Review of Relation Extraction Methods: What is New Out There?

Natalia Konstantinova

University of Wolverhampton,UK
n.konstantinova@wlv.ac.uk

Abstract. Relation extraction is a part of Information Extraction and an established task in Natural Language Processing. This paper presents an overview of the main directions of research and recent advances in the field. It reviews various techniques used for relation extraction including knowledge-based, supervised and self-supervised methods. We also mention applications of relation extraction and identify current trends in the way the field is developing.

Keywords: relation extraction, information extraction, natural language processing, review

1 Introduction

The modern world is rapidly developing and, in order to keep up-to-date, people must process large volume of information every day. Not only is the amount of this information is constantly increasing but the type of information is changing all the time. As a consequence of the sheer volume and heterogeneous nature of the information it is becoming impossible to analyse this data manually and new techniques are being used to automate this process. The field of Natural Language Processing (NLP) addresses this issue by analysing texts written in natural language and trying to understand them and extract valuable information. The problem of obtaining structured information from the text is dealt by Information Extraction (IE), a field of NLP. In this paper we mainly focus on one stage of IE – Relation Extraction (RE).

This paper is organised in the following way: Section 2 describes in more detail the field of Information Extraction and provides background on all its stages, Section 3 introduces the task of Relation Extraction, Section 4 presents knowledge-based methods, Section 5 describes supervised methods and Section 6 provides more details about the self-supervised approach. Section 7 introduces relation extraction as a part of joint modelling of several stages of IE. The paper finishes with Section 8 which summarises the material presented in the paper and also highlights possible future development of the field.

2 Information extraction

Information extraction (IE) is a field of computational linguistics which plays a crucial role in the efficient management of data. It is defined as "a process of getting structured data from unstructured information in the text" (Jurafsky and Martin, 2009). Grishman (1997) describes this process as "the identification of instances of a particular class of events or relationships in a natural language text, and the extraction of the relevant arguments of the event or relationship". After the information is structured and added to a database it can be used by a wide range of NLP applications, including information retrieval, question answering and many others.

Information extraction challenge has a long history and goes back to the late 1970s (Cowie and Lehnert, 1996); however the first commercial systems appeared only in the 1990s, e.g. JASPER (Andersen et al, 1992), specially built for Reuters. Later research was greatly inspired by a series of Message Understanding Conferences (MUC)[1], which were initiated and financed by the Defense Advanced Research Projects Agency (DARPA) to encourage the development of new methods in information extraction. The importance of the MUCs was not the conferences themselves, but the evaluations and evaluation competitions they proposed (Grishman and Sundheim, 1996). The organisers of these conferences defined tasks for all the participants, prepared the data and developed the evaluation framework for each task. Researchers had to address the task and find the best solution; therefore it added competition element to the research. In addition to all the above-mentioned advantages, these events were an opportunity to get comparable results and evaluate objectively the performance of different systems. MUCs were followed by several ACE (Automatic Content Extraction)[2] evaluations which also provided valuable feedback for researchers.

Usually IE, as many other NLP tasks, can be regarded as a pipeline process, where some kind of information is extracted at each stage. Jurafsky and Martin (2009) point out several different types of information that can be extracted:

- named entities (NE);
- temporal expressions;
- numeric values;
- relations between entities and expressions previously identified;
- events/template filling.

Generally IE starts with the detection and classification of proper names found in the text, which is usually referred to as Named Entity Recognition (NER). Most commonly IE systems search for names of people, companies and organisations, and geographical places. But the choice of the precise kind of NE to be extracted depends greatly on the task and system in mind. Sometimes the notion of Named Entities is extended to include items that are not really names or entities, but bear important information for analysing the texts; therefore, numeric values, such as measurements and prices, or temporal expressions can be included in this category. Extraction of such kinds of data is extremely important for correct analysis of texts and reasoning.

[1] http://www.itl.nist.gov/iaui/894.02/related_projects/muc/
[2] http://www.itl.nist.gov/iad/894.01/tests/ace/

Usually the next step in IE is coreference resolution, the identification of identity relations between Named Entities (Jurafsky and Martin, 2009). At this stage, mentions of the same Named Entity, which are expressed using different linguistic realisations, are found. The process of coreference resolution is crucial for getting more accurate results in IE and more details about this process are provided in the next section.

