Deriving a Large Scale Taxonomy from Wikipedia

Simone Paolo Ponzetto and **Michael Strube**
EML Research gGmbH
Schloss-Wolfsbrunnenweg 33
69118 Heidelberg, Germany
http://www.eml-research.de/nlp

Abstract

We take the category system in Wikipedia as a conceptual network. We label the semantic relations between categories using methods based on connectivity in the network and lexico-syntactic matching. As a result we are able to derive a large scale taxonomy containing a large amount of subsumption, i.e. *isa*, relations. We evaluate the quality of the created resource by comparing it with ResearchCyc, one of the largest manually annotated ontologies, as well as computing semantic similarity between words in benchmarking datasets.

Introduction

The availability of large coverage, machine readable knowledge is a crucial theme for Artificial Intelligence. While advances towards robust statistical inference methods (cf. e.g. Domingos et al. (2006) and Punyakanok et al. (2006)) will certainly improve the computational modeling of intelligence, we believe that crucial advances will also come from rediscovering the deployment of large knowledge bases.

Creating knowledge bases, however, is expensive and they are time-consuming to maintain. In addition, most of the existing knowledge bases are domain dependent or have a limited and arbitrary coverage – Cyc (Lenat & Guha, 1990) and WordNet (Fellbaum, 1998) being notable exceptions. The field of ontology learning deals with these problems by taking textual input and transforming it into a taxonomy or a proper ontology. However, the learned ontologies are small and mostly domain dependent, and evaluations have revealed a rather poor performance (see Buitelaar et al. (2005) for an extensive overview).

We try to overcome such problems by relying on a wide coverage online encyclopedia developed by a large number of users, namely Wikipedia. We use semi-structured input by taking the category system in Wikipedia as a conceptual network. This provides us with pairs of related concepts whose semantic relation is unspecified. The task of creating a subsumption hierarchy then boils down to distinguish between *isa* and *notisa* relations. We use methods based on connectivity in the network and lexico-syntactic patterns to label the relations between categories. As a result we are able to derive a large scale taxonomy.

Motivation

Arguments for the necessity of symbolically encoded knowledge for AI date back at least to McCarthy (1959). Such need has become clearer throughout the last decades, as it became obvious that AI subfields such as information retrieval, knowledge management, and natural language processing (NLP) all profit from machine accessible knowledge (see Cimiano et al. (2005) for a broader motivation). E.g., from a computational linguistics perspective, knowledge bases for NLP applications should be:

- *domain independent*, i.e. have a large coverage, in particular at the instance level;
- *up-to-date*, in order to process current information;
- *multilingual*, in order to process information in a language independent fashion.

The Wikipedia categorization system satisfies all these points. Unfortunately, the Wikipedia categories do not form a taxonomy with a fully-fledged subsumption hierarchy, but only a thematically organized thesaurus. As an example, the category CAPITALS IN ASIA[1] is categorized in the upper category CAPITALS (*isa*), whereas a category such as PHILOSOPHY is categorized under ABSTRACTION and BELIEF (*deals-with*?) as well as HUMANITIES (*isa*) and SCIENCE (*isa*). Another example is a page such as EUROPEAN MICROSTATES which belongs to the categories EUROPE (*are-located-in*) and MICROSTATES (*isa*).

Related Work

There is a large body of work concerned with acquiring knowledge for AI and NLP applications. Many NLP components can get along with rather unstructured, associative knowledge as provided by the cooccurence of words in large corpora, e.g., distributional similarity (Church & Hanks,

[1] We use Sans Serif for words and queries, CAPITALS for Wikipedia pages and SMALL CAPS for Wikipedia categories.

1990; Lee, 1999; Weeds & Weir, 2005, inter alia) and vector space models (Schütze, 1998). Such unlabeled relations between words proved to be as useful for disambiguating syntactic and semantic analyses as the manually assembled knowledge provided by WordNet.

