Deriving Concept Hierarchies from Text by Smooth Formal Concept Analysis

Philipp Cimiano, Steffen Staab, 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 texts based on Formal Concept Analysis. Our approach is based on the assumption that verbs pose 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. In particular, we focus on the question if smoothing techniques have an influence on the quality of the generated concept hierarchies. We evaluate our approach 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. 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 the sense that the more people are involved in it, the harder it is to actually agree on a certain conceptualization. On the other hand, ontologies are only useful if a larger community agrees on them. This trade-off seems definitely difficult (if not impossible) to overcome ([Pinto and Martins, 2003]).

In this paper we present a novel method to automatically acquire taxonomies from texts based on Formal Concept Analysis (FCA), a method mainly used for the analysis of data ([Ganter and Wille, 1999]). In this context, we address in particular the question whether the classical data sparseness problem one typically encounters when working with corpora ([Manning and Schuetze, 1999]) can be overcome by a smoothing technique based on the conceptual clustering algorithm presented in [Faure and Nedellec, 1998]. In particular, we cluster nouns and verbs based on the *distributional hypothesis*, i.e. the assumption that nouns and verbs are similar to the extent to which they share contexts. In this line, the context of a noun is represented as a vector containing all the verbs the noun appears as direct object of and the context of a verb is symmetrically represented by a vector containing all the nouns appearing as its direct objects. On the basis of such a representation, similar nouns and verbs are pairwise clustered together and the joint frequencies in the corpus are updated accordingly. As a result, some previously unseen events may yield a non-zero frequency. After this smoothing step, the significant verb/term pairs are considered for FCA as attribute/object pairs in order to yield a taxonomy.

The main benefits of our method are its adaptivity as it can be applied to arbitrary texts 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 ([Stojanovic *et al.*, 2002]). This is in line with the domain and corpus specific form of lexicon as envisioned in [Buitelaar, 2000].

2 The General 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 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

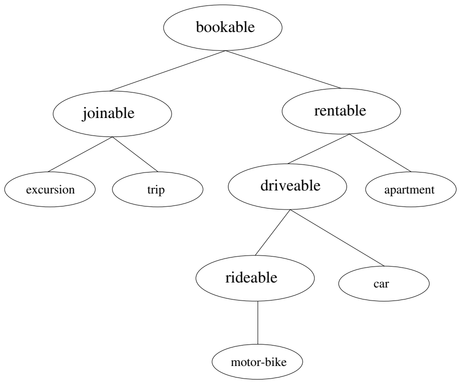

Figure 1: Hierarchy for the tourism example

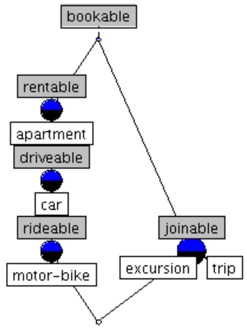

Figure 2: The tourism lattice

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 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*:

[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.

	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

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 :

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

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

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:

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

or which is equivalent:

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

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 computer 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 Smoothing FCA

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 head of the subcategorized objects into FCA objects and the corresponding verbs together with the postfix 'able' into FCA attributes.

As already mentioned before, we concentrated on the tourism domain. In particular, we used two rather small domain-specific corpora as well as one larger corpus of a general nature. The idea behind this decision is to get not only data about domain-specific terms but also make use of more general resources to get additional data about more general terms. The first domain-specific 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 domain-specific corpus is a collection of texts from *http://www.lonelyplanet.com/destinations/*. As general corpus we used the British National Corpus. The total size of the three corpora together was roughly over 100 million words. In order to acquire the verb-object dependencies from these corpora, we used LoPar, a trainable statistical left-corner parser ([Schmid, 2000]). From the parser's output we extracted the respective heads with *tgrep*[2] and lemmatized them. It is important to mention that by this method we also yield multi-word terms.

Regarding the postprocessed output of the parser, it has to be taken into account that on the one hand some verb/object dependencies can be spurious and that on the other hand not all the pairs 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 conditional probability that a certain (multi-word) term n appears as direct object of a certain verb v:

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

where $f(n,v)$ is the number of occurrences of a term n as object of a verb v and $f(v)$ is the number of occurrences of verb v with a direct object.