Relation extraction is a step further in analysing information in the texts and turning unstructured information into structured information. This stage involves identifying the links between Named Entities and deciding which ones are meaningful for the concrete application or problem.

The final stage of information extraction is template filling. Template filling involves extracting appropriate material to fill in the slots in templates for some stereotypical situations that recur quite often. For example, we can be interested in extracting information about some terrorist attack and this event can be treated as a template, which has predefined slots: place, date, number of people injured/killed, organisation who took responsibility for the terrorist act, etc.

3 Relation extraction

As mentioned in Section 2, relation extraction (RE) is one of the steps of information extraction. It typically follows named entity recognition and coreference resolution and aims to gather relations between NEs. Culotta et al (2006) define relation extraction as:

> "the task of discovering semantic connections between entities. In text, this usually amounts to examining pairs of entities in a document and determining (from local language cues) whether a relation exists between them."

Nowadays there are a lot of systems extracting relations from texts and there are different methods for dealing with this problem. Etzioni et al (2008) classify all the methods used for relation extraction into three classes:

- knowledge-based methods;
- supervised methods;
- self-supervised methods.

Each of these classes are explained in the remainder of this paper.

4 Knowledge-based methods

The first category of methods is used usually in domain-specific tasks, where the texts are similar and a closed set of relations needs to be identified. Systems which use these methods rely on pattern-matching rules manually crafted for each domain (Riloff and Jones, 1999; Pasca, 2004). However, not all the relations are domain-dependent and there are some domain-independent ones. Hearst (1992) describes the usage of lexico-syntactic patterns for extraction of hyponymy relations in an open domain. These patterns capture such hyponymy relations as between *"author"* and *"Shakespeare"*, *"wound"* and *"injury"*, *"England"* and *"European country"*. However, the author notes that this method does not work well for some other kinds of relations, for example, meronymy. This is explained by the fact that patterns do not tend to uniquely identify the given relation.

The systems which participated in MUC and deal with relation extraction also rely on rich rules for identifying relations (Fukumoto et al, 1998; Garigliano et al, 1998; Humphreys et al, 1998). Humphreys et al (1998) mention that they tried to add only those rules which were (almost) certain never to generate errors in analysis; therefore, they had adopted a low recall and high precision approach. However, in this case, many relations may be missed due to the lack of unambiguous rules to extract them.

To conclude, knowledge-based methods are not easily portable to other domains and involve too much manual labour. However, they can be used effectively if the main aim is to get results quickly in well-defined domains and document collections.

5 Supervised methods

Supervised methods rely on a training set where domain-specific examples have been tagged. Such systems automatically learn extractors for relations by using machine-learning techniques. The main problem of using these methods is that the development of a suitably tagged corpus can take a lot of time and effort. On the other hand, these systems can be easily adapted to a different domain provided there is training data.

There are different ways that extractors can be learnt in order to solve the problem of supervised relation extraction: kernel methods (Zhao and Grishman, 2005; Bunescu and Mooney, 2006), logistic regression (Kambhatla, 2004), augmented parsing (Miller et al, 2000), Conditional Random Fields (CRF) (Culotta et al, 2006).

In RE in general and supervised RE in particular a lot of research was done for IS-A relations and extraction of taxonomies. Several resources were built based on collaboratively built Wikipedia (YAGO – (Suchanek et al, 2007); DBpedia – (Auer et al, 2007); Freebase – (Bollacker et al, 2008); WikiNet – (Nastase et al, 2010)). In general, Wikipedia is becoming more and more popular as a source for RE, e.g. (Ponzetto and Strube, 2007; Nguyen et al, 2007a,b,c). Query logs are also considered a valuable source of information for RE and their analysis is even argued to give better results than other suggested methods in the field (Paşca, 2007, 2009).

