Learning Verbs on the Fly

Zornitsa Kozareva
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292-6695
kozareva@isi.edu

ABSTRACT
To answer the question *"What are the duties of a medical doctor?"*, one would require knowledge about verb-based relations. A lot of effort has been invested in developing relation learners, however to our knowledge there is no repository (or system) which can return all verb relations for a given term. This paper describes an automated procedure which can learn and produce such information with minimal effort. To evaluate the performance of our verb harvesting procedure, we have conducted two types of evaluations: (1) in the human based evaluation we found that the accuracy of the described algorithm is .95 at rank 100; (2) in the comparative study with existing relation learner and knowledge bases we found that our approach yields 12 times more verb relations.

KEYWORDS: verb harvesting, relation learning, information extraction, knowledge acquisition.

Proceedings of COLING 2012: Posters, pages 599–610,
COLING 2012, Mumbai, December 2012.

1 Introduction

To be able to answer the questions *"What causes ebola?"*, *"What are the duties of a medical doctor?"*, *"What are the differences between a terrorist and a victim?"*, *"Which are the animals that have wings but cannot fly?"* one requires knowledge about verb-based relations. Over the years, researchers have developed various relation learning algorithms. Some (Ravichandran and Hovy, 2002; Bunescu and Mooney, 2007) targeted specific relations like *BornInYear*, *CorporationAcquired*, others (Wu and Weld, 2010; Fader et al., 2011) extracted any phrase denoting a relation in an English sentence. (Banko, 2009) used labeled data to learn relations, (Suchanek et al., 2007) used information encoded in the structured Wikipedia documents, (Riloff and Jones, 1999) bootstrapped patterns. As a result various knowledge bases have been produced like TopicSignatures (Agirre and Lacalle, 2004), ConceptNet (Liu and Singh, 2004), Yago (Suchanek et al., 2007), NELL (Carlson et al., 2009) and ReVerb (Fader et al., 2011).

Despite the many efforts to date, yet there is no universal repository (or even a system), which for a given term it can immediately return all verb relations related to the term. However, one would still like to dispose of an automated procedure, which on the fly can accurately and quickly produce such information for any term. If available, such resource can aid different natural language processing tasks such as preposition sense disambiguation (Litkowski and Hargraves, 2007), selectional preferences (Resnik, 1996; Ritter et al., 2010), question answering (Ferrucci et al., 2010) and textual entailment (Szpektor et al., 2004).

The question we address in this paper is: *Is it possible to create a procedure which will go beyond existing techniques and learn in a semi-supervised manner for a given term all verb relations associated with it?*

The main contributions of the paper are:

- We develop an automatic procedure, which on the fly can learn a diverse set of *verb* and *verb-preposition* relations for a given term.
- We establish the effectiveness of our approach through human-based evaluation.
- We conduct a comparative study with the verb-based relation extraction system ReVerb (Fader et al., 2011) and show that our approach accurately extracts more verb-based relations.
- We also compare the verb relations produced by our system with those available in existing knowledge bases, and observe that despite their completeness these repositories lack many verb-based relations.

The rest of the paper is organized as follows. Next, we present related work. Section 3 outlines the verb-based relation learner. Section 4 describes the data collection process. Section 5 reports on the experimental results. Finally, we conclude in Section 6.

2 Related Work

Lots of attention has been payed on learning *is-a* and *part-of* relations (Hearst, 1992; Girju et al., 2003; Pasca, 2004; Etzioni et al., 2005; Kozareva et al., 2008; Pantel and Pennacchiotti, 2006; Carlson et al., 2009; Talukdar et al., 2008). Others (Ravichandran and Hovy, 2002; Bunescu and Mooney, 2007) have focused on learning specific relations like *BornInYear*, *EmployedBy* and *CorporationAcquired*. However to build a system that can learn a richer set of relations is not trivial, because often labeled training data is required (Kim and Moldovan, 1993; Soderland et al., 1999) and most methods do not scale to corpora where the number of relations is very large or when the relations are not specified in advance (Fader et al., 2011).

