LEARNING OPEN DOMAIN KNOWLEDGE FROM TEXT

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Gábor György Angeli
May 2016

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Christopher D. Manning) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Percy Liang)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Dan Jurafsky)

Approved for the Stanford University Committee on Graduate Studies

Preface

This thesis describes a new technique for learning open-domain knowledge from unstructured web-scale text corpora, making use of a probabilistic relaxation of natural logic – a logic which uses the syntax of natural language as its logical formalism. We begin by reviewing the theory behind natural logic, and propose a novel extension of the logic to handle propositional formulae.

We then show how to capture common sense facts: given a candidate statement about the world and a large corpus of known facts, is the statement likely to be true? This is treated as a search problem from the query statement to its appropriate support in the knowledge base over valid (or approximately valid) natural logical inference steps. This approach achieves a 4x improvement at retrieval recall compared to lemmatized lookup, maintaining above 90% precision.

We then extend the approach to handle longer, more complex premises by segmenting these utterance into a set of atomic statements entailed through natural logic. We evaluate this system in isolation by using it as the main component in an Open Information Extraction system, and show that it achieves a 3% absolute improvement in F1 compared to prior work on a competitive knowledge base population task.

Finally, we address how to elegantly handle situations where we could not find a supporting premise for our query. To address this, we create an analogue of an evaluation function in gameplaying search: a shallow lexical classifier is folded into the search program to serve as a heuristic function to assess how likely we would have been to find a premise. Results on answering 4th grade science questions show that this method improves over both the classifier in isolation, a strong IR baseline, and prior work.

Acknowledgements

Contents

List of Tables

List of Figures

Chapter 1

Introduction

At its core, machine learning driven natural language processing aims to immitate human intelligence by observing a human perform a given task repeatedly, and training from this data. For example, in order to train a system to recognize whether a given word is the name of a person, we would first collect a large set of words, labelled as either people or not. A system would then take this data, and learn a *model* that can then predict, on unseen words, whether it is a person or not. The great advantage of this framework is that it frees us from having to have a deep understanding of the underlying process by which humans perform the target task, instead allowing us to observe examples and use this to learn to replicate the task. In fact, this has been responsible for much of the progress in natural language processing in the past two decades. The field is finally at the point where many of the core NLP tasks can be done with high accuracy, and many of the higher-level tasks (relation extraction, sentiment analysis, question-answering, etc) have matured to the point of being useful as off-the-shelf components for both academia and industry.

With these advancements, I belive the next question should turn back to a relatively neglected topic: how do we begin to create programs that exhibit *general purpose* intelligence? In many ways, data driven natural language processing systems can be thought of as idiot savants: these systems perform at impressive accuracies at very narrow tasks – the tasks they were trained to repliate – but are incapable of either generalizing across tasks, or performing complex common-sense inferences. For example, we can list off some questions which are trivial for humans to answer, but are very difficult for a trained system without either (1) very specific, narrow, and deep training data, or (2) a very large amount of general-purpose background knowledge:

> ***I ate a bowl of soup, and put the bowl in the bathtub. Did it float?***
> Answering this question correctly requires not only a fairly complex bit of inference, but also a large amount of varried background knowledge: a bowl is concave, empty concave things float, if you eat soup the bowl becomes empty, bathtubs are full of water, etc.
>
> ***I left water in the freezer; what happened to it?***
> Here again, we need to know that freezers are cold (below freezing), that water turns to ice when it's below freezing, and that water turning to ice is more informative than the other things that also "happen" to it, such as it getting cold, or getting dark, or no longer sloshing, etc.
>
> ***The Congressman resigned to go back to governing his hometown. What is his new title?***
> To correctly answer "mayor," we would have to know that if someone resigns from a title, he no longer holds it, and that a mayor governs a town. Also, that a hometown is a city or other entity with a mayor – unlike, say, homework or downtown.

