A full knowledge cycle for semantic interoperability

Piek Vossen
Wauter Bosma
Faculty of Arts, VU University Amsterdam
De Boelelaan 1101, 1081HV, Amsterdam
p.vossen,w.bosma@let.vu.nl

Eneko Agirre
German Rigau
Aitor Soroa
Ixa Research Group, Universidad Del Pais Vasco
649 Posta kutza, 20080 Donostia, Basque Country
e.agirre,g.rigau,a.soroa@ehu.es

Abstract

In this paper, we describe the KYOTO project, which is a platform for establishing semantic interoperability across languages and cultures. Semantic interoperability is achieved in three ways: through a shared annotation format for representing text, through an interlinked repository of lexical resources and ontologies, and through a uniform system for mining text that enriches lexical resources and extracts new relations based on the ontology. Whereas most efforts focus on separate aspects of semantic interoperability, KYOTO implements a full knowledge cycle for sharing and exchanging knowledge by integrating these different operability layers. Sharing of knowledge as expressed in natural language is the genuine test for semantic interoperability.

1 Introduction

Standardization is essential for interchangeability of data and tools. Once a data format is accepted as a standard, tools can be developed and shared without much data conversion effort. A long-term goal of standardization is to achieve semantic interoperability of content and knowledge. For years, the Semantic Web Community has been working on the standardization of the representation of data (RDF), knowledge (OWL) and services[1] to achieve this. Less progress has been made however with semantic interoperability of natural language expressions, although this is essential for systems to interact with people that use natural language as their most intuitive interface for communication. Hardly any effort has been done to integrate data format standards, with knowledge representation standards, with conceptual standards. Nevertheless, it is this type of interoperability that is the ultimate goal for a uniform representation and interpretation of natural language text.

The European/Asian KYOTO project[2] aims at establishing semantic interoperability of both knowledge and language to express this knowledge. To achieve this, we anchor words and expressions in language to formal definitions of meaning and use this information to detect knowledge and facts in text. Semantic interoperability is achieved by mapping wordnets in each of these languages to a shared ontology, as proposed in the Global Wordnet Grid (Fellbaum and Vossen 2008), and by means of a common architecture for processing text. The former is explained in detail in Vossen and Rigau (2010fc) and the latter in Bosma et al (2009). In this paper, we focus on the complete knowledge cycle in KYOTO, which combines the two.

True semantic interoperability can be achieved only when linguistic and formal semantic layers are integrated with text processing and knowledge acquisition. KYOTO exemplifies that by separating the linguistic layers from the formal ontological layers and by separating the generic conceptual layers (represented in the vocabularies and ontologies) from the instantiation of these layers as found in text.

We will discuss the combination of these different aspects of semantic interoperability in an

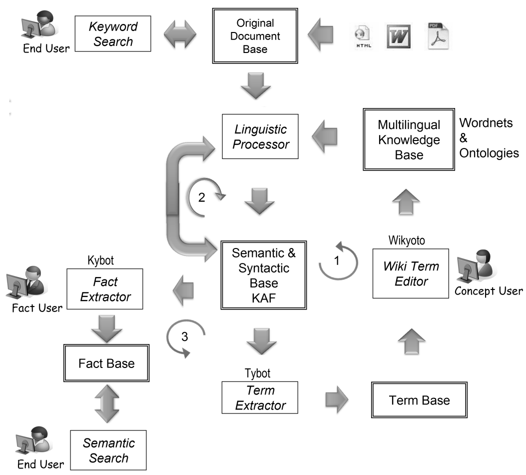
[1] www.w3.org/2002/ws/swsig

[2] www.kyoto-project.eu

overall framework. The paper is organized as follows. In the next section, we give an overview of the overall architecture of KYOTO and the knowledge sharing cycle through semantic interoperability. In the following sections, we work out a specific example illustrating the different interoperability issues.

2 The knowledge cycle in KYOTO

The KYOTO project allows users to model terms and concepts in their domain and to use this knowledge to apply text mining on documents. The knowledge cycle in the KYOTO system is outlined in Figure-1 (figure-1 is missing). It starts with a set of source documents produced by the community, such as PDFs and websites. Linguistic processors apply tokenization, segmentation, morpho-syntactic analysis and some semantic processing to the text in different languages. The semantic processing involves detection of named-entities (persons, organizations, places, time-expressions) and determining the meaning of words in the text using a given wordnet in a language. In the current system, there are processors for English, Dutch, Italian, Spanish, Basque, Chinese and Japanese. The process of word-sense-disambiguation is the same for all the languages (Agirre and Soroa, 2009).