However, recently developed OpenIE systems like TextRunner (Banko et al., 2007; Banko, 2009) and ReVerb (Fader et al., 2011) surmount the necessity of labeled data by extracting arbitrary phrases denoting relations in English sentences. (Banko et al., 2007; Banko, 2009) define relation to be any verb-prep, adj-noun construction. While such systems are great at learning general relations, they are not guided but simply gather in an undifferentiated way whatever happens to be contained in their input. In order to be able to extract all verb relations associated with a given term, such systems need to part-of-speech tag and parse a large document collection, then they have to extract all verb constructions and all arguments matching specific sets of patterns which were written by humans (or experts). Finally, they must filter out the information and retrieve only those verb relations that are associated with the specific term. Once compiled the repository is straightforward to query and use, however if a term is not present in the compiled repository, repeating the whole process on a new document collection becomes time consuming and unpractical. The main objective and contribution of our research is the development of a dynamic and flexible knowledge harvesting procedure, which for any given term can learn on the fly verb based relations associated with the term in a very fast and accurate manner.

3 Learning Verb-based Relations

3.1 Problem Formulation

We define our task as given a term, a relation expressed by a verb and a set of prepositions: (1) learn in bootstrapping fashion new relations (i.e. *verbs*) associated with the initial term and filter out erroneous extractions; (2) form triples of the term, the harvested verbs and the initial set of prepositions to learn additional relations (i.e. *verb-prepositions*) and their argument fillers.

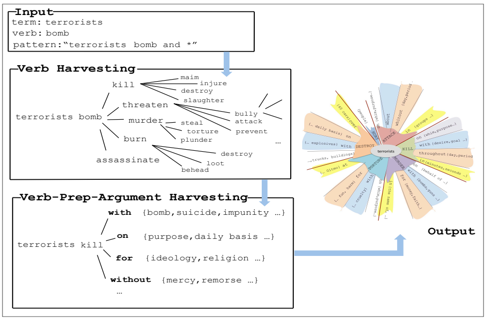

Figure 1: Verb-based Relation Learning.

Figure 1 shows an example for the input term *terrorists*, the verb relation *bomb* and the recursive pattern "*terrorists bomb and **". The algorithm learns on the * position verbs like *kill, murder, threaten, burn, assassinate*. We denote this phase as *verb* extraction. Then each learned verb is used to form triples of the type *term-verb-preposition* to learn new verb-preposition relations and their argument fillers. For instance, "*terrorists kill with **" extracts arguments like {*bombs, suicide, impunity*}. We denote this phase as *verb-preposition* extraction. Finally, the learned relations and arguments are ranked and arranged by their ranking score. The output of this harvesting procedure is triples of the kind "*terrorists kill people*", "*terrorists kill on purpose*", "*terrorists bomb buildings*" among others.

3.2 Algorithm Description

Because of their fixed nature, pattern-based methods often fail to extract information from small corpus or single document. However, nowadays we dispose of endless amount of data, which is easily accessible and is making it possible for such systems to work successfully by scanning billions of Web pages to extract the necessary information. Many of the existing and most accurate is-a relation learners rely on lexico-syntactic patterns (Hearst, 1992; Pasca, 2004; Etzioni et al., 2005), therefore we decided to use patterns for the verb extraction procedure.

PHASE1: Learning *Verb* Relations. The first phase of the algorithm focuses on verb extraction. We use (Kozareva et al., 2008) recursive DAP pattern for is-a relation learning and adapted it to verb extraction as follows: "*<seed-term>* *<seed-verb>* and *", where *<seed-term>* is any term (noun) given by the user or taken from an existing knowledge base, *<seed-verb>* is a seed relation expressed through a verb and * indicates the position on which new verbs will be extracted. The generated patterns are submitted to the search engine as a web query and all retrieved snippets are kept. The algorithm extracts on the position of the $*$ all verb constructions and if they were not previously explored by the algorithm, they are placed on the *<seed-verb>* position of DAP and used as seeds in the subsequent verb extraction iteration. The harvesting terminates when there are no more verbs to be explored. Following (Kozareva et al., 2008), we filter out erroneous extractions using graph ranking. We build a directed graph $G = (V, E)$, where each node $v \in V$ is an extracted verb candidate and $(u, v) \in E$ is an edge between two verb nodes indicating that the verb u lead to the extraction of the verb v. Each node u in the graph is ranked as $u = \sum_{\forall(u,v)\in E}(u, v)$. Confidence in u increases when u extracts more verbs.