A central tennant of this thesis is that, if we're in pursuit of general intelligence, we should be aiming to answer these sorts of questions not by collecting narrow deep training sets, but rather by developing techniques to collect and then leverage this sort of common-sense information at scale. Furthermore, the most promising way to collect this sort of common-sense knowledge is from text. Natural language is the de-facto standard for storing and transmitting information. Textual data stores information about people and places (e.g., *Obama was born in Hawaii*), facts about science and engineering (e.g., *Ice is frozen water*), or simply common-sense facts (e.g., *Some mushrooms are poisonous*). With the internet, we have unprecedented access to a huge – and growing – amount of text. This presents an immediate practical concern: it becomes infeasible for humans to digest and catalog this influx of information. Attempts at manually extracting knowledge (e.g., Freebase) have led to knowledge bases which are both woefully incomplete and quickly become outdated. In medicine, half of the medical school curriculum becomes obsolete within 5 years of graduation,[1] requiring constant updating. MEDLINE counts 800,000 biomedical articles published in 2013 alone.[2] In academia, studies show that up to 90% of papers are never cited, suggesting that many are never read. In this thesis, I will describe theoretically sound inference methods that can leverage unstructured text for knowledge extraction [1, 2].

A key challenge in this research direction is the ability to use a large corpus of plain text to query facts and answer questions which are not verbatim expressed in the text. For example, a statement "*the cat ate a mouse*" should support even lexically dissimilar queries like "*carnivores eat animals*" and reject logically contradicted queries (like "*no carnivores eat animals*"). Or, a long sentence from Wikipedia may include additional information besides the part that supports the query. This contrasts with information retrieval (IR), which simply retrieves lexically similar passages.

A natural formalism for addressing this challenge is *natural logic* – a proof theory over the syntax of natural language. The logic offers computational efficiency and eliminates the need for semantic parsing and domain-specific meaning representations, while still warranting most common language inferences (e.g., negation). Furthermore, the inferences warranted by the logic tend to be the same inferences that are cognitively easy for humans – that is, the inferences humans assume a reader will effortlessly make.

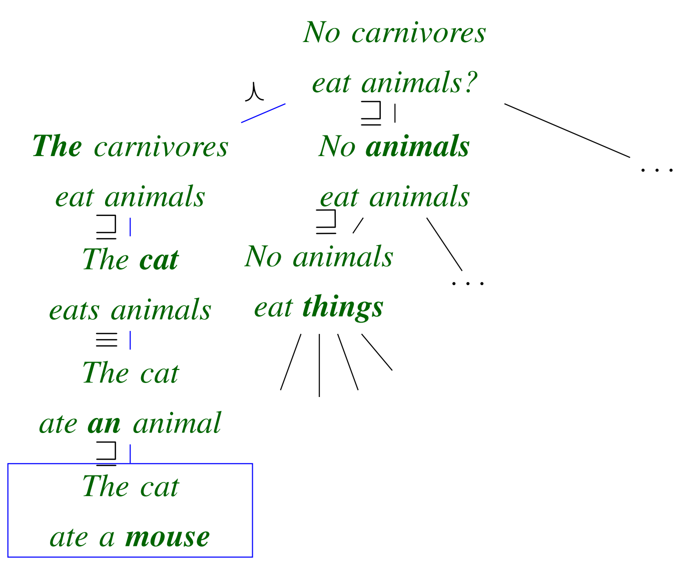
This thesis work explores how to leverage natural logic as a formalism for extracting knowledge not only when it is verbatim written in text, but also when it is implied by some statement in the text. In the subsequent chapters, we will review the theory behind natural logic (Chapter **??**), and then describee a system to (1) extract common-sense knowledge from a large corpus of unannotated text via a search procedure over a soft relaxation of natural logic; (2) simplify complex syntactic structures into maximally informative atomic statements, and (3) incorporate an entailment classifier into this search to serve as an informed backoff.

[1] http://uvamagazine.org/articles/adjusting_the_prescription/

[2] https://www.nlm.nih.gov/bsd/medline_cit_counts_yr_pub.html

In Chapter **??** we introduce our general framework inferring the truth or falsehood of common-sense facts from a very large knowledge base of statements about the world. For example, if a premise "*the cat ate a mouse*" is present in the knowledge base, we should conclude that a hypothesis "*no carnivores eat animals*" is false. The system constructs a search problem for each queried hypothesis over relaxed natural logic inferences: the surface form of the hypothesis is allowed to mutate until it matches one of the facts in the knowledge base. These mutations correspond to steps in a natural logic proof; a learned cost for each mutation corresponds to the system's confidence that the mutation is indeed logically valid (e.g., mutating to a hypernym has low cost, whereas nearest neighbors in vector space has high cost). This amounts to high-performance fuzzy theorem prover over an arbitrarily large premise sets. In my experiments, there are 270M premises in the knowledge base and the the system visits 1M candidates per second.

