Division of semantic labor in the Global WordNet Grid

Piek Vossen
De Boelelaan 1105,
1081HV Amsterdam
VU University Amsterdam
p.vossen@let.vu.nl

German Rigau
Donostia, Basque Country, Spain
IXA Group,
University of the Basque Country
g.rigau@ehu.es

Abstract

In this paper, we describe an implementation of the Global WordNet Grid in the KYOTO project that distinguishes 3 layers of knowledge: domain vocabularies, wordnets and a central ontology. The layers are distinguished according to the principle of the division of linguistic labor, as defined by Putnam (1975). Such a division is required to handle and structure the large quantities of domain vocabulary and its linguistic diversity. We define the relations between the layers and explain how they can be used for reasoning and inferencing.

1 Introduction

Since the introduction of the English WordNet (Fellbaum 1998), wordnets have been developed in many languages, more or less along the same basic principles. In the EuroWordNet project (Vossen 1998), wordnets in different languages have also been connected to each other, which since then has been followed in many other projects all over the world. The English WordNet has always been the connecting medium as the Inter-Lingual-Index or ILI. Through the years also other semantic frameworks have been used as language neutral representations of meanings that can be shared across wordnets, such as the EuroWordNet top-ontology (Vossen 1998), WordNet domains (Magnini 2002), etc. Most notably is SUMO (Niles and Pease 2002), which was mapped to the English WordNet but also to other wordnets such as Arabic, Chinese, Dutch, Spanish, Catalan or Basque. In most cases, the mapping to SUMO was carried over from English to the other languages, using the ILI.

A large ontology as a language independent representation of meaning holds many promises for future research and usage provided that it is tightly connected to these wordnets. Universalia and idiosyncracies of lexicalizations in language can be expressed in a systematic way, allowing language-independent reasoning over linguistically expressed knowledge. This has led to the idea of the Global Wordnet Grid (GWG), in which all wordnets are anchored to a shared ontology (Fellbaum and Vossen 2007, Pease, Fellbaum, Vossen 2008, Vossen and Fellbaum 2008).

The KYOTO project[1] can be seen as a first attempt to implement the GWG on a practical scale for specific domains. The goal is to develop a knowledge sharing and transition platform that can be used by communities in the world. The KYOTO platform operates as a Wiki for establishing semantic interoperability across languages for a specific domain by creating domain wordnets that get interlinked through a shared ontology. The resulting semantic knowledge base is further used to apply automatic fact mining on document collections. The platform allows for continuous updating and modeling of the vocabulary by the people in the community, while their domain wordnets remain anchored to a generic wordnet. If successful, the GWG can be built by the massive labor force of the Internet community and the results become available to the global community.

When applying the principle of GWG to a specific domain, one is confronted with numerous practical and fundamental problems to handle the domain data. First of all, existing background knowledge should be re-used to build the domain wordnet. Secondly, other new terms are automatically learned from the documents and web sites used in the community. Both background knowledge and domain terminology need to be aligned with existing generic wordnets to make

[1] KYOTO is an Asian-European project funded under project number 211423 in the 7th Frame Work in the area of Digital Libraries: FP7-ICT-2007-1, Objective ICT-2007.4.2: Intelligent Content and Semantics.

the domain wordnet interoperable with general concepts. The third aspect, is that any domain wordnet needs to be mapped to a shared domain ontology, which in itself is anchored to a common top and mid-level ontology.

In previous projects, plugin relations have been proposed to relate domain wordnets to generic wordnets. Similar relations can be defined for background vocabularies and wordnets. In addition, we need to define the semantics for the relations between the synsets and the ontology to separate the language specific properties from the language neutral properties. For instance, basic mapping relations have been defined to map SUMO to WordNet (Niles and Pease 2002, Vossen et al 2008) but none of these proposals provide an explicit semantic model for these relations. In fact, semantic information is duplicated in both wordnets and ontologies and it is not clear what knowledge should be expressed where and how this knowledge can be used.

In addition to the complex relations between the different knowledge repositories, we also have to deal with volume. Our experience is that vocabularies in domains are very large, covering millions of concepts. Representing and maintaining these vocabularies in domain wordnets and in the central ontology raises various problems in terms of maintenance and the kind of reasoning and inferencing that one might want to apply. In the case of the ontology, representing those amounts of concepts and applying reasoning is currently completely unfeasible.

A final issue is that background vocabularies are often maintained outside the wordnet community, without connecting their resources to the wordnet infrastructure.

