Scaling Natural Language Understanding via User-driven Ontology Learning

Berenike Loos

European Media Laboratory, GmbH

Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany

`firstname.lastname@eml-d.villa-bosch.de`

Abstract

Non-statistical natural language understanding components need world knowledge of the domain for which they are applied in a machine-readable form. This knowledge can be represented by manually created ontologies. However, as soon as new concepts, instances or relations are involved in the domain, the manually created ontology lacks necessary information, i.e. it becomes obsolete and/or incomplete. This means its "world model" will be insufficient to understand the user. The scalability of a natural language understanding system, therefore, essentially depends on its capability to be up to date. The approach presented herein applies the information provided by the user in a dialog system to acquire the knowledge needed to understand him or her adequately. Furthermore, it takes the position that the type of incremental ontology learning as proposed herein constitutes a viable approach to enhance the scalability of natural language systems.

1 Introduction

To let a computer system understand natural language one needs knowledge about objects and their relations in the real world. As the manual modeling and maintenance of such knowledge structures, i.e. ontologies, are not only time and cost consuming,

but also lead to not-scalable systems, there exists a demand to build and populate them automatically or at least semi automatically. This is possible by analyzing unstructured, semi-structured or fully structured data by various linguistic as well as statistical means and by converting the results into an ontological form.

In an open-domain scalable natural language understanding (NLU) system the automatic learning of ontological concepts and corresponding relations between them is essential, as a complete modeling of the world is neither practicable nor feasible, as the real world and its objects, models and processes are constantly changing along with their denotations.

This paper assumes that a viable approach to this challenging problem is to learn ontological concepts and relations relevant to a certain user in a given context by the dialog system at the time of the user's inquiry. My central hypothesis is that the information about terms that lack any mapping to the employed knowledge representation of the language understanding component can only be found in topical corpora such as the Web. With the help of this information one can find the right node in the ontology to append the concept corresponding to the unknown term in case it is a noun or to insert it as an instance in case it is a proper noun or another named entity.

The goal of the ontology learning component is to extend the knowledge base of the NLU system and therefore it will gradually adapt to the user's needs.

An example from the area of spoken dialog systems would be that of a user walking through the city of Heidelberg and asking: "How do I get to

the Auerstein". This would lead to the detection of *Auerstein* as being neither recognizable by the speech recognizer nor mappable to the knowledge representation of the system. Therefore, the corresponding hypernym of *Auerstein* has to be found on the internet by recourse to additional information about the context of the user. In this case, the additional information consists of the location of the user, namely *Heidelberg*. Once found, the hypernym is mapped to a corresponding concept, which already exists in the ontology. If there is no such corresponding concept, the concept for the hypernym thereof has to be determined. The formerly unknown term is mapped to a concept and is integrated into the system's ontology as a child of the concept for the found hypernym. In case the unknown term is a proper noun, it is integrated as an instance of the concept for the hypernym. So far, the research undertaken is related to nouns and proper nouns, also more generally referred to as *terms* in this paper.

In the following section, I will describe related work undertaken to solve the task of ontology learning, followed by some remarks of the distinction between ontology learning and natural language in Section 3. Thereafter, I will sketch out the minimal stages involved in the type of ontology learning proposed herein in Section 4.

2   Related Work

The capability to acquire knowledge exactly at the time it is needed can be regarded as an important stepping stone towards scalable natural language understanding systems. The necessity of scalability in NLU became more and more obvious in open-domain dialog systems, as the knowledge base integrated into those can never be complete. Before the emergence of open-domain systems, more or less complete ontologies were modeled manually for the domain needed in the NLU system and were therefore not scalable to additional domains, unless modeled in advance in a manual fashion or by means of off-line ontology learning. Nonetheless, numerous off-line ontology learning frameworks exist, which alleviate the work of an ontology engineer to construct knowledge manually (Maedche, 2002), (Schutz and Buitelaar, 2005), (Cimiano et al., 2005). Most of these frameworks apply hybrid methods to

optimize their learning results.