An illustration of a search from the query "*no carnivores eat animals*" is given below, with the appropriate natural logic relation annotated along the edges:



This framing of the problem has a number of advantages: unlike most approaches in textual entailment, it can scale to arbitrarily large knowledge bases. Unlike most approaches in relational inference (e.g., Markov Logic Networks), the runtime of the system decreases as the size of the knowledge base grows, since we can run a shallower search in expectation. From the other direction, unlike information retrieval approaches, we remain sensitive to a notion of entailment rather than simply similarity – for example, we can detect false facts in addition to true ones. In an empirical evaluation, we show that we can recover 50% of common sense facts from a subset of ConceptNet at 90% precision – 4x the recall of querying the knowledge base directly.

A common motif in extracting information from text is the value in converting a complete sentence into a set of atomic propositions. This is relevant not only as a standalone application, but also as a subcomponent in our reasoning engine: it will allow us to digest complex sentences from real-world data sources, and segment them into atomic facts. Chapter **??** describes our system to extract atomic propositions (e.g., "*Obama was born in Hawaii*") from longer, more syntactically difficult sentences (e.g., "*Born in Hawaii, Obama attended Columbia*") by recursively segmenting a dependency tree into a set of self-contained clauses expressing atomic propositions. These clauses are then maximally shortened to yield propositions which are logically entailed by the original sentence, and also maximally concise. For instance, the statement "*anchovies were an ideal topping for Italian sailors*" yields "*anchovies are a topping.*"

In addition to being a component in the reasoning engine described in the previous section, we can directly use this method for Open Information Extraction (Open IE) – a flavor of relation extraction where the relation, subject, and object are all allowed to be *open domain* plain-text strings. On a NIST-run knowledge base population task, we show that our system outperforms UW's 4^{th}generation Open IE system by 3 F_1. Despite not being developed for this task, our system achieves a score halfway between the median and top performing system, outperforming multiple purpose-built systems.

Lastly, a key property of natural logic is its ability to interface nicely with statistical models which featurize the surface form of a sentence. In addition to handling complex premises, we extend the inference system to allow for inexact matches against the premises in the knowledge base (Chapter **??**). This can be thought of as an evaluation function – akin to gameplaying search – which produces a score at each search state for how likely a simple statistical classifier thinks that the state is supported by any premise. This allows us to provide some judgment for every query hypothesis (in contrast to the 50% coverage in the original system over common-sense facts), while still adhering to logically valid inferences where possible and still detecting negation. We evaluate this complete system on 4^{th} grade science exams, and show that we outperform prior work, a strong information retrieval baseline, and a standalone version of the evaluation function. We can achieve a final score of 74% on our practice test, and 67% on unseen test questions.

Together, these contributions form a powerful reasoning engine for inferring open-domain knowledge from very large premise sets. Unlike traditional IR approaches, or shallow classification methods, the system maintains a notion of logical validity (e.g., proper handling of negation); unlike structured logical methods, the system is high-recall and robust to real-world language and fuzzy inferences. From here, the foundation is laid to *leverage* this sort of knowledge in a general way, in pursuit of systems that exhibit broad-domain intelligence. Returning to or examples from the beginning of the section, if we are faced with a question like:

> *I ate a bowl of soup, and put the bowl in the bathtub. Did it float?*

We have a method for finding out that a bowl is concave, empty concave things float, etc.; the remaining task for future work is putting these facts together to be able to reason about this and other complex questions.

Chapter 2

Related Work

2.1 Knowledge Base Population

Knowledge base population is the task of taking a large body of unstructured text, and extracting from it a structured knowledge base. Importantly, the knowledge base has a fixed schema of relations (e.g., *born in*, *spouse of*), usually with associated type signatures. These knowledge bases can then be used as a repository of knowledge for downstream applications – albeit restricted to the schema of the knowledge base itself. In fact, many downstream NLP applications do query large knowledge bases. Prominent examples include question answering systems [43], and semantic parsers [3, 20, 49, 51].

Prior work in this area can be categorized into a number of approaches. The most common of these are *supervised* relation extractors [11, 16, 40], *distantly supervised* relation extractors [10, 28, 39, 45], and rule based systems [9, 15, 38].

Relation extraction can be naturally cast as a supervised classification problem. A corpus of relation mentions is collected, and each mention x is annotated with the relation y, if any, it expresses. The classifier's output is then aggregated to decide the relations between the two entities.

