Merging Global and Specialized Linguistic Ontologies

Bernardo Magnini and Manuela Speranza

ITC-irst, Istituto per la Ricerca Scientifica e Tecnologica
I-38050 Povo (Trento), Italy
{magnini, manspera}@itc.it

Abstract

There is an increasing interest in linguistic ontologies (e.g. WordNet) for a variety of content-based tasks, including conceptual indexing, word sense disambiguation and cross-language information retrieval. A relevant contribution in this direction is represented by linguistic ontologies with domain specific coverage, which are a crucial topic for the development of concrete application systems.

This paper tries to go a step further in the direction of the interoperability of specialized linguistic ontologies, by addressing the problem of their integration with global ontologies. This scenario poses some simplifications with respect to the general problem of merging ontologies, since it enables to define a strong precedence criterion so that terminological information overshadows generic information whenever conflicts arise. We assume the EuroWordNet model and propose a methodology to "plug" specialized linguistic ontologies into global ontologies. Experimental data related to an implemented algorithm, which has been tested on a global and a specialized linguistic ontology for the Italian language, are provided.

1. Introduction

Ontologies have become an important topic in research communities across several disciplines in relation to the key challenge of making the Internet and the Web a more friendly and productive place by filling more meaning to the vast and continuously growing amount of data on the net. The surging interest in the discovery and automatic or semi-automatic creation of complex, multi-relational knowledge structures, in fact, converges with recent proposals from various communities to build a Semantic Web relying on the use of ontologies as a means for the annotation of Web resources.

There is also an increasing interest in linguistic ontologies, such as WordNet, for a variety of content-based tasks, such as conceptual indexing and semantic query expansion to improve retrieval performance. More recently, the role of linguistic ontologies is also emerging in the context of distributed agents technologies, where the problem of meaning negotiation is crucial. A relevant perspective in this direction is represented by linguistic ontologies with domain specific coverage, whose role has been recognized as one of the major topics in many application areas.

This paper tries to go a step further in the direction of the interoperability of specialized linguistic ontologies, by addressing the problem of their integration with global linguistic ontologies. The possibility of merging information at different levels of specificity seems to be a crucial requirement at least in the case of large domains where terminologies include both very specific terms and a significant amount of common terms that may be shared with global ontologies.

The global-specialized scenario poses some simplifications with respect to the general problem of merging ontologies at the same degree of specificity (Hovy, 1998); in particular, in the case of conflicting information, it is possible to define a strong precedence criterion according to which terminological information overshadows generic information. We assume the EuroWordNet model and propose a methodology to "plug" specialized linguistic ontologies into global ontologies. The formal apparatus to realize this is based on plug relations that connect *basic concepts* of the specialized ontology to corresponding concepts in the generic ontology. We provide experimental data to support our approach, which has been tested on a global and a specialized linguistic ontology for the Italian language.

The paper is structured as follows. Section 2 presents the main features and uses of linguistic ontologies as opposed to formal ontologies. Section 3 introduces specialized linguistic ontologies, i.e. linguistic ontologies with domain specific coverage, as opposed to global linguistic ontologies containing generic knowledge. Section 4 deals with the problem of the interoperability of linguistic ontologies and describes the relations and the procedure enabling an integrated access of pairs of global and specialized linguistic ontologies.

2. Linguistic ontologies versus formal ontologies

In the recent years the increasing interest in ontologies for many natural language applications has led to the creation of ontologies for different purposes and with different features; therefore, it is worth pointing out the distinction between two main kinds of existing ontologies, i.e. formal and linguistic ontologies.

Linguistic ontologies are large scale lexical resources that cover most words of a language, while at the same time also providing an ontological structure where the main emphasis is on the relations between concepts; linguistic ontologies can therefore be seen both as a particular kind of lexical database and as particular kind of ontology.

Linguistic ontologies mainly differ from formal ontologies as far as their degree of formalization is concerned. Linguistic ontologies, in fact, do not reflect all the inherent aspects of formal ontologies. As Guarino et al. (1999) point out, for instance, WordNet's upper level structure shows no distinction between types and roles, whereas most of the original Pangloss (Knight and Luk, 1994) nodes in the Sensus ontology are actually types; to give a further example, WordNet's hierarchical structure lacks information about mutual disjointness between concepts.

