

Fine-Grained Proper Noun Ontologies for Question Answering

Gideon S. Mann

Department of Computer Science
Johns Hopkins University
Baltimore, Maryland 21218
gsm@cs.jhu.edu

Abstract

The WordNet lexical ontology, which is primarily composed of common nouns, has been widely used in retrieval tasks. Here, we explore the notion of a fine-grained proper noun ontology and argue for the utility of such an ontology in retrieval tasks. To support this claim, we build a fine-grained proper noun ontology from unrestricted news text and use this ontology to improve performance on a question answering task.

1 Introduction

The WordNet lexical ontology (Miller, 1990) contains more than 100,000 unique noun forms. Most of these noun forms are common nouns (nouns describing non-specific members of a general class, e.g. “detective”). Only a small percentage¹ of the nouns in WordNet are proper nouns (nouns describing specific instances, e.g. “[the detective] Columbo”).

The WordNet ontology has been widely useful, with applications in information retrieval (Sussna, 1993), text classification (Scott and Matwin, 1998), and question answering (Pasca and Harabagiu, 2001). These successes have shown that common noun ontologies have wide applicability and utility.

There exists no ontology with similar coverage and detail for proper nouns. Prior work in proper noun identification has focused on ‘named entity’

¹A random 100 synset sample was composed of 9% proper nouns.

recognition (Chinchor et al., 1999), stemming from the MUC evaluations. In this task, each proper noun is categorized, for example, as a PERSON, a LOCATION, or an ORGANIZATION.

These coarse categorizations are useful, but more finely grained classification might have additional advantages. While Bill Clinton is appropriately identified as a PERSON, this neglects his identity as a president, a southerner, and a saxophone player. If an information request identifies the object of the search not merely as a PERSON, but as a **typed proper noun** (e.g. “a southern president”), this preference should be used to improve the search.

Unfortunately, building a proper noun ontology is more difficult than building a common noun ontology, since the set of proper nouns grows more rapidly. New people are born. As people change, their classification must change as well. A broad-coverage proper noun ontology must be constantly updated. Therefore, to propose a viable system, a method, however limited, must be presented to build a proper noun ontology.

In this paper, we explore the idea of a fine-grained proper noun ontology and its use in question answering. We build a proper noun ontology from unrestricted text using simple textual co-occurrence patterns (Section 3). This automatically constructed ontology is then used on a question answering task to give preliminary results on the utility of this information (Section 4).

2 Ontologies for Question Answering

Modern question answering systems rely heavily on the fact that questions contain strong preferences for

<p>The 1974 film ‘That’s Entertainment!’ was made from film clips from what Hollywood studio?</p> <p>What king of Babylonia reorganized the empire under the Code that bears his name?</p> <p>What rock ‘n’ roll musician was born Richard Penniman on Christmas Day?</p> <p>What is the oldest car company which still exists today?</p> <p>What was the name of the female Disco singer who scored with the tune ‘Dim All the Lights’ in 1979?</p> <p>What was the name of the first Russian astronaut to do a spacewalk?</p> <p>What was the name of the US helicopter pilot shot down over North Korea?</p> <p>Which astronaut did Tom Hanks play in ‘Apollo 13’?</p> <p>Which former Klu Klux Klan member won an elected office in the U.S.?</p> <p>Who’s the lead singer of the Led Zeppelin band?</p> <p>Who is the Greek goddess of retribution or vengeance?</p> <p>Who is the prophet of the religion of Islam?</p> <p>Who is the author of the book, ‘The Iron Lady: A Biography of Margaret Thatcher’?</p> <p>Who was the lead actress in the movie ‘Sleepless in Seattle’?</p>

Table 1: Questions Indicating a Typed Proper Noun Preference (Trivia and Trec-8/9 Questions)

the types of answers they expect. Kupiec (1993) observes that the WH word itself provides preferences (e.g. ‘‘Who’’ questions prefer PERSON answers). He further observes that questions also include type preferences in other parts of the question. Sometimes these preferences occur within the WH phrase (‘‘what **color**’’), and sometimes they are embedded elsewhere within the question (‘‘what is the **color** ...’’). In both, the question indicates a preference for colors as answers.

