In: *Proceedings of CoNLL-2000 and LLL-2000*, pages 167–175, Lisbon, Portugal, 2000.

Extracting a Domain-Specific Ontology from a Corporate Intranet

Jörg-Uwe Kietz and **Raphael Volz**
Swiss Life, Corporate Center, IT Research & Development, Zürich, Switzerland
uwe.kietz@swisslife.ch, ed.raphael@leahpar.de

Alexander Maedche
AIFB, Univ. Karlsruhe, D-76128 Karlsruhe, Germany
maedche@aifb.uni-karlsruhe.de

Abstract

This paper describes our actual and ongoing work in supporting semi-automatic ontology acquisition from a corporate intranet of an insurance company. A comprehensive architecture and a system for semi-automatic ontology acquisition supports processing semi-structured information (e.g. contained in dictionaries) and natural language documents and including existing core ontologies (e.g. GermaNet, WordNet). We present a method for acquiring a application-tailored domain ontology from given heterogeneous intranet sources.

1 Introduction

1.1 Need for focused access to knowledge

The amount of information available to corporate employees has grown drastically with the use of intranets. Unfortunally this growth of available information has made the *access* to useful or needed information not much easier as the access is usually based on keyword searching or even browsing. The focused access to knowledge resources like intranet documents plays a vital role in knowledge management and supports in general the shifting towards a *Semantic Web*.

Keyword searching results in a lot of irrelevant information as a term can have different meanings in distinct contents, e.g. "Zürich" as the name of a town and as the name of an insurance company. Presently it is quite difficult to provide this information to the search engine, e.g. exclude all the information about the town "Zürich" without losing information about the insurance company "Zürich".

The project *On-To Knowledge* builds an ontology-based tool environment to perform knowledge management dealing with large numbers of heterogeneous, distributed and semi-structured documents as found within large intranets and the World-Wide Web (Fensel et al., 2000). In this project ontologies play a key role by providing a common understanding of the domain. Semantically annotated documents are accessed using the vocabulary provided by a domain-specific ontology.

Providing the user with an access method based on ontological terms instead of keywords has several advantages. First, the abstraction given by the ontology provides that the user does not have to deal with document-specific representations. Second, by this abstraction robustness towards changes in content and format of the accessed documents is gained.

1.2 Semi-automatic vs. manual ontology construction

The required domain-specific ontologies for On-To-Knowledge are built manually. Ontologies are usually built using tools like *OntoEdit* (Staab and Maedche, 2000) or *Protege* (Grosso et al., 1999). Using such tools has simplified ontology construction. However, the wide-spread usage of ontologies is still hindered by the time-consuming and expensive manual construction task.

Within On-To-Knowledge our work evaluates semi-automatic ontology construction from texts as an alternative approach to ontology engineering. Based on the assumption that most concepts and conceptual structures of the domain as well the company terminology are described in documents, applying knowledge acquisition from text for ontology design seems to be promising. Therefore a number of proposals have been made to facilitate ontological engineering through automatic discovery from domain data, domain-specific natural language texts in particular (cf. (Byrd and Ravin, 1999; Faure and Nedellec, 1999; Hahn and Schnattinger, 1998; Morin, 1999; Resnik, 1993; Wiemer-Hastings et al., 1998)).

The extraction of ontologies from text can have additional benefits for On-To-Knowledge as the required semantic annotation of documents could be provided as a side effect of the extraction process.

1.3 The approach

Our approach is based on different heterogeneous sources: First, a generic core ontology is used as a top level structure for the domain-specific goal ontology. Second, domain-specific concepts are acquired using a dictionary that contains important corporate terms described in natural language. Third, we use a domain-specific and a general corpus of texts to remove concepts that were domain-unspecific. This task is accomplished using the heuristic that domain-specific concepts are more frequent in the domain-specific corpus. Eventually we learned relations between concepts by analyzing the aforementioned intranet documents. We used a multi-strategy approach in learning to level the specific advantages and drawbacks of different learning methods. Several methods were applied with the possibility to combine their results.