For example, the ontology population method OntoLearn (Navigli et al., 2004) is based on text mining and other machine learning techniques and starts with a generic ontology like WordNet and documents in a given domain. The result is a domain extended and trimmed version of the initial ontology. For this, the system applies three phases to learn concepts:

- First, a terminology extraction method, using shallow techniques that range from stochastic methods to more sophisticated syntactic approaches, is applied, which extracts a list of domain terms (mostly nouns and proper nouns) from a set of documents representative for a given domain.

- Second, a semantic interpretation takes place which makes use of a compositional interpretation and structural semantic interconnections.

- After these two phases the extending and trimming of the initial ontology takes place. With the help of the semantic interpretation of the terms they can be organized in sub-trees and appended under the appropriate node of the initial ontology applying linguistic rules.

The text understanding system SYNDICATE (SYNthesis of DIstributed Knowledge Acquired from Texts) uses an integrated ontology learning module (Hahn and Marko, 2002). In this approach new concepts are learned with the help of text understanding, which applies two different sources of evidence, namely, the prior knowledge of the topic domain of the texts and grammatical constructions in which unknown lexical items occur in the texts.

In an incremental process a given ontology is updated as new concepts are acquired from real-world texts. The acquisition process is centered on the linguistic and conceptual "quality" of various forms of evidence underlying the generation and refinement of concept hypotheses. On the basis of the quality of evidence, concept hypotheses are ranked according to credibility and the most credible ones are selected for assimilation into the domain knowledge base.

The project Disciple (Stanescu et al., 2003) builds agents which can be initially trained by a subject matter expert and a knowledge engineer, in

a way similar to how an expert would teach an apprentice. A Disciple agent applies two different methods for ontology learning, i.e. exception-based and example-based ontology learning. The exception-based learning approach consists of four main phases:

- First, a candidate discovery takes place, in which the agent analyzes a rule together with its examples, exceptions and the ontology and finds the most plausible types of extensions of the latter that may reduce or eliminate the rule's exceptions.

- In the second phase the expert interacts with the agent to select one of the proposed candidates.

- Afterwards the agent elicits the ontology extension knowledge from the expert and finally a rule refinement takes place, in which the agent updates the rule and eliminates its exceptions based on the performed ontology extension.

- When the subject matter expert has to specify a fact involving a new instance or new feature in the agent teaching process, the example-based learning method is invoked. In this process the agent tries to find example sentences of the words next to a new term through various heuristics. For instance, he finds out that X is member of Y, and consequently can ask the expert. If he affirms, the new term can be memorized.

All of the approaches described above exhibit theoretical as well as practical (in the light of the task undertaken herein) shortcomings. The theoretical problems that have not been resolved in a satisfactory manner by the works described above (as well as numerous others) are:

- a clear separation of the linguistic and ontological subtasks involved in the overall ontology learning endeavor

- systematic ways and methods for evaluating the individual learning results

- rigorously defined baselines against which to evaluate the ensuing learning approaches.

In the following I will describe how these issues can be addressed within the user-driven ontology learning framework proposed herein.

3 Natural Language versus Ontology Learning

Before describing the actual ontology learning process it is important to make a clear distinction between the two fields involved: This is on the one hand natural language and on the other hand ontology learning.

The corpora to extract knowledge from should come from the internet as this source provides the most up-to-date information. The natural language texts are rich in terms, which can be used as labels of concepts in the ontology and rich in semantic relations, which can be used as ontological relations (aka properties).