Moreover, what distinguishes linguistic ontologies from formal ontologies, is their size: linguistic ontologies are very large (WordNet, for instance, has several dozen thousand synsets), while formal ontologies are generally much smaller.

The duality characterizing linguistic ontologies is reflected in their most prominent features. If we consider the linguistic level, they are strongly language-dependent, like electronic dictionaries, glossaries and all other linguistic resources, which focus on the words used in one specific language (in the case of monolingual resources) or in two or more specific language (in the case of bilingual or multilingual resources). On the other hand, if we consider the semantic level, we can observe that concepts denotated by different words in different languages can be shared, as it happens with the concepts in formal ontologies. In fact it is possible, at least for the core Indo-European languages, to identify a common ontological backbone behind the lexical surface of different languages (Guarino et al., 1999).

WordNet (Fellbaum, 1998), the best-known linguistic ontology, is an electronic lexical database where each sense of a lemma belongs to a different synset, i.e. a set of synonyms. Synsets are organized hierarchically by means of hypernymy and hyponymy relations. In WordNet other kinds of semantic relations among synsets are defined (e.g. role relation, part-of relation and cause relation), so as to build a more rich and complex semantic net. WordNet thus offers two distinct services: a lexicon, which describes the various word senses, and an ontology, which describes the semantic relationships among concepts.

As a linguistic ontology, WordNet is strongly language-dependent, but as an ontology it could also be adapted to a cross-language environment using the EuroWordNet multilingual database (Vossen, 1998) and mapping synsets into the EuroWordNet InterLingual Index, i.e. the index that links monolingual wordnets for all the languages covered by EuroWordNet. There are several examples of monolingual wordnets for many other languages, such as Dutch, Spanish, Italian, German and Basque.

A formal ontology based on linguistic motivation is the Generalized Upper Model (GUM) knowledge base (Bateman et al., 1995), an ontology primarily developed for Natural Language Processing applications. An upper model is an abstract linguistically motivated ontology meeting two requirements at the same time: i) a sufficient level of abstraction in the semantic types employed, as to escape the idiosyncrasies of surface realization and ease interfacing with domain knowledge, and ii) a sufficiently close relationship to surface regularities as to permit interfacing with natural language surface components.

2.1. Uses of formal ontologies

Recently ontologies have been used in the context of the Semantic Web. Ontologies can be employed to associate meaning with data and documents found on the Internet thus boosting diverse applications of information-retrieval systems. For the retrieval of information from the Web, Luke et al. (1996) propose a set of simple HTML Ontology Extensions to manually annotate Web pages with ontology-based knowledge, which performs high precision but is very expensive in terms of time.

OntoSeek (Guarino et al., 1999) is also based on content, but uses ontologies to find user's data in a large classical database of Web pages. Erdmann and Studer (1999) use an ontology to access sets of distributed XML documents on a conceptual level. Their approach defines the relationship between a given ontology and a document type definition (DTD) for classes of XML document. Thus, they are able to supplement syntactical access to XML documents by conceptual access.
However, as pointed out by Guarino et al. (1999), the practical adoption of ontologies in information-retrieval systems is limited by their insufficiently broad coverage and their need to be constantly updated; linguistic ontologies encompass both ontological and lexical information thus offering a way to partly overcome these limitations.

2.2. Uses of linguistic ontologies

Linguistic ontologies, and WordNet in particular, are proposed for content-based indexing, where semantic information is added to the classic word-based indexing. As an example, *Conceptual Indexing* (Woods, 1997) automatically organizes words and phrases of a body of material into a conceptual taxonomy that explicitly links each concept to its most specific generalizations. This taxonomic structure is used to organize links between semantically related concepts, and to make connections between terms of a request and related concepts in the index.

Mihalcea and Moldovan (2000) designed an IR system which performs a combined word-based and sense-based indexing exploiting WordNet. The inputs to IR systems consist of a question/query and a set of documents from which the information has to be retrieved. They add lexical and semantic information to both the query and the documents, during a preprocessing phase in which the input question and the texts are disambiguated. The disambiguation process relies on contextual information, and identifies the meaning of the words using WordNet.