1.4 Organization

Section 2 describes the overall architecture of the system and explains our notion of ontologies. Section 3 discusses the methodology applied to acquire a domain-specific ontology. Section 4 highlights the applied learning mechanisms. Section 5 demonstrates some preliminary results. Eventually Section 6 points out further directions for our work and acknowledges other contributors to the work.

2 Architecture

A general architecture for the approach of semi-automatic ontology learning from natural language has been described in (Maedche and Staab, 2000b).

Our notion of ontologies is closely associated to the notion in OIL (Horrocks et al., 2000). From the expressive power it is equivalent to the CLASSIC-$\mathcal{ALN}$ (Borgida and Patel-Schneider, 1994) description logic. By that it combines three important aspects provided by Description Logics (providing formal semantics and efficient reasoning support), frame-based systems and web standards. XML is used as a serial syntax definition language, describing knowledge in terms of concepts and role restrictions (i.e. all- and cardinality-restrictions as in the DL $\mathcal{ALN}$). Also, relations can be introduced as an independent entity whose domain and range concepts can be restricted.

The system comprises of components for text management,information extraction, ontology learning, ontology storage, ontology visualisation and manual ontology engineering.

The text management& processing component supports efficient handling and processing of input sources. Multiple sources are supported like semi-structured information, natural language documents and existing ontologies.

The information extraction is provided by the system SMES (Saarbrücken Message Extraction System), a shallow text processor for German (cf. (Neumann et al., 2000; Neumann et al., 1997)). It performs syntactic analysis on natural language documents. SMES includes shallow parsing mechanisms and processes text at different layers: The first step tokenizes text and recognizes basic compound structures like numbers and dates. Next, the morphological structure of words is analyzed using a word stem lexicon containing 700.000 word stems. Phrasal structures are created in the third step using final state transducers that combine the morphological information. On top of these phrasal structures a dependency-based parser parser is applied. The results of each step are annotated in XML-tagged text and can be used independently.

Ontology learning operates on the extracted information and is used for three tasks. One task is the acquisition of new structures, the second task is the evaluation of given structures. Eventually this evaluation is used for pruning of domain-unspecific concepts. Several methods were implemented.

Pattern-based approaches The implemented heuristic methods are based on regular expressions, as the output of the information ex-