PHASE2: Learning *Verb-Preposition* Relations. In the second phase, the learned verbs are paired with an initial set of 17 prepositions to learn new relations and argument fillers. The prepositions were taken from the SemEval 2007 task on preposition disambiguation (Litkowski and Hargraves, 2007). To extract more relations, the algorithm uses the pattern "*<seed-term>* *<verb>* *<prep>* *", where *<seed-term>* is the initial term for which we want to learn verb-based relations, *<verb>* are the leaned verbs from the previous phase and * is the position of the argument fillers. Given the relation *kill* for the term *terrorists*, new relations like *terrorists kill on*, *terrorists kill with*, *terrorists kill for* and *terrorists kill without* are instantiated[1]. Similarly to the verb extraction phase, we rank terms by building a bipartite graph $G' = (V', E')$ with two types of nodes. One set represents the verbs and verb-prepositions V, and the other set represents the arguments A. An edge $e'(v, a) \in E'$ between $v \in V$ and $a \in A$ shows that the verb (or verb-prep) v extracted the argument a. Each argument is ranked as $a = \sum_{\forall(v,a)\in E'}(v, a)$. Confidence in a increases when a is extracted multiple times by different verbs.

4 Data Collection

It is impossible to collect and report results for *all* terms in the world. Still to evaluate the effectiveness of our verb-based relation learner, we have randomly selected 36 terms, which capture daily activities like going to a restaurant to unpleasant events like bombing. For the purpose of visualization, we have organized the terms into the following groups (topics): *Bombing*, *Diseases*, *Elections*, *Restaurants*, and *Animals*.

Table 1 shows the terms and seed verbs used to initiate the verb-based relation learning process, and summarizes the obtained results and the total number of iterations which were run to extract the verbs. *#Verbs Unique* shows the number of unique verbs after merging expressions

[1]Some verbs cannot be paired with all prepositions, we filter out those for which no results were found.

Seed Term	Seed Verb	#Verbs Learned	#Verbs Unique	#Iter.	#Args. Learned	#Args. with a >5
BOMBING						
authorities	say	3049	1805	14	7284	151
bomb	explodes	1020	705	11	13454	451
bombers	explode	265	224	19	9097	344
killers	kill	178	163	14	6906	217
soldiers	die	4588	2533	10	34330	1010
terrorists	kill	1401	941	10	13698	468
victims	suffer	1861	1263	13	21982	767
totalDomain	6	12362	7632	–	106751	3408
DISEASE						
bacteria	caused	1439	853	10	39573	1261
cancer	caused	1389	848	7	42640	1585
diseases	caused	792	582	12	38307	1387
doctors	cure	2700	1611	10	56935	1050
drugs	caused	1936	1242	9	60393	1890
nurses	help	1882	1167	8	39305	675
patient	lives	1631	923	9	78946	1668
virus	caused	1835	992	10	43481	1372
totalDomain	4	13604	8218	–	399580	9838
ELECTION						
candidates	vote	2116	1299	8	55009	1078
congressmen	say	92	86	9	5601	123
senators	vote	718	510	16	12385	340
presidents	run	717	535	11	18476	420
voters	vote	1400	935	13	38298	785
totalDomain	3	5043	3365	–	129769	2746
RESTAURANT						
drinks	tasted	881	591	11	39086	1088
food	tasted	984	664	8	74399	1740
meals	tasted	775	562	10	48474	1144
menu	looks	1479	870	11	51278	1041
restaurants	serve	711	532	8	36120	776
waiters	serve	123	107	9	8457	151
totalDomain	3	4953	3326	–	257814	5940
ANIMALS						
ants	eat	827	607	12	25046	753
birds	eat	3623	2064	8	62031	1465
dinosaurs	eat	544	386	11	11013	345
jellyfish	eat	12	11	4	1120	20
lice	eat	42	42	8	3330	131
mammals	eat	338	272	10	14224	527
otters	eat	190	159	8	5051	159
sharks	eat	697	500	12	16942	598
slugs	eat	60	60	11	5223	89
vultures	eat	36	36	5	2757	67
totalDomain	1	6369	4137	–	146737	4154