However, the availability of reliable preprocessing components like POS taggers, syntactic and semantic parsers allows the field to move towards higher level tasks, such as question answering, textual entailment, or complete dialogue systems which require to *understand* language. This lets researchers focus (again) on taxonomic and ontological resources. The manually constructed Cyc ontology provides a large amount of domain independent knowledge. However, Cyc cannot (and is not intended to) cope with most specific domains and current events. The emerging field of ontology learning tries to overcome these problems by learning (mostly) domain dependent ontologies from scratch. However, the generated ontologies are relatively small and the results rather poor (e.g., Cimiano et al. (2005) report an F-measure of about 33% with regard to an existing ontology of less than 300 concepts). It seems to be more promising to extend existing resources such as Cyc (Matuszek et al., 2005) or WordNet (Snow et al., 2006). The examples shown in these works, however, seem to indicate that the extension takes place mainly with respect to named entities, a task which is arguably not as difficult as creating a complete (domain-) ontology from scratch.

Another approach for building large knowledge bases relies on input by volunteers, i.e., on collaboration among the users of an ontology (Richardson & Domingos, 2003). However, the current status of the *Open Mind* and *MindPixel* projects[2] does indicate that they are largely academic enterprises. Similar to the *Semantic Web* (Berners-Lee et al., 2001), where users are supposed to explicitly define the semantics of the contents of web pages, they may be hindered by too high an entrance barrier. In contrast, Wikipedia and its categorization system feature a low entrance barrier achieving quality by collaboration. In Strube & Ponzetto (2006) we proposed to take the Wikipedia categorization system as a semantic network which served as basis for computing the semantic relatedness of words. In the present work we develop this idea a step further by automatically assigning *isa* and *notisa* labels to relations between the categories. That way we are able to compute the semantic similarity between words instead of their relatedness.

Methods

Since May 2004 Wikipedia allows for structured access by means of *categories*[3]. The categories form a graph which can be taken to represent a conceptual network with unspecified semantic relations (Strube & Ponzetto, 2006). We present here our methods to derive *isa* and *notisa* relations from these generic links.

[2] www.openmind.org and www.mindpixel.com

[3] Wikipedia can be downloaded at http://download.wikimedia.org. In our experiments we use the English Wikipedia database dump from 25 September 2006. This includes 1,403,207 articles, 99% of which are categorized.

Category network cleanup (1)

We start with the full categorization network consisting of 165,744 category nodes with 349,263 direct links between them. We first clean the network from meta-categories used for encyclopedia management, e.g. the categories under WIKIPEDIA ADMINISTRATION. Since this category is connected to many content bearing categories, we cannot remove this portion of the graph entirely. We remove instead all those nodes whose labels contain any of the following strings: wikipedia, wikiprojects, lists, mediawiki, template, user, portal, categories, articles, pages. This leaves 127,325 categories and 267,707 links still to be processed.

Refinement link identification (2)

The next preprocessing step includes identifying so-called *refinement links*. Wikipedia users tend to organize many category pairs using patterns such as Y X and X BY Z (e.g. MILES DAVIS ALBUMS and ALBUMS BY ARTIST). We label these patterns as expressing *is-refined-by* semantic relations between categories. While these links could be in principle assigned a full *isa* semantics, they represent meta-categorization relations, i.e., their sole purpose is to better structure and simplify the categorization network. We take all categories containing by in the name and label all links with their subcategories with an *is-refined-by* relation. This labels 54,504 category links and leaves 213,203 relations to be analyzed.

Syntax-based methods (3)

The first set of processing methods to label relations between categories as *isa* is based on string matching of syntactic components of the category labels.

Head matching. The first method labels pairs of categories sharing the same lexical head, e.g. BRITISH COMPUTER SCIENTISTS *isa* COMPUTER SCIENTISTS. We parse the category labels using the Stanford parser (Klein & Manning, 2003). Since we parse mostly NP fragments, we constrain the output of the head finding algorithm (Collins, 1999) to return a lexical head labeled as either a noun or a 3rd person singular present verb (this is to tolerate errors where plural noun heads have been wrongly identified as verbs). In addition, we modify the head finding rules to return both nouns for NP coordinations (e.g. both buildings and infrastructures for BUILDINGS AND INFRASTRUCTURES IN JAPAN). Finally, we label a category link as *isa* if the two categories share the same head lemma, as given by a finite-state morphological analyzer (Minnen et al., 2001).

Modifier matching. We next label category pairs as *notisa* in case the stem of the lexical head of one of the categories, as given by the Porter stemmer (Porter, 1980), occurs in non-head position in the other category. This is to rule out thematic categorization links such as CRIME COMICS and CRIME or ISLAMIC MYSTICISM and ISLAM.