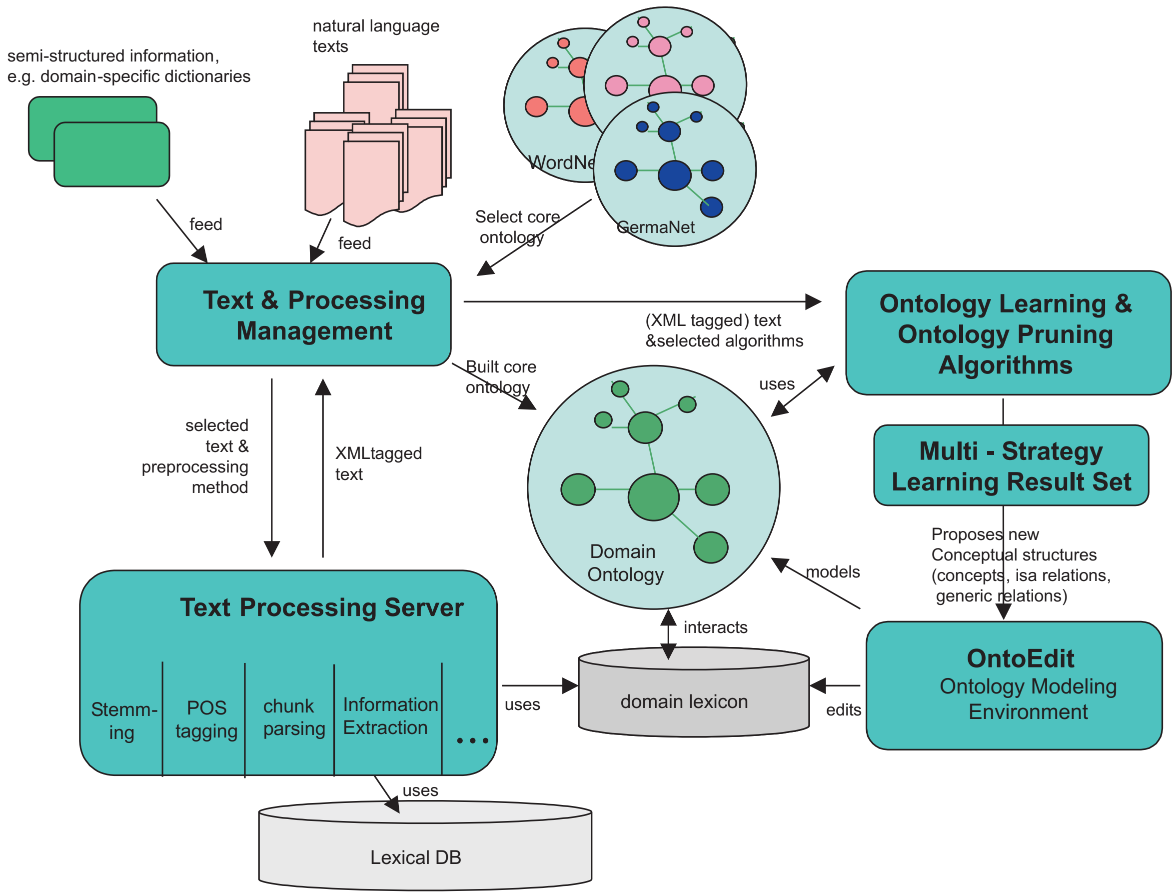

Figure 1: Architecture of the Ontology Learning Approach

traction component is regular. Patterns can be used to acquire taxonomic as well as conceptual relations. Also a very simple heuristic is used to acquire concepts from dictionaries, whereas dictionary entries are considered concepts.

Statistical approaches We implemented several statistical methods. One method retrieves concept frequencies from text (cf. (Salton, 1988)) and is used for the aforementioned pruning of concepts. A second method is used to acquire conceptual relations based on frequent couplings of concepts.

Combining results is enabled by the implementation of a common result structure for all learning methods. The complex task of ontology engineering is fitted better as it is possible to combine the results from different learning methods. (Michalski and Kaufmann, 1998) describe that multi-strategy learning architectures support balancing between advantages and disadvantages of different learning methods.

Ontology Engineering In our approach of semi-automatic ontology acquisition extensive support for ontology engineering is necessary. Manual ontology modeling & visualization is provided by the system OntoEdit[1]. It allows to edit and to browse existing as well as discovered ontological structures.

The acquired ontologies are stored in a relational database. To maximize portability only ANSI-SQL statements are used in the system. All data structures can be serialized into files, different formats like our internal XML-representation OXML, Frame-Logic (Kifer et al., 1995), RDF-Schema (W3C, 1999) and OIL (Horrocks et.al., 2000) are supported. An F-Logic inference engine described in further detail in (Decker, 1998) can be accessed from On-

[1] A comprehensive description of the ontology engineering system OntoEdit and the underlying methodology is given in (Staab and Maedche, 2000)

toEdit.

3 Methodology

Parallel to the architecture we developed an ontology acquisition methodology that emphasizes the semi-automatic manner of ontology construction. Figure 2 depicts the cyclic acquisition process.

The cycle starts with the selection of a generic core ontology (cf. subsection 4.1). Any large generic ontology (like CyC or Dahlgren's ontology), lexical-semantic nets (like WordNet, GermaNet or EuroWordNet) or domain-related ontologies (like TOVE) could start the process. After choosing the base ontology the user must identify the domain-specific texts that ought to be used for the extraction of domain-specific entities.

The next step is acquire domain-specific concept from the selected texts since the base ontology is domain-unspecific. The taxonomic embedding of all newly acquired concepts must also be established.