As mentioned already in the introduction, in order to overcome the typical data sparseness problem we make use of a word clustering method based on the conceptual clustering approach presented in [Faure and Nedellec, 1998]. There, nouns are clustered based on the assumption that they are similar to the extent to which they share contexts. In the approach presented here, we will consider as the context of a noun all the verbs v it appears with as direct object. As context for a verb we will consider all the nouns that appear as its direct objects. We will formalize these contexts respectively as sets C(n) for a noun n and C(v) for a verb v. As similarity measure we use the cosine metric, i.e. for two nouns n_1 and n_2 the similarity is calculated as follows:

$$sim(n_1,n_2) = \frac{\sum_{v \in C(n_1) \cap C(n_2)} P(n_1|v) * P(n_2|v)}{\sqrt{\sum_{v \in C(n_1)} P(n_1|v)^2 * \sum_{v \in C(n_2)} P(n_2|v)^2}}$$

For verbs v_1 and v_2 the similarity is symmetrically computed as follows:

$$sim(v_1,v_2) = \frac{\sum_{n \in C(v_1) \cap C(v_2)} P(n|v_1) * P(n|v_2)}{\sqrt{\sum_{n \in C(v_1)} P(n|v_1)^2 * \sum_{n \in C(v_2)} P(n|v_2)^2}}$$

Furthermore, for a noun n, the most similar noun n' is defined as follows:

$$similar(n) = max_{n'}\ sim(n,n')$$

and symmetrically also for verbs. As in [Faure and Nedellec, 1998], we also use an iterative clustering approach. In each iteration, the m most similar terms are clustered together, where only those pairs (n_1,n_2) are considered as potential candidates to be clustered which are mutually most similar to each other, i.e. $similar(n_1) = n_2$ and $similar(n_2) = n_1$. After each iteration - consisting in m pairwise clusterings - the joint frequencies for verbs and nouns are recalculated and the conditional probabilities are updated. Then, in a bootstrapping fashion, these updated probabilities are considered in the next clustering iteration. It is important to mention that in our approach nouns and verbs are clustered alternately at each iteration. In particular, odd iterations will correspond to noun clustering phases and even iterations will correspond to verb clustering phases. Furthermore, it is also important to mentioned that previously pairwise clustered pairs are stored such that no pair is clustered more than once.

Finally, we then only consider those verb/term pairs (v,t) as attribute-object pairs for FCA for which the conditional probability $P(t|v)$ is above some threshold t after the clustering iteration i. 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 below 10 seconds. Thus, the FCA processing time can certainly be neglected in comparison to the parsing and clustering 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 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 gold standard used in the study presented in [Maedche and Staab, 2002]. For this purpose, we translated the concepts of this gold standard from german to english thus yielding

[2] http://mccawley.cogsci.uiuc.edu/corpora/treebank3.html

a reference ontology consisting of 300 concepts with most of them (>95%) having also an english label. In what follows, we will refer to this reference ontology simply as *gold standard*.

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, 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.

$$LO(O_1, O_2) = \frac{|L_{O_1} \bigcap L_{O_2}|}{|L_{O_2}|}$$ [3]

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 the taxonomic overlap of two ontologies O_1 and O_2 is computed as follows:

$$\overline{TO}(O_1, O_2) = \frac{1}{|C_1|} \sum_{c \in C_1} TO(c, O_1, O_2)$$

where

$$TO(c, O_1, O_2) := \begin{cases} TO'(c, O_1, O_2) & if c \in C_2 \\ TO''(c, O_1, O_2) & if c \notin C_2 \end{cases}$$

and TO' and TO" are defined as follows:

$$TO'(c, O_1, O_2) := \frac{SC(c, \leq_C C_1) \cap SC(c, \leq_C C_2)}{SC(c, \leq_C C_1) \cup SC(c, \leq_C C_2)}$$

$$TO''(c, O_1, O_2) := max_{c' \in C_2} \frac{SC(c, \leq_C C_1) \cap SC(c', \leq_C C_2)}{SC(c, \leq_C C_1) \cup SC(c', \leq_C C_2)}$$

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, four 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.