To handle these practical and fundamental issues in the KYOTO project, we defined a three-layered model of semantic resources that are interconnected. Firstly, the **vocabulary layer** including background vocabulary and mined text terminology. Secondly, the **wordnet layer** integrating generic and domain wordnets. Finally, the **ontology layer** containing generic and domain ontologies. We also provide first definitions for the semantics of the mappings between these layers. Furthermore, we explain how different types of inferencing can be applied to each layer for different practical applications such as fact mining from textual repositories. For modeling the semantics, we use the division of labor principle from Putnam (1975), which we apply to the knowledge bases and computer systems that interact with human knowledge and language.

The structure of this paper is as follows. We first give some background information on the KYOTO project. Next, we describe the problems handling the knowledge resources in the domain of the environment. In section 4, we explain our three-layered model and, in section 5, we explain the different types of relations between the layers. Finally in section 6, we discuss how inferencing can be applied to each layer and how factual data can be extracted from text as an instantiation of the model.

2 The KYOTO project

The KYOTO project allows communities 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 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.

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 in separate layers with references across the structures. This makes it easier to harmonize the output of different linguistic processors for different languages and to add new semantic layers to the basic output, when needed (Bosma et al 2009). All modules in KYOTO draw their input from these structures. In fact, the word-sense-disambiguation process is carried out to the same KAF annotation in different languages and is therefore the same for all the languages (Agirre, Lopez de Lacalle & Soroa 2009). In the current system, there are processors for English, Dutch, Italian, Spanish, Basque, Chinese and Japanese.

The KYOTO system proceeds in 2 cycles (see Figure 1). In the 1st cycle, the **Tybot** (Term Yielding Robot) extracts the most relevant terms from the documents. The Tybot is another generic program that can do this for all the different languages in much the same way. The terms are organized as a structured hierarchy and, wherever possible, related to generic semantic databases, i.e. wordnets for each language. In Figure 1, italic terms occur in the text, and underlined terms are not found in wordnet. Straight terms are hyperonyms in wordnet that do not necessarily occur in the text but are linked to ontological classes. The domain experts can view the terms in the term database and edit them, i.e. adding or deleting terms, changing their meaning, adding definitions, changing relations, etc.

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.

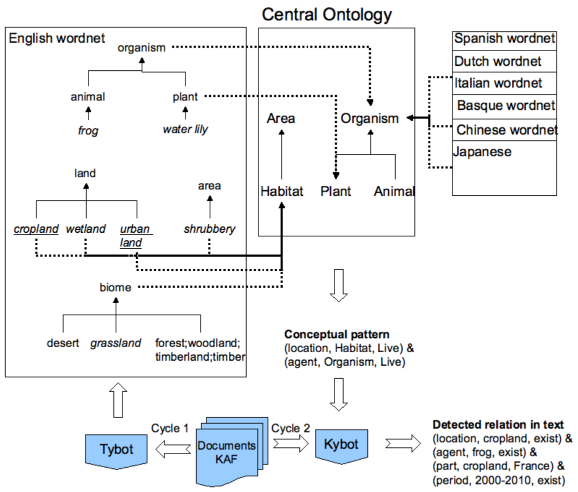

Figure 1: Two Cycles of processing in KYOTO

The 2nd cycle of the system involves the actual extraction of factual knowledge from the annotated documents by the **Kybots** (Knowledge Yielding Robots). Kybots use a collection of profiles that represent patterns 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. Since the semantics is defined through the ontology, it is possible to detect similar data across documents in different languages, even if expressed differently. In Figure 1, we give an example of a conceptual pattern that relates organisms that live in habitats. The Kybot can combine this pattern with words from the wordnet and morpho-syntactic structures. When a match is detected, the instantiation of the pattern is saved in a formal representation, either in KAF or in RDF. Since the wordnets in different languages are mapped to the same ontology and the text in these languages is represented in the same KAF, similar patterns can easily be applied to multiple languages.

