Machine Reading at the University of Washington

Hoifung Poon, Janara Christensen, Pedro Domingos, Oren Etzioni, Raphael Hoffmann,
Chloe Kiddon, Thomas Lin, Xiao Ling, Mausam, Alan Ritter, Stefan Schoenmackers,
Stephen Soderland, Dan Weld, Fei Wu, Congle Zhang

Department of Computer Science & Engineering
University of Washington
Seattle, WA 98195
{hoifung,janara,pedrod,etzioni,raphaelh,chloe,tlin,xiaoling,mausam,
aritter,stef,soderland,weld,wufei,clzhang}@cs.washington.edu

Abstract

Machine reading is a long-standing goal of AI and NLP. In recent years, tremendous progress has been made in developing machine learning approaches for many of its subtasks such as parsing, information extraction, and question answering. However, existing end-to-end solutions typically require substantial amount of human efforts (e.g., labeled data and/or manual engineering), and are not well poised for Web-scale knowledge acquisition. In this paper, we propose a unifying approach for machine reading by bootstrapping from the easiest extractable knowledge and conquering the long tail via a self-supervised learning process. This self-supervision is powered by joint inference based on Markov logic, and is made scalable by leveraging hierarchical structures and coarse-to-fine inference. Researchers at the University of Washington have taken the first steps in this direction. Our existing work explores the wide spectrum of this vision and shows its promise.

1 Introduction

Machine reading, or learning by reading, aims to extract knowledge automatically from unstructured text and apply the extracted knowledge to end tasks such as decision making and question answering. It has been a major goal of AI and NLP since their early days. With the advent of the Web, the billions of online text documents contain virtually unlimited amount of knowledge to extract, further increasing the importance and urgency of machine reading.

In the past, there has been a lot of progress in automating many subtasks of machine reading by machine learning approaches (e.g., components in the traditional NLP pipeline such as POS tagging and syntactic parsing). However, end-to-end solutions are still rare, and existing systems typically require substantial amount of human effort in manual engineering and/or labeling examples. As a result, they often target restricted domains and only extract limited types of knowledge (e.g., a pre-specified relation). Moreover, many machine reading systems train their knowledge extractors once and do not leverage further learning opportunities such as additional text and interaction with end users.

Ideally, a machine reading system should strive to satisfy the following desiderata:

End-to-end: the system should input raw text, extract knowledge, and be able to answer questions and support other end tasks;

High quality: the system should extract knowledge with high accuracy;

Large-scale: the system should acquire knowledge at Web-scale and be open to arbitrary domains, genres, and languages;

Maximally autonomous: the system should incur minimal human effort;

Continuous learning from experience: the system should constantly integrate new information sources (e.g., new text documents) and learn from user questions and feedback (e.g., via performing end tasks) to continuously improve its performance.

These desiderata raise many intriguing and challenging research questions. Machine reading research at the University of Washington has explored

a wide spectrum of solutions to these challenges and has produced a large number of initial systems that demonstrated promising performance. During this expedition, an underlying unifying vision starts to emerge. It becomes apparent that the key to solving machine reading is to:

1. Conquer the long tail of textual knowledge via a self-supervised learning process that leverages data redundancy to bootstrap from the head and propagates information down the long tail by joint inference;

2. Scale this process to billions of Web documents by identifying and leveraging ubiquitous structures that lead to sparsity.

In Section 2, we present this vision in detail, identify the major dimensions these initial systems have explored, and propose a unifying approach that satisfies all five desiderata. In Section 3, we reivew machine reading research at the University of Washington and show how they form synergistic effort towards solving the machine reading problem. We conclude in Section 4.