Both methods achieve a good coverage by identifying respectively 72,663 *isa* relations by head matching and 37,999 *notisa* relations by modifier matching.

1. *NP2,?* (such as|like|, especially) *NP* NP1*
   a stimulant such as caffeine
2. such *NP2* as *NP* NP1*
   such stimulants as caffeine
3. *NP1 NP** (and|or|,like) other *NP2*
   caffeine and other stimulants
4. *NP1*, one of *det_pl NP2*
   caffeine, one of the stimulants
5. *NP1*, *det_sg NP2 rel_pron*
   caffeine, a stimulant which
6. *NP2* like *NP* NP1*
   stimulants like caffeine

1. *NP2*'s *NP1*
   car's engine
2. *NP1* in *NP2*
   engine in the car
3. *NP2* with *NP1*
   a car with an engine
4. *NP2* contain(s|ed|ing) *NP1*
   a car containing an engine
5. *NP1* of *NP2*
   the engine of the car
6. *NP1* are? used in *NP2*
   engines used in cars
7. *NP2* ha(s|ve|d) *NP1*
   a car has an engine

Figure 1: Patterns for *isa* and *notisa* Detection

Connectivity-based methods (4)

The next set of methods employed relies on the structure and connectivity of the categorization network.

Instance categorization. Suchanek et al. (2007) show that *instance-of* relations in Wikipedia between entities (denoted by pages) and classes (denoted by categories) can be found heuristically with high accuracy by determining whether the head of the page category is plural, e.g. ALBERT EINSTEIN belongs to the NATURALIZED CITIZENS OF THE UNITED STATES category. We apply this idea to *isa* relation identification as follows. For each category c,

1. we find the page titled as the category or its lemma, for instance the page MICROSOFT for the category MICROSOFT;
2. we then collect all the page's categories whose lexical head is a plural noun $CP = \{c_1, c_2, \ldots c_n\}$;
3. for each c's supercategory sc, we label the relation between c and sc as *isa*, if the head lemma of sc matches the head lemma of at least one category $cp \in CP$.

For instance, from the page MICROSOFT being categorized into COMPANIES LISTED ON NASDAQ, we collect evidence that Microsoft is a company and accordingly categorize as *isa* the links between MICROSOFT and COMPUTER AND VIDEO GAME COMPANIES. The idea is to collect evidence from the instance describing the concept and propagate such evidence to the described concept itself.

Redundant categorization. This method labels pairs of categories which have at least one page in common. If users redundantly categorize by assigning two directly connected categories to a page, they often mark by implicature the page as being an instance of two different category concepts with different granularities, e.g. ETHYL CARBAMATE is both an AMIDE(S) and an ORGANIC COMPOUND(S). Assuming that the page is an instance of both conceptual categories, we can conclude by transitivity that one category is subsumed by the other, i.e. AMIDES *isa* ORGANIC COMPOUNDS.

The connectivity-based methods provide positive *isa* links in cases where relations are unlikely to be found in free text. Using instance categorization and redundant categorization we find 9,890 and 11,087 *isa* relations, respectively.

Lexico-syntactic based methods (5)

After applying methods (1-4) we are left with 81,564 unclassified relations. We next apply lexico-syntactic patterns to sentences in large text corpora to identify *isa* relations (Hearst, 1992; Caraballo, 1999). In order to reduce the amount of unclassified relations and to increase the precision of the *isa* patterns we also apply patterns to identify *notisa* relations. We assume that patterns used for identifying meronymic relations (Berland & Charniak, 1999; Girju et al., 2006) indicate that the relation is not an *isa* relation. The text corpora used for this step are the English Wikipedia (5×10^8 words) and the Tipster corpus (2.5×10^8 words; Harman & Liberman (1993)). In the patterns for detecting *isa* and *notisa* relations (Figure 1) *NP1* represents the hyponym, *NP2* the hypernym, i.e., we want to retrieve *NP1 isa NP2*; *NP** represents zero or more coordinated NPs.