The connection between the two areas which are working on similar topics but are using different terminology needs a distinction between the extraction of semantic information from natural language and the final process of integrating this knowledge into an ontology.
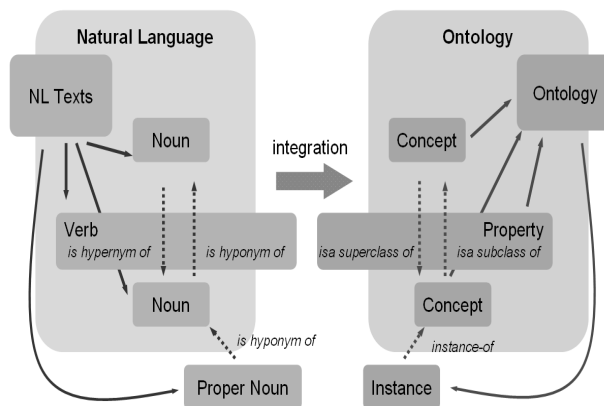

Figure 1: Natural Language versus Ontology Learning

Figure 1 shows the process of ontology learning from natural language text. On the left side relevant natural language terms are extracted. During a transformation process they are converted into labels of concepts and relations of an ontology. Proper nouns are transfered into instance labels in the ontology[1].

---

[1]In our understanding the term *ontology* denotes both the instance model as well as the ground ontology.

4 Scaling NLU via User-driven Ontology Learning

A user-driven ontology learning framework should be able to acquire knowledge at the run time of the NLU system. Therefore, terms which are not understood by the system have to be identified. In dialog systems this is true for all terms uttered or written by a user, which are not presently contained in the lexicon or can be derived by means of derivational or flexional morphology. In the following I will refer to these terms as *unknown terms*[2].

When a user of an open-domain spoken dialog system makes an utterance, it happens regularly, that the term is not represented in the system's lexicon. Since it is assumed, in this work, that the meaning of terms is represented by means of a formal ontology, a user-driven ontology learning framework is needed to determine the corresponding concepts for these terms, e.g., via a search on topical corpora. For instance, a term such as *Auerstein* could be employed to query a search engine. By applying natural language patterns, as proposed by Hearst (1992) and statistical methods, as proposed by Faulhaber et al. (2006) possible hypernyms or sets of hypernym candidates of the term can be extracted. For these a corresponding concept (or set of possible concepts) in the ontology employed by the dialog system need to be found. Last but not least the unknown term has to be inserted into the ontology as either an instance or a subclass of that concept. This process is described in greater detail in Section 5.4).

It is important to point out that terms often have more than one meaning, which can only be determined by recourse to the context in which it is uttered/found (Widdows, 2003), (Porzel et al., 2006). Therefore, information about this context needs to be added in order to make searching for the right hypernym feasible[3] as shown in Section 5.3. For example, the term *Lotus* can refer to a flower, a specific type of car or among copious other real world entities to a restaurant in Heidelberg. Therefore, a scalable ontology learning framework in a dialog system requires at least the following ingredients:

- A formal explicit model of a shared conceptualization of a specific domain of interest (Gruber, 1993), i.e. an ontology;

- processing methods which indicate the unknown terms;

- a corpus, as the starting point to retrieve hypernyms;

- methods for mapping hypernyms to concepts in the ontology;

- an evaluation framework;

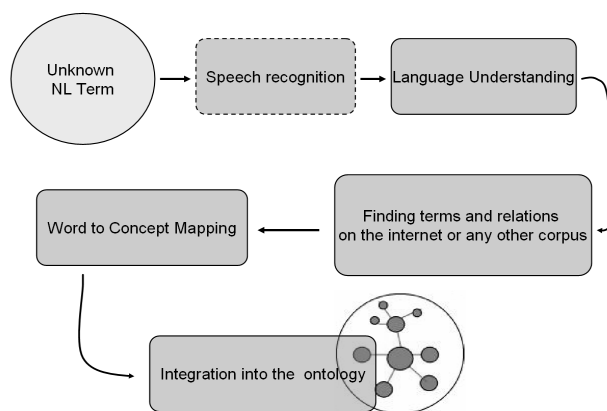Figure 2 shows the steps involved in on-demand ontology learning from the text to the knowledge side.