The output of this linguistic analysis is stored in an XML annotation format that is the same for all the languages, called the KYOTO Annotation Format (KAF, Bosma et al 2009). This format incorporates standardized proposals for the linguistic annotation of text but represents them in an easy to use layered structure. In this structure, words, terms, constituents and syntactic dependencies are stored as separate layers with references across the structures. All other modules in KYOTO draw their input from these structures.

The knowledge process proceeds in 2 cycles:

1. The acquisition, accumulation and integration of vocabularies and ontologies for a domain.
2. The automatic extraction of facts from text, based on the integrated knowledge.

In the 1st cycle, the Tybot (Term Yielding Robot) will extract the most relevant terms from the document collection. The Tybot is a generic program that can do this for all the different languages in much the same way due to the standardized format of KAF. The resulting terms are organized as a hierarchy with semantic relations and, wherever possible, related to generic semantic databases, i.e. wordnets for each language. The result is a domain wordnet in a specific language. Each new term can be seen as a possible proposal to also extend the ontology. Through the ontology, the domain experts can establish the similarities and differences across the languages and hence cultures. These users are called the concept users, since they are involved with the modeling of terms and concepts in their domain.

The second cycle of the system involves the actual extraction of data and factual knowledge from the annotated documents by the Kybots: Knowledge Yielding Robots. Kybots use a collection of profiles that represent the type of information of interest. In the profile, conceptual relations are expressed and their realization in a language is achieved through the domain wordnets and so-called expression rules. These patterns match different layers in the KAF annotation of text. Since the semantics are defined through the ontology, it is possible to detect similar data across documents in different languages, even if expressed differently.

3 The KYOTO global wordnet grid

The multilingual knowledge base plays an important role in the KYOTO project. It is designed as an implementation of the Global Wordnet Grid. The wordnets for 7 languages have been represented in the Wordnet-LMF format (Soria et al 2009) and stored in a DebVisDic server (Horak 2005). The DebVisDic server also contains the SUMO ontology and a first version of the KYOTO ontology in OWL-DL. The SUMO ontology is fully mapped to WordNet 3.0. The

KYOTO ontology currently consists of 786 classes divided over three layers. The basic layer is based on DOLCE (DOLCE-Lite-Plus version 3.9.7, Masolo et al 2003) and OntoWordNet. This layer of the ontology has been modified for our purposes (Herold and Hicks 2009).The second layer consists of concepts coming from the so-called Base Concepts in various wordnets (Vossen 2008; Izquierdo et al 2007). Examples of base concepts are: building, vehicle, animal, plant, change, move, size, weight. The Base Concepts are those synsets in WordNet3.0 that have the most relations with other synsets in the wordnet hierarchies and are selected in a way that ensures that each of the more specific concepts is connected to one of the Base Concepts as specific (sub-)hyponyms. This has been completed for the nouns (about 500 synsets) and is currently being carried out on verbs and adjectives in WordNet 3.0. Through the Base Concepts, we will ensure that any synset in the wordnets is mapped to some concept in the ontology either directly or indirectly.

The most specific layer of the ontology contains concepts representing species and regions relevant to the KYOTO domain. These concepts were provided by domain experts, and in certain cases, concepts have been added to link the domain specific terms to the ontology. This foundational ontology provides the basic building block for the domain experts to add their knowledge. The wordnets and the ontology play an important role for mining facts from text. They form the basis for the conceptual patterns of the Kybots. For resolving the constraints in these patterns, the Kybots need to apply some kind of reasoning over the available knowledge.

During the project, new terms and concepts will be added to the knowledge repository. Partly, these terms and concepts are learned from the domain corpus and partly they will be derived from existing background knowledge basis. Combining these resources and defining the semantics of the mappings across these resources presents a major knowledge integration task. For the domain of the environment, we specifically used the Species-2000 database: http://www.sp2000.org/. It contains 2.1 millions species structured according to a biological taxonomy: *Kingdom*, *Class*, *Order*, *Family*, *Genus*, *Species*. Each concept has at least a Latin name and often many alternative labels in different languages. An example of a Latin hierarchy is the following: Animalia -> Chordata -> Amphibia -> Anura -> Leptodactylidae -> Eleutherodactylus -> Eleutherodactylus augusti. To exploit the data, we converted it to SKOS format and published it in Virtuoso. The taxonomic relations are converted to skos:broader relations. Furthermore, we expanded the language labels by querying the DBPedia database for the Latin names. This increased the number of labels from 157,124 to 1,817,778. The concepts in the SKOS database were then aligned with WordNet3.0, using the hierarchical structure.