However, annotating supervised training data is generally expensive to perform at large scale. Although resources such as Freebase or the TAC KBP knowledge base have on the order of millions of training tuples over entities it is not feasible to manually annotate the corresponding mentions in the text. This has led to the rise of *distantly supervised*

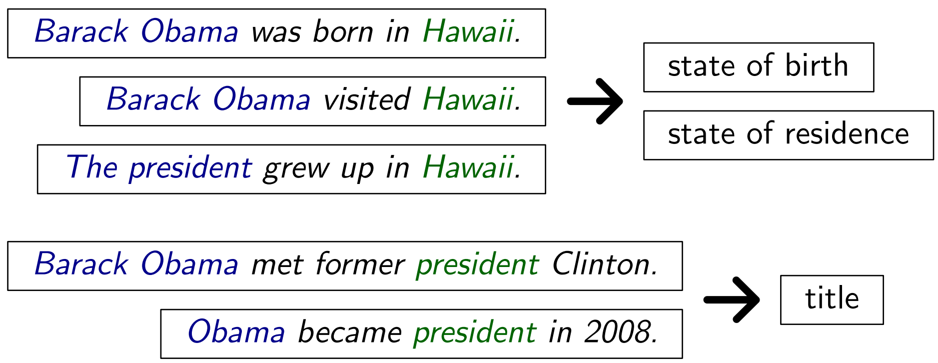


Figure 2.1: The relation extraction setup. For a pair of entities, we collect sentences which mention both entities. These sentences are then used to predict one or more relations between those entities. For instance, the sentences containing both *Barack Obama* and *Hawaii* should support the state of birth and state of residence relation.

methods, which make use of this indirect supervision, but do not necessitate mention-level supervision.

Traditional distant supervision makes the assumption that for every triple (e_1, y, e_2) in a knowledge base between a subject e_1, a relation y, and an object e_2, every sentence containing mentions for e_1 and e_2 express the relation y. For instance, taking Figure 2.1, we would create a datum for each of the three sentences containing BARACK OBAMA and HAWAII labeled with state of birth, and likewise with state of residence, creating 6 training examples overall. Similarly, both sentences involving *Barack Obama* and *president* would be marked as expressing the title relation.

While this allows us to leverage a large database effectively, it nonetheless makes a number of naïve assumptions. First – explicit in the formulation of the approach – it assumes that every mention expresses some relation, and furthermore expresses the known relation(s). For instance, the sentence *Obama visited Hawaii* would be erroneously treated as a positive example of the born in relation. Second, it implicitly assumes that our knowledge base is complete: entity mentions with no known relation are treated as negative examples.

The first of these assumptions is addressed by multi-instance multi-label (MIML) learning, which puts an intermediate latent variable for each sentence-level prediction, that then has to predict the correct knowledge base triples [41]. Min et al. [27] address the second assumption by extending the MIML model with additional latent variables, while **?**] allow feedback from a coarse relation extractor to augment labels from the knowledge base. These latter two approaches are compatible with but are not implemented in this work.

Lastly, there are approaches to inferring new facts in a knowledge base that do not make use of text at all. Chen et al. [8] and Socher et al. [35] use Neural Tensor Networks to predict unseen relation triples in WordNet and Freebase, following a line of work by Bordes et al. [5] and Jenatton et al. [19]. Yao et al. [47] and Riedel et al. [31] present a related line of work, inferring new relations between Freebase entities via inference over both Freebase and OpenIE relations. In contrast, this work runs inference over arbitrary text, without restricting itself to a particular set of relations, or even entities.

2.2 Open Information Extraction

One approach to broad-domain knowledge extraction is *open information extraction* (Open IE). Traditional relation extraction settings usually specify a domain of relations they're interested in (e.g., place of birth, spouse, etc.), and usually place restrictions on the types of arguments extracted (e.g., only people are born places). Open IE systems generalize this setting to the case where both the relation and the arguments are represented as plain text, and therefore can be entirely open-domain. For example, in the sentence *the president spoke to the Senate on Monday*, we might extract the following triples:

(*president*; *spoke to*; *Senate*)
(*president*; *spoke on*; *Monday*)

Open information extraction (open IE) has been shown to be useful in a number of NLP tasks, such as question answering [14] relation extraction [37], and information retrieval [12]. Open IE triples have been also been used in, for example, learning entailment graphs for new triples [4], and matrix factorization for unifying open IE and structured relations [31, 47]. Open-domain triples have also been used to improve knowledge base population, described in Section 2.1. Soderland et al. [37] submitted a system to KBP making use of open IE relations (see below) and an easily constructed mapping to KBP relations. Soderland et al. [36] use ReVerb extractions to enrich a domain-specific ontology. In each of these cases, the concise extractions provided by open IE allow for efficient symbolic methods for entailment, such as Markov logic networks or matrix factorization.