Figure 2: From text to knowledge

5 On-demand learning

From the cognitive point of view learning makes only sense when it happens on-demand. On-demand means, that it occurs on purpose and that activity is involved rather than passivity. As pointed out by Spitzer (2002) for human beings activity is necessary for learning. We cannot learn by drumming data into our brain through listening cassettes when sleeping or by similar fruitless techniques. The reason for this is, that we need active ways of structuring the data input into our brain. Furthermore, we try only to learn what we need to learn and are therefore quite economic with the "storage space" in our brain.

---

[2]This closely corresponds to what is termed *out-of-vocabulary (OOV) words* in the automatic speech recognition community.

[3]Of course, even in the same context a term can have more than one meaning as discussed in Section 5.7.

It makes not only for humans sense to simply learn whatever they need and what is useful for them. Therefore, I propose that ontology learning, as any other learning, is only useful and, in the end, possible if it is situated and motivated by the given context and the user needs. This can entail learning missing concepts relevant to a domain or to learn new concepts and instances which become necessary due to changes in a domain.

However, the fundamental ontological commitments should be adhered to. So, for example, the decision between a revisionary and a descriptive ontology should be kept in the hand of the knowledge engineer, as well as the choice between a multiplicative and a reductionist modeling[4]. As soon as the basic structure is given new knowledge can be integrated into this structure. Thus, for a reductionist ontology a concept such as *Hotel* should be appended only once, e.g. to an ontological concept as *PhysicalObject* rather than *NonPhysicalObject*.

In the following I will describe the various steps and components involved in on-demand ontology learning.

5.1 Unknown terms in dialog systems

In case the dialog system works with spoken language one can use the out-of-vocabulary (OOV) classification of the speech recognizer about all terms not found in the lexicon (Klakow et al., 2004). A solution for a phoneme-based recognition is the establishment of corresponding best-rated grapheme-chain hypotheses (Gallwitz, 2002). Those can be used for a search on the internet. In case the dialog system only works with written language it is easier to identify terms, which cannot be mapped to ontological concepts, at least if they are spelled correctly. To evaluate the framework itself adequately it is useful to apply only correctly written terms for a search.

Later on in both cases - i.e. in spoken and written dialog systems - a ranking algorithm of the best, say three, hypotheses should be selected to find the most adequate term. Here methods like the one of Google "Did you mean..." for spelling errors could be used.

5.2 Language Understanding

All correctly recognized terms of the user utterance can be mapped to concepts with the help of an analysis component. Frequently, production systems (Engel, 2002), semantic chunkers (Bryant, 2004) or simple word-to-concept lexica (Gurevych et al., 2003) are employed for this task. Such lexica assign corresponding natural language terms to all concepts of an ontology. This is especially important for a later semantic disambiguation of the unknown term (Loos and Porzel, 2004). In case the information of the concepts of the other terms of the utterance can help to evaluate results: When there is more than one concept proposal for an instance (i.e. on the linguistic side a proper noun like *Auerstein*) found in the word-to-concept lexicon, the semantic distance between each proposed concept and the other concepts of the user's question can be calculated[5].

5.3 Linguistic and Extra-linguistic Context

Not only linguistic but also extra linguistic context plays an important role in dialog systems. Thus, to understand the user in an open-domain dialog system it is important to know the extra-linguistic context of the utterances. If there is a context module or component in the system it can give information on the discourse domain, time and location of the user. This information can be used as a support for a search on the internet. E.g. the location of the user when searching for, say *Auerstein*, is advantageous, as in the context of the city *Heidelberg* it has a different meaning than in the context of another city (Bunt, 2000), (Porzel et al., 2006).

Part of the context information can be represented by the ontology as well as patterns for grouping a number of objects, processes and parameters for one distinctive context (Loos and Porzel, 2005).