Current question answering systems use ontologies when these type preferences are detected. One simple method is as follows: when a type preference is recognized, the preference is located within the WordNet ontology, and children of that synset are treated as potential answers. Given the question ‘‘In pool, what **color** is the eight ball?’’, and the ontology excerpt shown in Figure 1, the system can narrow down the range of choices. This approach has high precision: if the type preference can be located, and a candidate answer is found in a child node (in a suitable corpus context), then the candidate is likely to be the answer.

Harabagiu et al. (2000) proposes another method for using an ontology: WordNet subtrees are linked to types recognized by a named entity recognizer. Their system works as follows: given the question ‘‘What is the **wingspan** of a condor?’’, it locates ‘‘wingspan’’ in the WordNet ontology. It then detects that ‘‘wingspan’’ falls into the MAGNITUDE subtree

which is linked to the QUANTITY type. This links words in the MAGNITUDE subtree to numbers.

While the WordNet ontology is primarily composed of common nouns, it contains some proper nouns, typically those least likely to be ephemeral (e.g. countries, cities, and famous figures in history). These can be used as any other common nouns are used. Given the question ‘‘Which composer wrote ‘The Marriage of Figaro’?’’, the WordNet ontology will provide the fact that ‘‘Wolfgang Amadeus Mozart’’ is a composer.

Table 1 lists sample questions where a proper noun ontology would be useful. Some of the proper noun types are relatively static (Greek gods, kings of Babylonia). Other categories are more ephemeral (lead singers, British actresses). WordNet enumerates 70 Greek gods and 80 kings, but no lead singers and no British actresses.

Ravichandran and Hovy (2002) present an alternative ontology for type preference and describe a method for using this alternative ontology to extract particular answers using surface text patterns. Their proposed ontology is orders of magnitude smaller than WordNet and ontologies considered here, having less than 200 nodes.

3 Building a Proper Noun Ontology

In order to better answer the questions in Table 1, we built a proper noun ontology from approximately 1 gigabyte of AP news wire text. To do so, we tok-

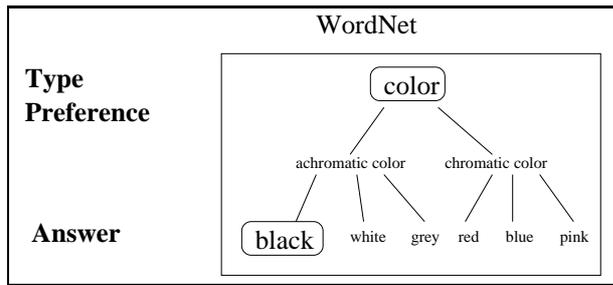


Figure 1: Using WordNet to Directly Provide Type Preferences

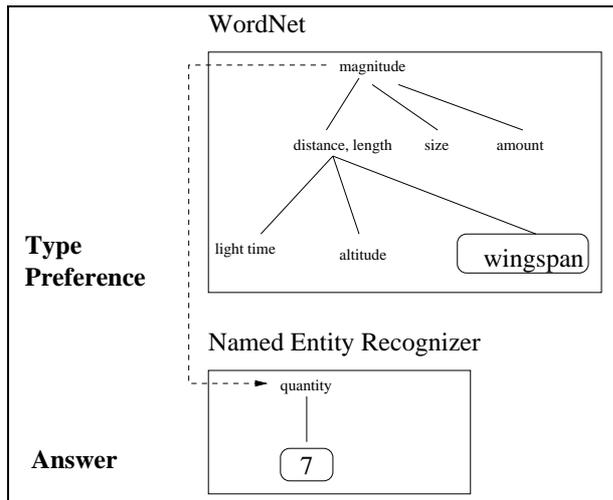


Figure 2: Linking WordNet subtrees to a Named Entity Recognizer

enized and part-of-speech tagged the text, and then searched for instances of a common noun followed immediately by a proper noun. This pattern detects phrases of the form '[the] automaker Mercedes Benz', and is ideally suited for proper nouns. In AP news wire text this is a productive and high precision pattern, generating nearly 200,000 unique descriptions, with 113,000 different proper nouns and 20,000 different descriptions. In comparison, the "such as" pattern (Section 5) occurs less than 50,000 times in the same size corpora. Table 2 shows the descriptions generated for a few proper nouns using this simple pattern.