The proble of sense disambiguation in the context of an IR task has been addressed, among the others, also by Gonzalo et al. (1998). In a preliminary experiment where disambiguation had been done manually, the vector space model for text retrieval gives better results if WordNet synsets are chosen as the indexing space, instead of word forms.

Desmontils and Jacquin (2001) present an approach where linguistic ontologies are used for information retrieval on the Internet. The indexing process is divided into four steps: i) for each page a flat index of terms is built; ii) WordNet is used to generate all candidate concepts which can be labeled with a term of the previous index; iii) each candidate concept of a page is studied to determine its representativeness of this page content; iv) all candidate concepts are filtered via an ontology, selecting the more representative for the content of the page.

More recently, the role of linguistic ontologies is also emerging in the context of distributed agents technologies, where the problem of meaning negotiation is crucial (Bouquet and Serafini, 2001).

3. Specialized linguistic ontologies

A particular kind of linguistic ontologies is represented by specialized linguistic ontologies, i.e. linguistic ontologies with domain specific coverage, as opposed to global linguistic ontologies, which contain generic knowledge. Focusing on one single domain, specialized linguistic ontologies often provide many sub-hierarchies of highly specialized concepts, whose lexicalizations tend to assume the shape of complex terms (i.e. multi-words); high level knowledge, on the other hand, tends to be simplified and domain oriented.

Many specialized linguistic ontologies have been developed, especially for practical applications, in domains such as art (see the Art and Architecture Getty Thesaurus), geography (see the Getty Thesaurus of Geographical Names), medicine (Gangemi et al., 1999), etc. and the importance of specialized linguistic ontologies is widely recognized in a number of works. The role of terminological resources for Natural Language Processing is addressed, for instance, by Maynard and Ananiadou (2000), who point out that high quality specialized resources such as dictionaries and ontologies are necessary for the development of hybrid approaches to automatic term recognition combining linguistic and contextual information with statistical information.

Buitelaar and Sacaleanu (2002) address the problem of tuning a general linguistic ontology such as WordNet or GermaNet to a specific domain (the medical domain, in the specific case). This involves both selecting the senses that are most appropriate for the domain and adding novel specific terms. Similarly, Turcato et al. (2000), describe a method for adapting a general purpose synonym database, like WordNet, to a specific domain (in this case, the aviation domain), adopting an eliminative approach based on the incremental pruning of the original database.

The use of domain terminologies also arises the problem of the (automatic) acquisition of thematic lexica and their mapping to a generic resource (Buitelaar and Sacaleanu, 2001; Vossen, 2001; Lavelli et al., 2002). As far as automatic term extraction is concerned, Basili et al. (2001) investigate whether syntactic context (i.e. structural information on local term context) can be used for determining "termhood" of given term candidates, with the aim of defining a weakly supervised "termhood" model suitably combining endogenous and exogenous syntactic information.

4. Merging global and specialized linguistic resources: the plug-in approach

One of the basic problems in the development of techniques for the Semantic Web is the integration of ontologies. Indeed the Web consists of a variety of information sources, and in order to extract information from such sources, their semantic integration is required.

Merging linguistic ontologies introduces issues concerning the amount of data to be managed (in the case of WordNet we have several dozen thousand synsets), which are typically neglected when upper levels are to be merged (Simov et al., 2001).

This paper tries to go a step further in the direction of the interoperability of linguistic ontologies, by addressing the problem of the integration of global and specialized linguistic ontologies. The possibility of merging information at different levels of specificity seems to be a crucial requirement at least in the case of domains, such as Economics or Law, that includes both very specific terms and a significant amount of common terms that may be shared by the two ontologies. We assume the EuroWordNet model and propose a methodology to "plug" specialized ontologies into global ontologies, i.e. to access them in conjunction through the construction of an integrated ontology.

4.1. Correspondences between global and specialized linguistic ontologies

A global linguistic ontology and a specialized one complement each other. The one contains generic knowledge without domain specific coverage, the other focuses on a specific domain, providing sub-hierarchies of highly specialized concepts. This scenario allows some significant simplifications when compared to the general problem of merging two ontologies. On the one hand, we have a specialized ontology, whose content is supposed to be more accurate and precise as far as specialized information is concerned; on the other hand, we can assume that the global ontology guarantees a more uniform coverage as far as high level concepts are concerned. These two assumptions provide us with a powerful precedence criterion for managing both information overlapping and inheritance in the integration procedure.