3 Knowledge integration

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 seven languages have been represented in the Wordnet-LMF format (Soria, Monachini and Vossen 2009) and stored in a DebVisDic server (Horak et al. 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 WordNet3.0. The KYOTO ontology (version 1) consists of 786 classes divided over three layers. The top 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 et. al 2009). The second layer consists of concepts coming from the so-called Base Concepts in various wordnets (Vossen 1998, Izquierdo et al 2007). Examples of base concepts are: *building, vehicle, animal, plant, change, move, size, weight*. The Base Concepts (BCs) 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 complete coverage of the nominal and verbal part of WordNet. This has been completed for the nouns (about 500 synsets) and is currently being carried out for verbs and adjectives in WordNet 3.0. Through the BCs, we will ensure that any synset in the wordnets is mapped to some concept in the ontology either directly or indirectly[2]. The most specific layer of the ontology contains concepts representing species and regions relevant to the KYOTO domain. These concepts were provided by the end users, and in certain cases, concepts have been added to link the domain specific terms to the ontology.

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, Kybots need to apply some kind of inferencing over the available knowledge.

[2] This set of BCs is more minimal than the BCs defined in EuroWordNet and BalkaNet. The original BC set contained too much redundancy and arbitrariness for our purposes.

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.

Concept mining systems for specific domains usually assume that the domain corpus provides the basis for building the vocabulary and, eventually, learning the associated domain ontology. However, modeling of the domain vocabulary requires to consider that:

- Every domain text contains general vocabulary in addition to domain terms;
- Every domain text contains references to named entities in the world;
- Every domain has large quantities of background concepts and terms, which are not all mentioned in the texts;

The KYOTO knowledge model assumes that the terminology from the domain text corpus is merged with a generic wordnet in a language so that the domain terms are anchored to more general terms and concepts. This requires that the term hierarchy for the domain is somehow disambiguated to match specific word meaning from the generic wordnet. Once the term hierarchy is aligned with a generic wordnet, existing mappings from wordnet to ontologies can be used to apply the ontological distinctions to the domain terms. Named entities are more likely to be found in other resources such as Wikipedia, DBPedia and GeoNames. This requires another alignment operation, where the concepts in the external sources need to be matched to wordnet as well and through wordnet to the ontology. The situation becomes more complex when existing domain thesauri and taxonomies are added to the knowledge base. Modeling the vocabulary and concepts in a domain is a complex knowledge integration problem.

The following knowledge repositories are relevant or the environment domain in KYOTO:

- Generic wordnets in each language ranging from 50,000 to 120,000 synsets.
- A term databases with about 500,000 terms extracted from about 1,000 documents in each language.
- Existing ontologies such as the EuroWordNet top-ontology (Vossen 1998), SUMO (Niles and Pease 2002) and DOLCE (Masolo et al 2003).
- Wikipedia: over 3 million articles in English and large volumes in other languages, by September 2009[3].
- DBPedia: 2.6 million things and 274 million pieces of information (RDF triples), by September 2009[4].
- GeoNames: 8 million geographical names and 6.5 million unique features whereof 2.2 million populated places and 1.8 million alternate names, by September 2009[5].
- The Species 2000 database with 2.1 million species, having taxonomic relations and labels in many different languages[6].

We will describe our approach to the problem of integrating these in a useful knowledge repository. We propose a solution with 3 layers of repositories with different types of links between them that support different types of inferencing.

4 Division of knowledge over different layers

The amount and complexity of the knowledge repository is enormous. The Global Wordnet Grid architecture suggests that the wordnets extended with the domain vocabulary are anchored through the domain extension of the ontology. In practice this means, that the ontology needs to be extended with millions of new concepts. For example, the KYOTO ontology needs to make a distinction between taxonomic groups and individual organisms. Instances of species are *members* of a taxonomic group and *instances* of an organism. Likewise, we can predict that if an instance of a *frog* ceases to exist, it is not implied that the taxonomic group *Anura* ceases to exist but only an instance of the organism *Anura*. The former is only the case when all members of *Anura* cease to exist. As a consequence, the ontology that represents all species in this domain should include all 2.1 million species twice (!), once as group and once as a type of organism.

Such a model leads to various practical problems. First of all, ontologies of that size cannot be loaded in any existing inferencing system. Inferences as the above can thus not be made be-

[3] http://www.wikipedia.org/

[4] http://dbpedia.org/About

[5] http://www.geonames.org/about.html

[6] http://www.sp2000.org/

cause of the size of such an ontology. Another problem is that the vocabularies are linguistically too complex and diverse. Whereas the species can be considered as rigid concepts, as defined by Welty and Guarino (2002), this is not the case for most of the terms that are learned from the document collection. In the environment domain, the documents typically include terms for roles of species rather than the species as such, e.g. *invasive species*, *migration species*, *threatened species*. For mining facts from documents, these non-rigid role terms have more information value than the defining properties of the species.