Table 1: *Tested Terms for Verb-based Relation Learning and Extracted Information.*

like (*were killed*, *are killed*, *killed*). For each domain, we also show the total number of verbs used to initiate the harvesting process and the total number of learned information. In total, we have submitted $\sim 101,559$ queries and we have collected 10.3GB snippets, which were cleaned, part-of-speech tagged (Schmid, 1994) and used for the extraction of the verb-based relations and arguments. In total for all terms the algorithm extracted $26,678$ candidate relations and $1,040,651$ candidate arguments of which $26,086$ have rank $a>5$.

5 Evaluation and Results

In this section, we evaluate the results of the verb-based relation learning procedure, which is extremely challenging because there is no universal knowledge repository against which one can compare performance in terms of precision and recall. To the extend to which it is possible, we conduct a human-based evaluation and we compare results to knowledge bases that have been extracted in a similar way (i.e., through pattern application over unstructured text).

5.1 Human-based Evaluation

Among the most common approaches on evaluating the correctness of the harvested information is by using human annotators (Pantel and Pennacchiotti, 2006; Navigli et al., 2011). Conducting such evaluations is very important, because the harvested information is often used by QA, machine reading and IE systems (Ferrucci et al., 2010; Freedman et al., 2011).

Since the evaluation of all $1,067,329$ harvested terms is time consuming and costly, we decided to annotate for each term 100 verb relations and argument fillers. We conducted two separate annotations for the verbs and arguments, which resulted in 7200 annotations. We used two annotators who were instructed to mark as incorrect verbs (and argument fillers) that do not correspond to the term. For instance, "*drugs affect*" is marked as correct, while "*drugs discuss*" is marked as incorrect. We compute *Accuracy* as the number of *Correct* terms, divided by the total number of terms used in the annotation. Table 2 shows the accuracy of each domain at different ranks. The overall performance of our relation learner is .95 at rank 100 for the learned verbs and argument fillers. Tables 3 and 4 show examples of the harvested information.

5.2 Comparison with Existing Knowledge Bases

In this evaluation, we measure the ability of our system to learn verb-based relations of a term with respect to already existing knowledge bases, which have been created in a similar way. However, such comparative evaluations are not always possible to perform, because researchers have not fully explored the same terms and relations we have studied. When we compared results against existing knowledge bases, we noticed that Yago (Suchanek et al., 2007) has more detailed information for the arguments of the verb relations rather than the verb relations themselves. Repositories like ConceptNet[2] (Liu and Singh, 2004) contain 1.6 million assertions, however they only belong to twenty relation types such as *is-a*, *part-of*, *made-of*, *effect-of* among others. The only repository that we found with a diverse set of verb relations is the never-ending language learner NELL[3] (Carlson et al., 2009). However, there were only 11 verb relations for *bomb* and 2 verb relations for *virus*. This analysis shows that despite their completeness and richness, existing knowledge repositories can be further enriched with verb-based relations produced by our learning procedure.