2 A Unifying Approach for Machine Reading

The core challenges to machine reading stem from the massive scale of the Web and *the long-tailed distribution of textual knowledge*. The heterogeneous Web contains texts that vary substantially in subject matters (e.g., finance vs. biology) and writing styles (e.g., blog posts vs. scientific papers). In addition, natural languages are famous for their myraid variations in expressing the same meaning. A fact may be stated in a straightforward way such as "kale contains calcium". More often though, it may be stated in a syntactically and/or lexically different way than as phrased in an end task (e.g., "calcium is found in kale"). Finally, many facts are not even stated explicitly, and must be inferred from other facts (e.g., "kale prevents osteoporosis" may not be stated explicitly but can be inferred by combining facts such as "kale contains calcium" and "calcium helps prevent osteoporosis"). As a result, machine reading must not rely on explicit supervision such as manual rules and labeled examples, which will incur prohibitive cost in the Web scale. Instead, it must be able to learn from indirect supervision.
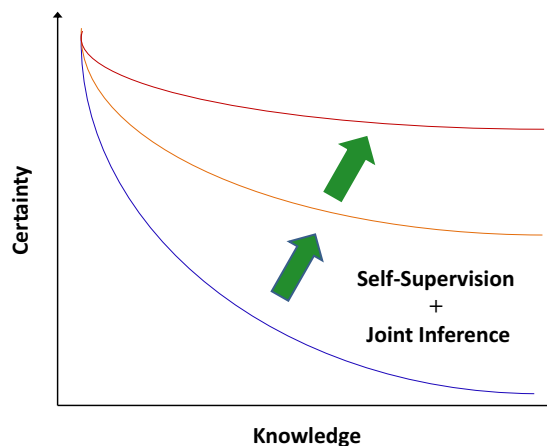

Figure 1: A unifying vision for machine reading: bootstrap from the head regime of the power-law distribution of textual knowledge, and conquer the long tail in a self-supervised learning process that raises certainty on sparse extractions by propagating information via joint inference from frequent extractions.

A key source of indirect supervision is meta knowledge about the domains. For example, the TextRunner system (Banko et al., 2007) hinges on the observation that there exist general, relation-independent patterns for information extraction. Another key source of indirect supervision is data redundancy. While a rare extracted fact or inference pattern may arise by chance of error, it is much less likely so for the ones with many repetitions (Downey et al., 2010). Such highly-redundant knowledge can be extracted easily and with high confidence, and can be leveraged for bootstrapping. For knowledge that resides in the long tail, explicit forms of redundancy (e.g., identical expressions) are rare, but this can be circumvented by joint inference. For example, expressions that are composed with or by similar expressions probably have the same meaning; the fact that kale prevents osteoporosis can be derived by combining the facts that kale contains calcium and that calcium helps prevent osteoporosis via a transitivity-through inference pattern. In general, joint inference can take various forms, ranging from simple voting to shrinkage in a probabilistic ontology to sophisticated probabilistic reasoning based on a joint model. Simple ones tend to scale better, but their capability in propagating information is limited. More sophisticated methods can uncover implicit redundancy and propagate much more in-

formation with higher quality, yet the challenge is how to make them scale as well as simple ones.

To do machine reading, a self-supervised learning process, informed by meta knowledge, stipulates what form of joint inference to use and how. Effectively, it increases certainty on sparse extractions by propagating information from more frequent ones. Figure 1 illustrates this unifying vision.

In the past, machine reading research at the University of Washington has explored a variety of solutions that span the key dimensions of this unifying vision: *knowledge representation, bootstrapping, self-supervised learning, large-scale joint inference, ontology induction, continuous learning.* See Section 3 for more details. Based on this experience, one direction seems particularly promising that we would propose here as our unifying approach for end-to-end machine reading:

Markov logic is used as the unifying framework for knowledge representation and joint inference;

Self-supervised learning is governed by a joint probabilistic model that incorporates a small amount of heuristic knowledge and large-scale relational structures to maximize the amount and quality of information to propagate;

Joint inference is made scalable to the Web by coarse-to-fine inference.

Probabilistic ontologies are induced from text to guarantee tractability in coarse-to-fine inference. This ontology induction and population are incorporated into the joint probabilistic model for self-supervision;

Continuous learning is accomplished by combining **bootstrapping** and **crowdsourced content creation** to synergistically improve the reading system from user interaction and feedback.