One line of work in this area are the early UW OpenIE systems: for example, TextRunner [48] and ReVerb [13]. In both of these cases, an emphasis is placed on *speed* – token-based surface patterns are extracted that would correspond to open domain triples. With the introduction of fast dependency parsers, [46] extracts triples from learned dependency patterns. Building upon this, Ollie [24] also learns patterns from dependency parses, but with the additional contributions of (1) allowing for extractions which are mediated by nouns or adjectives, not just verbs; and (2) considering context more carefully when extracting these triples. Exemplar [26] adapts the open IE framework to n-ary relationships similar to semantic role labeling, but without the expensive machinery.

In another line of work, The Never Ending Language Learning project (NELL) [7] iteratively learns more facts from the internet from a seed set of examples. In the case of NELL, the ontology is open-domain but fixed, and the goal becomes to learn all entries in the domain. For example, learning an extended hypernymy tree (*Post University* is a University); but also more general relations (*has aqcuired*, *publication writes about*, etc).

2.3 Common Sense Reasoning

The goal of tackling common-sense reasoning is by no means novel in itself either. Work by Reiter and McCarthy [25, 30] attempts to reason about the truth of a consequent in the absence of strict logical entailment. Similarly, Pearl [29] presents a framework for assigning confidences to inferences which can be reasonably assumed. Our approach differs from these attempts in part in its use of Natural Logic as the underlying inference engine, and more substantially in its attempt at creating a broad-coverage system. More recently, work by Schubert [34] and Van Durme et al. [42] approach common sense reasoning with *episodic logic*; we differ in our focus on inferring truth from an arbitrary query, and in making use of longer inferences.

2.4 Textual Entailment

This thesis is in many ways to work on recognizing textual entailment – e.g., Schoenmackers et al. [33], Berant et al. [4]. Textual Entailment is the task of determining if a given premise sentence entails a given hypothesis. That is, if without additional context, a human would infer that the hypothesis is true if the premise is true. For instance:

I drove up to San Francisco yesterday
I was in a car yesterday

Although the definition of entailment is always a bit fuzzy – what if I drove a train up to SF, or pehaps a boat? – nonetheless a reasonable person would assume that if you drove somewhere you were in a car. This sort of reasoning is similar to the goal of this thesis: given premises, infer valid hypothesis to claim as true. However, in RTE the premise set tends to be very small (1 or 2 premises), and the domain tends to have less of a focus on common-sense or broad domain facts.

For example, work by Lewis and Steedman [21] approach entailment by constructing a CCG parse of the query, while mapping questions which are paraphrases of each other to the same logical form using distributional relation clustering. Prior work has used natural logic for RTE-style textual entailment, as a formalism well-suited for formal semantics in neural networks, and as a framework for common-sense reasoning [1, 6, 23, 44]. We adopt the precise semantics of Icard and Moss [18]. Our approach of finding short entailments from a longer utterance is similar in spirit to work on textual entailment for information extraction [32].

2.5 Question Answering

Fader et al. [14] propose a system for question answering based on a sequence of paraphrase rewrites followed by a fuzzy query to a structured knowledge base. This work can be thought of as an elegant framework for unifying this two-stage process, while explicitly tracking the “risk” taken with each paraphrase step. Furthermore, our system is able to explore mutations which are only valid in one direction, rather than the bidirectional entailment of paraphrases, and does not require a corpus of such paraphrases for training.

Many systems make use of structured knowledge bases for question answering. Semantic parsing methods [22, 50] use knowledge bases like Freebase to find support for a complex question. Knowledge base completion (e.g., Chen et al. [8], Bordes et al. [5], or Riedel et al. [31]) can be thought of as entailment, predicting novel knowledge base entries from the original database. In contrast, this work runs inference over arbitrary text without needing a structured knowledge base. Open IE [24, 46] QA approaches – e.g., Fader et al. [14] are closer to operating over plain text, but still requires structured extractions.