In spite of the differences existing between the two ontologies, in fact, it is often possible to find a certain degree of correspondence between them. In particular, we have information *overlapping* when the same concept belongs to the global and to the specialized ontology, and *over-differentiation* when a terminological concept has two or more corresponding concepts in the global ontology or the other way round. Finally, some specific concepts referring to technical notions may have no corresponding concept in the global ontology, which means there is a *conceptual gap*; in such cases a correspondence to the global ontology can be found through a more generic concept.

The sections highlighted in the global and the specialized ontology represented in Figure 1 reflect the correspondences we typically find between the two kinds on ontologies.

As for the global ontology (the bigger triangle), area *B1* is highlighted since it corresponds to the sub-hierarchies containing the concepts belonging to the same specific domain of the specialized ontology (the smaller triangle). The middle part of the specialized ontology, which we call *B* area, is also highlighted and it corresponds to concepts which are representative of the specific domain but are also present in the global ontology.

When the two ontologies undergo the integration procedure, an integrated ontology is constructed (Figure 2). Intuitively, we can think of it as if the specialized ontology somehow shifts over the global. In the integrated ontology, the information of the generic is maintained, with the exclusion of the sub-hierarchies containing the concepts belonging to the domain of the specialized ontology, which are covered by the corresponding area of the specialized. The

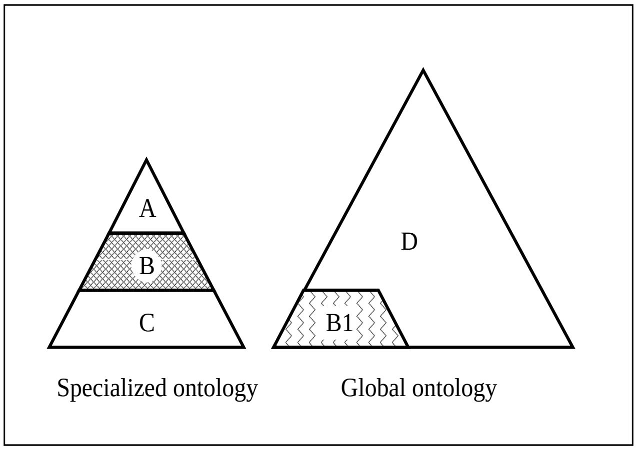

Figure 1: Separate specialized and global ontologies. Overlapping is represented in colored areas

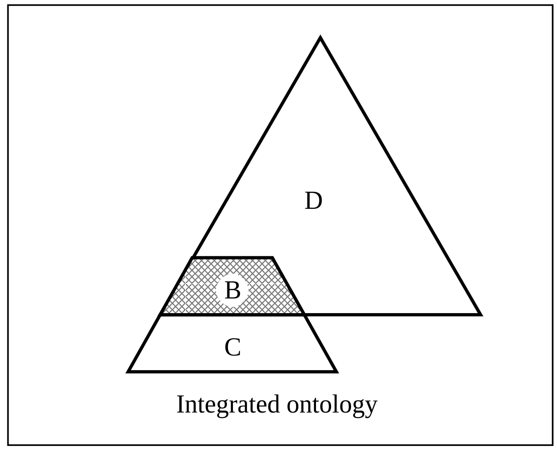

Figure 2: Integrated ontology. As to overlapping, precedence is given to the specialized ontology

integrated ontology also contains the most specific concepts of the specialized ontology (*C* area), which are not provided in the generic. What is excluded from the integrated ontology is the highest part of the hierarchy of the specialized ontology; it is represented by area *A* and contains generic concepts not belonging to a specialized domain, which are expected to be treated more precisely in the generic ontology.

4.2. Plug relations

The formal apparatus to realize an integrated ontology is based on the use of three different kinds of relations (plug-synonymy, plug-near-synonymy and plug-hyponymy) that connect basic concepts of the specialized ontology to the corresponding concepts in the global ontology, and on the use of eclipsing procedures that shadow certain concepts, either to avoid inconsistencies, or as a secondary effect of a plug relation.