A distinctive feature of this approach is its emphasis on using sophisticated joint inference. Recently, joint inference has received increasing interest in AI, machine learning, and NLP, with Markov logic (Domingos and Lowd, 2009) being one of the leading unifying frameworks. Past work has shown that it can substantially improve predictive accuracy in supervised learning (e.g., (Getoor and Taskar, 2007; Bakir et al., 2007)). We propose to build on these advances, but apply joint inference beyond supervised learning, with labeled examples supplanted by indirect supervision.

Another distinctive feature is that we propose to use coarse-to-fine inference (Felzenszwalb and McAllester, 2007; Petrov, 2009) as a unifying framework to scale inference to the Web. Essentially, coarse-to-fine inference leverages the sparsity imposed by hierarchical structures that are ubiquitous in human knowledge (e.g., taxonomies/ontologies). At coarse levels (top levels in a hierarchy), ambiguities are rare (there are few objects and relations), and inference can be conducted very efficiently. The result is then used to prune unpromising refinements at the next level. This process continues down the hierarchy until decision can be made. In this way, inference can potentially be sped up exponentially, analogous to binary tree search.

Finally, we propose a novel form of continuous learning by leveraging the interaction between the system and end users to constantly improve the performance. This is straightforward to do in our approach given the self-supervision process and the availability of powerful joint inference. Essentially, when the system output is applied to an end task (e.g., answering questions), the feedback from user is collected and incorporated back into the system as a bootstrap source. The feedback can take the form of explicit supervision (e.g., via community content creation or active learning) or indirect signals (e.g., click data and query logs). In this way, we can bootstrap an online community by an initial machine reading system that provides imperfect but valuable service in end tasks, and continuously improve the quality of system output, which attracts more users with higher degree of participation, thus creating a positive feedback loop and raising the machine reading performance to a high level that is difficult to attain otherwise.

3 Summary of Progress to Date

The University of Washington has been one of the leading places for machine reading research and has produced many cutting-edge systems, e.g., WIEN (first wrapper induction system for information extraction), Mulder (first fully automatic Web-scale question answering system), KnowItAll/TextRunner (first systems to do open-domain information extrac-

tion from the Web corpus at large scale), Kylin (first self-supervised system for Wikipedia-based information extraction), UCR (first unsupervised coreference resolution system that rivals the performance of supervised systems), Holmes (first Web-scale joint inference system), USP (first unsupervised system for semantic parsing).

Figure 2 shows the evolution of the major systems; dashed lines signify influence in key ideas (e.g., Mulder inspires KnowItAll), and solid lines signify dataflow (e.g., Holmes inputs TextRunner tuples). These systems span a wide spectrum in scalability (assessed by speed and quantity in extraction) and comprehension (assessed by unit yield of knowledge at a fixed precision level). At one extreme, the TextRunner system is highly scalable, capable of extracting billions of facts, but it focuses on shallow extractions from simple sentences. At the other extreme, the USP and LOFT systems achieve much higher level of comprehension (e.g., in a task of extracting knowledge from biomedical papers and answering questions, USP obtains more than three times as many correct answers as TextRunner, and LOFT obtains more than six times as many correct answers as TextRunner), but are much less scalable than TextRunner.

In the remainder of the section, we review the progress made to date and identify key directions for future work.

3.1 Knowledge Representation and Joint Inference

Knowledge representations used in these systems vary widely in expressiveness, ranging from simple ones like relation triples (<*subject, relation, object*>; e.g., in KnowItAll and TextRunner), to clusters of relation triples or triple components (e.g., in SNE, RESOLVER), to arbitrary logical formulas and their clusters (e.g., in USP, LOFT). Similarly, a variety forms of joint inference have been used, ranging from simple voting to heuristic rules to sophisticated probabilistic models. All these can be compactly encoded in Markov logic (Domingos and Lowd, 2009), which provides a unifying framework for knowledge representation and joint inference.