Likewise it is possible to match terms in the text either with terms in the term database or with terms in the Species2000 SKOS database. If a term is directly matched with WordNet3.0, we can access the WordNet3.0 hierarchy to obtain the relevant ontological label. If a term is not in WordNet3.0, we traverse the hierarchy in the term database or in SKOS up to the first parent concept that matches WordNet3.0.

In Vossen and Rigau (2010) and Rigau et al (2010) (missing references), we describe how these resources are aligned to wordnets in each KYOTO language. The result is a large wordnet grid repository, in which millions of concepts are directly or indirectly linked to wordnets in each language and through the wordnet to the central ontology (being DOLCE based or SUMO based).

4 Semantic interoperability illustrated

To illustrate the above model, we will explain an example taked from a report on the Humber estuary in mid-east England:

> "The Humber Estuary Low Tide Count Programme 2003-2004, was published in March 2005 by English Nature as "Research Report No 656 - Humber Estuary Low Tide Count Programme 2003-2004" (....) Notable trends include the recent recovery of the pinkfooted goose, avocet and black-tailed godwit populations. Shelduck and ringed plover are the most widely distributed birds across the estuary, while teal and wigeon are concentrated on the upper estuary, and golden plover use the middle/outer estuary. Birds like lapwing and golden plover use the estuarine habitat as a safe roost." *(EBB & FLOW, The newsletter of the Humber management scheme, No 5, Summer 2006, page 4)*

The example is typical for the type of text that we find in the environment domain. Most reports summarize trends or have a summarizing character. The information is very condensed.

4.1 KYOTO annotation format

The linguistic processing generates a shallow KAF structure consisting of the word tokens, the terms and at least the NPs to which the terms belong. Below is an example of the term structure in KAF which is enriched with WordNet3.0 synset identifiers by the automatic WSD system:

```
<term tid="t673" lemma="pinkfooted" pos="G" type="open">
<span><target id="w789"/></span>
</term>
<term tid="t674" lemma="goose" pos="N" type="open">
<span><target id="w790"/></span>
<externalReferences>
<externalRef resource="wn30g" reference="eng-30-01855672-n"
confidence="0.38"/>
<externalRef resource="wn30g" reference="eng-30-10157744-n"
confidence="0.31"/>
<externalRef resource="wn30g" reference="eng-30-07646821-n"
confidence="0.30"/>
</externalReferences>
</term>
<term tid="t675" lemma="avocet" pos="N" type="open">
<span><target id="w792"/></span>
<externalReferences/>
<externalRef resource="wn30g" reference="eng-30-02036711-n"
confidence="1"/>
</externalReferences>
</term>
<term tid="t676" lemma="and" pos="O" type="open">
<span><target id="w793"/></span>
</term>
<term tid="t677" lemma="black-tailed" pos="G" type="open">
<span><target id="w794"/></span>
</term>
<term tid="t678" lemma="godwit" pos="N" type="open">
<span><target id="w795"/></span>
<externalReferences>
<externalRef resource="wn30g" reference="eng-30-02034129-n"
confidence="1"/>
</externalReferences>
</term>
<term tid="t679" lemma="population" pos="N" type="open">
<span><target id="w796"/></span>
<externalReferences>
<externalRef resource="wn30g" reference="eng-30-06026276-n"
confidence="0.21"/>
<externalRef resource="wn30g" reference="eng-30-08178741-n"
confidence="0.21"/>
<externalRef resource="wn30g" reference="eng-30-01257969-n"
confidence="0.19"/>
<externalRef resource="wn30g" reference="eng-30-08179879-n"
confidence="0.19"/><externalRef resource="wn30g"
reference="eng-30-13779804-n" confidence="0.19"/>
</externalReferences>
</term>
```

You can see that the terms *pinkfooted goose* and *black-tailed godwit* are not recognized as a single term. Their headwords *goose* and *godwit* are represented separately and are matched with WordNet3.0 synsets, just as is the case for *avocet*. In the case of *goose* and also *population*, multiple synsets are ranked with different probabilities. The words *godwit* and *avocet* only have a single sense.