5.1 Weakly-supervised methods

Some supervised systems also use bootstrapping to make construction of the training data easier. These methods are also sometimes referred to as "weakly-supervised information extraction". Brin (1998) describes the *DIPRE* (Dual Iterative Pattern Relation Expansion) method used for identifying authors of the books. It uses an initial small set of seeds or a set of hand-constructed extraction patterns to begin the training process. After the occurrences of needed information are found, they are further used for recognition of new patterns. Regardless of how promising bootstrapping can seem, error propagation becomes a serious problem: mistakes in extraction at the initial stages generate more mistakes at later stages and decrease the accuracy of the extraction process. For example, errors that expand to named entity recognition, e.g. extracting incomplete proper names, result in choosing incorrect seeds for the next step of bootstrapping. Another problem that can occur is that of semantic drift. This happens when senses of the words are not taken into account and therefore each iteration results in a move from the original meaning. Some researchers (Kozareva and Hovy, 2010; Hovy et al, 2009; Kozareva et al, 2008) have suggested ways to avoid this problem and enhance the performance of this method by using doubly-anchored patterns (which include both the class name and a class member) as well as graph structures. Such patterns have two anchor seed positions "{type} such as {seed} and *" and also one open position for the terms to be learnt, for example, pattern "Presidents such as Ford and {X}" can be used to learn names of the presidents. Graphs are used for storing information about patterns, found words and links to entities they helped to find. This data is further used for calculating popularity and productivity of the candidate words. This approach helps to enhance the accuracy of bootstrapping and to find high-quality information using only a few seeds. Kozareva (2012) employs a similar approach for the extraction of cause-effect relations, where the pattern for bootstrapping has a form of "X and Y verb Z", for example, "* and virus cause * ". Human-based evaluation reports 89% accuracy on 1500 examples.

6 Self-supervised systems

Self-supervised systems go further in making the process of information extraction unsupervised. The KnowItAll Web IE system (Etzioni et al, 2005), an example of a self-supervised system, learns "to label its own training examples using only a small set of domain-independent extraction patterns". It uses a set of generic patterns to automatically instantiate relation-specific extraction rules and then learns domain-specific extraction rules and the whole process is repeated iteratively.

The Intelligence in Wikipedia (IWP) project (Weld et al, 2008) is another example of a self-supervised system. It bootstraps from the Wikipedia corpus, exploiting the fact that each article corresponds to a primary object and that many articles contain infoboxes (brief tabular information about the article). This system is able to use Wikipedia infoboxes as a starting point for training the classifiers for the page type. IWP trains extractors for the various attributes and they can later be used for extracting information from general Web pages. The disadvantage of IWP is that the amount of relations described in Wikipedia infoboxes is limited and so not all relations can be extracted using this method.

6.1 Open Information Extraction

Etzioni et al (2008) introduced the notion of Open Information Extraction, which is opposed to Traditional Relation Extraction. Open information extraction is "a novel extraction paradigm that tackles an unbounded number of relations". This method does not presuppose a predefined set of relations and is targeted at all relations that can be extracted.

The Open Relation extraction approach is relatively a new one, so there is only a small amount of projects using it. TextRunner (Banko and Etzioni, 2008; Banko et al, 2007) is an example of such a system. A set of relation-independent lexico-syntactic patterns is used to build a relation-independent extraction model. It was found that 95% of all relations in English can be described by only 8 general patterns, e.g. *"E1 Verb E2"*. The input of such a system is only a corpus and some relation-independent heuristics, relation names are not known in advance. Conditional Random Fields (CRF) are used to identify spans of tokens believed to indicate explicit mentions of relationships between entities and the whole problem of relation extraction is treated as a problem of sequence labelling. The set of linguistic features used in this system is similar to those used by other state of-the-art relation extraction systems and includes e.g. part-of-speech tags, regular expressions for detection of capitalization and punctuation, context words. At this stage of development this system "is able to extract instances of the four most frequently observed relation types: Verb, Noun+Prep, Verb+Prep and Infinitive". It has a number of limitations, which are however common to all RE systems: it extracts only explicitly expressed relations that are primarily word-based; relations should occur between entity names within the same sentence.

Banko and Etzioni (2008) report a precision of 88.3% and a recall of 45.2%. Even though the system shows very good results the relations are not specified and so there are difficulties in using them in some other systems. Output of the system consists of tuples stating there is some relation between two entities, but there is no generalization of these relations.