To assess the precision of the extractions, we took a sample of 100 patterns extracted from the AP-news text. From these 100, 79 of the items classified as named entities were in fact named entities, and out of those, 60 (75%) had legitimate descriptions.

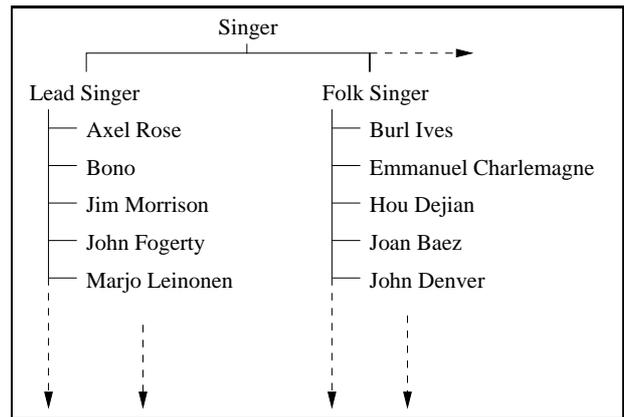


Figure 3: Subset of 'singer' subtree in the Induced Proper Noun Ontology

To build the complete ontology, first each description and proper noun forms its own synset. Then, links are added from description to each proper noun it appears with. Further links are put between descriptions "X Y" and "Y" (noun compounds and their heads). Clearly, this method is problematic in the cases of polysemous words or complex noun-noun constructions ("slalom king") and integrating this ontology with the WordNet ontology requires further study.

This proper noun ontology fills many of the holes in WordNet's world knowledge. While WordNet has no lead singer synset, the induced proper noun ontology detects 13 distinct lead singers (Figure 3). WordNet has 2 folk singers; the proper noun ontology has 20. In total, WordNet lists 53 proper nouns as singers, while the induced proper noun ontology has more than 900. While the induced ontology is not complete, it is more complete than what was previously available.

As can be seen from the list of descriptions generated by this pattern, people are described in a variety of different ways, and this pattern detects many of them. Table 3 shows the descriptions generated for a common proper noun ("Bill Gates"). When the descriptions are grouped by WordNet synsets and senses manually resolved, the variety of descriptions decreases dramatically (Figure 4). "Bill Gates" can be described by a few distinct roles, and a distribution over these descriptions provide an informative understanding: leader (.48), businessperson (.27), worker (.05), originator (.05), expert (.05), and rich

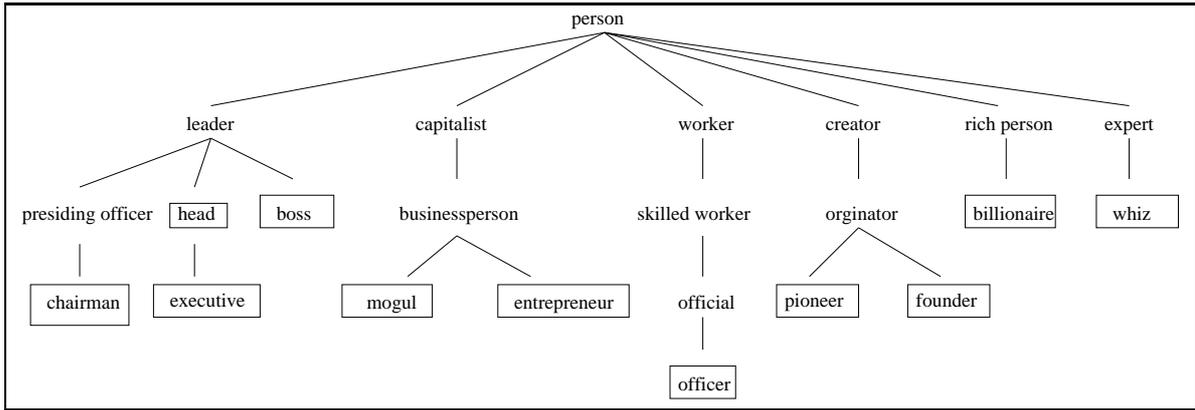


Figure 4: Descriptions of Bill Gates Organized into WordNet, observed descriptions boxed