4.2 Term extraction and concept alignment

The tybot extracts the terms from this structure, where multiword terms and compounds are grouped below the head of the term and single word terms and heads represent the tops of the term hierarchy. Some of these terms have WordNet3.0 synsets and some do not. In so far terms are matched with Wordnet3.0, which follows from the WSD output in KAF, the term hierarchy is extended with the hypernyms from WordNet3.0:

anseriform bird:1:eng-30-01845477-n
...goose:1:eng-30-01855672-n
......*pinkfooted goose*
...duck:1:eng-30-01846331-n
......*wigeon*:1:eng-30-01848648-n
......*teal*:1:eng-30-01848123-n
…...sheldrake:2: eng-30-01849466-n
.........*shelduck*:1:eng-30-01849676-n
shorebird:1:eng-30-02022684-n
...plover:1:eng-30-02023341-n
......*avocet*:1:eng-30-02036711-n
......*golden plover*:eng-30-02024479-n
......*ringed plover*
......*lapwing*:1:eng-30-02024763-n
...godwit:eng-30-02034129-n
......*black-tailed godwit*

The extracted terms are in italics. If they match with WordNet3.0 synsets, the synset identifier is added. The underscored terms do not occur in the text as terms but are WordNet3.0 synsets that can be derived from the terms that do occur. We can see here that WordNet3.0 provides a nice grouping of terms that are not morphologically related, such as *wigeon*, *teal* and *shelduck*. The co-occurrence of bird concepts in the text fragment has a strong disambiguating effect. Similarly, we have shown that disambiguating related words from the domain yields better WSD performance (Agirre et al. 2009).

We can further exploit the WordNet3.0 hierarchy which relates these concepts to the following chain of hypernym relations:

```
animal:1 (Base Concept: Animal in the ontology)
  chordate:1
    vertebrate:1
      bird:1
        aquatic bird:1
          water fowl:1
                anseriform bird:1
          wading bird:1
                shorebird:1
```

The synset for animal:1 is a base concept that is related to the ontology concept Animal. We can thus further enrich the term occurrences in KAF with ontology labels:

```
<term tid="t674" lemma="goose" pos="N" type="open">
<span><target id="w790"/></span>
<externalReferences>
```

```
<externalRef resource="k-ont-v2" reference="Animal"
confidence="0.38"/>
</externalReferences>
</term>
```

Likewise, we build a rich semantic representation of the text that is anchored to a common ontology, shared across different languages. This means that a text from a similar domain in a different language can be matched with this text on the basis of the same ontological labels. The next example is taken from a Dutch document on a similar estuary, the Westerschelde:

“Op de schorren in het oostelijke deel van de Westerschelde overwinteren veel eenden en ganzen, w.o. 2/3 van de Europese populatie Grauwe ganzen. De schorren vervullen ook een functie als hoogwatervluchtplaats voor steltlopers en eenden. Tenslotte zijn de schorren van belang als broedgebied voor soorten als Visdief en Tureluur. Ook vervullen schorren een belangrijke rol in de koolstof- en stikstofcyclus (vastlegging, filterwerking).”

English translation:
“On the salt marshes in the Eastern part of the Westerschelde, many ducks and geese overwinter, among which 2/3 of the European population of greylag geese. The salt mashes fullfil a purpose as high-tide shelter for stiltwalkers and ducks. Finally, the salt marshes are an important breeding area for species such as common stern and redshank The salt marshes also play an important role in the carbon and nitrogen cycle (absorption and filtering).” *(Waardering voor de Westerschelde, Rijksinstituut voor Kust en Zee/RIKZ, 2002, p.25)*

For this text, we will extract a similar term database from the KAF using the same Tybot program:

watervogel:1:d_n-15025 (water bird)
...*steltloper*:1/d_n-43044 (stiltwalker)
…...*tureluur*:1/n_n-501559 (redshank)
...zwemvogel:1:d_n_37777 (swimming bird)
…...*eend*:1 d_n-33460 (duck)
…...gans (goose)
…......*grauwe gans* (greylag goose)
visdief (common tern)

Although the coverage for the Dutch wordnet is not as high as the coverage for English, we can still see a similar conceptual structure that is derived, which is relevant to water birds.