These steps enriched the ontology with domain-specific concepts, but still many domain-unspecific concepts remain. The given ontology must be focused to the domain. Therefore all unspecific concepts must be removed from the ontology. Now the conceptual structure of the target ontology is established.

But what about conceptual relations ? In addition to the relations provided by the base ontology that survived the focusing step (as their domain/range-concepts still exist) new conceptual relations are induced in the next step by applying multiple learning methods to the selected texts.

The resulting domain-specific ontology can be further refined and improved by repeating the acquisition cycle. This approach acknowledges the *evolving nature* of domain-specific ontologies that have adopt to changes in their domain or their application such as described in (Erdmann et al., 2000).

4 Acquisition Process

4.1 Base Ontology

GermaNet We decided to choose a lexical-semantic net for the German language, called GermaNet (cf. (Hamp and Feldweg, 1997)) as our base ontology. GermaNet is the German counterpart to the well known WordNet. Presently it builds a lexical semantic network for 16.000 German words, where three different types of word classes are distinguished: nouns, verbs and adjectives. Words are grouped into sets of synonyms so called synsets. As in WordNet two kinds of relations are recognized: Lexical relations that hold between words (like antonym) and semantic relations that hold between synsets (like meronym).

Conversion to our ontology primitive Synsets are regarded as concepts. Synsets are converted to concepts if they have at least one hypernym or hyponym relation. Some semantic relations between synsets are converted to conceptual relations, as some semantic relations do not have the property of being inherited to subconcepts if they hold between superconcepts[2].

The taxonomy is established using the hyponym relations. If a synset (transitively) points to itself through its hyponym relations, the hyponym relation that causes the cycle is ignored. Unfortunally there were several cycles within the verb classes[3].

Every word in a synset is analyzed by the information extraction component to acquire its stem. The stem is assigned to the corresponding concept to get a link to the analyzed texts. This link must be unique for each stem, thus a 1:n relation between a concept and stems is established.

Sometimes the same stem is acquired from different synsets. The disambiguation can not be done without using the context of the word. As we do not have any relations to do the disambiguation yet, we introduce a new concept that gets the ambiguous stem and make all conflicting synsets subconcepts of the newly introduced concepts. The newly introduced concept is embedded into the taxonomy as a sub concept of the deepest common super concept of all conflicting synsets. The disambiguation will be enabled using the relations between concepts identified in the user query.

[2]This is one of the essential differences to ontologies, where relations must hold for all subconcepts

[3]The only correct interpretation of a cycle is to regard the synsets contained as synonyms, but this would be modelled in a different manner (one synset), therefore we consider cycles as bugs (and not existent in the final version of GermaNet)

