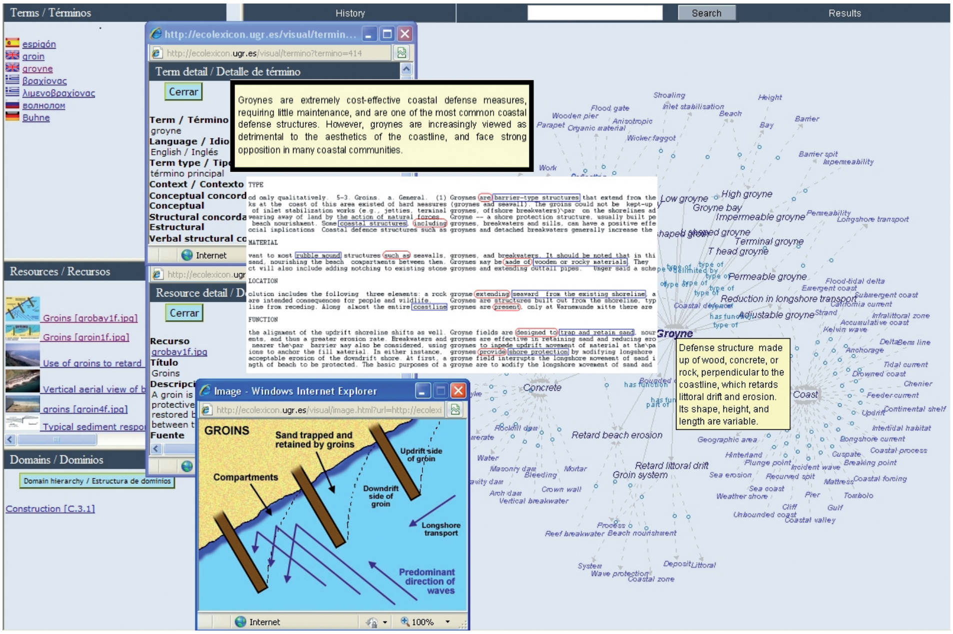


Figure 1. EcoLexicon user interface.

Under the tag 'Dominios' an ontological structure shows the exact position of the concept in the class hierarchy. GROYNE, for example, *is_a* construction (bottom-left corner of the window). The concept definition is shown when the cursor is placed on the concept. All definitions follow a category template (Faber et al., 2007) that constrains the definitional elements to be included. The definition for GROYNE, as a physical artificial object, is the linguistic expression of conceptual relations such as *is_a*, *made_of*, and *has_function*. Contexts (top window with black contour) and concordances (bottom window with black contour) appear when clicking on the terms, and inform different users about both conceptual and linguistic aspects. Graphical resources are displayed when clicking on the links in the box 'Recursos' (in the left-hand margin towards the middle), which are selected according to definitional information. At a more fine-grained level, conceptual relations are displayed in a dynamic network of related concepts (right-hand side of the window). The terminological units, under the tag 'Términos', designate the concept in English and Spanish: '*groyne*' and its variant '*groin*', and '*espigón*', respectively (top left-hand corner).

THE ENVIRONMENTAL EVENT

At a macrostructural level, all knowledge extracted from a specialized domain corpus has been organized in a frame-like structure or prototypical domain event, namely, the Environmental Event (EE; see Figure 2).

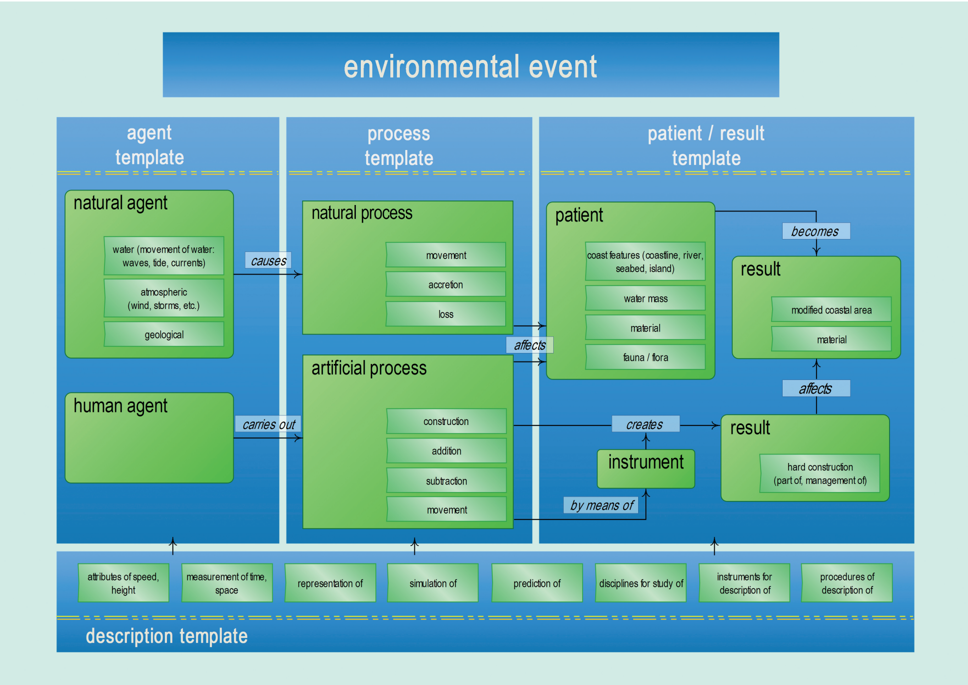


Figure 2. The Environmental Event (EE, Faber et al., 2005, 2006, 2007).