To improve the recall of applying these patterns, we use only the lexical head of the categories which were not identified as named entities. That is, if the lexical head of a category is identified by a Named Entity Recognizer (Finkel et al., 2005) as belonging to a named entity, e.g. Brands in YUM! BRANDS, we use the full category name, else we simply use the head, e.g. albums in MILES DAVIS ALBUMS. In order to ensure precision in applying the patterns, both the Wikipedia and Tipster corpora were preprocessed by a pipeline consisting of a trigram-based statistical POS tagger (Brants, 2000) and a SVM-based chunker (Kudoh & Matsumoto, 2000), to identify noun phrases (NPs).

The patterns are used to provide evidence for semantic relations employing a majority voting strategy. We positively label a category pair with *isa* in case the number of matches of positive patterns is greater than the number of matches of negative ones. In addition, we use the patterns to filter the *isa* relations created by the connectivity-based methods (4). This is due to instance categorization and redundant categorization giving results which are not always reliable,

e.g. we incorrectly find that CONSONANTS *isa* PHONETICS. We use the same majority voting scheme, except that this time we mark as *notisa* those pairs with a number of negative matches greater than the number of positive ones. This ensures better precision by leaving the recall basically unchanged. These methods create 15,055 *isa* relations and filter out 3,277 previously identified positive links.

Inference-based methods (6)

The last set of methods propagate the previously found relations by means of multiple inheritance and transitivity. We first propagate all *isa* relations to those superclasses whose head lemma matches the head lemma of a previously identified *isa* superclass. E.g., once we found that MICROSOFT *isa* COMPANIES LISTED IN NASDAQ we can infer also that MICROSOFT *isa* MULTINATIONAL COMPANIES.

We then propagate all *isa* links to those superclasses which are connected through a path found along the previously discovered subsumption hierarchy. E.g., given that FRUIT *isa* CROPS and CROPS *isa* EDIBLE PLANTS, we can infer that FRUITS *isa* EDIBLE PLANTS.

Evaluation

We evaluate the coverage and quality of the semantic relations extracted automatically. This is because the size of the induced taxonomy is very large – up to 105,418 generated *isa* semantic links – and also to avoid any bias in the evaluation method.

Comparison with ResearchCyc

We first compute the amount of *isa* relations we correctly extracted by comparing with ResearchCyc[4], the research version of the Cyc knowledge base (Lenat & Guha, 1990) including (as of version 1.0) more than 300,000 concepts and 3 millions assertions. For each category pair, we first map each category to its Cyc concept using Cyc's internal lexeme-to-concept denotational mapper. Concepts are found by querying the full category label (e.g. Alan Turing). In case no matching concept is found, we fall back to querying its lexical head (hardware for IBM hardware).

We evaluate only the 85% of the pairs which have corresponding concepts in Cyc. These pairs are evaluated by querying Cyc whether the concept denoted by the Wikipedia subcategory is either an instance of (`#$isa`) or is generalized by (`#$genls`) the concept denoted by its superclass[5]. We then take the result of the query as the actual (*isa* or *notisa*) semantic class for the category pair and use it to evaluate the system's response. This way we are able to compute standard measures of precision, recall and balanced F-measure. Table 1 shows the results obtained by taking the syntax-based methods (i.e. head matching) as baseline and incrementally augmenting them with different sets of methods, namely our connectivity and pattern based methods. All differences in performance are statistically significant at $p < 0.001$. We test for statistical significance by performing a McNemar test.

	R	P	F_1
baseline (methods 1-3)	73.7	100.0	84.9
+ connectivity (methods 1-4, 6)	80.6	91.8	85.8
+ pattern-based (methods 1-3, 5-6)	84.3	91.5	87.7
all (methods 1-6)	89.1	86.6	87.9

Table 1: Comparison with Cyc

Discussion and Error Analysis. The simple methods employed for the baseline work suprisingly good with perfect precision and somewhat satisfying recall. However, since only categories with identical heads are connected, we do not create a single interconnected taxonomy but many separate taxonomic islands. In practice we simply find that HISTORICAL BUILDINGS are BUILDINGS. The extracted information is trivial.

By applying the connectivity-based methods we are able to improve the recall considerably. The drawback is a decrease in precision. However, a closer look reveals that we now in fact created a interconnected taxonomy where concepts with quite different linguistic realization are connected. We observe the same trend by applying the pattern-based methods in addition to the baseline. They improve the recall even more, but they also have a lower precision. The best results are obtained by combining all methods.