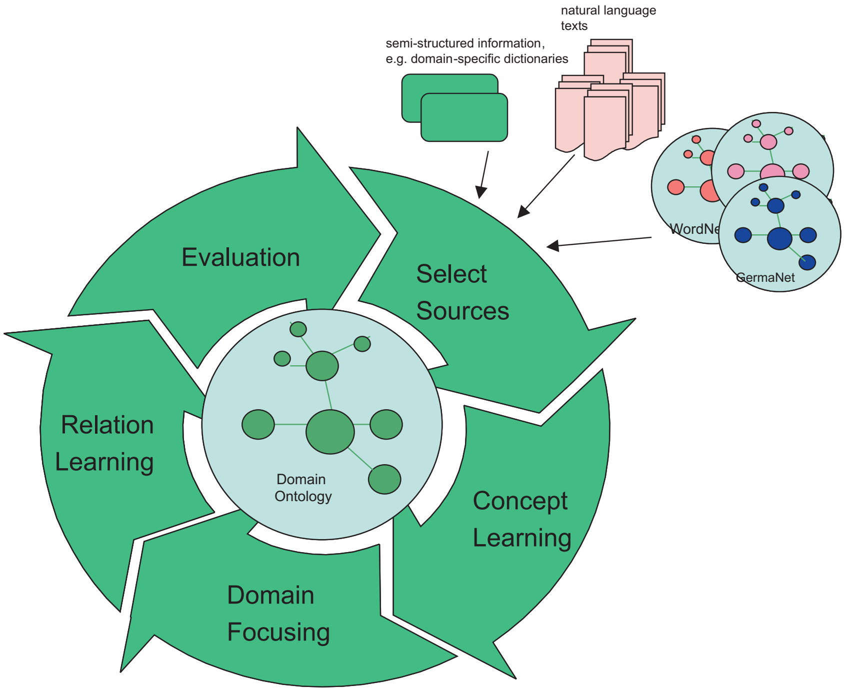

Figure 2: Semi-Automatic Ontology Acquisition Process

4.2 Acquisition of concepts

4.2.1 Getting concepts

As already mentioned we used a dictionary of corporate terms to acquire domain specific concepts. Figure 3 shows an example.

> **A.D.T.**
> **Automatic Debit Transfer**
>
> Electronic service arising from a debit authorization of the Yellow Account holder for a recipient to debit bills that fall due direct from the account.
> Cf. also *direct debit system*.

Figure 3: An example entry

This dictionary is multi-lingual to provide that all employees use common translations of domain terms. We converted the headwords of all German entries to concepts. Some entries share descriptions, as shown in the example. Those entries are joined and the headwords are regarded as synonyms. Some entries show references to other entries (e.g. "direct debit system'" is referenced by "A.D.T.'") we converted these links to conceptual relations.

Again, every headword is analyzed by the information extraction component (SMES) to acquire the word stem. This stem is assigned to the newly created concept, if not already existent in the ontology. If this stem exists, we need to do find out whether or not the dictionary entry describes the same concept as contained in the ontology.

4.2.2 Resolving conflicts

We apply several heuristics to solve this problem automatically. Table 1 shows the applied heuristics. In general dictionary entries are considered domain-specific and thus more important than existing concepts. The algorithm uses the information included within the dictionary entry and its description to find out whether the entry denotes the existing concept or induces a new concept.

First, the algorithm checks whether the conflicting dictionary head word denotes an acronym (e.g. *ALE* is an acronym for *unemployment benefits* in German. Unfortunally the stem reference contained in the ontology points to the concept *ale*, which is a sub concept of *alcoholic beverage*), in this case the stem reference is reassigned to the dictionary concept.

Propery	Automatic resolution
Word is acronym	Remove stem reference in ontology
Dictionary entry has no description	Do not import the entry and keep concept in ontology
Dictionary entry and ontology entry have a common super concept	Do not import the entry and keep concept in ontology
else	ask the user to resolve the conflict

Table 1: Dictionary: Resolution for stem conflicts

If this doesn't help, the algorithm checks whether further information is contained in the dictionary description by trying to find the super concept using the taxonomy acquisition method explained in the next section. If the found super concept is also a super concept of the concept in the ontology, the dictionary entry and the concept in the ontology are considered equal.

If no descriptions are contained in the dictionary entry and the entry is not an acronym , the concept in the ontology is kept. Last but not least, if no heuristics could be applied, the user is asked to resolve the conflict.

4.2.3 Getting the taxonomy

We used a pattern matching method to acquire the taxonomic embedding for the newly created concepts. The entires are aligned into the taxonomy using their descriptions. Based on the extracted syntactic information, several heuristic patterns for extracting taxonomic relations are applied. This works quite well, as the information extraction component supplies regular output.

(Hearst, 1992) motivated the acquisition of hyponyms by applying pattern matching to text. This approach was also applied in (Morin, 1999). Our patterns acquire stem references from the given texts using backward references inside the defined patterns. The stem references are used to get the concept contained in the ontology. Patterns can also reference the dictionary headwords, figure 4 depicts a very successful pattern.

Since regular expressions are not easy to read and much testing is required until expressions work, a pattern definition workbench was implemented to be able to document, categorize and test the defined patterns.

Pattern

1. *lexicon entry* :: (NP_1, NP_2, NP_i, *and / or* NP_n)
2. for all NP_i, $1 <= i <= n$ hypernym(NP_i, *lexicon entry*)

Result hypernym("electronic service", "A.D.T.")