5.4 Finding the appropriate hypernym on the internet

For this, the unknown term as well as an appropriate context term (if available) needs to be applied for searching possible hypernyms on the Web. As mentioned before an example could be the unknown term *Auerstein* and the context term *Heidelberg*.

---

[4]More information on these and other ontological choices can be found summarized in (Cimiano et al., 2004)

[5]E.g. with the single-source shortest path algorithm of Dijkstra (Cormen et al., 2001).

For searching the internet different encyclopedias and search engines can be used and the corresponding results can be compared. After a distinction between different types of unknown terms, the search methods are described.

Global versus local unknown terms: In the case of generally familiar proper nouns like stars, hotel chains or movies (so to say global unknown terms), a search on a topical encyclopedia can be quite successful. In the case of proper nouns, only common in a certain country region, such as Auerstein (Restaurant), Bierbrezel (Pub) and Lux (Cinema), which are local unknown terms, a search in an encyclopedia is generally not fruitful. Therefore, one can search with the help of a search engine.

As one can not know the kind of unknown terms beforehand, the encyclopedia search should be executed before the one using the search engine. If no results are produced, the latter will deliver them (hopefully). In case results are retrieved by the former, the latter can still be used to test those.

Encyclopedia Search: The structure of Encyclopedia entries is generally pre-assigned. That means, a program can know, where to find the most suitable information beforehand. In the case of finding hypernyms the first sentence in the encyclopedia description is often found to be the most useful. To give an example from Wikipedia[6], here is the first sentence for the search entry *Michael Ballack*:

(1) *Michael Ballack* (born September 26, 1976 in Grlitz, then East Germany) IS A German **football player**.

With the help of lexico-syntactic patterns, the hypernym can be extracted. These so-called Hearst patterns (Hearst, 1992) can be expected to occur frequently in lexicons for describing a term. In example 1 the pattern *X is a Y* would be matched and the hypernym *football player* of the term *Michael Ballack* could be extracted.

Title Search: To search only in the titles of web pages might have the advantage, that results can be

generated relatively fast. This is important as real-time performance is an important usability factor in dialog systems. When the titles contain the hypernym it still is to be expected that they might not consist of full sentences, Hearst patterns (Hearst, 1992) are, therefore, unlikely to be found. Alternatively, only the nouns in the title could be extracted and their occurrences counted. The noun most frequently found in all the titles could then be regarded as the most semantically connected term. To aid such frequency-based approaches stemming and clustering algorithms can be applied to group similar terms.

Page Search: For a page search Hearst patterns as in the encyclopedia search can almost certainly be applied. In contrast to encyclopedia entries the recall of those patterns is not so high in the texts from the web pages.
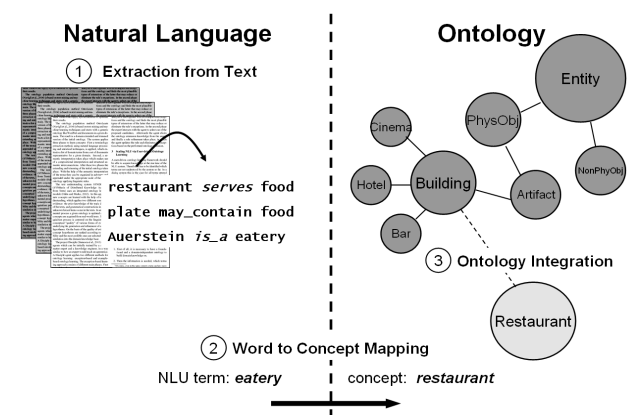

Figure 3: Tasks for the evaluation of ontology learning

The text surrounding the unknown term is searched for nouns. Equal to the title search the occurrence of nouns can then be counted. With the help of machine learning algorithms a text mining can be done to ameliorate the results.