The EE provides a basic template applicable to all levels of information structuring. The Environmental Event (EE) is conceptualised as a dynamic process that is initiated by an agent (either natural or human), affects a specific kind of patient (an environmental entity), and produces a result in a geographical area. These macro-categories (agentà process à patient/result, and location) are the semantic roles characteristic of this specialized domain, and the EE provides a model to represent their interrelationships at a more specific level.

CONCEPTUAL RELATIONS

From a more fine-grained view, concepts appear in dynamic networks linking them to all related concepts by means of a closed inventory of semantic relations especially conceived for the environmental domain. Figure 3 shows the network of GROYNE, associated with other concepts in a two-level hierarchy through both vertical (*type_of*, *part_of*, etc.) and horizontal relations (*has_function*, *located_at*, etc.).

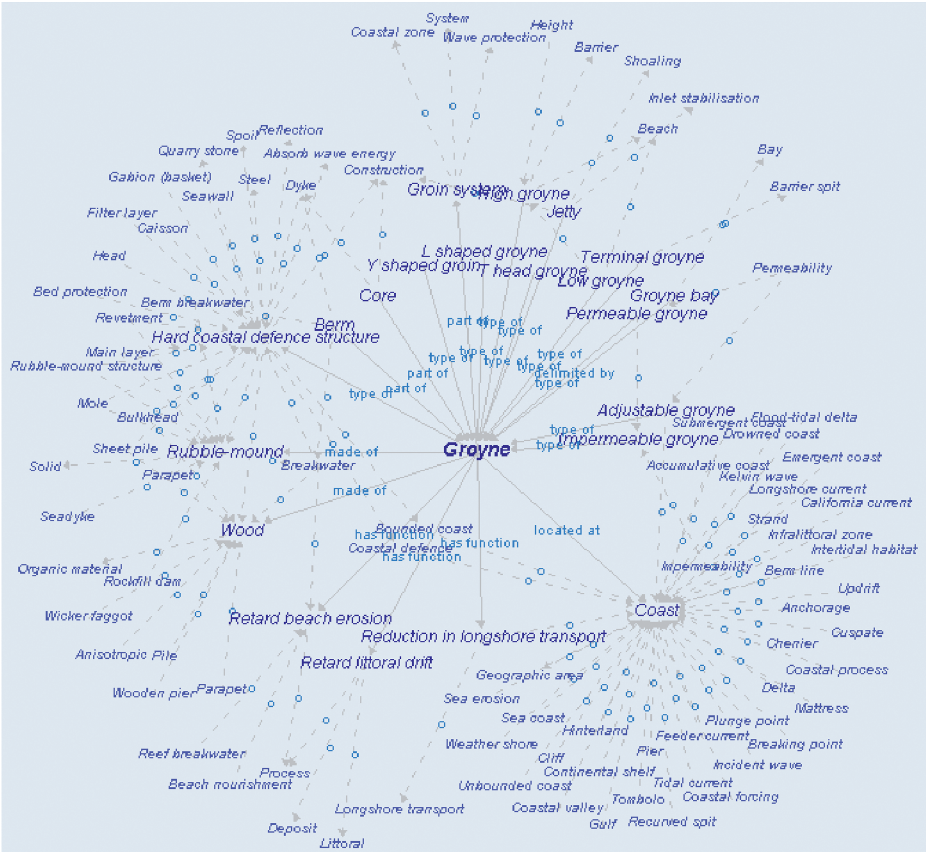


Figure 3. Conceptual network of groyne.

According to our corpus data, conceptual relations depend on concept types and their relational power. Table 1 shows our relation types associated with the elements they can link in each conceptual proposition (León Araúz, 2009; León Araúz & Faber, 2010).

Table 1. Relation types

Conceptual relations	**Concept 1**	**Concept 2**	**Examples**
Type_of	Physical entity Mental entity Process	Physical entity Mental entity Process	Masonry dam *type_of* dam
Part_of	Physical entity Mental entity	Physical entity Mental entity	Main layer *part_of* breakwater microbiology *part_of* biology
Phase_of	Process	Process	pumping *phase_of* dredging
Made_of	Physical entity	Physical entity	air *made_of* gas
Located_at	Physical entity	Physical entity	jetty *located_at* canal
Takes_place_at	Process	Process	Littoral transport *takes_place_at* sea
Delimited_by	Physical entity	Physical entity	estratosfera *delimited_by* estratopausa
Result_of	Process	Process	agraddation *result_of* sedimentation
Causes	Physical entity	Process	water *causes* erosion
Affects	Physical entity Mental entity Physical entity Mental entity Process Process	Process Entity Entity Process	groyne *affects* littoral transport pesticide *affects* water wave *affects* groyne precipitation *affects* erosion
Has_function	Entity	Process	aquifer *has_function* human supply
Attribute_of	Property Property	Entity Process	abyssal *attribute_of* plain anthropic *attribute_of* process

Apart from those reflected in Table 1, some of the relations have their own hierarchy. For example, *has_function* and *affects* include more specific knowledge, which is codified through domain-specific verbs: *studies*, *represents*, *measures*, *effected_by* (as functions of mental entities or instruments), or *erodes*, *changes_state_of*, etc (for processes or entities that affect others in a more concrete way).

According to the above-mentioned criteria, concept nature alone determines the potential activation of certain semantic relations, but at the same time, semantic relations determine which kind of concepts can be part of the same conceptual proposition. This gives rise to all these possible combinations (Figure 4).