Figure 4: Pattern Definition

4.3 Removal of concepts

We motivated the removal of generic concepts in section 3. In order to prune domain unspecific concepts, concept frequencies are determined from the selected domain-specific documents (see (Salton, 1988)). Concept frequencies are also determined from a second corpus that contains generic documents (as found in reference corpora like CELEX). We used the publicly available archive of a well-known German newspaper (http://http://www.taz.de/) as generic corpus.

All concept frequencies are propagated to super concepts, by summing the frequencies of sub concepts. The frequencies of both corpora are compared. All existing concepts that are more frequent within the domain-specific corpus remain in the ontology. Additionally a minimum concept frequency can be defined. All remaining concepts must satisfy this minimum frequency. Eventually the user can specify that concepts acquired from the dictionary (as well as their super concepts) must remain in the ontology.

4.4 Acquisition of Conceptual Relations

4.4.1 Statistical approach

This approach is founded on the idea that frequent couplings of concepts in sentences can be regarded as relevant relations between concepts. We adopted an algorithm based on association rules (see (Skrikant and Agrawal, 1995)) to find frequent correlations between concepts. Linguistically processed texts as input, where coupling of concepts within sentences are retrieved, are processed by our algorithm. Consult (Volz, 2000a; Maedche and Staab, 2000a) for a detailed description of this approach.

Two measures denote the statistical data derived by the algorithm: *Support* measures the quota of a specific coupling within the total number of couplings. *Confidence* denotes the part of all couplings supporting both domain and range concepts within the number of couplings that support the same domain concept. The retrieved measures are propagated to super concepts using the background knowledge provided by the taxonomy. This strategy is used to emphasize the couplings in higher levels of the taxonomy.

For instance, the linguistic processing may find that the word "policy" frequently co-occurs with each of the words "policy owner" and "insurance salesman". From this statistical linguistic data our approach derives correlations at the conceptual level, viz. between the concept *Policy* and the concepts, *PolicyOwner* and *InsuranceSalesman*. The discovery algorithm determines support and confidence measures for the relationships between these three pairs, as well as for relationships at higher levels of abstraction, such as between *Policy* and *Person*. In a final step, the algorithm determines the level of abstraction most suited to describe the conceptual relationships by pruning appearingly less adequate ones. Here, the relation between *Policy* and *Person* may be proposed for inclusion in the ontology.

Results are presented to the user, if the measures of a coupling satisfy specific minimum values provided by the user. Also, the input structures can be restricted to a set of certain concepts (whereas at least one element of every coupling must be in the given set) to be able to do a more focused way of relation acquisition.

We have to stress that this method can only be used to retrieve suggestions that are presented to the user. Manual labour is still needed to select and name the relations. To simplify user access results are conveniently displayed in the common result structure. The correctness towards the inheritance property of the taxonomy is automatically determined.

4.4.2 Pattern based approach

We extended the pattern based approach mentioned above towards the acquisition of named, non-taxonomic conceptual relations from text. We have defined some patterns on top of the information extracted by phrase-level processing of documents. The name of a relation can be assigned using the aforementioned backward references within a pattern. In contrast to the patterns defined to acquire taxonomic relations no dictionary headwords can be assigned.

5 Results

We can only present partial results at the moment, as our work is still ongoing. We converted only one part of speech class of the GermaNet lexical-semantic net. Our version contained 18509 synsets for the noun class. 18056 synsets have been converted, as all synsets that were not embedded in the taxonomy (by having neither super nor sub concepts) were skipped. 3565 new super concepts were introduced due to our disambiguation strategy. Therefore 21621 concepts were created.

The acquisition of concepts using the dictionary led to 1054 new concepts. 62 of the concepts in the dictionary were already included in the ontology. Only 518 concepts remain after the acquisition of the taxonomic embedding.