5.5 Mapping text to knowledge by term narrowing and widening

As soon as an appropriate hypernym is found in a text the corresponding concept name should be determined. For term narrowing, the term has to be stemmed to its most general form. For the term widening, this form is used to find synonyms. Those

are, in turn, used for searching ontological concept names in the ontology integration phase. If the hypernym found is in a language other than the one used for the ontology, a translation of the terms has to take place as well.

5.6 Integration into an ontology

After the mapping phase newly learned concepts, instances or relations can be integrated into any domain-independent or even foundational ontology. If no corresponding concept can be found the next more general concept has to be determined by the techniques described above.

5.7 Evaluation

An evaluation of such a system can be divided into two types: one for the performance of the algorithms before the deployment of the system and one, which can be performed by a user during the run time of the system.

Methodological evaluation Before integrating the framework into a dialog system or any other NLU system an evaluation of the methods and their results should take place. Therefore, a representative baseline has to be established and a *gold-standard* (Grefenstette, 1994) created, depending on the task which is in the target of the evaluation. The ensuing steps in this type of evaluation are shown in Figure 3 and described here in their order:

1. The extraction of hypernyms of unknown words from text and the extraction of semantic relations (other than *is-a*) between NLU terms.

2. The mapping of a linguistic term to an ontological concept.

3. The integration of ontological concepts, instances and relations into the system's ontology.

Depending on the three steps the most adequate baseline method or algorithm for each of them has to be identified. In step 1 for the extraction of hypernyms a chance baseline as well as a majority class baseline will not do the job, because their performance would be too poor. Therefore, a well established algorithm which, for example applies a set of

standard Hearst patterns (Hearst, 1992) would constitute a potential candidate. For the mapping from text to knowledge (see step 2) the baseline could be a established by standard stemming combined with string similarity metrics. In case of different source and goal languages an additional machine translation step would also become necessary. For the baseline of ontology evaluation a task-based framework as proposed by (Porzel and Malaka, 2005) could be employable.

Evaluation by the user As soon as the framework is integrated into a dialog system the only way to evaluate it is by enabling the user to browse the ontological additions at his or her leisure and to decide whether terms have been understood correctly or not. In case two or more hypernyms are scored with the same – or quite similar – weights, this approach could also be quite helpful. An obvious reason for this circumstance is, that the term in question has more than one meaning in the same context. Here, only a further inquiry to the user can help to disambiguate the unknown term. In the *Auerstein* example a question like "Did you mean the hotel or the restaurant?" could be posed. Even though the system would show the user that it did not perfectly understand him/her, the user might be more contributory and less annoyed than with a question like "What did you mean?". The former question could also be posed by a person familiar with the place, to disambiguate the question of someone in search for *Auerstein* and would therefore mirror a human-human dialogs, which in turn would furthermore lead to more natural human-computer dialogs.

6 Concluding Remarks

In this paper I have shown, that the scalability of non-statistical natural language understanding systems essentially depends on its capability to be up to date when it comes to understand language. Furthermore, I took the position that ontology learning is viable, when it happens incrementally and in a context-sensitive fashion. Future work will focus on implementation and evaluation within a running multi-modal dialog system. Additionally, a tight integration with automatic lexicon and grammar learning is of paramount importance.

References

John Bryant. 2004. Scalable construction-based parsing and semantic analysis. In *Proceedings of the 2nd International Workshop on Scalable Natural Language Understanding (ScaNaLU 2004) at HLT-NAACL 2004*.

Harry Bunt. 2000. Dialogue pragmatics and context specification. In H.C. Bunt and W.J. Black, editors, *Computational Pragmatics, Abduction, Belief and Context; Studies in Computational Pragmatics*, pages 81–150. John Benjamins, Amsterdam.

Philipp Cimiano, Andreas Eberhart, Daniel Hitzler, Pascal Oberle, Steffen Staab, and Rudi Studer. 2004. The SmartWeb foundational ontology. *SmartWeb Project Report*.