For a knowledge sharing system as modeled by the Global Wordnet Grid, it is thus more important to precisely define what the roles and processes are in which species participate than to provide the defining properties of the species as such. Likewise, we propose a model of division of knowledge along the lines of the division of linguistic labor defined by Putnam (1975). Putnam argues that linguistic communities rely on the fact that experts know the defining properties of natural kind terms such as *gold* and can thus determine which instances of matter are gold and which are not. Most natural language users therefore have a shallow definition of what gold is and can still use this definition to communicate valuable information on gold, such as for trading gold or buying jewelry.

Along the same lines, we propose a digital version of this principle, where we state that a computer does not need to know the defining properties of each rigid concept but can rely on the capacity of the domain expert to determine what the instances are of, for example, a particular species. Vast amounts of words for rigid concepts can likewise remain in the vocabularies as long as we indicate their status as rigid concepts.

More useful is to properly represent the roles and processes in which the rigid concepts participate. These need to be represented both in the vocabularies and in the ontology to be able to process information in a proper way and to carry out the necessary inferencing.

In addition, terms from the term database are mapped to the most specific synset as well. In the example shown in Figure-2, we see typical role concepts as terms. For these role concepts, we infer that they do not represent rigid subtypes but can be used to refer to instances of concepts that play a specific role. The role relation to the process needs to be defined more specifically through a mapping relation with the ontology. To properly define the semantics of this model, we need to define the precise relations between the concepts represented in the different repositories. This will be discussed in the next section.

5 Relations between the different layers

We thus have three different types of repositories: vocabularies, wordnets and ontologies. Each repository has internal relations and also there are relations from vocabularies to wordnets and from wordnets to the ontology. Thus, wordnet can be seen as conceptual bridge between the vocabulary and the ontology.

Following the DOLCE model, the KYOTO ontology has major hierarchies for **endurants** (e.g. things such as 'plant', 'highway'), **perdurants** (processes such as 'migration', 'obstruction'), and **qualities** (e.g. properties such as 'endemic', 'poisonus'). Endurants include both types and roles such as 'frog' or 'EndangeredRole'. Events, processes and states are classified under Perdurants. Properties are classified under Quality. The following relations are used within the ontology:

- subClassOf, equivalentTo, generic-constituent relations between Endurant:Endurant, Perdurant:Perdurant, Quality:Quality.
- playedBy relation between Role:Endurant.
- hasRole[7] relation between Perdurant:Role.

For example, the Endurant concepts ‘plant’ and ‘animal’ have a subClassOf relation to ‘organism’ and the Endurant ‘highway’ is a subClassOf ‘physical-object’, Perdurant ‘AnimalMigrationProcess’ is a subClassOf ‘MigrationProcess’. Endurants ‘MigrationRole’ and ‘BreedingRole’ both have a subClassOf relation to ‘AnimalRole’ and ‘AnimalRole’ has a playedBy relation to the Endurant 'Animal'. Finally, a 'MigrationRole' playedBy 'Animal' is part of the Perdurant 'MigrationProcess' through the hasRole relation.

The ontology is used to model the shared and language-neutral concepts and relations in the domain. Instances are excluded from the ontology. Instances will be detected in the documents and will be mapped to the ontology through instance to ontology relations (see below). There are two relations that we need for this: instanceOf from instances to Endurant, Perdurant, or Quality and instancePlay from instances to Role. Specific entities in discourse, such as an

[7] The hasRole relation is compliant to the participant relation in DOLCE. Whereas participant is between Perdurant and Endurant, hasRole is more specific: between Perdurant and Role.

animal identified as *Duck1*, are then instances of a class in the type hierarchy of objects, e.g. Duck1 *instanceOf* Duck and can play roles, e.g. Duck1 *instancePlay* BreedingRole. The latter states that Duck1 could cease being a breeder while the former states that he cannot cease being a duck. Likewise, we will get a clear separation between the ontological model and the instantiation of the model as described in the text.