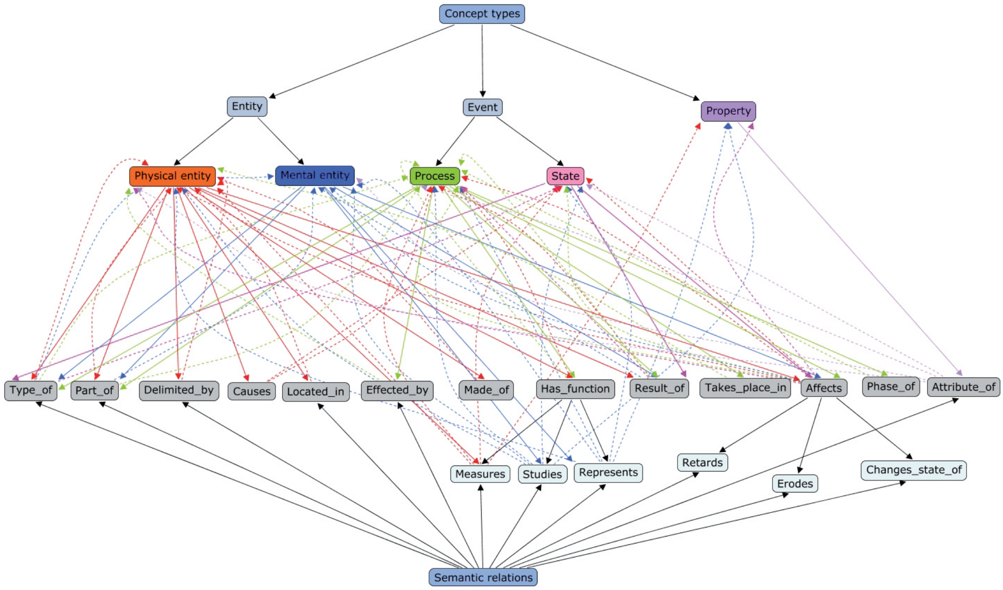


Figure 4. Combinatorial potential.

This combinatorial potential represents certain constraints associated with the natural aspect of concepts. For instance, a process may activate the relation *affected_by*, but only if it is associated with a physical entity. However, if it activates *affects*, it can be linked to entities, events and properties.

THE DOMAIN ONTOLOGY

Data in our TKB are primarily hosted in a relational database (RDB). This widespread modeling allowed for a quick deployment of the platform and fed the system from very early stages. Nevertheless, relational modeling has some limitations. One of the biggest ones is its limited capability to represent real-world entities, since natural human implicit knowledge cannot be inferred. This is why ontologies arose as a powerful representational model, but in our approach, we emphasize the importance of storing 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 legacy system.

Upper-level classes in our ontology correspond to the basic semantic roles described in the EE (agent-process-patient-result-location). As shown in Figure 5, all classes constitute a general knowledge hierarchy derived from each of them. This structure enables users to gain a better understanding of the complexity of environmental events, since they give a process-oriented general overview of the domain:

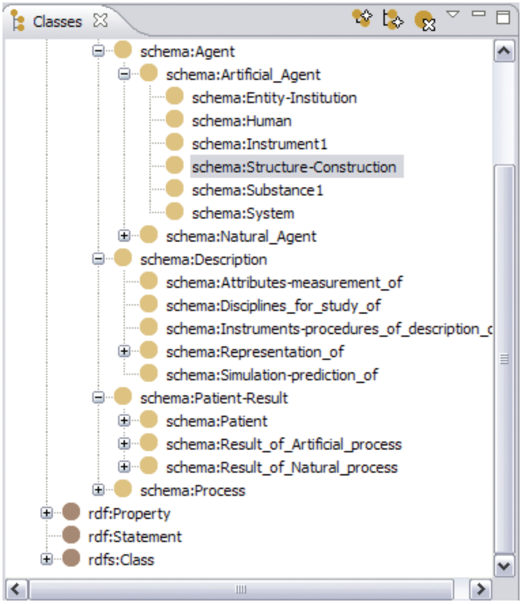


Figure 5. Ontological classes.

Those conceptual relations, specifically conceived for our Environmental TKB, can be enhanced by an additional degree of OWL semantic expressiveness provided by property characteristics. This is one of the main advantages of ontologies, making reasoning and inferences possible. For example, *part_of* relations can benefit from transitivity, as shown in Figure 6.

In this figure, a SPARQL query is made in order to retrieve which concepts are *part_of* Concept 3262, which refers to the concept SEWER. On the right side, DRAINAGE SYSTEM is retrieved as a direct *part-of* relation, whereas SEWAGE COLLECTION AND DISPOSAL SYSTEM and SEWAGE DISPOSAL SYSTEM are implicitly inferred through the Jena reasoner.