Philipp Cimiano, Günter Ladwig, and Steffen Staab. 2005. Gimme' the context: Context-driven automatic semantic annotation with C-PANKOW. In *Proceedings of the 14th World Wide Web Conference*. ACM Press.

Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. 2001. Section 24.3: Dijkstra's algorithm. In *Introduction to Algorithms, Second Edition*, pages 595–601. MIT Press and McGraw-Hill.

Ralf Engel. 2002. SPIN: Language understanding for spoken dialogue systems using a production system approach. In *Proceedings of the International Conference on Speech and Language Processing 2002*, Denver, USA.

Arndt Faulhaber, Berenike Loos, Robert Porzel, and Rainer Malaka. 2006. Open-class named entity classification in multiple domains. In *Proceedings of the Ontolex Workshop*. Genua, Italy.

Florian Gallwitz. 2002. *Integrated Stochastic Models for Spontaneous Speech Recognition*. Logos, Berlin.

Gregory Grefenstette. 1994. *Explorations in Automatic Thesaurus Discovery*. Kluwer Academic Publishers, USA.

Thomas Gruber. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition (5)*.

Iryna Gurevych, Rainer Malaka, Robert Porzel, and Hans-Peter Zorn. 2003. Semantic coherence scoring using an ontology. In *Proc. of the HLT/NAACL 2003*, page (in press), Edmonton, CN.

Udo Hahn and Kornl G. Marko. 2002. Ontology and lexicon evolution by text understanding. In *Proceedings of the ECAI 2002 Workshop on Machine Learning and Natural Language Processing for Ontology Engineering (OLT'2002)*. Lyon, France.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of COLING 92*, Nantes, France.

Dietrich Klakow, Georg Rose, and Xavier Aubert. 2004. Oov-detection in a large vocabulary system using automatically defined word-fragments as filler. In *Proceedings of EUROSPEECH'99*, Budapest, Hungary.

Berenike Loos and Robert Porzel. 2004. Resolution of lexical ambiguities in spoken dialogue systems. In *Proceedings of the 5th Workshop on Diascourse and Dialogue*, Cambridge, Massachusetts, USA.

Berenike Loos and Robert Porzel. 2005. Towards ontology-based pragmatic analysis. In *Proceedings of DIALOR'05*, Nancy, France.

Alexander Maedche. 2002. *Ontology Learning for the Semantic Web*. Kluwer Academic Publishers, USA.

Roberto Navigli, Paola Velardi, Alessandro Cucchiarelli, and Francesca Neri. 2004. Extending and enriching WordNet with OntoLearn. In *Proceeeings of Integrated Approach for Web Ontology Learning and Engineering*. IEEE Computer.

Robert Porzel and Rainer Malaka. 2005. A task-based framework for ontology learning, population and evaluation. *Ontology Learning from Text: Methods, Evaluation and Applications Frontiers in Artificial Intelligence and Applications Series*, 123.

Robert Porzel, Iryna Gurevych, and Rainer Malaka. 2006. In context: Integrating domain- and situation-specific knowledge. *SmartKom Foundations of Multimodal Dialogue Systems, Springer, Cognitive Technologies*.

Alexander Schutz and Paul Buitelaar. 2005. RelExt: A tool for relation extraction in ontology extension. In *Proceedings of the 4th International Semantic Web Conference*. Galway, Ireland.

Manfred Spitzer. 2002. *Lernen*. Spektrum Akademischer Verlag.

Bogdan Stanescu, Cristina Boicu, Gabriel Balan, Marcel Barbulescu, Mihai Boicu, and Gheorghe Tecuci. 2003. Ontologies for learning agents: Problems, solutions and directions. In *Proceedings of the 2003 IEEE International Conference on Systems, Man and Cybernetics, Volume: 3*. Washington D.C.

Dominic Widdows. 2003. A mathematical model for context and word-meaning. In *International and Interdisciplinary Conference on Modeling and Using Context*. Stanford, California.