Past work at Washington has shown that in supervised learning, joint inference can substantially improve predictive performance on tasks related to machine reading (e.g., citation information extraction (Poon and Domingos, 2007), ontology induction (Wu and Weld, 2008), temporal information extraction (Ling and Weld, 2010)). In addition, it has demonstrated that sophisticated joint inference can enable effective learning without any labeled information (UCR, USP, LOFT), and that joint inference can scale to millions of Web documents by leveraging sparsity in naturally occurring relations (Holmes, Sherlock), showing the promise of our unifying approach.

Simpler representations limit the expressiveness in representing knowledge and the degree of sophistication in joint inference, but they currently scale much better than more expressive ones. A key direction for future work is to evaluate this tradeoff more thoroughly, e.g., for each class of end tasks, to what degree do simple representations limit the effectiveness in performing the end tasks? Can we automate the choice of representations to strike the best tradeoff for a specific end task? Can we advance joint inference algorithms to such a degree that sophisticated inference scales as well as simple ones?

3.2 Bootstrapping

Past work at Washington has identified and leveraged a wide range of sources for bootstrapping. Examples include Wikipedia (Kylin, KOG, IIA, WOE, WPE), Web lists (KnowItAll, WPE), Web tables (WebTables), Hearst patterns (KnowItAll), heuristic rules (TextRunner), semantic role labels (SRL-IE), etc.

In general, potential bootstrap sources can be broadly divided into domain knowledge (e.g., patterns and rules) and crowdsourced contents (e.g., linguistic resources, Wikipedia, Amazon Mechanical Turk, the ESP game).

A key direction for future work is to combine bootstrapping with crowdsourced content creation for continuous learning. (Also see Subsection 3.6.)

3.3 Self-Supervised Learning

Although the ways past systems conduct self-supervision vary widely in detail, they can be divided into two broad categories. One uses heuristic rules that exploit existing semi-structured resources to generate noisy training examples for use by supervised learning methods and with cotraining (e.g.,