5.1 Results

We generated different concept hierarchies with the approach described in section 4 by using different values for the threshold parameter t. In particular, we used the values $t \in \{0.005, 0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9\}$. Furthermore, we experimented with up to 11 clustering iterations with the maximal number of terms being clustered at each of them set to 5. The reason for using up to 11 iterations is that at the 12th iteration no more mutually similar pairs were found. Figure 3 shows the nouns and verbs which were pairwise clustered at each of the iterations. The odd numbers correspond to noun clustering phases while the even numbers correspond to verb clustering phases. The values in parenthesis indicate the average mutual distance (sim) between the two nouns. Obviously, there are also spurious pairs, but in general the clustering seems reasonable. We then compared these 88 (11x8) automatically generated taxonomies against the taxonomy of the *gold standard* in terms of an F-Measure balancing *lexical overlap* (LO) and *taxonomic overlap* $\overline{TO}$ against each other:

$$F = \frac{2 * \overline{TO} * LO}{\overline{TO} + LO}$$

Figures 4-13 show the results of this comparison against the baseline without clustering in terms of the *F-Measure* for the different number of iterations as well as the different threshold values for t. It is interesting to observe that there are increases in F-Measure only after iterations with an odd number, i.e. iterations corresponding to a noun clustering phase. However, it can not be concluded from this observation that the verb clustering has no influence on the results. In fact, as the joint frequencies of verb and nouns are updated after each iteration, the influence of the verb clustering could be an indirect one. Further experiments seem necessary to clarify this question. Table 2 presents the best results of each iteration. It shows that the best results are increasing nonmonotonically with each clustering iteration. Overall, the results with clustering are almost 1.5% better than the baseline without clustering.

	$\overline{TO}$	LO	F	t
Baseline	18.78%	19.44%	19.11%	0.7
Iteration 1	18.78%	19.44%	19.11%	0.7
Iteration 2	18.78%	19.44%	19.11%	0.7
Iteration 3	17.40%	22.25%	19.52%	0.3
Iteration 4	17.40%	22.25%	19.52%	0.3
Iteration 5	17.94%	21.41%	19.52%	0.5
Iteration 6	17.94%	21.41%	19.52%	0.5
Iteration 7	18.27%	21.80%	19.88%	0.5
Iteration 8	18.27%	21.80%	19.88%	0.5
Iteration 9	18.47%	22.65%	20.35%	0.5
Iteration 10	18.47%	22.65%	20.35%	0.5
Iteration 11	18.50%	23.00%	20.50%	0.5

Table 2: Best result of each iteration

5.2 Discussion of Results

It has become clear that the described smoothing method can in fact improve the overall results of the FCA-based approach to the automatic acquisition of concept hierarchies presented in this paper. In order to compare the results of the approach presented in this paper 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 a set of 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 very far away from.

On the other hand, the assumption of a human LO of 100%

[3] As the terms to be ordered hierarchically are given there is no need to measure the lexical precision.

Iteration 1:
whirlpool solarium (0.87)
ballroom lounge (0.72)
cinema theatre (0.76)
badminton billiard (0.82)
swimming kitchenette (0.92)

Iteration 2:
know think (0.70)
leave enter (0.77)
love like (0.66)
work go (0.79)
keep open (0.75)

Iteration 3:
yacht boat (0.25)
club party (0.11)
telephone terrace (0.11)
pension musical (0.12)
drier radio (0.14)

Iteration 4:
put hold (0.44)
provide offer (0.55)
watch saw (0.65)
move turn (0.45)
follow hear (0.64)

Iteration 5:
person family (0.09)
tent caravan (0.06)
park kiosk (0.09)
masseur basilica (0.09)
animation booking (0.09)

Iteration 6:
start enjoy (0.37)
require involve (0.43)
contain include (0.35)
consider describe (0.40)

Iteration 7:
gallery shower (0.04)
period daytime (0.06)
time distance (0.06)
autumn cure (0.05)
agreement situation (0.05)

Iteration 8:
have remember (0.63)

Iteration 9: night bank (0.03)
organization room (0.02)
group place (0.04)
view law (0.02)
city sight (0.02)