Wu and Weld (2010) combine the idea of Open Relation Extraction and the use of Wikipedia infoboxes and produce systems called $\text{WOE}^{\text{parse}}$ and WOE^{pos}. $\text{WOE}^{\text{parse}}$ improves TextRunner dramatically but it is 30 times slower than TextRunner. However, WOE^{pos} does not have this disadvantage and still shows an improved F-measure over TextRunner between 15% to 34% on three corpora.

Fader et al (2011) identify several flaws in previous works in Open Information Extraction: "the learned extractors ignore both "holistic" aspects of the relation phrase (e.g., is it contiguous?) as well as lexical aspects (e.g., how many instances of this relation are there?)". They target these problems by introducing syntactic constraints (e.g., they require the relation phrase to match the POS tag pattern) and lexical constraints. Their system ReVerb achieves an AUC which is 30% better than WOE (Wu and Weld, 2010) and TextRunner (Banko and Etzioni, 2008).

Nakashole et al (2012a) approach this problem from another angle. They try to mine for patterns expressing various relations and organise then in hierarchies. They explore binary relations between entities and employ frequent itemset mining (Agrawal et al, 1993; Srikant and Agrawal, 1996) to identify the most frequent patterns. Their work results in a resource called PATTY which contains 350.569 pattern synsets and subsumption relations and achieves 84.7% accuracy. Unlike ReVerb (Fader et al, 2011) which constrains patterns to verbs or verb phrases that end with prepositions, PATTY can learn arbitrary patterns. The authors employ so called syntactic-ontologic-lexical patterns (SOL patterns). These patterns constitute a sequence of words, POS-tags, wildcards, and ontological types. For example, the pattern "persons [adj] voice * song" would match the strings Amy Winehouses soft voice in Rehab and Elvis Presleys solid voice in his song All shook up. Their approach is based on collecting dependency paths from the sentences where two named entities are tagged (YAGO2 (Hoffart et al, 2011) is used as a database of all NEs). Then the textual pattern is extracted by finding the shortest paths connecting two entities. All of these patterns are transformed into SOL (abstraction of a textual pattern). Frequent itemset technique is used for this: all textual patterns are decomposed into n-grams (n consecutive words). A SOL pattern contains only the n-grams that appear frequently in the corpus and the remaining word sequences are replaced by wildcards. The support set of the pattern is described as the set of pairs of entities that appear in the place of the entity placeholders in all strings in the corpus that match the pattern. The patterns are connected in one synset (so are considered synonymous) if their supporting sets coincide. The overlap of the supporting sets is also employed to identify subsumption relations between various synsets.

6.2 Distant learning

Mintz et al (2009) introduce a new term "distant supervision". The authors use a large semantic database Freebase containing 7,300 relations between 9 million named entities. For each pair of entities that appears in Freebase relation, they identify all sentences containing those entities in a large unlabeled corpus. At the next step textual features to train a relation classifier are extracted. Even though the 67,6% of precision achieved using this method has room for improvement, it has inspired many researchers to further investigate in this direction.

Currently there are a number of papers trying to enhance "distant learning" in several directions. Some researchers target the heuristics that are used to map the relations in the databases to the texts, for example, (Takamatsu et al, 2012) argue that improving matching helps to make data less noisy and therefore enhances the quality of relation extraction in general.

Yao et al (2010) propose using an undirected graphical model for relation extraction which employs "distant learning" but enforces selectional preferences. Riedel et al (2010) reports 31% error reduction compared to (Mintz et al, 2009).

Another problem that has been addressed is language ambiguity (Yao et al, 2011, 2012). Most methods cluster shallow or syntactic patterns of relation mentions, but consider only one possible sense per pattern. However, this assumption is often violated in reality. Yao et al (2011) uses generative probabilistic models, where both entity type constraints within a relation and features on the dependency path between entity mentions are exploited. This research is similar to DIRT (Lin and Pantel, 2001) which explores distributional similarity of dependency paths in order to discover different representations of the same semantic relation. However, Yao et al (2011) employ another approach and apply LDA (Blei et al, 2003) with a slight modification: observations are relation tuples and not words. So as a result of this modification instead of representing semantically related words, the topic latent variable represents a relation type. The authors combine three models: Rel-LDA, Rel-LDA1 and Type-LDA. In the third model the authors split the features of a tuple into relation level features and entity level features. Relation level features include the dependency path, trigger, lexical and POS features; entity level features include the entity mention itself and its named entity tag. These models output clustering of observed relation tuples and their associated textual expressions. The evaluation shows that the use of these resulting clusters helps to improve distant learning and results in 12% better performance.