[2] http://web.media.mit.edu/h̃ugo/conceptnet/#overview
[3] Comparison done in March 2012 with http://rtw.ml.cmu.edu/rtw/kbbrowser/

Term	Accuracy Verbs			Accuracy Arguments		
	@10	@50	@100	@10	@50	@100
BOMBING						
authorities	1	1	1	1	1	.90
soldiers	1	1	1	1	1	.97
killers	1	.98	.99	1	1	.96
Av.Domain	1	.98	.98	1	1	.97
DISEASE						
diseases	1	.98	.95	1	1	.94
virus	1	.94	.93	1	1	.93
drugs	1	.92	.94	1	1	.93
Av.Domain	.99	.97	.96	1	1	.93
ELECTION						
candidates	1	1	1	1	1	1
voters	1	1	1	1	1	1
senators	1	1	.95	1	1	.97
Av.Domain	1	.99	.95	1	1	.96
RESTAURANT						
food	1	1	.93	1	1	.94
restaurants	1	.94	.89	1	1	.98
menu	1	.92	.89	1	1	.95
Av.Domain	1	.94	.89	1	1	.95
ANIMALS						
otters	1	1	.96	1	1	.94
mammals	1	1	.95	1	1	.95
sharks	1	1	.98	1	1	1
Av.Domain	1	.99	.96	1	1	.92

Table 2: *Accuracy of the Harvested Information.*

Term	Learned Verbs
diseases	spread, develop, treat, come, kill, mutate, diagnose, evolve, are caught, survive, grow, occur, carry, cause, are cured, affect, are identified, start, prevent, propagate, are transmitted, thrive, sicken, change, flourish
meals	are prepared, are served, are cooked, are delivered, are planned, are eaten, are tasted, are provided, look, are made, are consumed, are offered, are created, are frozen, are bought, are packed, are paid, smell, are designed, are purchased, are sold, are produced, are prepped, are shared, are catered
soldiers	kill, shoot, beat, fought, fell, destroyed, fired, attacked, are trained, died, took, said, laughed, kicked, die, were humiliating, cheered, mocked, raised, drummed, captured, looted, ran, arrested, buried, defended

Table 3: *Examples of Learned Verbs.*

5.3 Comparison with Existing Relation Learner

For our comparative study with existing systems, we used ReVerb[4] (Fader et al., 2011), which similarly to our approach was specifically designed to learn verb-based relations from unstructured texts. Currently, ReVerb has extracted relations from ClueWeb09[5] and Wikipedia, which have been freely distributed to the public. ReVerb learns relations by taking as input any document and applies POS-tagging, NP-chunking and a set of rules over all sentences in the document to generate triples containing the verbs and the arguments associated with them. According to (Fader et al., 2011) ReVerb outperforms TextRunner (Banko et al., 2007) and the open Wikipedia extractor WOE (Wu and Weld, 2010) in terms of the quantity and quality of the learned relations. For comparison, we took five terms from our experiment: *ant, bomb, president, terrorists, virus* and collected all verbs found by ReVerb in the ClueWeb09 and Wikipedia triples.

Table 5 summarizes the total number of unique verb extractions found by ReVerb in ClueWeb09 since the Wikipedia ones had low coverage. We have also manually validated the correctness of the verbs found by ReVerb and have seen that their accuracy is 100%. With respect to our extractions ReVerb has lower recall.

[4]http://reverb.cs.washington.edu/
[5]http://lemurproject.org/clueweb09.php/

Term-Verb	Preposition	Learned Arguments
terrorists communicate	**through**	violence, micro technology, orkut secure channels, email, internet, internet networks, cellphones
	with	their contacts, each other, the world, other terrorists, US citizens, Korea, governments, America
	in	brief, code, VW, Russian, French, various ways, secret, English
	by	mail, phone, fax, email
	without	detection, tapping calls
birds fly	**above**	earth, castles, our heads, trees, lake, field, river, cloud, city
	through	air, night, sky, park, country club, wind, storm, region, city
	around	her, fish, house, my head, bird feeder, home, your city, ruins, place
	across	sky, gulf, screen, rainbow, sunset, horizon, african savanna, our path, street, hometown
	into	windows, walls, power lines, towers, sun, sea, darkness, mist, house
killers kill	**for**	power, thrill, sexual reasons, money, fun, the sake, rush, sport, cash, fame
	in	ridiculous ways, patterns, cold blood, silence, groups, conflict with, series, certain periods, captivity, sequence
	with	some criteria, knife, brutality, hands, motive, intention, impunity, stealth, purpose, violence
	to	relieve themselves, symbolize, show others, make a statement, just kill, gain money, gain identity, gain control, gain material
	over	a period, time, robberies, course, many months, multiple time