In addition to these terms, we also detect named entities in the text. This is a separate process, where terms are matched against GeoNames and disambiguated using density of location and country names in the text.

```
<location lid="l14">
  <span><target id="t1873"/></span>
<externalReferences>
<externalRef resource="GeoNames"
        reference="2635503"/>
</externalReferences>
<geoInfo>
<place name="River Trent" countryCode="GB" countryName="United Kingdom" latitude="53.7" longitude="-0.7"
fname="stream" timezone="Europe/London"/>
</geoInfo>
</location>
```

This example shows the named entity River Trent. It is represented as a separate layer in KAF and points back to(multiple) term occurrences in the text. The entity includes identifiers and properties from the GeoNames database.
In addition to the terms referring to species, other less strict concepts are mentioned in both text fragments:

habitat:1:eng-30-08580583-n
    estuarine habitat
shelter:2:
    roost:1:eng-30-04107984-n
        safe roost
body of water:1
    estuary:1:eng-30-09274500-n
        middle/outer estuary
        upper estuary
gebied:1/d_n-22485 (area)
    schor:1:d_n-12265 (salt marshes)

These mainly refer to location types that are ultimately mapped to WordNet3.0 as well. The remainder of the text contains processes and properties that need to be matched to the above entities. This is done through the text mining module that we will discuss in the next section.

4.3 Relation detection

The Kybot module uses so-called profiles to match conceptual schemes with textual patterns. Kybot profiles consist of three different components: Expression Rules, Semantic Conditions and the Output Template. Once the Kybot profile has been checked and compiled, the resulting Kybot can be applied to the analysed text (KAF file). Thus, for each analyzed sentence a Kybot is applied using the following rule:

IF (Expression-Rules match AND Semantic-Conditions hold)
THEN generate the Output-Template

For the above text fragments, two different strategies can be applied. In the case of the English text, we see that rather general and abstract constructions are used:

recent recovery
use as a safe roost

concentrated
widely distributed

These words indicate *occurrence* of birds in locations and habitats and the *recovery* of the population as a whole. For the occurrence relation a very generic profile can be used that states that any concept that is a subject or object of a verb and that is followed by a location expression, implies that the concept is located in the location:

e1(subj)+V+e2(location)
=>
(located, e1, e2)

This pattern will over-generate many relations because the location relation is not directly expressed but the precision of the implication can still be correct.

In the case of the concept of *population*, the matching is a bit more complex. First, it is matched with the ontology as a biological group consisting of species that live in a habitat. Another profile will look for possible species in the surrounding text that may be a member. Within the same sentence, the candidates are *pinkfooted goose*, *avocet* and *black-tailed godwit*, resulting in the relations:

(member, pinkfooted goose, population)
(member, avocet, population)
(member, black-tailed godwit, population)

The concept of *recovery* is a subjective indicator that is ignored in the analysis here. Still, a profile could be built that assigns a quality property to the concept population.

The Dutch example is richer. It refers to the following processes and properties:

overwinteren (overwinter)
hoogwatervluchtplaats (high tide shelter)
broedgebied (breeding area)
vastlegging (absorption)
filterwerking (filtering)

Obviously, we can make a profile for each specific verb, but again, through wordnet these verbs can be mapped to ontological concepts as well. It then depends on the processes and roles that are defined in the ontology. The current ontology is extended with such processes for the environment domain. For example, the process of breeding is defined in terms of the possible result and the involved agents (BreederRole restricted to Organisms). A specific terms such as *hoogwatervluchtplaats* (high tide shelter) is related to the process of shelter through a ShelterLocationRole. Profiles that include these processes can match surrounding concepts to the respective roles. It means that process verbs or terms that imply a process, can be saturated with arguments that fit the associated role pattern. Applying this to the Dutch text will generate relations as the following:

(playRole, salt marsh, ShelterRole)
(playRole, stiltwalker, ShelteredRole)
(playRole, duck, ShelteredRole)

(playRole, salt marsh, BreedingLocationRole)
(playRole, common stern, BreederRole)
(playRole, redshank, BreederRole)

(playRole, salt marsh, OverwinterLocationRole)
(playRole, duck, OverwintererRole)
(playRole, goose, OverwintererRole)

The more semantics we add to the KAF representation, by adding synset identifiers and/or ontology labels to the terms, the more general and language neutral can the profiles can be. Specific morpho-syntactic patterns may add precision, but many patterns may be needed to maximize the recall. Drop of precision depends on the ambiguity of role fillers in the context.