Iteration 10:
no mutually similar verbs found

Iteration 11:
bicycle bed (0.01)
station service (0.01)

Figure 3: Results of the clustering

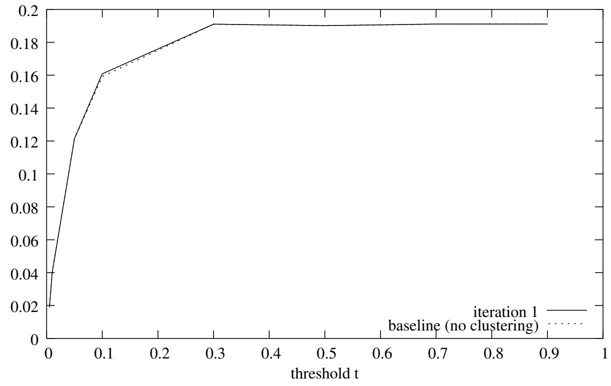

Figure 4: Iteration 1

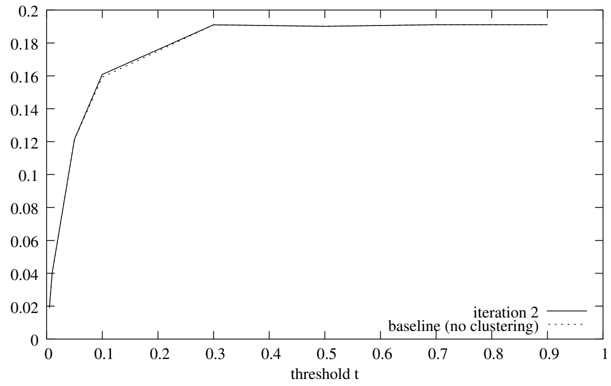

Figure 5: Iteration 2

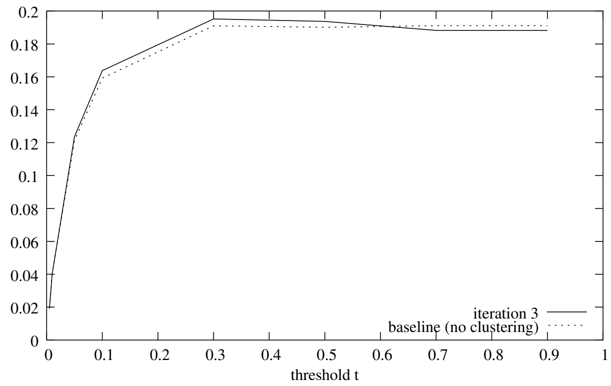

Figure 6: Iteration 3

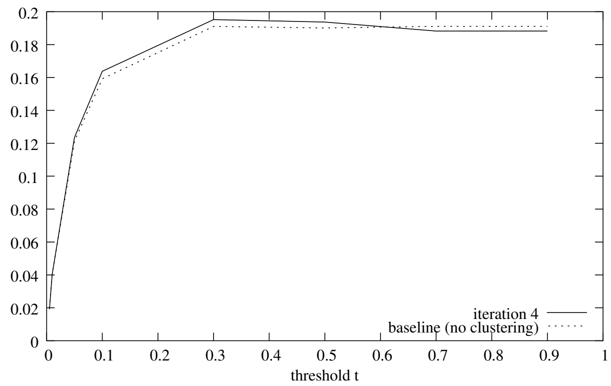

Figure 7: Iteration 4

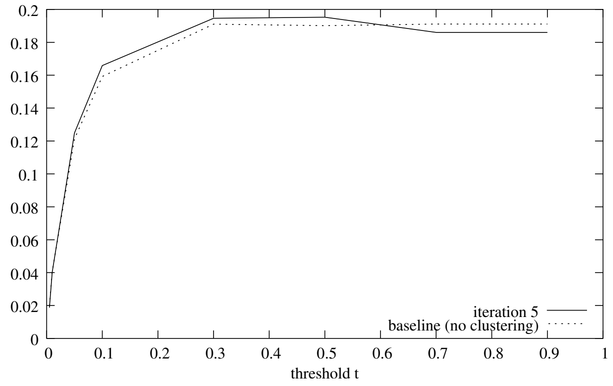

Figure 8: Iteration 5

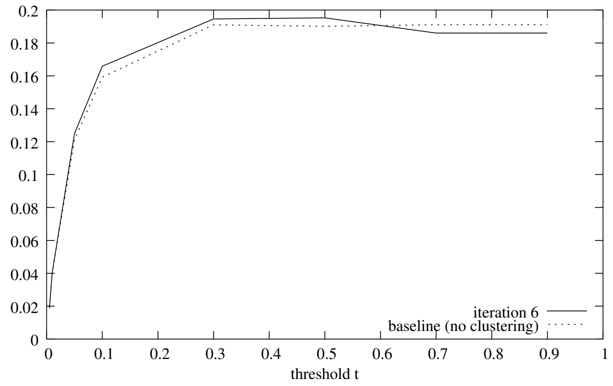

Figure 9: Iteration 6

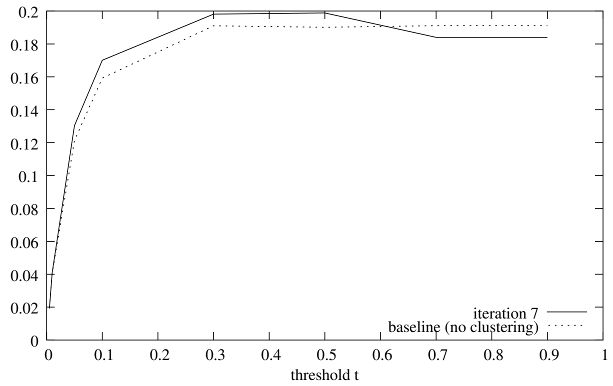

Figure 10: Iteration 7

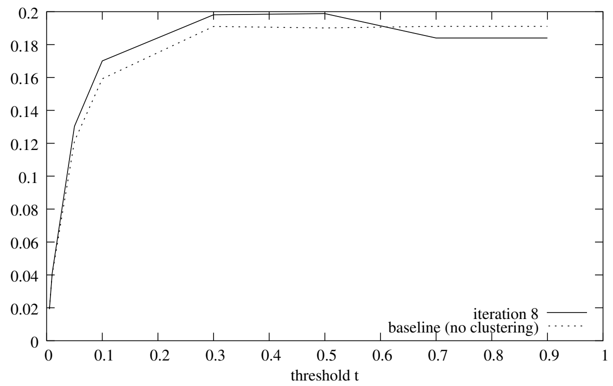

Figure 11: Iteration 8

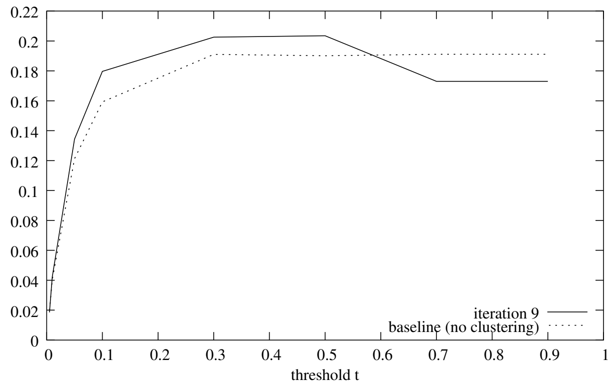

Figure 12: Iteration 9

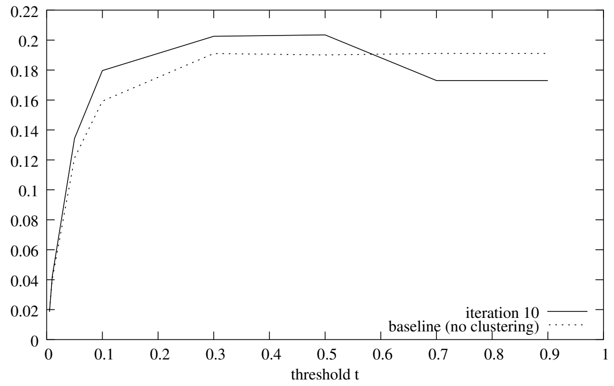

Figure 13: Iteration 10

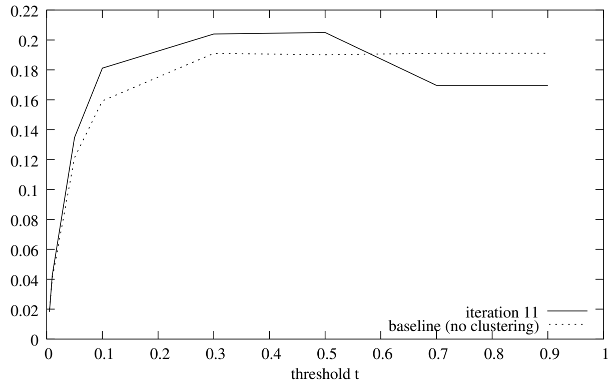

Figure 14: Iteration 11

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 bio-medicine. 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 as ours 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.

[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. Thus it is doubtful that her approach would be scalable to full text articles such as considered here.

[Pereira *et al.*, 1993] present a top-down clustering approach to build an unlabeled hierarchy of nouns. As in our approach, they also make use of verb-object relations to represent the context of a certain noun. In contrast to Pereira et al.'s work, in our approach the clustering is only used as a method in order to overcome data sparseness, while the hierarchy is then built at a second step on the basis of FCA.

