Automatic Acquisition of Taxonomies from Text: FCA meets NLP

Philipp Cimiano, Steffen Staab and Julien Tane
Institute AIFB
University of Karlsruhe
{cimiano,staab,tane}@aifb.uni-karlsruhe.de

Abstract

We present a novel approach to the automatic acquisition of taxonomies or concept hierarchies from domain-specific texts based on Formal Concept Analysis (FCA). Our approach is based on the assumption that verbs pose more or less strong selectional restrictions on their arguments. The conceptual hierarchy is then built on the basis of the inclusion relations between the extensions of the selectional restrictions of all the verbs, while the verbs themselves provide intensional descriptions for each concept. We formalize this idea in terms of FCA and show how our approach can be used to acquire a concept hierarchy for the tourism domain out of texts. We then evaluate our method by considering an already existing ontology for this domain.

1 Introduction

Taxonomies or conceptual hierarchies are crucial for any knowledge-based system, i.e. any system making use of declarative knowledge about the domain it deals with. Within natural language processing (NLP), information extraction or retrieval systems for example can profit from a taxonomy to provide information at different levels of detail. Machine-learning based IE systems such as described in (Ciravegna, 2001) could for example identify concepts at different levels of abstraction with regard to a given concept hierarchy. Furthermore, the integration of a concept hierarchy would also enable such systems to generalize semantically over words and thus produce more compact and concise extraction rules. In information retrieval (IR), the availability of an ontology for a certain domain allows to replace traditional keyword-based approaches by more sophisticated ontology-based search mechanisms such as the one proposed in (Guarino et al., 1999). In general it is clear that any form of syntax-semantics interface will become more transparent and concise if a conceptual hierarchy is available.

However, it is also well known that every knowledge-based system suffers from the so called *knowledge acquisition bottleneck*, i.e. the difficulty to actually model the knowledge relevant for the domain in question. In particular ontology development is known to be a hard and time-consuming task.

In this paper we present a novel method to automatically acquire taxonomies from domain-specific texts based on Formal Concept Analysis (FCA), a method mainly used for the analysis of data (Ganter and Wille, 1999). The main benefits of our method are its adaptivity as it can be applied to arbitrary corpora and domains, its speed (compared to the process of hand-coding an ontology) as well as its robustness in the sense that it will not fail due to social aspects as present in traditional ontology development projects. Furthermore, if the corpora are updated regularly, it is also possible to let the ontology evolve according to the changes in the corpus. This is in line with the corpus and domain-specific form of lexicon as envisioned in (Buitelaar, 2000).

2 The Underlying Idea

An ontology is a formal specification of a conceptualization (Gruber, 1993). A conceptualization can be understood as an abstract representation of the world or domain we want to model for a certain purpose. The ontological model underlying this work is based on the one in (Bozsak et al., 2002):

Definition 1 (Ontology)
An ontology is a structure $O := (C, \leq_C, R, \sigma, \leq_R)$ *consisting of (i) two disjoint sets* C *and* R *called concept identifiers and relation identifiers respectively, (ii) a partial order* $\leq_C$ *on* C *called concept hierarchy or taxonomy, (iii) a function* $\sigma : R \rightarrow C^+$ *called signature and (iv) a partial order* $\leq_R$ *on* R *called relation hierarchy, where* $r_1 \leq_R r_2$ *implies* $|\sigma(r_1)| = |\sigma(r_2)|$ *and* $\pi_i(\sigma(r_1)) \leq_C \pi_i(\sigma(r_2))$ *for each* $1 \leq i \leq |\sigma(r_1)|$.