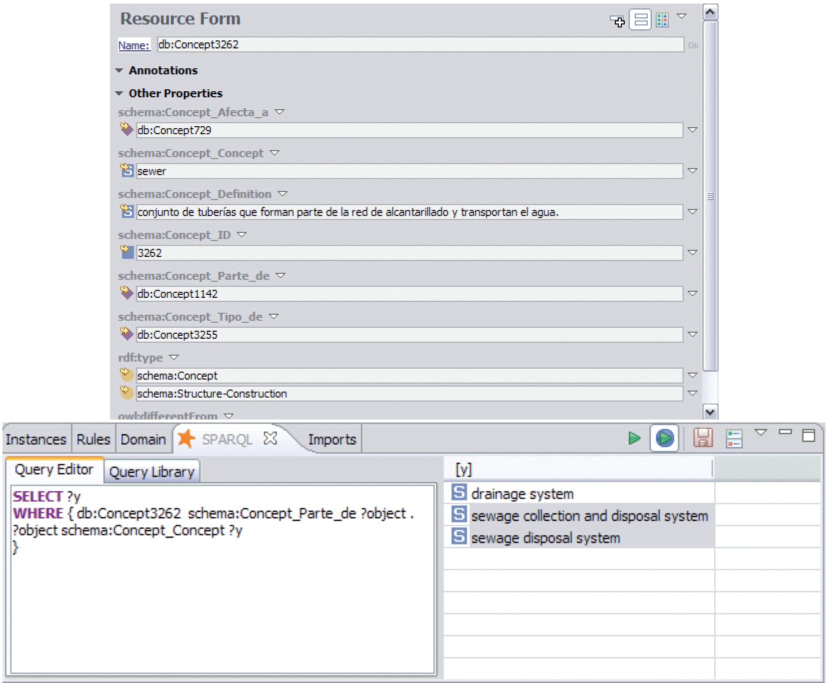


Figure 6. Concept SEWER in the ontology and inferred transitivity.

DEFINITIONS

In EcoLexicon definitions of concepts are elaborated following the constraints imposed by the EE and the inventory of conceptual relations. We group certain similar concepts in different templates according to category membership. For example, the definitional statement of GROYNE (Figure 7) is based on the number and type of conceptual relations defined for the category template HARD COASTAL DEFENCE STRUCTURE.

All coordinate concepts of GROYNE make use of the same template. As functional agentive entities, all HARD COASTAL DEFENCE STRUCTURES need the following information for an overall description: (1) the *is_a* relation marking category membership; (2) the material they are *made_of*, completed with the values of the CONSTRUCTION MATERIAL class; (3) their location, since a GROYNE is not a GROYNE if it is not *located_at* the SEA; and (4) especially the purpose for which they are built.

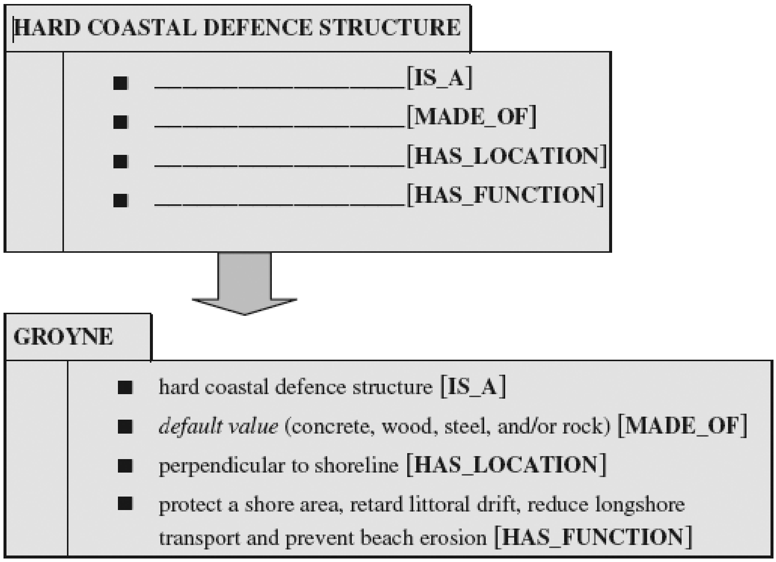


Figure 7. Activation of the HARD COASTAL DEFENCE template in the definition of GROYNE.

LINGUISTIC AND GRAPHICAL INFORMATION

Apart from concepts, conceptual networks, definitions and terms, EcoLexicon provides the user with additional information: linguistic contexts, concordances and images.

Linguistic contexts help the user achieve a level of understanding of a specialized domain. The linguistic contexts included in the TKB go beyond the relations expressed in the definition. In Table 2, for example, GROYNE is not only defined as a COASTAL DEFENCE STRUCTURE. Other relevant information is included as well: they are cost-effective and many coastal communities prefer other solutions.

Table 2. Linguistic context of GROYNE.

Groynes are extremely cost-effective coastal defense measures, requiring little maintenance, and are one of the most common coastal defense structures. However, groynes are increasingly viewed as detrimental to the aesthetics of the coastline, and face strong opposition in many coastal communities.

Three types of concordances are included in each entry of EcoLexicon: conceptual, phraseological and verbal. These concordances allow the users to widen their knowledge from different perspectives. Conceptual concordances show the activation of conceptual relations in the real use of terms. Phraseological concordances help the user in acquiring specialized discourse. Thirdly, verbal concordances highlight the most frequent verbal collocations, which offer, again, both linguistic and conceptual information.