The COGEX system [**?**] incorporates a theorem prover into a QA system, boosting overall performance. Similarly, Watson [**?**] incorporates logical reasoning components. This work follows a similar vein, but both the theorem prover and lexical classifier operate over text, without requiring either the premises or axioms to be in logical forms.

On the Aristo corpus we evaluate on in Chapter **??**, Hixon et al. [17] proposes a dialog system to augment a knowledge graph used for answering the questions. This is in a sense an oracle measure, where a human is consulted while answering the question; although, they show that their additional extractions help answer questions other than the one the dialog was collected for.

Bibliography

[1] Gabor Angeli and Christopher D. Manning. Naturalli: Natural logic inference for common sense reasoning. In *EMNLP*, 2014.

[2] Gabor Angeli, Melvin Johnson Premkumar, and Christopher D. Manning. Leveraging linguistic structure for open domain information extraction. In *ACL*, 2015.

[3] J. Berant and P. Liang. Semantic parsing via paraphrasing. In *Association for Computational Linguistics (ACL)*, 2014.

[4] Jonathan Berant, Ido Dagan, and Jacob Goldberger. Global learning of typed entailment rules. In *Proceedings of ACL*, Portland, OR, 2011.

[5] Antoine Bordes, Jason Weston, Ronan Collobert, Yoshua Bengio, et al. Learning structured embeddings of knowledge bases. In *AAAI*, 2011.

[6] Samuel R. Bowman, Christopher Potts, and Christopher D. Manning. Recursive neural networks can learn logical semantics. *CoRR*, (arXiv:1406.1827), 2014.

[7] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka Jr, and Tom M Mitchell. Toward an architecture for never-ending language learning. In *AAAI*, 2010.

[8] Danqi Chen, Richard Socher, Christopher D Manning, and Andrew Y Ng. Learning new facts from knowledge bases with neural tensor networks and semantic word vectors. *arXiv preprint arXiv:1301.3618*, 2013.

[9] Zheng Chen, Suzanne Tamang, Adam Lee, Xiang Li, Wen-Pin Lin, Matthew Snover, Javier Artiles, Marissa Passantino, and Heng Ji. CUNY-BLENDER. In *TAC-KBP*, 2010.

[10] Mark Craven and Johan Kumlien. Constructing biological knowledge bases by extracting information from text sources. In *AAAI*, 1999.

[11] George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie Strassel, and Ralph M Weischedel. The automatic content extraction (ACE) program–tasks, data, and evaluation. In *LREC*, 2004.

[12] Oren Etzioni. Search needs a shake-up. *Nature*, 476(7358):25–26, 2011.

[13] Anthony Fader, Stephen Soderland, and Oren Etzioni. Identifying relations for open information extraction. In *EMNLP*, 2011.

[14] Anthony Fader, Luke Zettlemoyer, and Oren Etzioni. Open question answering over curated and extracted knowledge bases. In *KDD*, 2014.

[15] Ralph Grishman and Bonan Min. New York University KBP 2010 slot-filling system. In *Proc. TAC 2010 Workshop*, 2010.

[16] Zhou GuoDong, Su Jian, Zhang Jie, and Zhang Min. Exploring various knowledge in relation extraction. In *ACL*, 2005.

[17] Ben Hixon, Peter Clark, and Hannaneh Hajishirzi. Learning knowledge graphs for question answering through conversational dialog. *NAACL*, 2015.

[18] Thomas Icard, III and Lawrence Moss. Recent progress on monotonicity. *Linguistic Issues in Language Technology*, 2014.

[19] Rodolphe Jenatton, Nicolas L Roux, Antoine Bordes, and Guillaume R Obozinski. A latent factor model for highly multi-relational data. In *NIPS*, 2012.

[20] Tom Kwiatkowski, Eunsol Choi, Yoav Artzi, and Luke Zettlemoyer. Scaling semantic parsers with on-the-fly ontology matching. 2013.

[21] Mike Lewis and Mark Steedman. Combined distributional and logical semantics. *TACL*, 1:179–192, 2013.

[22] P. Liang, M. I. Jordan, and D. Klein. Learning dependency-based compositional semantics. In *ACL*, 2011.

[23] Bill MacCartney and Christopher D Manning. An extended model of natural logic. In *Proceedings of the eighth international conference on computational semantics*, 2009.