[Schulte im Walde, 2000] approaches the clustering of verbs based on syntactic subcatgorization frames. Her model allows to account for polysemy as verbs can be assigned to different clusters. Interestingly, her clustering results get worse when adding selectional restrictions to the subcategorization frames.

[Basili *et al.*, 1993] also address the problem of clustering verbs with the aim of building a verb hierarchy. For this purpose they make use of an incremental conceptual clustering algorithm which does not only consider syntactic relations but manually assigned selectional restrictions. Their algorithm is able to consider multiple instances of a given verb such that it can in principle account for polysemous verbs.

[Bisson *et al.*, 2000] present an interesting framework and a corresponding workbench - Mo'K - allowing users to design conceptual clustering methods to assist them in an ontology building task. The framework is general enough to integrate different clustering methods. Thus, it would be certainly interesting to clarify if the approach presented in this paper is compatible with their framework.

[Velardi *et al.*, 2001] present the OntoLearn system which discovers i) the domain concepts relevant for a certain domain, i.e. the relevant terminology, ii) named entities, iii) 'vertical' (is-a or taxonomic) relations as well iv) as certain relations between concepts based on specific syntactic relations. In their approach a 'vertical' relation is established between a term t_1 and a term t_2, i.e. is-a(t_1,t_2), if the head of t_2 matches the head of t_1 and additionally the former is additionally modified in t_1. Thus, a 'vertical' relation is for example established between the term 'international credit card' and the term 'credit card', i.e. is-a(international credit card,credit card). This approach is certainly very simple and could be complemented by the one presented in this paper.