Because we did not expect such a big drop in precision – and only moderate improvement over the baseline in F-measure – we closely inspected a random sample of 200 false positives, i.e., the cases which led to the low precision score. Three annotators labeled these cases as true if judged to be in fact correct *isa* relations, false otherwise. It turned out that about 50% of the false positives were indeed labeled correctly as *isa* relations by the system, but these relations could not be found in Cyc. This is due to (1) Cyc missing the required relations (e.g. BRIAN ENO *isa* MUSICIANS) or (2) missing the required concepts (e.g., we correctly find that BEE TRAIN *isa* ANIMATION STUDIOS, but since Cyc provides only the TRAIN-TRANSPORTATION-DEVICE and STUDIO concepts, we query: *"is train a studio?"* which leads to a false positive.

Computing semantic similarity using Wikipedia

In Strube & Ponzetto (2006) we proposed to use the Wikipedia categorization as a conceptual network to compute the semantic relatedness of words. However, we could not compute semantic similarity, because approaches to measuring semantic similarity that rely on lexical resources use paths based on *isa* relations only. These are only available in the present work.

We perform an extrinsic evaluation by computing semantic similarity on two commonly used datasets, namely Miller & Charles' (1991) list of 30 noun pairs (M&C) and the 65 word synonymity list from Rubenstein & Goodenough (1965, R&G). We compare the results obtained by using

[4] `http://research.cyc.com/`

[5] Note that our definition of *isa* is similar to the one found in WordNet prior to version 2.1. That is, we do not distinguish hyponyms that are classes from hyponyms that are instances (cf. Miller & Hristea (2006)).

		M&C				R&G			
		pl	wup	lch	res	pl	wup	lch	res
WordNet	all	0.72	0.77	**0.82**	0.78	0.78	0.82	**0.86**	0.81
Wikirelate!	all	**0.60**	0.53	0.58	0.30	0.62	0.63	**0.64**	0.34
	non-missing	**0.65**	0.61	**0.65**	0.41	0.66	0.69	**0.70**	0.42
Wikirelate! *isa*-only	all	0.67	0.65	0.67	**0.69**	0.67	0.69	**0.70**	0.66
	non-missing	0.71	0.70	0.72	**0.74**	0.70	**0.73**	**0.73**	0.70
Wikirelate! PageRank filter	all	0.68	**0.74**	0.73	0.62	0.67	**0.74**	0.73	0.58
	non-missing	0.72	**0.79**	0.78	0.68	0.70	**0.79**	0.77	0.63
Wikirelate! *isa* + PageRank	all	0.73	0.79	0.78	**0.81**	0.69	0.75	0.74	**0.76**
	non-missing	0.76	0.84	0.82	**0.86**	0.72	0.79	0.77	**0.80**

Table 2: Results on correlation with human judgements of similarity measures

Wikipedia with the ones obtained by using WordNet, which is the most widely used lexical taxonomy for this task. Following the literature on semantic similarity, we evaluate performance by taking the Pearson product-moment correlation coefficient r between the similarity scores and the corresponding human judgements. For each dataset we report the correlation computed on all pairs (*all*). In the case of word pairs where at least one of the words could not be found the similarity score is set to 0. In addition, we report the correlation score obtained by disregarding such pairs containing missing words (*non-missing*).

Table 2 reports the scores obtained by computing semantic similarity in WordNet as well as in Wikipedia using different scenarios and measures (Rada et al. (1989, *pl*), Wu & Palmer (1994, *wup*), Leacock & Chodorow (1998, *lch*), Resnik (1995, *res*)). We first take as baseline the Wikirelate! method outlined in Strube & Ponzetto (2006) and extend it by first computing only paths based on *isa* relations. Since experiments on development data[6] revealed a performance improvement far lower than expected, we performed an error analysis. This revealed that many dissimilar pairs received a score higher than expected, because of coarse-grained overconnected categories containing a large amount of dissimilar pages, e.g. mound and shore were directly connected through LANDFORMS though they are indeed quite different according to human judges.