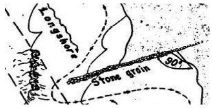
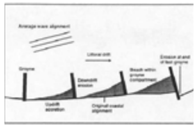
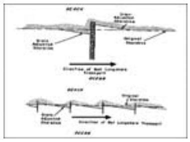
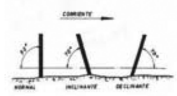
Figure 8 shows the conceptual concordances in the entry of GROYNE. Linguistic markers such as *designed to* and *provide* explicitly relate the concept to its function, *shore protection* and *trap and retain sand*.

```
TYPE

od only qualitatively.  5-3. Groins.  a. General.  (1) Groynes are barrier-type structures that extend from the
ks at the  coast of this area existed of hard measures (groynes and seawall). The groins could not be  kept-up
 of inlet stabilization works (e.g., jetties, terminal groynes, offshore breakwaters)\par  on the shorelines ad
wearing away of land by the action of natural forces.  Groyne -- a shore protection structure, usually built pe
 beach nourishment. Some coastal structures, including groynes, breakwaters and sills, can have a positive effe
ocial implications  Coastal defence structures such as groynes and detached breakwaters generally increase the

MATERIAL

vant to most rubble mound structures such as seawalls, groynes, and breakwaters. It should be noted that in thi
sand, nourishing the beach  compartments between them. Groynes may be made of wooden or rocky materials. They
ct will also include adding notching to existing stone groynes and extending outfall pipes.  Unger said a sche

LOCATION

olution includes the following  three elements: a rock groyne extending seaward  from the existing shoreline, a
are intended consequences for people and wildlife.     Groynes are structures built out from the shoreline, typ
line from receding. Along  almost the entire coastline groynes are present, only at Warnemunde mitte there are

FUNCTION

the alignment of the updrift shoreline shifts as well. Groyne fields are designed to trap and retain sand, nour
ents, and thus a greater erosion rate. Breakwaters and groynes are effective in retaining sand and reducing ero
 nearer the\par  barrier may also be considered, using groynes to impede updrift movement of material at the\pa
ions to anchor the fill material. In either instance,  groynes provide shore protection by modifying longshore
acceptable erosion of the downdrift shore. At first, a groyne field interrupts the longshore movement of sand i
ngth of beach to be protected. The basic purposes of a groyne are to modify the longshore movement of sand and
```

Figure 8. Conceptual concordances in the entry of GROYNE.

Finally, the third type of contextual information added to the entry are images. These images are selected according to their most salient functions (Anglin et al., 2004; Faber et al., 2007) or in terms of their relationship with the real-world entity that they represent to illustrate the relations a concept can express. Table 3 shows an example of how several images are explicitly related to the conceptual relations expressed in the definition of GROYNE.

Table 3. The convergence of linguistic and graphic descriptions of GROYNE.

GROYNE		
Formal role	• hard coastal defence structure **[is_a]**	
Constitutive role	• *default value* (concrete, wood, steel, and/or rock) **[made_of]**	
Formal role	• perpendicular to shoreline **[has_location]**	 
Telic role	• protect a shore area, retard littoral drift, reduce longshore transport and prevent beach erosion **[has_function]**	 

OVERINFORMATION

In knowledge representation, concepts are very often classified according to different facets or dimensions. This phenomenon is widely known as *multidimensionality* (Kageura, 1997). The representation of multidimensionality enhances knowledge acquisition providing different points of view in the same conceptual system. However, not all dimensions can always be represented at the same time, since their activation is context-dependent. This is the case of certain versatile concepts involved in a myriad of events, such as WATER. In EcoLexicon this has led to a great deal of information overload (see Figure 9), which jeopardizes knowledge acquisition.

Yeh & Barsalou (2006) state that when situations are not ignored, but incorporated into a cognitive task, processing becomes more tractable. In the same way, any specialized domain reflects different situations in which certain conceptual dimensions become more or less salient. As a result, a more believable representational system should account for reconceptualization according to the situated nature of concepts. Rather than being decontextualized and stable, conceptual representations should be dynamically contextualized to support diverse courses of goal pursuit (Barsalou, 2005: 628). In EcoLexicon, overloaded concepts are reconceptualised according to two contextual factors: domain membership and semantic role.

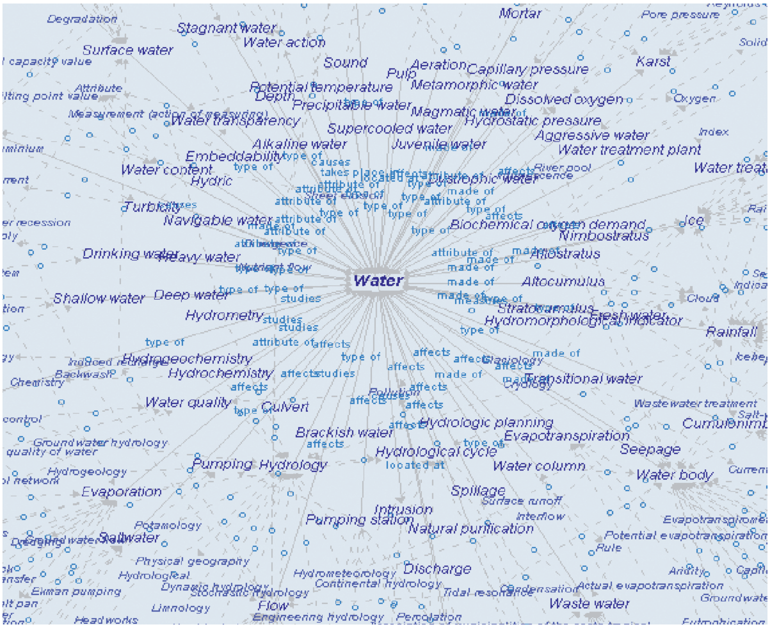


Figure 9. Information overload in the network of WATER.