Distant learning and other types of relation extraction are, as we have already seen, based on several assumptions. Another assumption that is often used in this field is that a pair of entities can have only one relation. However, if we examine the following examples –"Steve Jobs founded Apple" and "Steve Jobs is CEO of Apple" – we can see that this assumption is rather restrictive.

Hoffmann et al (2011) identified this problem with previous RE systems and try to address this issue by employing Multi-Instance Multi-label (MIML) approach. They employ distant learning with Multi-Instance learning with overlapping relations (where two same instances may be in two different relations). The resulting system MultiR achieves competitive or higher precision over all ranges of recall.

Surdeanu et al (2012) tackle the same problem. They identify two main problems of distant learning: (1) some training examples obtained through this heuristic are not valid (they report 31%), (2) the same pair of entities can have several relations. Therefore they try to improve distant learning by taking into account Multi-instance Multi-label settings and using Bayesian framework (they call their system MIML-RE) which can capture dependencies between labels and learn in the presence of incorrect and incomplete labels.

When using "distant supervision" in its original version we are limited by the schema imposed by the database that is used for mapping. Yao et al (2013) suggest several ways how it can be overcome. They suggest using raw texts in addition to distant supervision, therefore relations in the text and pre-existing structured databases can be employed together. Riedel et al (2013) address this problem by using matrix factorisation and collaborative filtering. Previously, matrix factorisation was employed by Nickel et al (2012) in order to predict new relations (triples) in terms of YAGO2. All relations between entities are presented as a matrix where there is an indication whether there is a relation or not. Riedel et al (2013) use three models: (1) latent feature model, which is a generalised PCA (Collins et al, 2001); (2) neighbourhood model, neighbour based approach (Koren, 2008); (3) entity model, which learns a latent entity representation from data. The authors also present a combined model that incorporates all three models with various weights. In order to overcome the lack of negative examples, they employ the technique of implicit feedback (Rendle et al, 2009), where observed true facts are given higher scores than unobserved (true or false) facts. The authors report competitive results of evaluation and also mention computational efficiency of their methods which is an important aspect for such systems. They also discuss the fact that this approach is not merely a tool for information extraction and that the same technique can be used for integrating databases with different schemata.

7 Joint Prediction

The joint modelling of several levels of Information extraction is also explored by several research papers. In their position paper Mccallum and Jensen (2003) propose to use "unified, relational, undirected graphical models for information extraction and data mining". This common inference procedure can help to improve all the stages, so that each component is able to make up for the weaknesses of the other and therefore improve the performance of both.

Finkel et al (2006) explore the idea of joint modeling as well. They present a novel architecture, which models pipelines as Bayesian networks. Each low level task corresponds to a variable in the network, and then an approximate inference is performed to find the best labelling. This approach is tested on two tasks: semantic role labelling and recognizing textual entailment.

Roth and Yih (2007) employ the same idea when they combine two stages on Information Extraction: named entity recognition and relation extraction. However, Singh et al (2013) go even further and include coreference resolution as well. So they propose a single, joint graphical model that represents the various dependencies between the tasks (entity tagging, relation extraction, and coreference). Their joint modelling approach helps to avoid cascading errors. The joint model obtains 12% error reduction on tagging over the isolated models.

8 Conclusions

This paper introduced a field of Information Extraction and provided more details about recent developments in its subfield, Relation Extraction. We have presented the main approaches to this task and also outlined some challenges. All the methods described above have advantages and disadvantages and the choice depends greatly on the task in mind and the accuracy needed. Relation extraction has a lot of uses in NLP and can be beneficial for: semantic search, machine reading, question answering, knowledge harvesting, paraphrasing, building thesauri etc. (Nakashole et al, 2012b, 2013).

The field appears to be becoming more and more interdisciplinary and methods from data mining and recommendation systems are currently used to assist in the task of relation extraction (Cergani and Miettinen, 2013; Riedel et al, 2013; Nakashole et al, 2012a). Also modeling all stages of Information Extraction as a single task is another recent trend in the field (Mccallum and Jensen, 2003; Finkel et al, 2006; Roth and Yih, 2007; Singh et al, 2013). The research in this area reports significant improvement in all tasks when they are modeled jointly, it also helps to avoid error propagation which is a frequent problem in the pipeline approach.