Proper Noun	Count	Description
Axel Rose	3	singer
	2	lead singer
	2	vocalist
Emma Thompson	3	actress
Mercedes-Benz	4	Luxury car maker
	4	car maker
	3	automaker
	2	family
	2	luxury
	1	gold
	1	service
	1	subsidiary

Table 2: Proper Noun Descriptions Extracted from News Corpora

person (.02). Steve Jobs, who has a career path similar to Bill Gates, has a similar but distinct signature: originator (.6), expert (.4).

One immediate observation is that some of the descriptions may be more relevant than others. Is Gates' role as an 'office worker' as important as his role as a 'billionaire'? The current system makes no decision and treats all descriptions as equally relevant and stores all of them. There is no need to reject descriptions since there is no human cost in superfluous or distracting descriptions (unlike in summarization tasks). It is important that no invalid descriptions are added.

The previous examples have focused on proper nouns which are people's names. However, this method works for many organizations as well, as

Proper Noun	Count	Description
Bill Gates	15	chairman
	9	mogul, tycoon, magnate
	2	officer
	2	whiz, genius
	1	pioneer
	1	head
	1	founder
	1	executive
	1	entrepreneur
	1	boss
	1	billionaire

Table 3: Bill Gates Descriptions in AP Newswire, grouped by WordNet synset

the data in Table 2 show. However, while description extraction for people is high quality (84% correct descriptions in a 100 example sample), for non-people proper names, the quality of extraction is poorer (47% correct descriptions). This is a trend which requires further study.

4 Using a Proper Noun Ontology in a Question Answering Task

We generated the above ontology and used it in a sentence comprehension task: given a question and a sentence which answers the question, extract the minimal short answer to the question from the sentence. The task is motivated by the observation that extracting short answers is more difficult than extracting full sentence or passage length ones. Fur-

Ontology	Correct	Total Answered	Precision
WordNet	127	169	75.1
IPNO	46	67	68.6
WN + IPNO	145	194	74.7

Table 4: Performance on a Test Corpus when an Induced Proper Noun Ontology (IPNO) is combined with Wordnet

thermore, retrieving answers from smaller document spaces may be more difficult than retrieving answers from larger ones, if smaller spaces have less redundant coverage of potential answers. In this sentence comprehension task, there is virtually no redundancy. To generate data for this task, we took trivia games, which, along with the question, had a full sentence explanation (Mann, 2002).

Baseline experiments used the WordNet ontology alone. From a semantic type preference stated in the question, a word was selected from the sentence as an answer if it was a child of the type preference. ‘Black’ would be picked as an answer for a ‘color’ type preference (Figure 1).

To utilize the induced proper noun ontology, we took the raw data and selected the trailing noun for each proper noun and for each description. Thus, for an extraction of the form “computer mogul Bill Gates”, we added a pattern of the form “Gates mogul”. We created an ontology from these instances completely separate from the WordNet ontology.

We put this induced proper noun ontology into the pipeline as follows: if WordNet failed to find a match, we used the induced proper noun ontology. If that ontology failed to find a match, we ignored the question. In a full system, a named entity recognizer might be added to resolve the other questions.

We selected 1000 trivia game questions at random to test out the new two-ontology system. Table 4 shows the results of the experiments. The boost is clear: improved recall at slightly decreased precision. Gains made by inducing an ontology from an unrestricted text corpus (newstext) and applying it to a unmatched test set (trivia games), suggests that a broad-coverage general proper noun ontology may be useful.

It is further surprising that this improvement

comes at such a small cost. The proper noun ontology wasn’t trimmed or filtered. The only disadvantage of this method is simply that its coverage is small. Coverage may be increased by using ever larger corpora. Alternatively, different patterns (for example, appositives) may increase the number of words which have descriptions. A rough error analysis suggests that most of the errors come from mis-tagging, while few come from correct relationships in the ontology. This suggests that attempts at noise reduction might be able to lead to larger gains in performance.