Role-based Reconceptualization

Role-based relational constraints are applied to individual concepts according to their own perspective in a given proposition. For example, in WATER CYCLE *affects* WATER, WATER is a patient. However, if a role-based domain was to be associated with WATER CYCLE, this would require the application of agent-based constraints. Role-based constraints apply for non-hierarchical relations since hierarchical ones are always activated, whether concepts are agents or patients (León Araúz and Faber, 2010). Moreover, this kind of constraints can only be applied to the first hierarchical level, since they are focused on a particular concept and not its whole conceptual proposition. In the next figures, the overloaded network of WATER (Figure 10) is restricted according to the agent role (Figure 11).

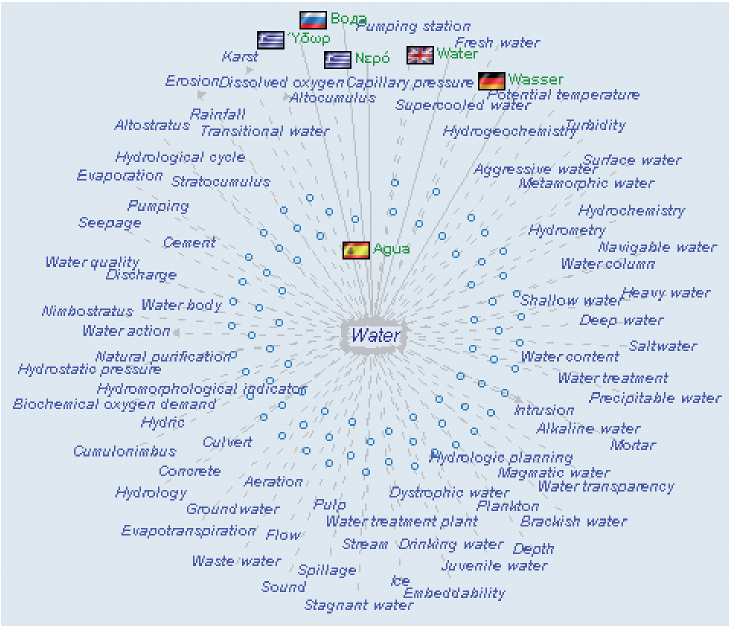


Figure 10. Role-free network of WATER.

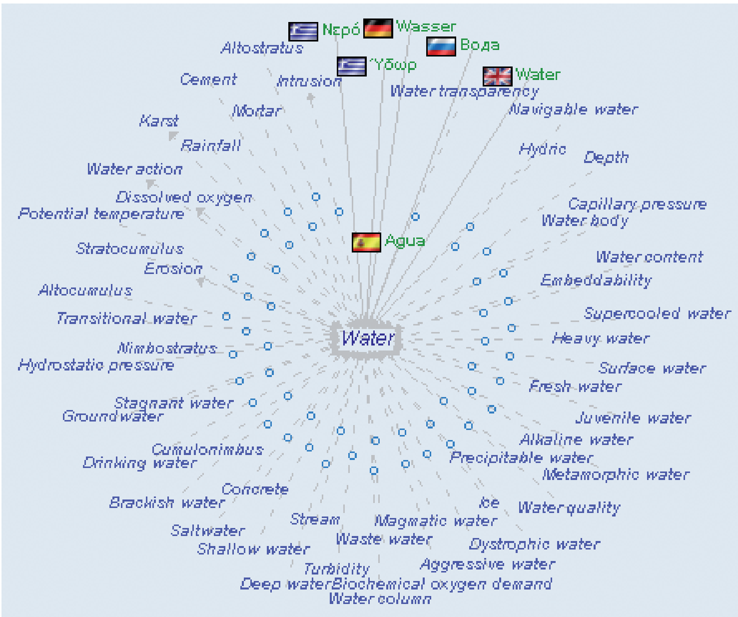


Figure 11. Agent-based network of WATER.

Actually, role-based domains by themselves are not sufficient to reconceptualize knowledge in a meaningful way. In the role-free network, WATER appears linked to 72 concepts, whereas in the role-based one, WATER is related to 50. Despite the difference, the concept still appears overloaded, especially once the second hierarchical level is displayed. However, contextual domains, although usually dominated by one role, restrict relational power of versatile concepts in a more quantitative way.

Domain-based Reconceptualization

We have divided the environmental field in different contextual domains according to corpus information and expert collaboration: HYDROLOGY, GEOLOGY, METEOROLOGY, BIOLOGY, CHEMISTRY, CONSTRUCTION/ENGINEERING, WATER TREATMENT/SUPPLY, COASTAL PROCESSES and NAVIGATION.

Our contextual domains have been allocated similarly to the European General Multilingual Environmental Thesaurus, whose structure is based on themes and descriptors, reflecting a systematic, category or discipline-oriented perspective (GEMET, 2004). They provide the clues to simplify the background situations in which concepts can occur in reality.

Domain membership restricts concepts' relational behaviour according to how their referents interact in the real world. Contextual constraints are neither applied to individual concepts nor to individual relations, since one concept can be activated in different contexts or use the same relations but with different values. Constraints are instead applied to conceptual propositions (León Araúz et al., 2009). For instance, CONCRETE is linked to WATER through a *part_of* relation. Nevertheless, this proposition is irrelevant if users only want to know how WATER naturally interacts with the landscape or how it is purified of contaminants. Consequently, the proposition WATER *part_of* CONCRETE only appears if users select the CONSTRUCTION/ENGINEERING context.