Research in terms of relation extraction has still room for improvement, however, it targets a very difficult problem where language ambiguity is a significant obstacle. The majority of research in the field is done for English language, therefore targeting other languages and exploring further multilingual information extraction and possibility of aligning resources in various languages can be the future direction of Relation Extraction.

Bibliography

Agrawal R, Imieliński T, Swami A (1993) Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, SIGMOD '93, pp 207–216

Andersen PM, Hayes PJ, Huettner AK, Schmandt LM, Nirenburg IB, Weinstein SP (1992) Automatic extraction of facts from press releases to generate news stories. In: Proceedings of the Third Conference on Applied Natural Language Processing, pp 170–177

Auer S, Bizer C, Kobilarov G, Lehmann J, Ives Z (2007) DBpedia: A nucleus for a web of open data. In: Proccedings of 6th Intl Semantic Web Conference, Springer, Busan, Korea, pp 11–15

Banko M, Etzioni O (2008) The tradeoffs between open and traditional relation extraction. In: Proceedings of ACL-08: HLT, pp 28–36

Banko M, Cafarella MJ, Soderland S, Broadhead M, Etzioni O (2007) Open information extraction from the web. In: Veloso MM (ed) Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, pp 2670–2676

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. Journal of Machine Learning Research 3:993–1022

Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J (2008) Freebase: A collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, SIGMOD '08, pp 1247–1250

Brin S (1998) Extracting patterns and relations from the world wide web. In: Proceedings of the First International Workshop on the Web and Databases, pp 172–183

Bunescu R, Mooney R (2006) Subsequence kernels for relation extraction. In: Weiss Y, Schölkopf B, Platt J (eds) Advances in Neural Information Processing Systems 18, MIT Press, Cambridge, MA, pp 171–178

Cergani E, Miettinen P (2013) Discovering relations using matrix factorization methods. In: He Q, Iyengar A, Nejdl W, Pei J, Rastogi R (eds) CIKM, ACM, pp 1549–1552

Collins M, Dasgupta S, Schapire RE (2001) A generalization of principal component analysis to the exponential family. In: Advances in Neural Information Processing Systems, MIT Press

Cowie J, Lehnert W (1996) Information extraction. Communications of the ACM 39(1):80–91

Culotta A, McCallum A, Betz J (2006) Integrating probabilistic extraction models and data mining to discover relations and patterns in text. In: Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Association for Computational Linguistics, New York, New York, pp 296–303

Etzioni O, Cafarella M, Downey D, Popescu AM, Shaked T, Soderland S, Weld DS, Yates A (2005) Unsupervised named-entity extraction from the web: an experimental study. In: Artificial Intelligence, Elsevier Science Publishers Ltd., Essex, UK, vol 165, pp 91–134

Etzioni O, Banko M, Soderland S, Weld DS (2008) Open information extraction from the web. Communications of the ACM 51:68–74

Fader A, Soderland S, Etzioni O (2011) Identifying relations for open information extraction. In: Proceedings of EMNLP 2011, UK, Edinburgh

Finkel JR, Manning CD, Ng AY (2006) Solving the problem of cascading errors: Approximate bayesian inference for linguistic annotation pipelines. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Stroudsburg, PA, USA, EMNLP '06, pp 618–626

Fukumoto J, Masui F, Shimohata M, Sasaki M (1998) Oki electric industry: Description of the Oki system as used for MUC-7. In: Proceedings of the 7th Message Understanding Conference

Garigliano R, Urbanowicz A, Nettleton DJ (1998) University of Durham: Description of the LOLITA system as used in MUC-7. In: Proceedings of the 7th Message Understanding Conference

Grishman R (1997) Information extraction: techniques and challenges. In: In Information Extraction (International Summer School SCIE-97, Springer-Verlag, pp 10–27

Grishman R, Sundheim B (1996) Message Understanding Conference-6: a brief history. In: Proceedings of the 16th conference on Computational linguistics, Association for Computational Linguistics, Morristown, NJ, USA, pp 466–471