1997   2001   2004   2005   2007   2008   2009   2010

ShopBot   WIEN   Mulder   KnowItAll   Opine   LEX   RESOLVER   TextRunner   REALM   Kylin   Holmes   UCR   SNE   AuContraire   KOG   WebTables   USP   IIA   Sherlock   SRL-IE   PrecHybrid   LOFT   LDA-SP   WOE   WPE   OLPI
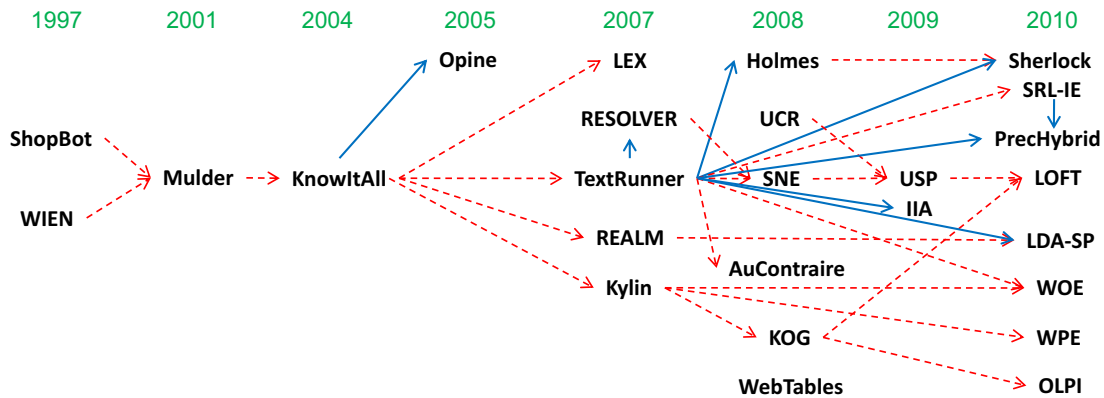
Figure 2: The evolution of major machine reading systems at the University of Washington. Dashed lines signify influence and solid lines signify dataflow. At the top are the years of publications. **ShopBot** learns comparison-shopping agents via self-supervision using heuristic knowledge (Doorenbos et al., 1997); **WIEN** induces wrappers for information extraction via self-supervision using joint inference to combine simple atomic extractors (Kushmerick et al., 1997); **Mulder** answers factoid questions by leveraging redundancy to rank candidate answers extracted from multiple search query results (Kwok et al., 2001); **KnowItAll** conducts open-domain information extraction via self-supervision bootstrapping from Hearst patterns (Etzioni et al., 2005); **Opine** builds on KnowItAll and mines product reviews via self-supervision using joint inference over neighborhood features (Popescu and Etzioni, 2005); **Kylin** populates Wikipedia infoboxes via self-supervision bootstrapping from existing infoboxes (Wu and Weld, 2007); **LEX** conducts Web-scale name entity recognition by leveraging collocation statistics (Downey et al., 2007a); **REALM** improves sparse open-domain information extraction via relational clustering and language modeling (Downey et al., 2007b); **RESOLVER** performs entity and relation resolution via relational clustering (Yates and Etzioni, 2007); **TextRunner** conducts open-domain information extraction via self-supervision bootstrapping from heuristic rules (Banko et al., 2007); **AuContraire** automatically identifies contradictory statements in a large web corpus using functional relations (Ritter et al., 2008); **HOLMES** infers new facts from TextRunner output using Markov logic (Schoenmackers et al., 2008); **KOG** learns a rich ontology by combining Wikipedia infoboxes with WordNet via joint inference using Markov Logic Networks (Wu and Weld, 2008), shrinkage over this ontology vastly improves the recall of Kylin's extractors; **UCR** performs state-of-the-art unsupervised coreference resolution by incorporating a small amount of domain knowledge and conducting joint inference among entity mentions with Markov logic (Poon and Domingos, 2008b); **SNE** constructs a semantic network over TextRunner output via relational clustering with Markov logic (Kok and Domingos, 2008); **WebTables** conducts Web-scale information extraction by leveraging HTML table structures (Cafarella et al., 2008); **IIA** learns from infoboxes to filter open-domain information extraction toward assertions that are interesting to people (Lin et al., 2009); **USP** jointly learns a semantic parser and extracts knowledge via recursive relational clustering with Markov logic (Poon and Domingos, 2009); **LDA-SP** automatically infers a compact representation describing the plausible arguments for a relation using an LDA-Style model and Bayesian Inference (Ritter et al., 2010); **LOFT** builds on USP and jointly performs ontology induction, population, and knowledge extraction via joint recursive relational clustering and shrinkage with Markov logic (Poon and Domingos, 2010); **OLPI** improves the efficiency of lifted probabilistic inference and learning via coarse-to-fine inference based on type hierarchies (Kiddon and Domingos, 2010). **SHERLOCK** induces new inference rules via relational learning (Schoenmackers et al., 2010); **SRL-IE** conducts open-domain information extraction by bootstrapping from semantic role labels, **PrecHybrid** is a hybrid version between SRL-IE and TextRunner, which given a budget of computation time does better than either system (Christensen et al., 2010); **WOE** builds on Kylin and conducts open-domain information extraction (Wu and Weld, 2010); **WPE** learns 5000 relational extractors by bootstrapping from Wikipedia and using Web lists to generate dynamic, relation-specific lexicon features (Hoffmann et al., 2010).

TextRunner, Kylin, KOG, WOE, WPE). Another uses unsupervised learning and often takes a particular form of relational clustering (e.g., objects associated with similar relations tend to be the same and vice versa, as in REALM, RESOLVER, SNE, UCR, USP, LDA-SP, LOFT, etc.).