[Caraballo, 1999] also uses clustering methods to derive an unlabeled hierarchy of nouns by using data on conjunctions of nouns and appositive constructs collected from the Wall Street Journal corpus. Interestingly, at a second step she also labels the abstract concepts of the hierarchy by considering the Hearst patterns in which the children of the concept in question appear as hyponyms. The most frequent hypernym is then chosen in order to label the concept. Furthermore, at a further step she also compresses the produced ontological tree by eliminating internal nodes without a label. The final ontological tree is then evaluated by presenting a random choice of clusters and the corresponding hypernym to three human judges for validation.

[Sanderson and Croft, 1999] describe an interesting approach to automatically derive a hierarchy by considering the document a certain term appears in as context. In particular, they present a document-based definition of subsumption according to which a certain term t_1 is more special than a term t_2 if t_2 also appears in all the documents in which t_1 appears.

[Hahn and Schanttinger, 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.

[Maedche *et al.*, 2002] present a classification approach based on a combination of the *k nearest neighbor* (kNN) method and the tree-ascending classification algorithm in order to learn the appropriate class in a given thesaurus for a certain term. In [Maedche and Staab, 2000], they moreover describe an approach to learning generic binary relations between terms with regard to given a certain taxonomy by computing association rules.

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. From this point of view all the above approaches are definitely related to ours.