Hearst MA (1992) Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th conference on Computational linguistics, Association for Computational Linguistics, Morristown, NJ, USA, pp 539–545

Hoffart J, Suchanek FM, Berberich K, Lewis-Kelham E, de Melo G, Weikum G (2011) Yago2: Exploring and querying world knowledge in time, space, context, and many languages. In: Proceedings of the 20th International Conference Companion on World Wide Web, ACM, New York, NY, USA, WWW '11, pp 229–232

Hoffmann R, Zhang C, Ling X, Zettlemoyer L, Weld DS (2011) Knowledge-based weak supervision for information extraction of overlapping relations. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, Association for Computational Linguistics, Stroudsburg, PA, USA, HLT '11, pp 541–550

Hovy E, Kozareva Z, Riloff E (2009) Toward completeness in concept extraction and classification. In: EMNLP '09: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Morristown, NJ, USA, pp 948–957

Humphreys K, Gaizauskas R, Azzam S, Huyck C, Mitchell B, Cunningham H, Wilks Y (1998) University of Sheffield: Description of the LaSIE-II System as Used for MUC-7. In: Proceedings of MUC-7

Jurafsky D, Martin JH (2009) Speech and language processing an introduction to natural language processing, computational linguistics, and speech recognition, 2nd edn. Prentice-Hall, Inc.
Kambhatla N (2004) Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In: Proceedings of the ACL 2004 on Interactive poster and demonstration sessions, Association for Computational Linguistics, Morristown, NJ, USA, p 22
Koren Y (2008) Factorization meets the neighborhood: A multifaceted collaborative filtering model. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA, KDD '08, pp 426–434
Kozareva Z (2012) Cause-effect relation learning. In: Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing, Association for Computational Linguistics, Stroudsburg, PA, USA, TextGraphs-7 '12, pp 39–43
Kozareva Z, 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, Association for Computational Linguistics, Uppsala, Sweden, pp 1482–1491
Kozareva Z, Riloff E, Hovy E (2008) Semantic class learning from the web with hyponym pattern linkage graphs. In: Proceedings of ACL-08: HLT, Association for Computational Linguistics, Columbus, Ohio, pp 1048–1056
Lin D, Pantel P (2001) DIRT - discovery of inference rules from text. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA, KDD '01, pp 323–328
Mccallum A, Jensen D (2003) A note on the unification of information extraction and data mining using conditional-probability, relational models. In: In Proceedings of the IJCAI-2003 Workshop on Learning Statistical Models from Relational Data
Miller S, Fox H, Ramshaw L, Weischedel R (2000) A novel use of statistical parsing to extract information from text. In: Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference, Morgan Kaufmann Publishers Inc., Seattle, Washington, pp 226–233
Mintz M, Bills S, Snow R, Jurafsky D (2009) Distant supervision for relation extraction without labeled data. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2, Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '09, pp 1003–1011
Nakashole N, Weikum G, Suchanek F (2012a) Patty: A taxonomy of relational patterns with semantic types. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics, Stroudsburg, PA, USA, EMNLP-CoNLL '12, pp 1135–1145
Nakashole N, Weikum G, Suchanek FM (2012b) Discovering and exploring relations on the web. PVLDB 5(12):1982–1985

Nakashole N, Weikum G, Suchanek FM (2013) Discovering semantic relations from the web and organizing them with patty. SIGMOD Record 42(2):29–34

Nastase V, Strube M, Boerschinger B, Zirn C, Elghafari A (2010) Wikinet: A very large scale multi-lingual concept network. In: Calzolari N, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Rosner M, Tapias D (eds) LREC, European Language Resources Association

Nguyen DP, Matsuo Y, Ishizuka M (2007a) Exploiting syntactic and semantic information for relation extraction from Wikipedia. In: Proceedings of the IJCAI Workshop on Text-Mining and Link- Analysis, TextLink07

Nguyen DP, Matsuo Y, Ishizuka M (2007b) Relation extraction from Wikipedia using subtree mining. In: Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, AAAI Press, Vancouver, British Columbia, Canada, pp 1414–1420