In addition to the ontology, we will have a wordnet for each language in the domain. In addition to the regular synset to synset relations in the wordnet, we will have a specific set of relations for mapping the synsets to the ontology, which are all prefixed with *sc_* standing for synset-to-concept. For rigid synsets, we have an *sc_equivalenceOf* or *sc_subclassOf* relation to Endurant, Perdurant or Quality. For non-rigid synsets, we have an *sc_domainOf* between synsets and Endurants, and an *sc_playRole* relation between synset and Roles. For each of these relations, the logical implications are defined as follows:

- **sc_equivalenceOf**: the synset is fully equivalent to the ontology Type & inherits all properties; the synset is *Rigid*
- **sc_ subclassOf**: the synset is a proper subclass of the ontology Type & inherits all properties; the synset is *Rigid*
- **sc_domainOf**: the synset is not a proper subclass of the ontology Type & is not disjoint (therefore orthogonal) with other synsets that are mapped to the same Type either through *sc_subclassOf* or s*c_domainOf*; the synset is *non-Rigid* but still inherits all properties of the target ontology Type; the synset is also related to a Role with a *sc_playRole* relation
- **sc_playRole**: the synset denotes instances for which the context of the Role applies for some period of time but this is not essential for the existence of the instances, i.e. if the context ceases to exist then the instances may still exist (Mizoguchi et al. 2007).

Only the *sc_equivalenceOf* and *sc_subclassOf* relations are used in the SUMO to Wordnet mapping, represented by the symbols '=' and '+' respectively. The SUMO-Wordnet mapping likewise does not systematically distinguish rigid from non-rigid concepts. In our model, we separate the linguistically and culturally specific vocabularies from the shared ontology while using the ontology as a point of interface for the concepts used by the various communities.

The lexicalization of the concepts can differ considerably across languages. Consider the following examples of different lexicalizations that can now be elegantly modeled:

{meat}Noun, English
-> sc_domainOf ***Cow, Sheep, Pig***
-> sc_playRole ***EatenRole***
{名 肉, 食物, 餐 }Noun, Chinese
-> sc_domainOf ***Cow, Sheep, Pig, Rat, Dog***
-> sc_playRole ***EatenRole***
{ طعام, لحم, غذاء}Noun, Arabic
-> sc_domainOf ***Cow, Sheep***
-> sc_playRole ***EatenRole***

In these examples, we see that words for *meat* in English, Chinese and Arabic are defined by the same role relation but have different ranges of domains, indicating what animals are considered as food. Similar cultural differences can be represented in this way.

6 Inferencing over the different layers

In text mining, there is a tight connection between the computational model for representing knowledge and the inferencing capabilities supported by the model. However, current logic based reasoning systems do not scale to the amount of information and the setting that is required for KYOTO to match text with the semantic model. For instance, state of the art machinery like formal reasoners such as Pellet or Fact++ are unable to deal with large and complex ontologies as the ones the KYOTO project is currently envisaging. Thus, a knowledge representation and reasoning infrastructure must be designed and built that can scale and can be flexibly adapted to the varying capabilities required by the different modules of the whole KYOTO System.

New approaches to the problem follow a rather loose definition of inference, mainly relying on the use of large amounts of automatically acquired informal and inaccurate knowledge and approximate inferences (Agirre and Soroa 2009). For KYOTO, we can combine such loose approaches with more strict reasoning, each being applied to the different layers. Large amounts of named entities like those appearing in YAGO (Suchanek et al 2008) or DBpedia (Auer et al 2007) or Species2000, etc. are stored in advanced XML databases such as Virtuoso. Wordnets are stored in relational or XML databases (DebVisDic, Horak et al 2005), but for inferencing a more complex graph representation is required (Agirre and Soroa 2009; Laparra and Rigau 2009). Finally, formal ontologies are stored in standard OWL-DL.

Each type of knowledge repository allows different inferencing capabilities with its own benefits and drawbacks. For instance, Virtuoso allows to store millions of instances but only have a minimal inferencing ability when querying on SPARQL, while OWL-DL allows to perform complex logical operations on the stored data (like consistency, etc.) but it scales poorly. However, the three knowledge repositories are connected by different relationships, which allows computer programs to use different representational layers and different inferencing capabilities. SPARQL queries on Virtuoso provide in simple lookup and relation tracking facility until a match with a wordnet synset is found. Within the wordnet knowledge base more complex operations can be applied such as measuring distances and similarities in a graph-structure. The ontological structures applied to the wordnet can be used to perform formal inferencing over a limited set of fundamental implications (Álvez et al. 2008).