Another potential method for improving coverage is by bootstrapping descriptions. Our test corpus contained a question whose answer was “Mercedes-Benz”, and whose type preference was “car company”. While our proper noun ontology contained a related link (Mercedes-Benz automaker), it did not contain the exact link (Mercedes-Benz car company). However, elsewhere there existed the links (Opel automaker) and (Opel car company). Potentially these descriptions could be combined to infer (Mercedes-Benz car company). Formally :

$$(B \text{ Y}) \text{ and } (A \text{ Y}) \text{ and } (A \text{ Z}) \Rightarrow (B \text{ Z})$$

$$\begin{aligned} &(\text{Mercedes-Benz automaker}) \text{ and } (\text{Opel} \\ &\text{automaker}) \text{ and } (\text{Opel car company}) \Rightarrow \\ &(\text{Mercedes-Benz car company}) \end{aligned}$$

Expanding descriptions using a technique like this may improve coverage. Still, care must be taken to ensure that proper inferences are made since this rule is not always appropriate. Bill Gates is a ten-billionaire; Steve Jobs isn’t.

5 Prior Work in Building Ontologies

There has been considerable work in the past decade on building ontologies from unrestricted text. Hearst (1992) used textual patterns (e.g. “such as”) to identify common class members. Carballo and Charniak (1999) and Carballo (1999) augmented these lexical patterns with more general lexical co-occurrence statistics (such as relative entropy). Berland and Charniak (1999) use Hearst style techniques to learn meronym relationships (part-whole) from corpora. There has also been work in building ontologies from structured

Correct Answer	Question
(Debbie) Reynolds	What actress once held the title of ‘Miss Burbank’?
(Jim) Lovell	Which astronaut did Tom Hanks play in ‘Apollo 13’?
Xerxes	Which Persian king moved an invasion force across the Hellespont on a bridge of ships?
(Donna) Summer	What was the name of the female Disco singer who scored with the tune ‘Dim All the Lights’ in 1979?
MGM	The 1974 film ‘That’s Entertainment!’ was made from film clips from what Hollywood studio ?

Table 5: Successes of the Proper Noun Ontology for the Question Answering task

text, notably in the AQUILEX project (e.g. Copestake, 90) which builds ontologies from machine readable dictionaries.

The most closely related work is (Girju, 2001), which describes a method for inducing a domain-specific ontology using some of the techniques described in the previous paragraph. This induced ontology is then potential useful for a matched question domain. Our paper differs in that it targets proper nouns, in particular people, which are overlooked in prior work, have broad applicability, and can be used in a cross-domain fashion. Furthermore, we present initial results which attempt to gauge coverage improvement as a result of the induced ontology.

Another related line of work is word clustering. In these experiments, the attempt is made to cluster similar nouns, without regard to forming a hierarchy. Pereira et al. (1993) presented initial work, clustering nouns using their noun-verb co-occurrence information. Riloff and Lehnert (1993) build semantic lexicons using extraction pattern co-occurrence. Lin and Pantel (2001) extend these methods by using many different types of relations and exploiting corpora of tremendous size.

The important difference for this work between the hierarchical methods and the clustering methods is that clusters are unlabelled. The hierarchical methods can identify that a “Jeep Cherokee” is a type of car. In contrast, the clustering methods group together related nouns, but exactly what the connection is may be difficult to distinguish (e.g. the cluster “Sierra Club”, “Environmental Defense Fund”, “Natural Resources Defense Council”, “Public Citizen”, “National Wildlife Federation”). Generating

labels for proper noun clusters may be another way to build a proper noun ontology.

The method we use to build the fine-grained proper name ontology also resembles some of the work done in coarse-grained named entity recognition. In particular, Collins and Singer (1999) present a sophisticated method for using bootstrapping techniques to learn the coarse-classification for a given proper noun. Riloff and Jones (1999) also present a method to use bootstrapping to create semantic lexicons of proper nouns. These methods may be applicable for use in fine-grained proper noun ontology construction as well.

Schiffman et al. (2001) describe work on producing biographical summaries. This work attempts to synthesize one description of a person from multiple mentions. This summary is an end in itself, as opposed to general knowledge collected. These descriptions also attempt to be parsimonious in contrast to the rather free associations extracted by the method presented above.