Furthermore, for each ontology O we will define a lexicon L_O as well as a mapping $F_O : C \rightarrow 2^{L_O}$ by which each concept is mapped to its possible lexical realizations. In addition, we will also consider the inverse function $F_O^{-1} : L_O \rightarrow 2^C$. Thus in our model a concept can be expressed through different expressions (*synonyms*) and one expression can refer to different concepts, i.e. expressions can be *polysemous*.
The aim of the approach presented in this paper is now to automatically acquire the partial order $\leq_C$ between a given set of concepts C. The general idea underlying our approach can be best illustrated with an example. In the context of the tourism domain, we all have for example the knowledge that things like a *hotel*, a *car*, a *bike*, a *trip* or an *excursion* can be booked. Furthermore, we know that we can rent a *car*, a *bike* or an *apartment* and that we can drive a *car* or a *bike*, but only ride a *bike*. Moreover, we know that we can join an *excursion* or a *trip*. We can now represent this knowledge in form of a matrix as depicted in table 1. On the basis of this knowledge, we could intuitively build a conceptual hierarchy as depicted in figure 1. If we furthermore

	bookable	rentable	driveable	rideable	joinable
apartment	x	x			
car	x	x	x		
motor-bike	x	x	x	x	
excursion	x				x
trip	x				x

Table 1: Tourism domain knowledge as matrix

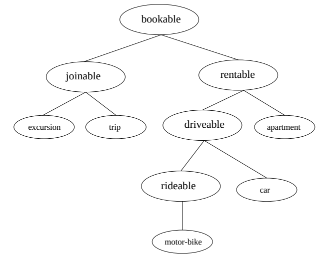

Figure 1: Hierarchy for the tourism example

reflect about the intuitive method we have used to construct this conceptual hierarchy, we would come to the conclusion that we have basically mapped the inclusion relations between the sets of the verbs' arguments to a partial order and furthermore have used the verbs itself to provide an intensional description of the abstract or non-lexical concepts we have created to group together certain 'lexicalized' concepts. In the next section we introduce Formal Concept Analysis and show how it can be used to formalize the intuitive method described above.

3 Formal Concept Analysis

Formal Concept Analysis (FCA) is a method mainly used for the analysis of data, i.e. for investigating and processing explicitly given information. Such data are structured into units which are formal abstractions of concepts[1] of human thought allowing meaningful comprehensible interpretation (Ganter and Wille, 1999). Central to FCA is the notion of a *formal context*:

Definition 2 (Formal Context)
A triple (G,M,I) is called a **formal context** *if* G *and* M *are sets and* $I \subseteq G \times M$ *is a binary relation between* G *and* M*. The elements in* G *are called* **objects***, those in* M **attributes** *and* I *the* **incidence** *of the context.*

For $A \subseteq G$ and dually for $B \subseteq M$, we define :

$$A' := \{m \in M | (g, m) \in I \; \forall g \in A\}$$

$$B' := \{g \in G | (g, m) \in I \; \forall m \in B\}$$

[1]Throughout this paper we will use the notion *concept* in the sense of *formal concept* as used in FCA (see below) as well in the ontological sense as defined in section 2. The meaning should in any case be clear from the context.

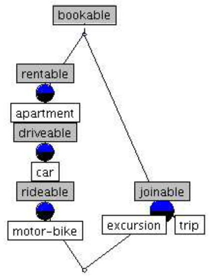

Figure 2: The tourism lattice

Intuitively speaking, A' is the set of all the attributes common to the objects in A, while B' is respectively the set of all the objects which have in common with each other the attributes in B. Furthermore, we define what a *formal concept* is:

Definition 3 (Formal Concept)
A pair (A,B) is a **formal concept** *of (G,M,I) if and only if*

$$A \subseteq G, B \subseteq M, A' = B \wedge A = B'$$

In other words, (A,B) is a **formal concept** if and only if the set of all attributes shared by the objects in A is identical with B and on the other hand A is also the set of all the objects which have in common with each other the attributes in B. A is then called the **extent** and B the **intent** of the concept (A,B). The concepts of a given context are naturally ordered by the **subconcept-superconcept relation** as defined by:

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 (\Leftrightarrow B_2 \subseteq B_1)$$