[24] Mausam, Michael Schmitz, Robert Bart, Stephen Soderland, and Oren Etzioni. Open language learning for information extraction. In *EMNLP*, 2012.

[25] John McCarthy. Circumscription—a form of non-monotonic reasoning. *Artificial intelligence*, 1980.

[26] Filipe Mesquita, Jordan Schmidek, and Denilson Barbosa. Effectiveness and efficiency of open relation extraction. In *EMNLP*, 2013.

[27] Bonan Min, Ralph Grishman, Li Wan, Chang Wang, and David Gondek. Distant supervision for relation extraction with an incomplete knowledge base. In *NAACL-HLT*, 2013.

[28] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. Distant supervision for relation extraction without labeled data. In *ACL*, 2009.

[29] Judea Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. *Principles of Knowledge Representation and Reasoning*, 1989.

[30] Raymond Reiter. A logic for default reasoning. *Artificial intelligence*, 13(1):81–132, 1980.

[31] Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M Marlin. Relation extraction with matrix factorization and universal schemas. In *NAACL-HLT*, 2013.

[32] Lorenza Romano, Milen Kouylekov, Idan Szpektor, Ido Dagan, and Alberto Lavelli. Investigating a generic paraphrase-based approach for relation extraction. *EACL*, 2006.

[33] Stefan Schoenmackers, Oren Etzioni, Daniel S Weld, and Jesse Davis. Learning first-order horn clauses from web text. In *EMNLP*, 2010.

[34] Lenhart Schubert. Can we derive general world knowledge from texts? In *HLT*, 2002.

[35] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning with neural tensor networks for knowledge base completion. In *NIPS*, 2013.

[36] Stephen Soderland, Brendan Roof, Bo Qin, Shi Xu, Oren Etzioni, et al. Adapting open information extraction to domain-specific relations. *AI Magazine*, 2010.

[37] Stephen Soderland, John Gilmer, Robert Bart, Oren Etzioni, and Daniel S. Weld. Open information extraction to KBP relations in 3 hours. In *Text Analysis Conference*, 2013.

[38] Stephen G Soderland. *Learning text analysis rules for domain-specific natural language processing*. PhD thesis, University of Massachusetts, 1997.

[39] Ang Sun, Ralph Grishman, Wei Xu, and Bonan Min. New York University 2011 system for KBP slot filling. In *Proceedings of the Text Analytics Conference*, 2011.

[40] Mihai Surdeanu and Massimiliano Ciaramita. Robust information extraction with perceptrons. In *ACE07 Proceedings*, 2007.

[41] Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D. Manning. Multi-instance multi-label learning for relation extraction. In *EMNLP*, 2012.

[42] Benjamin Van Durme, Phillip Michalak, and Lenhart K Schubert. Deriving generalized knowledge from corpora using wordnet abstraction. In *EACL*, 2009.

[43] Ellen M Voorhees. Question answering in TREC. In *Proceedings of the tenth international conference on Information and knowledge management*, 2001.

[44] Yotaro Watanabe, Junta Mizuno, Eric Nichols, Naoaki Okazaki, and Kentaro Inui. A latent discriminative model for compositional entailment relation recognition using natural logic. In *COLING*, 2012.

[45] Fei Wu and Daniel S Weld. Autonomously semantifying wikipedia. In *Proceedings of the sixteenth ACM conference on information and knowledge management*. ACM, 2007.

[46] Fei Wu and Daniel S Weld. Open information extraction using wikipedia. In *ACL*. Association for Computational Linguistics, 2010.

[47] Limin Yao, Sebastian Riedel, and Andrew McCallum. Probabilistic databases of universal schema. In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction*, 2012.

[48] Alexander Yates, Michael Cafarella, Michele Banko, Oren Etzioni, Matthew Broadhead, and Stephen Soderland. TextRunner: Open information extraction on the web. In *ACL-HLT*, 2007.

[49] John M. Zelle and Raymond J. Mooney. Learning to parse database queries using inductive logic programming. In *AAAI/IAAI*, Portland, OR, 1996.

[50] Luke S. Zettlemoyer and Michael Collins. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *UAI*. AUAI Press, 2005.

[51] Luke S. Zettlemoyer and Michael Collins. Online learning of relaxed CCG grammars for parsing to logical form. In *EMNLP-CoNLL*, 2007.