7 Conclusion and Further Work

We have presented a method for the automatic acquisition of taxonomies out of text which is in line with the idea of a dynamic and corpus-specific lexicon as presented in [Buitelaar, 2000]. In future work, we would like to apply our approach to other domains as well as other languages, in particular german. On the other hand, our approach suffers from the fact that it completely neglects polysemy. Further work will thus address the question if polysemy can be taken into account in some way. Further work will also consider other similarity measures as well as clustering techniques. 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.

Acknowledgments This work was carried out within the IST-dot.kom project (http://www.dot-kom.org), sponsored by the European Commission as part of the framework V, (grant IST-2001-34038). dot.kom involves the University of Sheffield (UK), ITC-Irst (I), Ontoprise (D), the Open University (UK), Quinary (I) and the University of Karlsruhe (D) . Its objectives are to develop Knowledge Management and Semantic Web methodologies based on Adaptive Information Extraction from Text. We would also like to thank Martin Kavalec for kindly providing the Lonely Planet Corpus.

References

[Basili *et al.*, 1993] R. Basili, M.T. Pazienza, and P. Velardi. Hierarchical clustering of verbs. In *Proceedings of the Workshop on Acquisition of Lexical Knowledge from Text*, 1993.

[Bisson *et al.*, 2000] G. Bisson, C. Nedellec, and L. Canamero. Designing clustering methods for ontology building - The Mo'K workbench. In *Proceedings of the ECAI Ontology Learning Workshop*, 2000.

[Bozsak *et al.*, 2002] E. Bozsak, M. Ehrig, S. Handschuh, A. Hotho, A. Maedche, 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. Kaon - towards a large scale semantic web. In *Proceedings of the Third International Conference on E-Commerce and Web Technologies (EC-Web 2002)*. Springer Lecture Notes in Computer Science, 2002.

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

[Caraballo, 1999] S.A. Caraballo. Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguists*, pages 120–126, 1999.

[Faure and Nedellec, 1998] D. Faure and C. Nedellec. 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*, 1998.

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

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

[Hahn and Schanttinger, 1998] U. Hahn and K. Schanttinger. Towards text knowledge engineering. In *AAAI'98/IAAI'98 Proceedings of the 15th National Conference on Artificial Intelligence and the 10th Conference on Innovative Applications of Artificial Intelligence*, 1998.

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

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

[Maedche and Staab, 2000] A. Maedche and S. Staab. Mining ontologies from text. In *Proceedings of the European Conference on Artificial Intelligence (ECAI 2000)*, 2000.

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

[Maedche *et al.*, 2002] A. Maedche, V. Pekar, and S. Staab. Ontology learning part one - on discovering taxonomic relations from the web. In *Web Intelligence*. Springer, 2002.

[Manning and Schuetze, 1999] C. Manning and H. Schuetze. *Foundations of Statistical Language Processing*. MIT Press, 1999.

[Pereira *et al.*, 1993] F. Pereira, N. Tishby, and L. Lee. Distributional clustering of english words. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pages 183–190, 1993.

[Pinto and Martins, 2003] H.S. Pinto and J.P. Martins. Ontologies: How can they be built? *Knowledge and Information Systems*, 2003.

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

[Sanderson and Croft, 1999] M. Sanderson and B. Croft. Deriving concept hierarchies from text. In *Research and Development in Information Retrieval*, pages 206–213. 1999.

[Schmid, 2000] Helmut Schmid. Lopar: Design and implementation. In *Arbeitspapiere des Sonderforschungsbereiches 340*, number 149. 2000.

[Schulte im Walde, 2000] S. Schulte im Walde. Clustering verbs semantically according to their alternation behaviour. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING-00)*, 2000.

[Stojanovic *et al.*, 2002] L. Stojanovic, N. Stojanovic, and A. Maedche. Change discovery in ontology-based knowledge management systems. In *Proceedings of the 2nd International Workshop on Evolution and Change in Data Management (ECDM 2002)*, 2002.

[Velardi *et al.*, 2001] P. Velardi, P. Fabriani, and M. Missikoff. Using text processing techniques to automatically enrich a domain ontology. In *Proceedings of the ACM International Conference on Formal Ontology in Information Systems*, 2001.