The learning method finds 679 is-a relations, but 161 is-a relations were determined to be wrong by empiric user evaluation. Thus 76,29 percent of all discovered relations were regarded correct by the user. Notably 49,14 percent of all dictionary entries could be imported automatically.

The evaluation of the acquisition of relations, as well as the pruning step is not yet done. Results regarding the acquisition of relations using our statistical approach in a different domain (tourism) have been presented in (Maedche and Staab, 2000a). Using a minimum support value of 0.04 and a minimum confidence of 0.01 best results were reached. 98 relations were discovered using an ontology that contained 284 concepts and 88 conceptual relations. 11% of the discovered relations were already modeled before, thus 13% of the hand-modelled relations were discovered by the learning algorithm.

6 Conclusions & Further Work

In this paper we have described our recent and ongoing work in semi-automatic ontology acquisition from a corporate intranet. Based on our comprehensive architecture a new approach for supporting the overall process of engineering ontologies from text is described. It is mainly based on a given core ontology, which is extended with domain specific concepts. The resulting ontology is pruned and restricted to a specific application using a corpus-based mechanism for ontology pruning. On top of the ontology two approaches supporting the difficult task of determining non-taxonomic conceptual relationships are applied.

In the future much work remains to be done. First, several techniques for evaluating the acquired ontology have to be developed. In our scenario we will apply ontology cross comparison techniques such as described in (Maedche and Staab, 2000a). Additionally, applying the ontology on top of the intranet documents (e.g. a information retrieval scenario, a semantic document annotation scenario such as described in (Erdmann et al., 2000)) will allow us an application-specific evaluation of the ontology using standard measures such as precision and recall. Second, our approach for multi-strategy learning is still in an early stage. We will have to elaborate how the results of different learning algorithms will have to be assessed and combined in the multi-strategy learning set. Nevertheless, an approach combing different resources on which different techniques are applied, seems promising for supporting the complex task of ontology learning from text.

Acknowledgements: This work has been partially found by the European Union and the Swiss Government under the contract-no "BBW Nr.99.0174" as part of the European commission Research Project "IST-1999-10132" (On-to-Knowledge). We thank the DFKI, language technology group, in particular Günter Neumann, who generously supported us in using their SMES system.

References

A. Borgida and P. Patel-Schneider. 1994. A semantics and complete algorithm for subsumption in the CLASSIC description logic. *Journal of Artificial Intelligence Research*, 1:277–308.

Roy J. Byrd and Yael Ravin. 1999. Identifying and extracting relations from text. In *NLDB'99 — 4th International Conference on Applications of Natural Language to Information Systems.*

S. Decker. 1998. On domain-specific declarative knowledge representation and database languages. In *Proc. of the 5th Knowledge Representation meets Databases Workshop (KRDB98)*, pages 9.1–9.7.

M. Erdmann, A. Maedche, H.-P. Schnurr, and Steffen Staab. 2000. From manual to semi-automatic semantic annotation: About ontology-based text annotation tools. In *P. Buitelaar & K. Hasida (eds). Proceedings of the COLING 2000 Workshop on Semantic Annotation and Intelligent Content*, Luxembourg, August.

David Faure and Claire Nedellec. 1999. Knowledge acquisition of predicate-argument structures from technical texts using machine learning. In *Proc. of Current Developments in Knowledge Acquisition, EKAW-99.*

D. Fensel, F. van Harmelen, H. Akkermans, M. Klein, J. Broekstra, C. Fluyt, J. van der Meer, H.-P. Schnurr, R. Studer, J. Davies, J. Hughes, U. Krohn, R. Engels, B. Bremdahl, F. Ygge, U. Reimer, and I. Horrocks. 2000. Ontoknowledge: Ontology-based tools for knowledge management. In *Proceedings of the eBusiness and eWork 2000 Conference (EMMSEC 2000)*, Madrid, Spain, To appear October.