Some distinctive types of self-supervision include shrinkage based on an ontology (KOG, LOFT, OLPI), probabilistic inference via hand-crafted or learned inference patterns (Holmes, Sherlock), and cotraining using relation-specific and relation-independent (open) extraction to reinforce semantic coherence (Wu et al., 2008).

A key direction for future work is to develop a unifying framework for self-supervised learning by combining the strengths of existing methods and overcoming their limitations. This will likely take the form of a new learning paradigm that combines existing paradigms such as supervised learning, relational clustering, semi-supervised learning, and active learning into a unifying learning framework that synergistically leverages diverse forms of supervision and information sources.

3.4 Large-Scale Joint Inference

To apply sophisticated joint inference in machine reading, the major challenge is to make it scalable to billions of text documents. A general solution is to identify and leverage ubiquitous problem structures that lead to sparsity. For example, order of magnitude reduction in both memory and inference time can be achieved for relational inference by leveraging the fact that most relational atoms are false, which trivially satisfy most relational formulas (Singla and Domingos, 2006; Poon and Domingos, 2008a); joint inference with naturally occurring textual relations can scale to millions of Web pages by leveraging the fact that such relations are approximately functional (Schoenmackers et al., 2008).

More generally, sparsity arises from hierarchical structures (e.g., ontologies) that are naturally exhibited in human knowledge, and can be leveraged to do coarse-to-fine inference (OLPI).

The success of coarse-to-fine inference hinges on the availability and quality of hierarchical structures. Therefore, a key direction for future work is to automatically induce such hierarchies. (Also see next subsection.) Moreover, given the desideratum of

continuous learning from experience, and the speedy evolution of the Web (new contents, formats, etc.), it is important that we develop *online methods* for self-supervision and joint inference. For example, when a new text document arrives, the reading system should not relearn from scratch, but should identify only the relevant pieces of knowledge and conduct limited-scoped inference and learning accordingly.

3.5 Ontology Induction

As mentioned in previous subsections, ontologies play an important role in both self-supervision (shrinkage) and large-scale inference (coarse-to-fine inference). A distinctive feature in our unifying approach is to induce probabilistic ontologies, which can be learned from noisy text and support joint inference. Past systems have explored two different approaches to probabilistic ontology induction. One approach is to bootstrap from existing ontological structures and apply self-supervision to correct the erroneous nodes and fill in the missing ones (KOG). Another approach is to integrate ontology induction with hierarchical smoothing, and jointly pursue unsupervised ontology induction, population and knowledge extraction (LOFT).

A key direction for future work is to combine these two paradigms. As case studies in ontology integration, prior research has devised probabilistic schema mappings and corpus-based matching algorithms (Doan, 2002; Madhavan, 2005; Dong et al., 2007), and has automatically constructed mappings between the Wikipedia infobox "ontology" and the Freebase ontology. This latter endeavor illustrated the complexity of the necessary mappings: a simple attribute in one ontology may correspond to a complex relational view in the other, comprising three join operations; searching for such matches yields a search space with billions of possible correspondences for just a single attribute.

Another key direction is to develop general methods for inducing multi-facet, multi-inheritance ontologies. Although single-inheritance, tree-like hierarchies are easier to induce and reason with, naturally occurring ontologies generally take the form of a lattice rather than a tree.

3.6 Continuous Learning

Early work at Washington proposed to construct knowledge bases by mass collaboration (Richardson and Domingos, 2003). A key challenge is to combine inconsistent knowledge sources of varying quality, which motivated the subsequent development of Markov logic. While this work did not do machine reading, its emphasis on lifelong learning from user feedback resonates with our approach on continuous learning.

Past work at Washington has demonstrated the promise of our approach. For example, (Banko and Etzioni, 2007) automated theory formation based on TextRunner extractions via a lifelong-learning process; (Hoffmann et al., 2009) show that the pairing of Kylin and community content creation benefits both by sharing Wikipedia edits; (Soderland et al., 2010) successfully adapted the TextRunner open-domain information extraction system to specific domains via active learning.