Table 4: *Examples of Learned Arguments.*

Term	ClueWeb (ReVerb)	Web (DAP)
ants	32	607
bomb	46	535
presidents	32	705
terrorists	96	941
virus	128	992

Table 5: *Comparison of Verb-based Relation Learners.*

6 Conclusion

Our key contribution is the development of a semi-supervised procedure, which starts with a term and a verb to learn from Web documents a large and diverse set of verb relations. We have conducted an experimental evaluation with 36 terms and have collected $26,678$ unique candidate verbs and $1,040,651$ candidate argument fillers. We have evaluated the accuracy of our approach using human based evaluation and have compared results against the ReVerb (Fader et al., 2011) system and existing knowledge bases like NELL (Carlson et al., 2009), Yago (Suchanek et al., 2007) and ConceptNet (Liu and Singh, 2004). Our study showed that despite their completeness these resources lack verb-based information and there is plenty of room for improvement since they can be further enriched with verbs using our harvesting procedure. In the future, we would like to test the usefulness of the generated resources in NLP applications.

Acknowledgements

We would like to thank Ed Hovy for initial comments on the work and the anonymous reviewers.

References

Agirre, E. and Lacalle, O. L. D. (2004). Publicly available topic signatures for all wordnet nominal senses.

Alfonseca, E., Pasca, M., and Robledo-Arnuncio, E. (2010). Acquisition of instance attributes via labeled and related instances. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 58–65.

Banko, M. (2009). Open information extraction from the web. In *Ph.D. Dissertation from University of Washington*.

Banko, M., Cafarella, M. J., Soderl, S., Broadhead, M., and Etzioni, O. (2007). Open information extraction from the web. In *In IJCAI*, pages 2670–2676.

Buitelaar, P., Cimiano, P., and Magnini, B., editors (2005). *Ontology Learning from Text: Methods, Evaluation and Applications*, volume 123 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, Amsterdam.

Bunescu, R. and Mooney, R. (2007). Learning to extract relations from the web using minimal supervision. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 576–583.

Carlson, A., Betteridge, J., Jr., E. R. H., and Mitchell, T. M. (2009). Coupling semi-supervised learning of categories and relations. In *Proceedings of the NAACL HLT 2009 Workskop on Semi-supervised Learning for Natural Language Processing*.

Cuadros, M. and Rigau, G. (2008). KnowNet: A Proposal for Building Highly Connected and Dense Knowledge Bases from the Web. In *Semantics in Text Processing. STEP 2008 Conference Proceedings*, volume 1 of *Research in Computational Semantics*, pages 71–84.

Etzioni, O., Cafarella, M., Downey, D., Popescu, A.-M., Shaked, T., Soderland, S., Weld, D. S., and Yates, A. (2005). Unsupervised named-entity extraction from the web: an experimental study. *Artificial Intelligence*, 165(1):91–134.

Fader, A., Soderland, S., and Etzioni, O. (2011). Identifying relations for open information extraction. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011*, pages 1535–1545.

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., Lally, A., Murdock, J. W., Nyberg, E., Prager, J., Schlaefer, N., and Welty, C. (2010). Building watson: An overview of the deepqa project. *AI Magazine*, 31(3):59–79.

Freedman, M., Ramshaw, L. A., Boschee, E., Gabbard, R., Kratkiewicz, G., Ward, N., and Weischedel, R. M. (2011). Extreme extraction - machine reading in a week. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011*, pages 1437–1446.

Girju, R., Badulescu, A., and Moldovan, D. (2003). Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 1–8.

Girju, R., Nakov, P., Nastaste, V., Szpakowicz, S., Turney, P., and Yuret, D. (2007). SemEval-2007 task 04: Classification of semantic relations between nominals. In *SemEval 2007*.

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics*, pages 539–545.