Thus, formal concepts are partially ordered with regard to inclusion of their extents or (which is equivalent) to inverse inclusion of their intent.
Thus, table 1 represents the incidence I of the formal context in form of a matrix. The corresponding sub-/superconcept partial order computed by FCA is depicted in figure 2 in form of a lattice. The representation makes use of **reduced labeling** as described in (Ganter and Wille, 1999) such that each object and each attribute is entered only once in the diagram. Finally, it is just left to clarify how we obtain a concept hierarchy, i.e. our partial order $\leq_C$ out of a lattice such as depicted in figure 2. We accomplish this by creating for each node in the lattice a concept labeled with the intent of the node as well as a subconcept of this concept for each element in the extent of that node. Furthermore, we also remove the bottom element of the lattice and preserve the other nodes and edges. Thus, in general we yield a partial order which can be represented as a DAG. In particular, for the lattice in figure 2 we yield the partial order in figure 1.

4 FCA meets NLP

The decisive question is now where to get from the objects as well as the corresponding attributes in order to create a taxonomy by using Formal Concept Analysis. A straightforward idea is to extract verb-object dependencies from texts and turn the objects' head into FCA objects and the corresponding verbs together with the postfix 'able' into attributes.
As already mentioned before, we concentrated on texts related to the tourism domain. In particular, we used two rather small corpora. The first corpus was acquired from *http://www.all-in-all.de/english/*, a web-page containing information about the history, cultural events, accommodation facilities, etc. of *Mecklenburg Vorpommern*, a region in north-east Germany. The second corpus is a collection of texts from *http://www.lonelyplanet.com/destinations/*. The total size of both corpora together was roughly about a million words.
In order to acquire the verb-object dependencies from these corpora, we used LoPar [2],a trainable and statistical left-corner parser. LoPar was trained on the corpora before actually parsing them. LoPar's output was then post-processed with tgrep to actually yield the desired dependencies, i.e. the verbs and the heads of the objects they subcategorize. It is important to mention that with our method we are also able to get multi-word terms. Furthermore, we use a simple method to lemmatize the extracted terms and verbs by looking up the lemma of each word in the lexicon provided with LoPar.
Regarding the output of the parser, it has to be taken into account that on the one hand it can be erroneous

[2]http://www.ims.uni-stuttgart.de/projekte/gramotron/SOFTWARE/LoPar-en.html

and on the other hand not all the verb-object dependencies produced are significant from a statistical point of view. Thus an important issue is actually to reduce the 'noise' produced by the parser before feeding the output into FCA. For this purpose, we calculate the overall probability P(n) that a certain (multi-word) term n appears as direct object of a verb, the overall probability P(v) for a certain verb v, the probability P(n,v) that n occurs as the head of v's object as well as the probability P($n|v$) that given a certain verb v, n appears in object position. Here are the corresponding formulas:

$$P(n) = \frac{f(n)}{\sum_{n' \in N} f(n')}$$

$$P(v) = \frac{f(v)}{\sum_{v' \in V} f(v')}$$

$$P(v,n) = \frac{f(v,n)}{\sum_{v' \in V} f(v')}$$

$$P(n|v) = \frac{f(n,v)}{f(v)}$$

where N and V are respectively the set of all terms appearing as direct objects of a verb and the set of all verbs with a direct object, $f(n)$ and $f(v)$ are respectively the number of occurrences of a term $n \in N$ as direct object and a verb $v \in V$ and $f(v,n)$ is the number of times that n occurs in the object position of v.