A plug relation directly connects pairs of corresponding concepts, one belonging to the global ontology and the other to the specialized ontology. The main effect of a plug relation is the creation of one or more "plug concepts", which substitute the connected concepts, i.e. those directly involved in the relation. To describe the relations inherited by a plug concept, the following classification, adapted from Hirst and St-Onge (1998) is used: *up-links* of a concept are those whose target concept is more general (i.e. hypernymy and instance-of relations), *down-links* are those whose target is more specific (i.e. hyponymy and has-instance relations) and *horizontal-links* include all other relations (i.e. part-of relations, cause relations, derivation, etc.).

Plug-synonymy is used when overlapping concepts are found in the global ontology (hereafter *GO*) and in the specialized ontology (hereafter *SO*). The main effect of establishing a relation of plug-synonymy between concept *C* belonging to the global ontology (indicated as C^{GO}) and $C1^{SO}$ (i.e. concept *C1* belonging to the specialized ontology) is the creation of a plug concept $C1^{PLUG}$. The plug concept gets its linguistic forms (i.e. synonyms) from *SO*, up-links from GO, down-links from *SO* and horizontal-links from *SO* (see Table 1). As a secondary effect, the up relations of $C1^{SO}$ and the down relations of C^{GO} are eclipsed.

	$C1^{PLUG}$
Up links	*GO*
Down links	*SO*
Horizontal links	*GO* + *SO*

Table 1: Merging rules for plug-synonymy and plug-near-synonymy.

Plug-near-synonymy is used in two cases: (i) over-differentiation of the *GO*, i.e. when a concept in the *SO* has two or more corresponding concepts in the *GO*; this happens, for instance, when regular polysemy is represented in the GO but not in the SO; (ii) over-differentiation of the *SO*, i.e. when a concept in the *GO* corresponds to two or more concepts in the *SO*; this situation may happen as a consequence of subtle conceptual distinctions made by domain experts, which are not reported in the global ontology. Establishing a plug-near-synonymy relation has the same effect of creating a plug-synonymy (see Table 1).

Plug-hyponymy is used to connect concepts of the specialized ontology to more generic concepts in the case of conceptual gaps. The main effect of establishing a plug-hyponymy relation between C^{GO} (i.e. concept *C* of the global ontology) and $C1^{SO}$ (i.e. concept *C* of the specialized ontology) is the creation of the two plug concepts C^{PLUG} and $C1^{PLUG}$ (see Table 2). C^{PLUG} gets its linguistic forms from the *GO*, up-links from the *GO*, down-links are the hyponyms of C^{GO} plus the link to $C1^{PLUG}$ and horizontal-links from the *GO*. The other plug node, $C1^{PLUG}$, gets its linguistic form from the *SO*, C^{PLUG} as hypernym, down links from the *SO* and horizontal links from the *SO*. As a secondary effect, the hypernym of $C1^{SO}$ is eclipsed.

Eclipsing is a secondary effect of establishing a plug re-

	C^{PLUG}	$C1^{PLUG}$
Up links	*GO*	C^{PLUG}
Down links	*GO* + $C1^{PLUG}$	*SO*
Horizontal links	*GO*	*SO*

Table 2: Merging rules for plug-hyponymy

lation and is also an independent procedure used to avoid the case that pairs of overlapping concepts placed inconsistently in the taxonomies are included in the merged ontology; this could happen, for instance, when ”whale” is placed under a ”fish” sub-hierarchy in a common sense ontology, while also appearing in the mammal taxonomy of a scientific ontology.

4.3. Integration procedure

The plug-in approach described in the previous subsection has been realized by means of a semi-automatic procedure with the following four main steps.

(1) Basic concepts identification. The domain expert identifies a preliminary set of ”basic concepts” in the specialized ontology. These concepts are highly representative of the domain and are also typically present in the global ontology. In addition, it is required that basic concepts are disjoint among each other and that they assure a complete coverage of the specialized ontology, i.e. it is required that all terminal nodes have at least one basic concept in their ancestor list.