A way to model the categories' connectivity is to compute their authoritativeness, i.e. we assume that overconnected, semantically coarse categories will be the most authoritative ones. This can be accomplished for instance by computing the centrality scores of the Wikipedia categories. Since the Wikipedia categorization network is a directed acyclic graph, link analysis algorithms such as PageRank (Brin & Page, 1998) can be easily applied to automatically detect and remove these coarse categories from the categorization network. We take the graph given by all the categories and the pages that point to them and apply the PageRank algorithm. PageRank scores are computed recursively for each category vertex v by the formula

$$PR(v) = (1-d) + d \sum_{v' \in I(v)} \frac{PR(v')}{|O(v')|}$$

where $d \in (0,1)$ is a dumping factor (we set it to the standard value of .85), $I(v)$ is the set of nodes linked to v and $|O(v')|$ the number of outgoing links of node v'. This gives a ranking of the most authoritative categories, which in our case happen to be the categories in the highest regions of the categorization network – i.e., the top-ranked categories are FUNDAMENTAL, SOCIETY, KNOWLEDGE, PEOPLE, SCIENCE, ACADEMIC DISCIPLINES and so on.

The third experimental setting of Table 2 shows the results obtained by computing relatedness using the method from Strube & Ponzetto (2006) and removing the top 200 highest ranked PageRank categories[7]. Finally, we present results of using both *isa* and PageRank filtering. The results indicate that using both *isa* relations and applying PageRank filtering work better than the simple Wikirelate! baseline. This is because in both cases we are able to filter out categories and category relations which decrease the similarity scores, i.e. coarse-grained categories using PageRank, and *notisa* (e.g. meronymic, antonymic) semantic relations. The two methods are indeed complementary, which is shown by the best results being obtained by applying them together. Using PageRank filtering together with paths including only *isa* relations yields results which are close to the ones obtained by using WordNet.

The results indicate that Wikipedia can be successfuly used as a taxonomy to compute the semantic similarity of words. In addition, our application of PageRank for filtering out coarse-grained categories highlights that, similarly to the connectivity-based methods used to identify *isa* relations, the internal structure of Wikipedia can be used to generate semantic content, being based on a meaningful set of conventions the users tend to adhere.

Conclusions

We described the automatic creation of a large scale domain independent taxonomy. We took Wikipedia's categories as

[6] In order to perform a blind test evaluation, we developed the system for computing semantic similarity using a different version of Wikipedia, namely the database dump from 19 February 2006.

[7] The optimal threshold value was established again by analyzing performance on the development data.

concepts in a semantic network and labeled the relations between these concepts as *isa* and *notisa* relations by using methods based on the connectivity of the network and on applying lexico-syntactic patterns to very large corpora. Both connectivity-based methods and lexico-syntactic patterns ensure a high recall while decreasing the precision. We compared the created taxonomy with ResearchCyc and via semantic similarity measures with WordNet. Our Wikipedia-based taxonomy proved to be competitive with the two arguably largest and best developed existing ontologies. We believe that these results are caused by taking already structured and well-maintained knowledge as input.

Our work on deriving a taxonomy is the first step in creating a fully-fledged ontology based on Wikipedia. This will require to label the generic *notisa* relations with particular ones such as *has-part*, *has-attribute*, etc.

Acknowledgements. This work has been funded by the Klaus Tschira Foundation, Heidelberg, Germany. The first author has been supported by a KTF grant (09.003.2004). We thank our colleagues Katja Filippova and Christoph Müller for useful feedback.

References

Berland, M. & E. Charniak (1999). Finding parts in very large corpora. In *Proc. of ACL-99*, pp. 57–64.

Berners-Lee, T., J. Hendler & O. Lassila (2001). The semantic web. *Scientific American*, 284(5):34–43.

Brants, T. (2000). TnT – A statistical Part-of-Speech tagger. In *Proc. of ANLP-00*, pp. 224–231.

Brin, S. & L. Page (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117.

Buitelaar, P., P. Cimiano & B. Magnini (Eds.) (2005). *Ontology Learning from Text: Methods, Evaluation and Applications.* Amsterdam, The Netherlands: IOS Press.

Caraballo, S. A. (1999). Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proc. of ACL-99*, pp. 120–126.

Church, K. W. & P. Hanks (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29.

Cimiano, P., A. Pivk, L. Schmidt-Thieme & S. Staab (2005). Learning taxonomic relations from heterogenous sources of evidence. In P. Buitelaar, P. Cimiano & B. Magnini (Eds.), *Ontology Learning from Text: Methods, Evaluation and Applications*, pp. 59–73. Amsterdam, The Netherlands: IOS Press.