Now in order to weigh the significance of a certain verb-object pair (v,n), we used three different measures: a standard measure based on the conditional probability, the mutual information measure used in (Hindle, 1990), as well as a measure based on Resnik's *selectional preference strength* of a predicate (Resnik, 1997). Here are the formulas:

Standard: $P(n|v)$
Hindle: $log\ \frac{P(v,n)}{P(v)\ P(n)}$
Resnik: $P(n|v) * S_R(v)$

where the *selectional preference strength of a verb* is defined according to (Resnik, 1997):

$$S_R(v) = \sum_{n \in N} P(n|v)\ log\ \frac{P(n|v)}{P(n)}$$

Thus, the selectional preference of a verb is stronger the less frequent the nouns are that appear as its direct objects. In our approach, we then only consider those verb-term pairs as attribute-object pairs for which the values of the above measures are above some threshold t. For the Formal Concept Analysis we use the *Concepts* tool downloadable from *http://www.fcahome.org.uk/*. The processing time for building the lattice was for all thresholds and measures in the worst case 12 seconds. Thus, the FCA processing time can certainly be neglected in comparison to the parsing time.

5 Evaluation

Before actually presenting the results of our evaluation, we first have to describe the task we are evaluating against. Basically, the task can be described as follows: given a set of m concepts relevant for a certain domain, order these concepts hierarchically in form of a taxonomy. Certainly, this is not a trivial task and as shown in (Maedche and Staab, 2002) human agreement on such a task has also its limits.

In order to evaluate our automatically generated taxonomies, we compare them with the tourism domain ontology developed within the GETESS project[3]. This ontology was primarily designed for the analysis of german texts, but also english labels are available for many of the concepts. Moreover, we manually added the english labels for those concepts whose german label has an english counterpart. As a result we yielded an ontology consisting of 1026 concepts with most of them (>95%) having an english label.

Certainly, it is not clear how two ontologies (as a whole) can be compared to each other in terms of similarity. In fact, the only work in this direction the authors are aware of is the one in (Maedche and Staab, 2002). There, ontologies are seen as a semiotic sign system and compared at a syntactic, i.e. lexical, as well as semantic level. In this line,

[3] http://www.getess.de/index_en.html

we present a comparison based on lexical overlap as well as taxonomic similarity between ontologies. Lexical overlap (LO) of two ontologies O_1 and O_2 will be measured as the recall of the lexicon L_{O_1} compared to the lexicon L_{O_2}, i.e.

$$LO(O_1, O_2) = \frac{|L_{O_1} \bigcap L_{O_2}|}{|L_{O_2}|}$$

In order to compare the taxonomy of the ontologies, we use the *semantic cotopy* (SC) presented in (Maedche and Staab, 2002). The semantic cotopy of a concept is defined as the set of all its super- and subconcepts:

$$SC(c_i, \leq_C) := \{c_j | c_i \leq_C c_j \vee c_j \leq_C c_i\},$$

where $c_j, c_i \in C$. Now, we also extend the definition to operate on sets of concepts:

$$SC(C', \leq_C) := \bigcup_{c' \in C'} SC(c', \leq_C)$$

On the basis of this definition we introduce the *common lexical cotopy* (CLC) of a lexical entry $l \in L_{O_{1,2}} = L_{O_1} \cap L_{O_2}$ with regard to O_1 and O_2 as follows:

$$CLC(l, O_1, O_2) = \bigcup_{c \in SC(F^{-1}_{O_1}(l))} F_{O_1}(c) \cap L_{O_{1,2}}$$

Taxonomic overlap $\overline{TO}$ is now defined as follows[4]:

$$\overline{TO}(O_1, O_2) = \frac{1}{|L_{O_{1,2}}|} \sum_{l \in L_{O_{1,2}}} TO(l, O_1, O_2),$$

$$TO(l, O_1, O_2) = \frac{|CLC(l,O_1,O_2) \bigcap CLC(l,O_2,O_1)|}{|CLC(l,O_1,O_2) \bigcup CLC(l,O_2,O_1)|}$$ [5]

In order to compare the performance of our approach with the human performance on the task, we will interpret our results with regard to the study presented in (Maedche and Staab, 2002). In this study, five subjects were asked to model a taxonomy on the basis of 310 lexical entries relevant for the tourism domain. The taxonomic overlap ($\overline{TO}$) between the manually engineered ontologies reached from 47% to 87% with an average of 56.35%. Thus, it is clear that any automatic approach to derive a conceptual hierarchy between a set of concepts has definitely its limits.