Our approach also resonates with the never-ending learning paradigm for "Reading the Web" (Carlson et al., 2010). In future work, we intend to combine our approach with related ones to enable more effective continuous learning from experience.

4 Conclusion

This paper proposes a unifying approach to machine reading that is end-to-end, large-scale, maximally autonomous, and capable of continuous learning from experience. At the core of this approach is a self-supervised learning process that conquers the long tail of textual knowledge by propagating information via joint inference. Markov logic is used as the unifying framework for knowledge representation and joint inference. Sophisticated joint inference is made scalable by coarse-to-fine inference based on induced probabilistic ontologies. This unifying approach builds on the prolific experience in cutting-edge machine reading research at the University of Washington. Past results demonstrate its promise and reveal key directions for future work.

5 Acknowledgement

References

G. Bakir, T. Hofmann, B. B. Schölkopf, A. Smola, B. Taskar, S. Vishwanathan, and (eds.). 2007. *Predicting Structured Data*. MIT Press, Cambridge, MA.

M. Banko and O. Etzioni. 2007. Strategies for lifelong knowledge extraction from the web. In *Proceedings of the Sixteenth International Conference on Knowledge Capture*, British Columbia, Canada.

Michele Banko, Michael J. Cafarella, Stephen Soderland, Matt Broadhead, and Oren Etzioni. 2007. Open information extraction from the web. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*, pages 2670–2676, Hyderabad, India. AAAI Press.

Michael J. Cafarella, Jayant Madhavan, and Alon Halevy. 2008. Web-scale extraction of structured data. *SIGMOD Record*, 37(4):55–61.

Andrew Carlson, Justin Betteridge, Richard C. Wang, Estevam R. Hruschka Jr., and Tom M. Mitchell. 2010. Coupled semi-supervised learning for information extraction. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*.

Janara Christensen, Mausam, Stephen Soderland, and Oren Etzioni. 2010. Semantic role labeling for open information extraction. In *Proceedings of the First International Workshop on Formalisms and Methodology for Learning by Reading*.

Anhai Doan. 2002. *Learning to Map between Structured Representations of Data*. Ph.D. thesis, University of Washington.

Pedro Domingos and Daniel Lowd. 2009. *Markov Logic: An Interface Layer for Artificial Intelligence*. Morgan & Claypool, San Rafael, CA.

Xin Dong, Alon Halevy, and Cong Yu. 2007. Data integration with uncertainty. *VLDB Journal*.

Robert B. Doorenbos, Oren Etzioni, and Daniel S. Weld. 1997. A scalable comparison-shopping agent for the world-wide web. In *Proceedings of AGENTS-97*.

Doug Downey, Matthew Broadhead, and Oren Etzioni. 2007a. Locating complex named entities in web text. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*.

Doug Downey, Stefan Schoenmackers, and Oren Etzioni. 2007b. Sparse information extraction: Unsupervised language models to the rescue. In *Proceedings of the Forty Fifth Annual Meeting of the Association for Computational Linguistics*.

Doug Downey, Oren Etzioni, and Stephen Soderland. 2010. Analysis of a probabilistic model of redundancy in unsupervised information extraction. In *Artificial Intelligence*.

Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. 2005. Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, 165(1):91–134.

Pedro F. Felzenszwalb and David McAllester. 2007. The generalized A* architecture. *Journal of Artificial Intelligence Research*, 29.

Lise Getoor and Ben Taskar, editors. 2007. *Introduction to Statistical Relational Learning*. MIT Press, Cambridge, MA.

Raphael Hoffmann, Saleema Amershi, Kayur Patel, Fei Wu, James Fogarty, and Daniel S. Weld. 2009. Amplifying community content creation using mixed-initiative information extraction. In *Proceedings of CHI-09*.

Raphael Hoffmann, Congle Zhang, and Daniel S. Weld. 2010. Learning 5000 relational extractors. In *submission*.

Chloe Kiddon and Pedro Domingos. 2010. Ontological lifted probabilistic inference. In *submission*.