4.4 Fact extraction

After the relations are extracted, we can combine the pieces of information into facts. The important difference here is that facts are independent from the textual representation and order of expressions. To begin with, each fact should be embedded in time and place. This knowledge is provided by the named entities for time and location that are added separately to the KAF representation. We have seen above that these entities refer back to the terms in the KAF. The same holds for any relation that is extracted above. Aggregating facts then involves grouping all relations within a time frame and a regional boundary and deciding on what relations belong to the same process or property. In the case of the English document, the direct context only mentions the Humber Estuary and the dates 2003, 2004 and 2005. In the case of the Dutch text, the location is Westerschelde and there is no mention of a date in the direct context. If information is missing in the direct context, the system will fall back on more global contexts (preceding or following pages or the complete document).

Obviously, users of the database of facts can decide on the relevant time frame and the region of interest, which results in further lumping or splitting facts.

Facts are then represented in a separate KAF layer, just as the named entities, independently of the textual occurrences and order but with references to where the elements of the facts are mentioned. For the English text, we can extract many partial facts for the different birds like the following, where we combine the knowledge on the location and time with the basic implication expressed by the relation. Below is an example of such a fact that can be extracted from the English text:

```
<fact fid="f1">
<!-- recover →
<process eid="e1"/>
<!--pinkfooted goose population -->
<arg tid="t674" role="experiencer"/>
<! estuary -->
<arg tid="t567" role="location"/>
<!-- 2003, 2004 and 2005 -->
<timex3 texid="timex3"/>
<!-- Humber Estuary -->
<locationx4 texid="lid711"/>
</fact>
```

This fact is amalgamated from the recovery clause and other text fragments that contained expressions of dates and locations. Similarly, we can derive from the disjunction of the populations separate facts for *avocet population* and *black-tailed godwit populatio*, which represent a case of splitting of an expression into multiple facts. In that case, we assign a different event identifier for each fact that points to the same recovery expression in the text:

```
<fact fid="f2">
<!-- recover →
<process eid="e2"/>
<!--avocet population -->
<arg tid="t671" role="experiencer"/>
<! estuary -->
<arg tid="t567" role="location"/>
<!-- 2003, 2004 and 2005 -->
<timex3 texid="timex3"/>
<!-- Humber Estuary -->
<locationx4 texid="lid711"/>
</fact>
```

The textual representation is thus a condensed representation of different facts and processes. As a general rule, we group all facts that denote processes or qualities of the same type together, in so far as they involve the same region and time period. The same fact can thus come from different documents and sources and from different languages. Obviously, also the opposite occurs in language. Through ellipses or coordination multiple facts are packed in a single expression.

The basic structure for fact representation is taken from the SemAf proposal (reference). However, whereas SemAf represents events at the sentence level, our fact representation completely abstracts from the textual level of expressing semantic knowledge. The fact representation can easily be converted to RDF and can be labelled with words from any language or even the ontology labels itself, regardless of the language from which it is extracted.

The representation of facts is still in development in the project. We will closely collaborate with the ISO SemAf working group to further refine the specification.

5 Conclusions

Semantic interoperability is a core issue in the design of an integrated semantic software environment, although current approaches typically regard isolated aspects such as annotation of dependency relations or annotation of events. In KYOTO, we take advantage of this insight by building a system around interoperability between languages and cultures at different levels of operation. Central to this approach is KAF, a multi-layer annotation format which allows cross-references between annotation layers. Ultimately, knowledge is anchored to natural language expressions.

The KYOTO project is still in progress. Empirical validation of the system and our ideas is on its way. First versions of the various modules have been completed and are being tested on the available document sets for the environment domain. Demos are available on the project website (http://www.kyoto-project.eu/) and integrated and validated results will be made available there soon.

6 Acknowledgements

The KYOTO project is co-funded by EU - FP7 ICT Work Programme 2007 under Challenge 4 - Digital libraries and Content, Objective ICT-2007.4.2 (ICT-2007.4.4): Intelligent Content and Semantics (challenge 4.2). The Asian partners from Tapei and Kyoto are funded from national funds.