[4] Here we assume that $|L_{O_{1,2}}| > 0$; otherwise $\overline{TO}$ will be 0.

[5] The reader is referred to (Maedche and Staab, 2002) for some concrete examples for these measures.

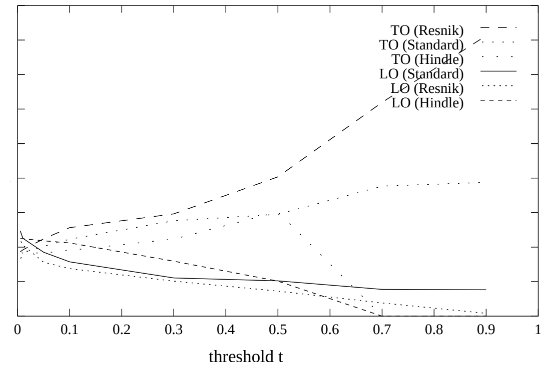

Figure 3: LO and $\overline{TO}$ values over threshold t

5.1 Results

We generated different taxonomies with the approach described in section 4 by using the measures *Standard*, *Resnik* and *Hindle* as well as different values for the threshold t. In particular, we used the values $t \in \{0.005, 0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9\}$. The size of these ontologies ranged from 0 to 7874 concepts. We then compared these 24 (=3x8) automatically generated taxonomies against the taxonomy of the GETESS ontology in terms of *lexical overlap* (LO) and *taxonomic overlap* $\overline{TO}$. The results of this comparison are depicted in figure 3. The best results in terms of $\overline{TO}$ are certainly achieved by the *Resnik* measure having always the highest $\overline{TO}$ with a small LO between 22.02% ($t = 0.005$) and 0.8% ($t = 0.9$). Interestingly, the *Standard* measure shows a worse $\overline{TO}$ than the *Resnik* measure, but a more stable LO between 24.71% ($t = 0.005$) and 7.65% ($t = 0.9$) The performance of the *Hindle* measure in terms of LO and $\overline{TO}$ is certainly the worst due to the fact that LO falls down to 0 at $t = 0.7$. Though these results don't allow much conclusions to be drawn, three interesting observations can be made. First, the lexical overlap seems to be very low in general, a problem which seems definitely necessary to overcome in order to yield also higher $\overline{TO}$ values. The second observation is that the higher the threshold t, the lower the lexical overlap and the higher the taxonomic overlap get. This is on the one hand clearly due to the fact that the higher the threshold t is, the less pairs (n, v) we feed into FCA and thus the amount of lexical categories in the taxonomy decreases. On the other hand it seems

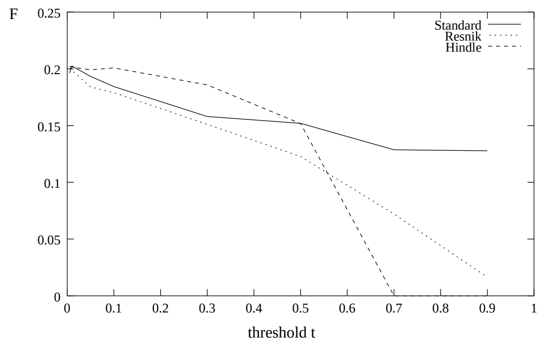

Figure 4: F-Measure (LoPar)