E. Grosso, H. Eriksson, R. W. Fergerson, S. W. Tu, and M. M. Musen. 1999. Knowledge modeling at the millennium — the design and evolution of Protégé-2000. In *Proc. the 12th International Workshop on Knowledge Acquisition, Modeling and Mangement (KAW'99), Banff, Canada, October 1999.*

Udo Hahn and Klemens Schnattinger. 1998. Towards text knowledge engineering. In *AAAI '98 - Proceedings of the 15th National Conference on Artificial Intelligence. Madison, Wisconsin, July 26-30, 1998*, pages 129–144, Cambridge/Menlo Park. MIT Press/AAAI Press.

B. Hamp and H. Feldweg. 1997. Germanet - a lexical-semantic net for german. In *Proceedings of* ACL *workshop Automatic Information Extraction and Building of Lexical Semantic Resources for* NLP *Applications, Madrid.*

M.A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th International Conference on Computational Linguistics. Nantes, France.*

I. Horrocks, D. Fensel, J. Broekstra, S. Decker, M. Erdmann, C. Goble, F. van Harmelen, M. Klein, S. Staab, and R. Studer. 2000. The ontology inference layer oil, on-to-knowledge euist-10132 project deliverable no. otk-d1. Technical report, Free University Amsterdam, Division of Mathematics and Computer Science, Amsterdam, NL.

I. Horrocks et.al. 2000. The ontology interchange language oil: The grease between ontologies. Technical report, Dep. of Computer Science, Univ. of Manchester, UK/ Vrije Universiteit Amsterdam, NL/ AIdministrator, Nederland B.V./ AIFB, Univ. of Karlsruhe, DE. http://www.cs.vu.nl/˜dieter/oil/.

M. Kifer, G. Lausen, and J. Wu. 1995. Logical foundations of object-oriented and frame-based languages. *Journal of the ACM*, 42.

A. Maedche and S. Staab. 2000a. Discovering conceptual relations from text. In *Proceedings of ECAI-2000*. IOS Press, Amsterdam.

A. Maedche and S. Staab. 2000b. Semi-automatic engineering of ontologies from text. In *Proceedings of the 12th Internal Conference on Software and Knowledge Engineering. Chicago, USA*. KSI.

R. Michalski and K. Kaufmann. 1998. Data mining and knowledge discovery: A review of issues and multistrategy approach. In *Machine Learning and Data Mining Methods and Applications*. John Wiley, England.

E. Morin. 1999. Automatic acquisition of semantic relations between terms from technical corpora. In *Proc. of the Fifth International Congress on Terminology and Knowledge Engineering - TKE'99*.

G. Neumann, R. Backofen, J. Baur, M. Becker, and C. Braun. 1997. An information extraction core system for real world german text processing. In *ANLP'97 — Proceedings of the Conference on Applied Natural Language Processing*, pages 208–215, Washington, USA.

G. Neumann, C. Braun, and J. Piskorski. 2000. A divide-and-conquer strategy for shallow parsing of german free texts. In *Proceedings of ANLP-2000, Seattle, Washington*.

P. Resnik. 1993. *Selection and Information: A Class-based Approach to Lexical Relationships*. Ph.D. thesis, University of Pennsylania.

G. Salton. 1988. *Automatic Text Processing*. Addison-Wesley.

R. Skrikant and R. Agrawal. 1995. Mining generalized association rules. In *Proceedings of VLDB 1995*, pages 407–419.

S. Staab and A. Maedche. 2000. Ontology engineering beyond the modeling of concepts and relations. In *Proceedings of the ECAI'2000 Workshop on Application of Ontologies and Problem-Solving Methods*.

Raphael Volz. 2000a. Discovering conceptual relations from text. Studienarbeit, University of Karlsruhe (TH), Karlsruhe - Germany. in German.

Raphael Volz. 2000b. From texts to ontologies using machine learning and lexical-semantic networks [working title]. Diploma thesis, University of Karlsruhe (TH), Karlsruhe - Germany. in German.

W3C. 1999. Rdf schema specification. http://www.w3.org/TR/PR-rdf-schema/.

P. Wiemer-Hastings, A. Graesser, and K. Wiemer-Hastings. 1998. Inferring the meaning of verbs from context. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*.