Igo, S. and Riloff, E. (2009). Corpus-based semantic lexicon induction with web-based corroboration. In *Proceedings of the Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics*.

Jain, A. and Pantel, P. (2010). Factrank: Random walks on a web of facts. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 501–509.

Katz, B. and Lin, J. (2003). Selectively using relations to improve precision in question answering. In *In Proceedings of the EACL-2003 Workshop on Natural Language Processing for Question Answering*, pages 43–50.

Kim, J. and Moldovan, D. (1993). Acquisition of semantic patterns for information extraction from corpora. In *Proceedings of Ninth IEEE Conference on Artificial Intelligence for Applications*, page 17176.

Kozareva, Z. and Hovy, E. (2010). Learning arguments and supertypes of semantic relations using recursive patterns. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 1482–1491.

Kozareva, Z., Riloff, E., and Hovy, E. (2008). Semantic class learning from the web with hyponym pattern linkage graphs. In *Proceedings of ACL-08: HLT*, pages 1048–1056.

Lin, C.-Y. and Hovy, E. (2000). The automated acquisition of topic signatures for text summarization. In *Proceedings of the 18th conference on Computational linguistics - Volume 1*, COLING '00, pages 495–501.

Lin, D. and Pantel, P. (2002). Concept discovery from text. In *Proceedings of the 19th international conference on Computational linguistics*, pages 1–7.

Litkowski, K. C. and Hargraves, O. (2007). Semeval-2007 task 06: Word-sense disambiguation of prepositions. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 24–29.

Liu, H. and Singh, P. (2004). Focusing on ConceptNet's natural language knowledge representation. In *Commonsense Reasoning in and over Natural Language Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES 2004)*, pages 71–84.

Navigli, R., Velardi, P., and Faralli, S. (2011). A graph-based algorithm for inducing lexical taxonomies from scratch. In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 1872–1877.

Pantel, P. and Pennacchiotti, M. (2006). Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, ACL 2006*.

Pasca, M. (2004). Acquisition of categorized named entities for web search. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pages 137–145.

Ravichandran, D. and Hovy, E. (2002). Learning surface text patterns for a question answering system. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 41–47.

Resnik, P. (1996). Selectional constraints: an information-theoretic model and its computational realization.

Riloff, E. (1996). Automatically generating extraction patterns from untagged text. In *Proceedings of the thirteenth national conference on Artificial intelligence - Volume 2*, AAAI'96, pages 1044–1049.

Riloff, E. and Jones, R. (1999). Learning dictionaries for information extraction by multi-level bootstrapping. In *AAAI '99/IAAI '99: Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence*.

Ritter, A., Mausam, and Etzioni, O. (2010). A latent dirichlet allocation method for selectional preferences. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL 2010, pages 424–434.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees.

Sekine, S. (2006). On-demand information extraction. In *Proceedings of the COLING/ACL on Main conference poster sessions*, COLING-ACL '06, pages 731–738.

Snow, R., Jurafsky, D., and Ng, A. Y. (2006). Semantic taxonomy induction from heterogenous evidence. In *Proceedings of 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, ACL*.

Soderland, S., Cardie, C., and Mooney, R. (1999). Learning information extraction rules for semi-structured and free text. In *Machine Learning*, pages 233–272.

Suchanek, F. M., Kasneci, G., and Weikum, G. (2007). Yago: a core of semantic knowledge. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 697–706.

Szpektor, I., Tanev, H., Dagan, I., and Coppola, B. (2004). Scaling web-based acquisition of entailment relations. In *Proc. Empirical Methods in Natural Language Processing (EMNLP)*.

Talukdar, P. P., Reisinger, J., Pasca, M., Ravichandran, D., Bhagat, R., and Pereira, F. (2008). Weakly-supervised acquisition of labeled class instances using graph random walks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP 2008*, pages 582–590.

Widdows, D. (2003). Unsupervised methods for developing taxonomies by combining syntactic and statistical information. In *Proceedings of HLT-NAACL*.

Wu, F. and Weld, D. S. (2010). Open information extraction using wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 118–127.