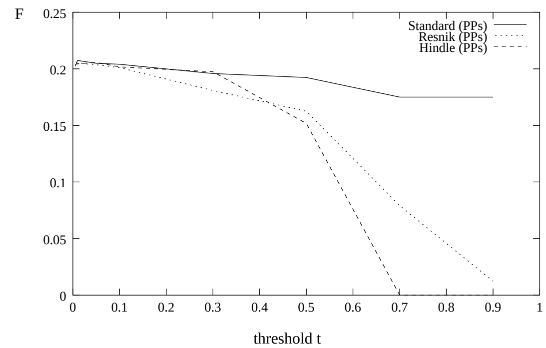

Figure 5: F-Measure (LoPar with PPs)

intuitive that it is easier to organize hierarchically fewer classes, which would explain the increasing $\overline{TO}$ values. This leads us to the third observation, i.e. that it is important to balance $\overline{TO}$ and LO against each other. In line with the information extraction (IE) task, in which *precision* and *recall* have to be balanced against each other, we compute also the F-Measure balancing LO and $\overline{TO}$, i.e. $F = \frac{2*\overline{TO}*LO}{\overline{TO}+LO}$. The results of the F-Measures for the three measures and the different threshold values are depicted in figure 4. The three measures achieve more or less the same top results. However, the *Standard* measure definitely shows the most stable behavior. It achieves a $\overline{TO}$ of 18.39% and a LO of 22.54% and thus a F-Measure of F=20.25% at $t = 0.01$. The best result of the *Resnik measure* (F=20.18%) corresponds to a $\overline{TO}$ 18.62% of and a LO of 22.02% at $t = 0.005$. The *Hindle* measure achieves an F-Measure of F=20.22% with a $\overline{TO}$ of 18.40% and a LO of 22.44% at $t = 0.005$. Certainly, the major bottleneck of the approach seems to be the low lexical overlap between the ontologies.

5.2 Adding Prepositional Phrases (PPs)

The results in the previous section lead us to investigate if we could increase the lexical overlap of the automatically generated taxonomies by extracting additionally verb-PP pairs and feeding them into FCA in the same way as the verb-object pairs with the only exception that instead of adding to the verb the postfix “able” we add the postfix “_<PREPOSITION>” The results in terms of F-Measures over the different thresholds in fact show that there was a slight increase in overall performance (compare figure 5). The best result (F=20.74%) is again achieved by the *Standard* measure with a $\overline{TO}$ of 16.57% and a LO of 27.71% at a threshold of $t = 0.01$. The second best result was reached by the *Resnik* measure with a $\overline{TO}$ of 16.61% and a LO of 27.20% (and thus F=20.63%) at a threshold of $t = 0.005$. The *Hindle* measure reached a F-Measure of F=20.56% corresponding to a $\overline{TO}$ of 17.07% and a LO of 25.85% at $t = 0.05$. It seems that we have in fact increased the lexical overlap at the cost of a decrease in taxonomic overlap. However, the overall results as indicated by the F-Measures have definitely increased.

5.3 Chunking

As final experiment we substituted LoPar by Steven Abney’s chunker CASS (Abney, 1996) in the hope that due to its robustness the lexical overlap would increase even more. For this purpose we used a straightforward heuristic and interpreted every NP and PP following a verb respectively as its direct object and PP-complement. The resulting F-measures are depicted in figure 6. Interestingly, the *Standard* and *Resnik* measures seem to perform more or less the same. The best result (F=20.21%) is achieved by the *Resnik* measure with a $\overline{TO}$ 14.08% of and a LO of 34.33% at a threshold $t = 0.05$. The *Standard* measure achieves the second best result of F=20.17% at $t = 0.1$ corresponding to a $\overline{TO}$ of 14.89% and an LO of 31.13%. The best result of the *Hindle* measure is F=20.16% with a $\overline{TO}$ of 14.93% and an LO of 31.02% at $t = 0.5$. So the results are more or less comparable to those produced by LoPar without taking into account PPs with the difference