Nguyen DPT, Matsuo Y, Ishizuka M (2007c) Subtree mining for relation extraction from Wikipedia. In: Sidner CL, Schultz T, Stone M, Zhai C (eds) Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, Rochester, New York, USA, The Association for Computational Linguistics, pp 125–128

Nickel M, Tresp V, Kriegel HP (2012) Factorizing YAGO: Scalable machine learning for linked data. In: Proceedings of the 21st International Conference on World Wide Web, ACM, New York, NY, USA, WWW '12, pp 271–280

Paşca M (2007) Organizing and searching the world wide web of facts – step two: harnessing the wisdom of the crowds. In: WWW '07: Proceedings of the 16th international conference on World Wide Web, ACM, New York, NY, USA, pp 101–110

Paşca M (2009) Outclassing Wikipedia in open-domain information extraction: weakly-supervised acquisition of attributes over conceptual hierarchies. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, Athens, Greece, pp 639–647

Pasca M (2004) Acquisition of categorized named entities for web search. In: Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management, ACM, New York, NY, USA, CIKM '04, pp 137–145

Ponzetto SP, Strube M (2007) Deriving a large scale taxonomy from Wikipedia. In: Proceedings of the 22nd Conference on the Advancement of Artificial Intelligence, Vancouver, B.C., Canada, pp 1440–1445

Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L (2009) BPR: Bayesian personalized ranking from implicit feedback. In: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, AUAI Press, Arlington, Virginia, United States, UAI '09, pp 452–461

Riedel S, Yao L, McCallum A (2010) Modeling relations and their mentions without labeled text. In: Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD '10)

Riedel S, Yao L, McCallum A, Marlin BM (2013) Relation extraction with matrix factorization and universal schemas. In: Vanderwende L, III HD, Kirchhoff K (eds) HLT-NAACL, The Association for Computational Linguistics, pp 74–84

Riloff E, Jones R (1999) Learning dictionaries for information extraction by multi-level bootstrapping. In: Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference, American Association for Artificial Intelligence, Menlo Park, CA, USA, AAAI '99/IAAI '99, pp 474–479

Roth D, Yih Wt (2007) Global inference for entity and relation identification via a linear programming formulation. In: Getoor L, Taskar B (eds) Introduction to Statistical Relational Learning, The MIT press

Singh S, Riedel S, Martin B, Zheng J, McCallum A (2013) Joint inference of entities, relations, and coreference. In: Proceedings of the 3rd Workshop on Automated Knowledge Base Construction

Srikant R, Agrawal R (1996) Mining sequential patterns: Generalizations and performance improvements. In: Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology, Springer-Verlag, London, UK, UK, EDBT '96, pp 3–17

Suchanek FM, Kasneci G, Weikum G (2007) YAGO: A core of semantic knowledge. In: Proceedings of WWW-07, pp 697–706

Surdeanu M, Tibshirani J, Nallapati R, Manning CD (2012) Multi-instance multi-label learning for relation extraction. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics, Stroudsburg, PA, USA, EMNLP-CoNLL '12, pp 455–465

Takamatsu S, Sato I, Nakagawa H (2012) Reducing wrong labels in distant supervision for relation extraction. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1, Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '12, pp 721–729

Weld DS, Wu F, Adar E, Amershi S, Fogarty J, Hoffmann R, Patel K, Skinner M (2008) Intelligence in Wikipedia. In: Proceedings of the 23rd AAAI Conference, Chicago, USA

Wu F, Weld DS (2010) Open information extraction using Wikipedia. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '10, pp 118–127

Yao L, Riedel S, McCallum A (2010) Collective cross-document relation extraction without labelled data. In: EMNLP, ACL, pp 1013–1023

Yao L, Haghighi A, Riedel S, McCallum A (2011) Structured relation discovery using generative models. In: Proceedings of the Conference on Empirical methods in natural language processing (EMNLP '11), pp 1456–1466

Yao L, Riedel S, McCallum A (2012) Unsupervised relation discovery with sense disambiguation. In: ACL, The Association for Computer Linguistics, pp 712–720

Yao L, Riedel S, McCallum A (2013) Universal schema for entity type prediction. In: Proceedings of the 3rd Workshop on Automated Knowledge Base Construction

Zhao S, Grishman R (2005) Extracting relations with integrated information using kernel methods. In: ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, Morristown, NJ, USA, pp 419–426