(2) Alignment. This step consists in aligning each basic concept with the more similar concept of the global ontology, on the basis of the linguistic form of the concepts. Then, for each pair a plug-in configuration is selected among those described in Section 4.2.

(3) Merging. For each plug-in configuration an integration algorithm reconstructs the corresponding portion of the integrated ontology. If the integration algorithm detects no inconsistencies, the next plug-in configuration is considered, otherwise step 4 is called.

(4) Resolution of inconsistencies. An inconsistency occurs when the implementation of a plug-in configuration is in contrast with an already realized plug-in. In this case the domain expert has to decide which configuration has the priority and consequently modify the other configuration, which will be passed again to step 2 of the procedure.

5. Experiments

The integration procedure described in Section 4.3 has been tested within the SI-TAL project [1] to connect a global wordnet and a specialized wordnet that have been created independently. ItalWordNet (IWN) (Roventini et al., 2000), which was created as part of the EuroWordNet project (Vossen, 1998) and further developed through the introduction of adjectives and adverbs, is the lexical database involved in the plug-in as a generic resource and consists of about 45,000 lemmas. Economic-WordNet (ECOWN) is a specialized wordnet for the economic domain and consists of about 5,000 lemmas distributed in about 4,700 synsets. Table 3 summarizes the quantitative data of the two resources considered.

	Specialized	Generic
Synsets	4,687	49,108
Senses	5,313	64,251
Lemmas	5,130	45,006
Internal Relations	9,372	126,326
Variants/synsets	1.13	1.30
Senses/lemmas	1.03	1.42

Table 3: IWN and ECOWN quantitative data

As a first step, about 250 basic synsets (5.3% of the resource) of the specialized wordnet were manually identified by a domain expert, including, for instance “azione” (“share”), and excluding less informative synsets, such as “azione” (“action”). Alignment with respect to the generic wordnet (step 2 of the procedure) is carried out with an algorithm that considers the match of the variants. Candidates are then checked by the domain expert, who also chooses the proper plug relation. In the case of gaps, a synset with a more generic meaning was selected and a plug-hyponymy relation was chosen.

At this point the merging algorithm takes each plug relation and reconstructs a portion of the integrated wordnet. In total, 4,662 ECOWN synsets were connected to IWN: 577 synsets (corresponding to area *B* in Figure 2) substitute the synsets provided in the global ontology to represent the corresponding concepts (*B1* area in Figure 1); 4085 synsets, corresponding to the most specific concepts of the domain (*C* area in Figure 2) are properly added to the database. 25 high level ECOWN synsets (*A* area in Figure 1) were eclipsed as the effect of plug relations. The number of plug relations established is 269 (92 plug-synonymy, 36 plug-near-synonymy and 141 plug-hyponymy relations), while 449 IWN synsets with an economic meaning were eclipsed, either as a consequence of plug relations (when the two taxonomic structures are consistent) or through the independent procedure of eclipsing (when the taxonomies are inconsistent). Each relation connects on average 17,3 synsets.

6. Conclusions

After discussing the main features and uses of linguistic ontologies as opposed to formal ontologies, we have addressed the problem of the interoperability between linguistic ontologies. We have presented a methodology for the integration of a global and a specialized linguistic ontology. The global-specialized situation allows to define a strong precedence criterion to solve cases of conflicting information. The advantage of the approach is that a limited number of plug relations allows to connect a large amount of concepts (i.e. synsets) in the two ontologies.

[1] Si-TAL (Integrated System for the Automatic Treatment of Language) is a National Project devoted to the creation of large linguistic resources and software for Italian written and spoken language processing.

7. References

R. Basili, M.T. Pazienza, and F.M. Zanzotto. 2001. Modelling syntactic context in automatic term extraction. In *Proc. of Recent Advances in Natural Language Processing (RANLP '01)*, Tzigov Chark, Bulgaria, September.

J.A. Bateman, B. Magnini, and G. Fabris. 1995. The generalized upper model knowledge base: Organization and use. In *Proc. of International Conference on Bulding and Sharing of Very Large-Scale Knowledge Bases*, Twente, The Netherlands, April.

P. Bouquet and L. Serafini. 2001. Two formalizations of a context: a comparison. In *Proc. of Third International Conference on Modeling and Using Context*, Dundee, Scotland, July.