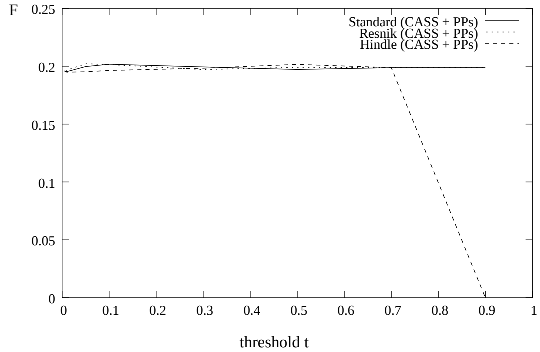

Figure 6: F-Measure (CASS)

that the lexical overlap has increased at the cost of a lower taxonomic overlap. This is certainly due to erroneous verb-object or verb-PP pairs yielded by postproceesing the chunker's output with the above mentioned heuristic.

5.4 Discussion of Results

The experiments have shown that the best results are achieved by using a parser such as LoPar to extract verb-object and verb-PP pairs as well as the *Standard* or *Resnik* measures to select the most significant pairs. However, is has also become clear that a chunker like CASS is an attractive alternative to using a parser as it leads to only slightly worse results but is actually much faster. Overall, the best result achieved is a F-Measure of 20.74% corresponding to a $\overline{TO}$ of 16.57% and an LO of 27.71%. In order to compare these results against human performance on the task, we first have to assess human performance in terms of the F-Measure. The assumption will be that if humans have to order hierarchically m terms they will always succeed in ordering all of them. In this sense, assuming for humans a LO of 100% and considering the average human agreement of 56.35% in terms of $\overline{TO}$ as given above, we yield a human average performance on the task of F=72.08% which we are still quite far away from. On the other hand, when comparing the manually engineered ontologies in (Maedche and Staab, 2002) with the GETESS ontology, we get F-measure values between F = 32.03% and F= 34.07%. In the light of these numbers, our results seem definitely more promising. Furthermore, the assumption of a human LO of 100% as well as the above average agreement may hold for relatively trivial domains such as tourism, but are certainly too optimistic for more technical domains such as for example biomedicine. Thus, we believe that the more specific and technical the underlying corpus is, the closer our approach will get to human performance. In the future, we hope to support this claim with further experiments on more technical domains.

6 Discussion of Related Work

In this section, we discuss some work related to the automatic acquisition of taxonomies out of texts. The early work of (Hindle, 1990) on noun classification from predicate-argument structures is very related to the approach presented here. Hindle's work is based on the distributional hypothesis, i.e. that nouns are similar to the extent that they share contexts. The central idea of his approach is that nouns may be grouped according to the extent to which they appear in similar verb frames. In particular, he takes into account nouns appearing as subjects and objects of verbs, but does not distinguish between them in his similarity measure. Our approach goes one step further in the sense that we do not only group nouns together, but also derive a hierarchical order between them.

Also very related to the work presented here is the approach of (Faure and Nedellec, 1998). Their work is also based on the distributional hypothesis and they present an iterative bottom-up clustering approach of nouns appearing in similar contexts. At each step, they cluster together the two most similar extents of some argument position of two verbs. However, their approach requires manual validation after each clustering step so that in our view it can not be called *unsupervised* or *automatic* anymore.

(Hahn and Schnattinger, 1998) aim at learning the correct ontological class for unknown words. For this purpose, when encountering an unknown word in a text they initially create one 'hypothesis space' for each concept the unknown word could actually belong to. These initial hypothesis spaces are then iteratively refined on the basis of evidence extracted from the linguistic context the unknown word appears in. In their approach, evidence is formalized in the form of quality labels attached to each hypothesis space. At the end the hypothesis space with maximal evidence with regard to the qualification calculus used is chosen as the correct ontological concept for the word in question.

Finally, (Hearst, 1992) aims at the acquisition of hyponym relations from Grolier's American Academic Encyclopedia. In order to identify these relations, she makes use of lexico-syntactic patterns manually acquired from her corpus. Hearst's approach is characterized by a high precision in the sense that the quality of the learned relations is very high. However, her approach suffers from a very low recall which is due to the fact that the patterns are very rare.