Stanley Kok and Pedro Domingos. 2008. Extracting semantic networks from text via relational clustering. In *Proceedings of the Nineteenth European Conference on Machine Learning*, pages 624–639, Antwerp, Belgium. Springer.

Nicholas Kushmerick, Daniel S. Weld, and Robert Doorenbos. 1997. Wrapper induction for information extraction. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*.

Cody Kwok, Oren Etzioni, and Daniel S. Weld. 2001. Scaling question answering to the web. In *Proceedings of the Tenth International Conference on World Wide Web*.

Thomas Lin, Oren Etzioni, and James Fogarty. 2009. Identifying interesting assertions from the web. In *Proceedings of the Eighteenth Conference on Information and Knowledge Management*.

Xiao Ling and Daniel S. Weld. 2010. Temporal information extraction. In *Proceedings of the Twenty Fifth National Conference on Artificial Intelligence*.

Jayant Madhavan. 2005. *Using known schemas and mappings to construct new semantic mappings*. Ph.D. thesis, University of Washington.

Slav Petrov. 2009. *Coarse-to-Fine Natural Language Processing*. Ph.D. thesis, Department of Computer Science, University of Berkeley.

Hoifung Poon and Pedro Domingos. 2007. Joint inference in information extraction. In *Proceedings of the Twenty Second National Conference on Artificial Intelligence*, pages 913–918, Vancouver, Canada. AAAI Press.

Hoifung Poon and Pedro Domingos. 2008a. A general method for reducing the complexity of relational inference and its application to mcmc. In *Proceedings of the Twenty Third National Conference on Artificial Intelligence*, pages 1075–1080, Chicago, IL. AAAI Press.

Hoifung Poon and Pedro Domingos. 2008b. Joint unsupervised coreference resolution with Markov logic. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 649–658, Honolulu, HI. ACL.

Hoifung Poon and Pedro Domingos. 2009. Unsupervised semantic parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1–10, Singapore. ACL.

Hoifung Poon and Pedro Domingos. 2010. Unsupervised ontological induction from text. In *submission*.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of the Joint Conference on Human Language Technology and Empirical Methods in Natural Language Processing*.

Matt Richardson and Pedro Domingos. 2003. Building large knowledge bases by mass collaboration. In *Proceedings of the Second International Conference on Knowledge Capture*, pages 129–137, Sanibel Island, FL. ACM Press.

Alan Ritter, Doug Downey, Stephen Soderland, and Oren Etzioni. 2008. It's a contradiction – no, it's not: A case study using functional relations. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*.

Alan Ritter, Mausam, and Oren Etzioni. 2010. A latent dirichlet allocation method for selectional preferences. In *submission*.

Stefan Schoenmackers, Oren Etzioni, and Daniel S. Weld. 2008. Scaling textual inference to the web. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*.

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

Parag Singla and Pedro Domingos. 2006. Memory-efficient inference in relational domains. In *Proceedings of the Twenty First National Conference on Artificial Intelligence*.

Stephen Soderland, Brendan Roof, Bo Qin, Mausam, and Oren Etzioni. 2010. Adapting open information extraction to domain-specific relations. In *submission*.

Fei Wu and Daniel S. Weld. 2007. Automatically semantifying wikipedia. In *Proceedings of the Sixteenth Conference on Information and Knowledge Management*, Lisbon, Portugal.

Fei Wu and Daniel S. Weld. 2008. Automatically refining the wikipedia infobox ontology. In *Proceedings of the Seventeenth International Conference on World Wide Web*, Beijing, China.

Fei Wu and Daniel S. Weld. 2010. Open information extraction using wikipedia. In *submission*.

Fei Wu, Raphael Hoffmann, and Daniel S. Weld. 2008. Information extraction from Wikipedia: Moving down the long tail. In *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Las Vegas, NV.

Alexander Yates and Oren Etzioni. 2007. Unsupervised resolution of objects and relations on the web. In *Proceedings of Human Language Technology (NAACL)*.