Reference

Agirre, E., & Soroa, A. (2009) Personalizing PageRank for Word Sense Disambiguation. Proceedings of the 12th conference of the European chapter of the Association for Computational Linguistics (EACL-2009). Athens, Greece. http://ixa.si.ehu.es/ukb

Agirre, E., Lopez de Lacalle, O., & Soroa, A. (2009) Knowledge-based WSD and specific domains: performing over supervised WSD. Proceedings of IJCAI. Pasadena, USA.

Álvez J., Atserias J., Carrera J., Climent S., Laparra E., Oliver A. and Rigau G. (2008) Complete and Consistent Annotation of WordNet using the Top Concept Ontology. Proceedings of the 6th international conference on Language Resources and Evaluation, LREC'08, Marrakesh, Morroco. 2008.

Bosma, W., Vossen, P., Soroa, A. , Rigau, G., Tesconi, M., Marchetti, A., Monachini, M., & Apiprandi, C. (2009) KAF: a generic semantic annotation format. In Proceedings of the GL2009 Workshop on Semantic Annotation, Sept 17-19, 2009, Pisa, Italy.

Fellbaum C., & Vossen, P. (2007) Connecting the Universal to the Specific: Towards the Global Grid,In: Proceedings of The First International Workshop on Intercultural Collaboration (IWIC 2007), Kyoto, Japan, January 25-26, 2007, also in LNCS Vol.4568, Springer-Verlag, 2007.

Fellbaum, C. (Ed.) (1998) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

Gangemi A., Guarino N., Masolo C., Oltramari A., Schneider L. (2002) Sweetening Ontologies with DOLCE. Proceedings of EKAW. 2002

Herold, A., & Hicks, A., (2009). Evaluating Ontologies with Rudify Knowledge. Proceedings of the International Conference on Knowledge Engineering and Ontology Development, Madeira, Portugal, October, 2009.

Herold, A., Hicks, A., Rigau, G., & Laparra, E. (2009) KYOTO Deliverable D6.2: Central Ontology Version – 1 www.kyoto-project.eu.

Horak, Ales - Pala, Karel - Rambousek, Adam - Povolny, Martin. (2005) *DEBVisDic - First Version of New Client-Server Wordnet Browsing and Editing Tool*. In Proceedings of the Third International WordNet Conference - GWC 2006. Brno, Czech Republic: Masaryk University, 2005. pp. 325-328. ISBN 80-210-3915-9.

Izquierdo R., Suárez A. & Rigau G. Exploring the Automatic Selection of Basic Level Concepts. Proceedings of the International Conference on Recent Advances on Natural Language Processing (RANLP'07), Borovetz, Bulgaria. September, 2007.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N. & Oltramari, A. (2003) WonderWeb Deliverable D18: Ontology Library, ISTC-CNR, Trento, Italy.

Miller, G. A. (1995) WordNet: A Lexical Database for English. Communications of the ACM 38:39-4.

Mizoguchi R., Sunagawa E., Kozaki K. & Kitamura Y. (2007 A Model of Roles within an Ontology Development Tool: Hozo. Journal of Applied Ontology, Vol.2, No.2, 159-179.

Niles, I. & Pease, A. (2001) Formal Ontology in Information Systems. Proceedings of the international Conference on Formal Ontology in Information Systems – Vol. 2001 Ogunquit, Maine, USA

Putnam, H., (1975) The Meaning of “Meaning”. In Philosophical papers: Volume 2. Mind, language and reality, Cambridge University Press, 215-271.

Soria C., M. Monachini, P. Vossen: "Wordnet-LMf: fleshing out a standardized format for wordnet interoperability", in: Proceedings of IWIC2009, Stanford, USA, February 20-21, 2009.

Tudorache, T., & Noy, N. (2007) Collaborative Protégé - Stanford Medical Informatics, Stanford, CA 94305, USA - presented at Social and Collaborative Construction of Structured Knowledge Workshop at World Wide Web Conference 2007

Vossen, P. (Ed.) (1998) EuroWordNet: a multilingual database with lexical semantic networks for European Languages. Kluwer, Dordrecht.

Vossen, P. & G. Rigau (2010fc) Division of semantic labor in the Global Wordnet Grid, in: Proceedings of the 5th Global Wordnet Conference, Mumbai