7 Conclusion and Further Work

We have presented a method for the automatic acquisition of taxonomies out of domain-specific text which is in line with the idea of a dynamic as well as corpus- and domain-specific lexicon as presented in (Buitelaar, 2000). The method presented is certainly adaptive as it only relies on generic NLP tools. Our approach would definitely profit from some form of smoothing. We are currently experimenting with an approach to cluster the verbs into classes before actually feeding the verb-object or verb-PP pairs into FCA. Moreover, we would like to apply our approach to larger corpora, other domains as well as other languages, in particular german. In order to overcome the low lexical recall of our approach, we think that crawling texts containing the appropriate words from the WWW as presented in (Agirre et al., 2000) is a promising option. Finally, we also aim at learning other relations than taxonomic ones. For this purpose we envision an approach as described in (Resnik, 1997) in order to learn relations at the right level of abstraction with regard to our automatically acquired taxonomy. In general, we believe that a combination approach of different methodologies is the key towards automatically generating acceptable and reasonable taxonomies which can be used as a starting point for applications and then be refined during their life-cycle.

Acknowledgments Philipp Cimiano is currently supported by the IST-Dot.Kom project (http://www.dot-kom.org), sponsored by the EC as part of the framework V, (grant IST-2001-34038). We would like to acknowledge Martin Kavalec for kindly providing the Lonely Planet corpus. Thanks also to the two anonymous reviewers of the ATEM Workshop for their comments.

References

Steven Abney. 1996. Partial parsing via finite-state cascades. In *Proceedings of the ESSLLI '96 Robust Parsing Workshop*.

E. Agirre, O. Ansa, E. Hovy, and D. Martinez. 2000. Enriching very large ontologies using the WWW. In *Proceedings of the Workshop on Ontology Construction of the ECAI*.

E. Bozsak, M. Ehrig, S. Handschuh, A. Hotho, A. Mädche, B. Motik, D. Oberle, C. Schmitz, S. Staab, L. Stojanovic, N. Stojanovic, R. Studer, G. Stumme, Y. Sure, J. Tane, R. Volz, and V. Zacharias. 2002. Kaon - towards a large scale semantic web. In *Proceedings of the Third International Conference on E-Commerce and Web Technologies (EC-Web 2002)*. Springer LNCS.

Paul Buitelaar. 2000. Semantic lexicons. In K. Simov and A. Kiryakov, editors, *Ontologies and Lexical Knowledge Bases*, pages 16–24.

F. Ciravegna. 2001. Adaptive information extraction from text by rule induction and generalization. In *Proceedings of tht 17th International Joint Conference on Artificial Intelligence (IJCAI 2001)*.

D. Faure and C. Nedellec. 1998. A corpus-based conceptual clustering method for verb frames and ontology. In P. Verlardi, editor, *Proceedings of the LREC Workshop on Adapting lexical and corpus resources to sublanguages and applications*.

B. Ganter and R. Wille. 1999. *Formal Concept Analysis – Mathematical Foundations*. Springer Verlag.

T.R. Gruber. 1993. Toward principles for the design of ontologies used for knowledge sharing. In *Formal Analysis in Conceptual Analysis and Knowledge Representation*. Kluwer.

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

U. Hahn and K. Schnattinger. 1998. Ontology engineering via text understanding. In *Proceedings of the 15th World Computer Congress 'The Global Information Society on the Way to the Next Millenium' (IFIP'98)*.

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

D. Hindle. 1990. Noun classification from predicate-argument structures. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 268–275.

A. Maedche and S. Staab. 2002. Measuring similarity between ontologies. In *Proceedings of the European Conference on Knowledge Acquisition and Management (EKAW)*. Springer.

Philip Resnik. 1997. Selectional preference and sense disambiguation. In *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How?*