As a result, when constraints are applied, WATER only shows relevant dimensions for each contextual domain. In Figure 12 WATER is just linked to propositions belonging to the context of ENGINEERING/CONSTRUCTION:

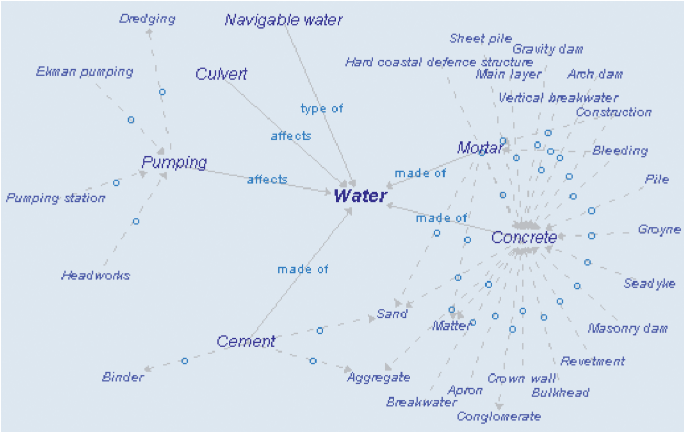


Figure 12. WATER in the ENGINEERING/CONSTRUCTION contextual domain.

However, in Figure 13 the GEOLOGY context shows WATER in a new structure with other concepts and relations:

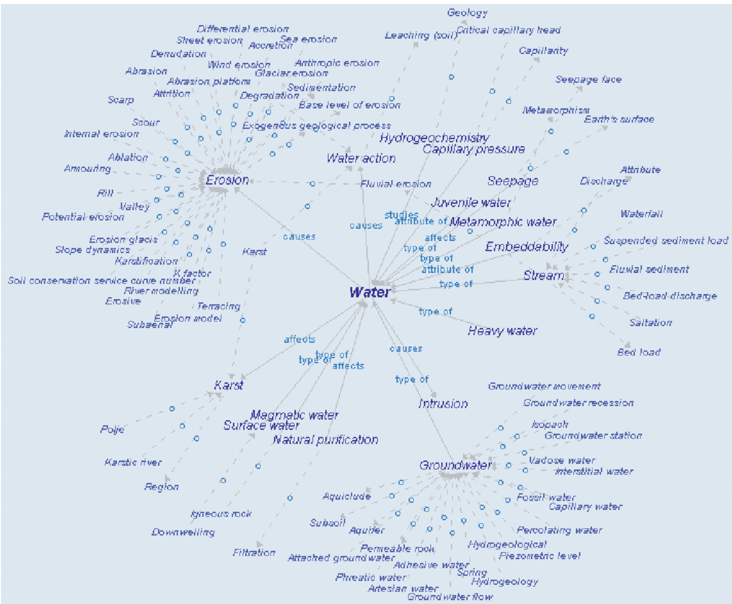


Figure 13. WATER in the GEOLOGY contextual domain.

The number of conceptual relations changes from one network to another, as WATER is not equally relevant in all contextual domains. Furthermore, relation types also differ, which also highlights the changing nature of WATER'S internal structure in each case. For example, in the ENGINEERING/CONSTRUCTION context domain, most relations are *made_of* and *affects*, whereas in the GEOLOGY domain, *causes* and *type_of* stand out. *Affects* is also shared by the GEOLOGY domain, but the arrow direction shows a different perspective: in geological contexts WATER is a much more active agent than in ENGINEERING/CONSTRUCTION, where the concept is more subject to changes (patient). Finally, WATER is not always related to the same concept types. In ENGINEERING/CONSTRUCTION, WATER is only linked to artificial entities or processes (PUMPING, CONCRETE, CULVERT), while in GEOLOGY it is primarily related to natural ones (EROSION, GROUNDWATER, SEEPAGE).

Intersection of Role- and Domain-Based Reconceptualization

A new reconceptualization can take place with the intersection of role-based constraints and contextual domains. For example, WATER can be framed as an AGENT (Figure 14) or a PATIENT (Figure 15) or even both (Figure 16) within the HYDROLOGY context.

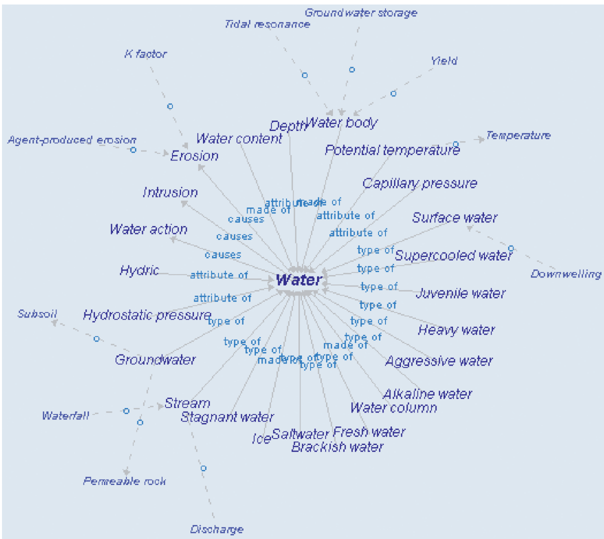


Figure 14. WATER as an AGENT in HYDROLOGY.