P. Buitelaar and B. Sacaleanu. 2001. Ranking and selecting synsets by domain relevance. In *Proc. of NAACL Workshop WordNet and Other Lexical Resources: Applications, Extensions and Customizations*, Pittsburgh, June. held in conjunction with NAACL2001.

P. Buitelaar and B. Sacaleanu. 2002. Extending synsets with medical terms. In *Proc. of the First Global WordNet Conference*, Mysore, India, January.

E. Desmontils and C. Jacquin. 2001. Indexing a web site with a terminology oriented ontology. In *Proc. of SWWS International Semantic Web Working Symposium*, Stanford University, USA, July, August.

M. Erdmann and R. Studer. 1999. Ontologies as conceptual models for XML documents. In *Proc. of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management (KAW '99)*, Voyager Inn, Banff, Alberta, Canada, October.

C. Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, US.

A. Gangemi, D.M. Pisanelli, and G. Steve. 1999. Overview of the ONIONS project: Applying ontologies to the integration of medical terminologies. *Data and Knowledge Engineering*, 31.

J. Gonzalo, F. Verdejio, Chugur, and J. Cigarran. 1998. Indexing with WordNet synsets can improve text retrieval. In S. Harabagiu, editor, *Proceeding of the Workshop "Usage of WordNet in Natural Language Processing Systems"*, Montreal, Quebec, Canada, August.

N. Guarino, C. Masolo, and G. Vetere. 1999. OntoSeek: Contet-based access to the web. *IEEE Intelligent Systems and Their Application*, 14(3):70–80.

G. Hirst and D. St-Onge. 1998. Lexical chains representations of context for the detection and correction of malapropisms. In C. Fellbaum, editor, *WordNet. An Electronic Lexical Database*. The MIT Press.

E. Hovy. 1998. Combining and standardizing large-scale, pratical ontologies for machine translation and other uses. In *Proc. of the First International Conference on Language Resources and Evaluation*, Granada, Spain, August.

K. Knight and S. Luk. 1994. Building a large knowledge base for machine translation. In *Proceedings of the American Association of Artificial Intelligence Conference AAAI-94*, Seattle, WA.

A. Lavelli, B. Magnini, and F. Sebastiani. 2002. Building thematic lexical resources by bootstrapping and machine learning. In *Proc. of the Workshop "Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data", Workshop at LREC-2002*. to appear.

S. Luke, L. Spector, and D. Rager. 1996. Ontology-based knowledge discovery on the world-wide-web. In *Proc. of the AAAI1996 Workshop on Internet-based Information Systems*, Portland, Oregon, August.

D. Maynard and S. Ananiadou. 2000. Creating and using domain-specific ontologies for terminological applications. In *Proc. of Second International Conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece, May, June.

R. Mihalcea and D. Moldovan. 2000. Semantic indexing using WordNet senses. In *Proc. of the ACL workshop on Recent Advances in Natural Language Processing and Information Retrieval*, Hong Kong, October.

A. Roventini, A. Alonge, F. Bertagna, B. Magnini, and N. Calzolari. 2000. ItalWordNet: a large semantic database for Italian. In *Proc. of the Second International Conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece, May, June.

K. I. Simov, K. Kiryakov, and M. Dimitrov. 2001. OntoMap - the guide to the upper-level. In *Proc. of SWWS International Semantic Web Working Symposium*, Stanford University, USA, July, August.

D. Turcato, F. Popowich, J. Toole, D. Fass, D. Nicholson, and G. Tisher. 2000. Adapting a synonym database to specific domains. In *Proc. of Workshop on Information Retrieval and Natural Language Processing*, Hong-Kong, October. held in conjunction with ACL2000.

P. Vossen, editor. 1998. *EuroWordNet: a Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers.

P. Vossen. 2001. Extending, trimming and fusing WordNet for technical documents. In *Proc. of NAACL Workshop WordNet and Other Lexical Resources: Applications, Extensions and Customizations*, Pittsburgh, June. held in conjunction with NAACL2001.

W.A. Woods. 1997. Conceptual indexing: A better way to organize knowledge. Technical report, SUN Technical Report TR-97-61.