6 Conclusions

In this paper we have motivated the use of a proper noun ontology for question answering. We described a method for inducing pieces of this ontology, and then showed preliminary methods can be useful. Prior work on proper nouns has focused on classifying them into very coarse categories (e.g. PERSON, LOCATION). As this paper has shown, these coarse classifications can be refined fortuitously, especially for the PERSON type.

This paper demonstrates that inducing a general ontology improves question answering performance. Previous work examined ontology induction

for a specialized domain. It is somewhat surprising that an ontology built from unrestricted text can lead to improvement on unmatched questions.

The experiments we performed demonstrated that though the precision of the ontology is high, the crucial problem is increasing coverage. Tackling this problem is an important area of future work. Finally, this work opens up a potential new avenue for work on inducing proper noun ontologies. There are doubtlessly many more ways to extract descriptions and to improve coverage.

References

- Matthew Berland and Eugene Charniak. 1999. Finding parts in very large corpora. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 57–64.
- S. Caraballo and E. Charniak. 1999. Determining the specificity of nouns from text.
- Sharon Caraballo. 1999. Automatic acquisition of a hypernym-labeled noun hierarchy from text. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*.
- N. Chinchor, E. Brown, L. Ferro, and P. Robinson. 1999. 1999 named entity recognition task definition. *Tech Report*.
- Michael Collins and Yoram Singer. 1999. Unsupervised models for named entity classification. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing*.
- Ann Copestake. 1990. An approach to building the hierarchical element of a lexical knowledge base from a machine readable dictionary. In *First International Workshop on Inheritance in NLP*.
- Roxana Girju. 2001. Answer fusion with on-line ontology development. In *Student Research Workshop Proceedings at The 2nd Meeting of the North American Chapter of the Association for Computational Linguistics*.
- S. Harabagiu, D. Moldovan, M. Pasca, R. Mihalcea, M. Surdea nu, R. Bunescu, R. Girju, V. Rus, and P. Mor. 2000. Falcon : Boosting knowledge for answer engines. *Proc. of TREC-9*.
- Marti Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. *Proceedings of the Fourteenth International Conference on Computational Linguistics (COLING-92)*.
- J. Kupiec. 1993. Murax: A robust linguistic approach for question answering using an on-line encyclopedia. In *ACM-SIGIR'93*.
- Dekang Lin and Patrick Pantel. 2001. Induction of semantic classes from natural language text. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 317–322.
- Gideon S. Mann. 2002. Learning how to answer questions using trivia games. In *Proceedings of the Nineteenth International Conference on Computational Linguistics (COLING 2002)*.
- G. Miller. 1990. Wordnet: An On-line Lexical Database. *International Journal of Lexicography*, 3(4):235–312.
- Marius Pasca and Sanda Harabagiu. 2001. The informative role of wordnet in open-domain question answering. In *Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations*, pages 138–143. Association for Computational Linguistics.
- Fernando C. N. Pereira, Naftali Tishby, and Lillian Lee. 1993. Distributional clustering of english words. In *Meeting of the Association for Computational Linguistics*, pages 183–190.
- Deepak Ravichandran and Eduard Hovy. 2002. Learning surface text patterns for a question answering system. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*.
- Ellen Riloff and Rosie Jones. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pages 1044–1049.
- E. Riloff and W. Lehnert. 1993. Automated Dictionary Construction for Information Extraction from Text. In *Proceedings of the Ninth IEEE Conference on Artificial Intelligence for Applications*, pages 93–99, Los Alamitos, CA. IEEE Computer Society Press.
- Barry Schiffman, Inderjeet Mani, and Kristian J. Conception. 2001. Producing biographical summaries: Combining linguistic knowledge with corpus statistics. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*.
- Sam Scott and Stan Matwin. 1998. Text classification using WordNet hypernyms. In Sanda Harabagiu, editor, *Use of WordNet in Natural Language Processing Systems: Proceedings of the Conference*, pages 38–44. Association for Computational Linguistics, Somerset, New Jersey.
- M. Sussna. 1993. Word sense disambiguation for free-text indexing using a massive semantic network. In *Proceedings of CIKM '93*.