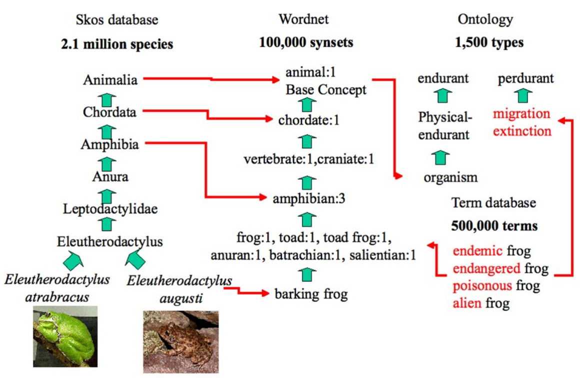

Figure 2: Division of knowledge over 3 layers

In Figure 2, we show an example of the three layers of the KYOTO model. We include in the vocabulary vast quantities of species represented as a SKOS hierarchy in Virtuoso database[8]. The species hierarchy is partially linked to a generic wordnet. SPARQL queries can be used to extract the hierarchical relations to find the most specific matching wordnet synset. The wordnet synset hierarchy can be traversed to find the most specific Base Concept that is matched to the ontology. In this way, we can infer for all species in the vocabulary that they are both *members* of a taxonomic group and *rigid subtypes* of organism.

[8]http://virtuoso.openlinksw.com/

For instance, expressions such as "migration of Hirundo rustica" can be semantically processed to obtain an appropriate interpretation. Querying the vocabulary database for "Hirundo rustica", we obtain a Species2000 entry corresponding to the WordNet3.0 synset 01594787 <barn_swallow, chimney_swallow, Hirundo_rustica>. Although this synset has not a direct connection to the KYOTO ontology, following the hypernym hierarchy, we find <bird> which is connected to the Endurant type Bird in the KYOTO ontology, which in turn is a subClassOf Animal. Our graph-based Word Sense Disambiguation algorithm can also assign the synset 07312616 to "migration". This synset is directly connected to AnimalMigrationProcess in the ontology, which is a subclassOf the Perdurant MigrationProcess type. MigrationProcess hasRole MigrationRole, which is playedBy the type Animal. Thus, different inferencing mechanisms can be applied to each knowledge repository in order to obtain the most appropriate interpretation for "migration of Hirundo rustica".

7 Conclusions

We described a three-layered model for representing vast and diverse amounts of knowledge in the Global Wordnet Grid. We defined the relations between these layers and the ways of inferencing on each layer. This model gives a more precise definition of linguistic and ontological knowledge and a more realistic implementation. The KYOTO model allows more flexibility to divide the burden of semantics to different layers and different communities. The KYOTO model can be used to represent wordnet families both in ILI style and in GWG style. It also allows a gradual transition from ILI to GWG representation of data.

8 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.

References

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.

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

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 LREC'08, Marrakesh, Morroco. 2008.

Auer A., C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak and Z. Ives. DBpedia: A Nucleus for a Web of Open Data. In Proceedings of the International Semantic Web Conference (ISWC), volume 4825 of Lecture Notes in Computer Science, pages 722-735. 2007.

Black, W., Elkateb. S., Rodriguez, H., Alkhalifa, M., Vossen, P.,Pease, A., Bertran, M., & Fellbaum, C. (2006). The Arabic WordNet Project. In: Proceedings of the Conference on Lexical Resources in the European Community. Genoa, Italy.

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 5th International Conference on Generative Approaches to the Lexicon 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.

Freitag, D. (1998) Information extraction from html: Application of a general machine learning approach. In Proceedings of the Fifteenth National Conference on Artificial Intelligence, 1998.

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 RANLP'07, Borovetz, Bulgaria. September, 2007.

Laparra E. and Rigau G. Integrating WordNet and FrameNet using a knowledge-based Word Sense Disambiguation algorithm. Proceedings of RANLP'09. Borovets, Bulgaria, September, 2009.

Magnini B. & Cavaglia, G. (2000) Integrating Subject Field Codes into WordNet. In Gavrilidou M., Crayannis G., Markantonatu S., Piperidis S. & Stainhaouer G. (Eds.) Proceedings of LREC-2000. Athens, Greece, 31 May- 2 June 2000

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

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

Pease A., C. Fellbaum, P. Vossen, (2008) Building the Global WordNet Grid. In Proceedings of the 18th International Congress of Linguists (CIL18), Seoul, Republic of Korea, July 21-26, 2008.

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.

Suchanek F., G. Kasneci and G. Weikum. YAGO - A Large Ontology from Wikipedia and WordNet. Journal of Web Semantics 6(39). 2008

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