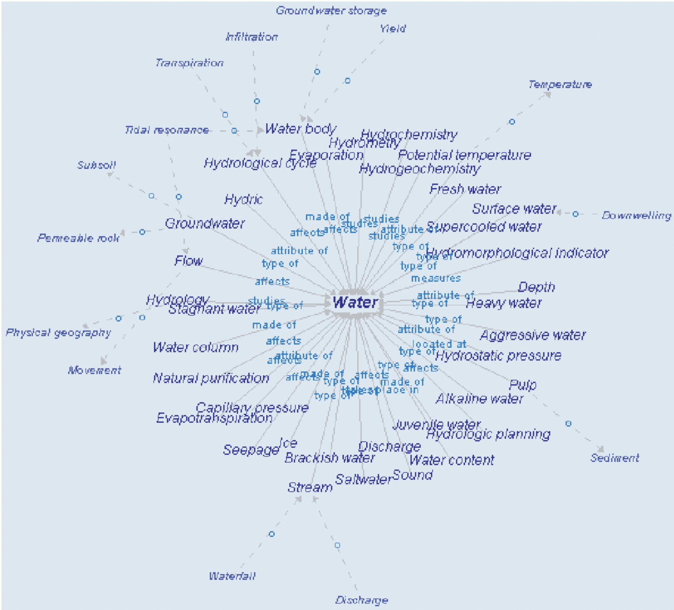


Figure 15. WATER as a PATIENT in HYDROLOGY.

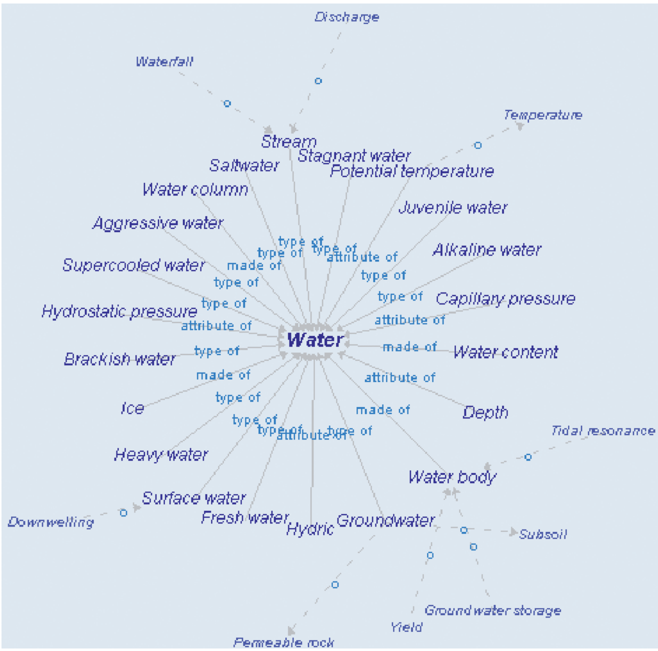


Figure 16. WATER as an AGENT and PATIENT in HYDROLOGY

Now, the first level appears constrained according to different roles in a particular contextual domain, which at the same time applies for the second level. It is worth noting that Figure 16 only shows hierarchical relations (*type_of, attribute_of*, *made_of*), because these are the only ones shared by concepts that can be agents or patients. In Figure 14, however, the representation adds the relation *causes*, typical of agents, and in Figure 15, it adds propositions where WATER is *affected_by*, *measured*, *studied* or *located_at*.

CONCLUSIONS

In this paper we have presented EcoLexicon from several points of view. We have briefly explained the methodology we apply for knowledge representation, and we have shown how all this information is presented to the end user. The internal coherence at all levels of a dynamic knowledge representation shows that even complex domains can be represented in a user-friendly way. EcoLexicon combines the advantages of a relational database, allowing for a quick deployment and feeding of the platform, and an ontology, enhancing user queries. Reconceptualization provides a way of representing the dynamic and multidimensional nature of concepts and terms. It offers a qualitative criterion for the representation of specialized concepts in line with the workings of the human conceptual system. Moreover, it is a quantitative solution to the problem of information overload, as it significantly reduces irrelevant context-free information.

REFERENCES

Anglin, G.; Vaez, H. & Cunningham, K. L. (2004). Visual representations and learning: the role of static and animated graphic. *Visualization and Learning, 33*, 865-917.

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

Faber, P.; Montero Martínez, S.; Castro Prieto, M.C.; Senso Ruiz, J.; Prieto Velasco, J.A.; León Araúz, P.; Márquez Linares, C.F. & Vega Expósito, M. (2006). Process-oriented terminology management in the domain of Coastal Engineering. *Terminology, 12(*2), 189-213.

Faber, P.; Márquez Linares, C. & Vega Expósito, M. (2005). Framing Terminology: A Process-Oriented Approach. *META 50*(4): CD-ROM.

GEMET. (2004). About GEMET. General Multilingual Environmental Thesaurus. http://www.eionet.europa.eu/gemet/about

Kageura, K. (1997). Multifaceted/Multidimensional concept systems. In S.E. Wright & G. Budin. *Handbook of Terminology Management*: *Basic Aspects of Terminology Management* (pp. 119-32). Ámsterdam/Philadelphia: John Benjamins.

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

León Araúz, P. (2009). *Representación multidimensional del conocimiento especializado: el uso de marcos desde la macroestructura hasta la microestructura*. PhD Thesis. University of Granada, Spain.

León Araúz, P.; Magaña Redondo, P. & Faber, P. (2009). Managing inner and outer overinformation in Ecolexicon: an environmental ontology. In *Proceedings of the 8th International Conference on Terminology and Artificial Intelligence*. Toulouse, France.

Yeh, W. & Barsalou, L. W. (2006). The situated nature of concepts. *American Journal of Psychology, 119*, 349-384.