Collins, M. (1999). *Head-Driven Statistical Models for Natural Language Parsing.*, (Ph.D. thesis). University of Pennsylvania.

Domingos, P., S. Kok, H. Poon, M. Richardson & P. Singla (2006). Unifying logical and statistical AI. In *Proc. of AAAI-06*, pp. 2–7.

Fellbaum, C. (Ed.) (1998). *WordNet: An Electronic Lexical Database.* Cambridge, Mass.: MIT Press.

Finkel, J. R., T. Grenager & C. Manning (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proc. of ACL-05*, pp. 363–370.

Girju, R., A. Badulescu & D. Moldovan (2006). Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1):83–135.

Harman, D. & M. Liberman (1993). *TIPSTER Complete.* LDC93T3A, Philadelphia, Penn.: Linguistic Data Consortium.

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proc. of COLING-92*, pp. 539–545.

Klein, D. & C. D. Manning (2003). Fast exact inference with a factored model for natural language parsing. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in Neural Information Processing Systems 15 (NIPS 2002)*, pp. 3–10. Cambridge, Mass.: MIT Press.

Kudoh, T. & Y. Matsumoto (2000). Use of Support Vector Machines for chunk identification. In *Proc. of CoNLL-00*, pp. 142–144.

Leacock, C. & M. Chodorow (1998). Combining local context and WordNet similarity for word sense identification. In C. Fellbaum (Ed.), *WordNet. An Electronic Lexical Database*, Chp. 11, pp. 265–283. Cambridge, Mass.: MIT Press.

Lee, L. (1999). Measures of distributional similarity. In *Proc. of ACL-99*, pp. 25–31.

Lenat, D. B. & R. V. Guha (1990). *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project.* Reading, Mass.: Addison-Wesley.

Matuszek, C., M. Witbrock, R. C. Kahlert, J. Cabral, D. Schneider, P. Shah & D. Lenat (2005). Searching for common sense: Populating Cyc from the web. In *Proc. of AAAI-05*, pp. 1430–1435.

McCarthy, J. (1959). Programs with common sense. In *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, pp. 75–91.

Miller, G. A. & W. G. Charles (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1):1–28.

Miller, G. A. & F. Hristea (2006). WordNet nouns: Classes and instances. *Computational Linguistics*, 32(1):1–3.

Minnen, G., J. Carroll & D. Pearce (2001). Applied morphological processing of English. *Natural Language Engineering*, 7(3):207–223.

Porter, M. (1980). An algorithm for suffix stripping. *Program*, 14(3):130–137.

Punyakanok, V., D. Roth, W. Yih & D. Zimak (2006). Learning and inference over constrained output. In *Proc. of IJCAI-05*, pp. 1117–1123.

Rada, R., H. Mili, E. Bicknell & M. Blettner (1989). Development and application of a metric to semantic nets. *IEEE Transactions on Systems, Man and Cybernetics*, 19(1):17–30.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *Proc. of IJCAI-95*, Vol. 1, pp. 448–453.

Richardson, M. & P. Domingos (2003). Building large knowledge bases by mass collaboration. In *Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP 2003).* Sanibel Island, Fl., October 23–25, 2003, pp. 129–137.

Rubenstein, H. & J. Goodenough (1965). Contextual correlates of synonymy. *Communications of the ACM*, 8(10):627–633.

Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.

Snow, R., D. Jurafsky & A. Y. Ng (2006). Semantic taxonomy induction from heterogeneous evidence. In *Proc. of COLING-ACL-06*, pp. 801–808.

Strube, M. & S. P. Ponzetto (2006). WikiRelate! Computing semantic relatedness using Wikipedia. In *Proc. of AAAI-06*, pp. 1419–1424.

Suchanek, F. M., G. Kasneci & G. Weikum (2007). YAGO: A core of semantic knowledge. In *Proc. of WWW-07*.

Weeds, J. & D. Weir (2005). Co-occurrence retrieval: A flexible framework for lexical distributional similarity. *Computational Linguistics*, 31(4):439–475.

Wu, Z. & M. Palmer (1994). Verb semantics and lexical selection. In *Proc. of ACL-94*, pp. 133–138.