TEG-REP: A Corpus of Textual Entailment Graphs based on Relation Extraction Patterns

Kathrin Eichler, Feiyu Xu, Hans Uszkoreit, Leonhard Hennig, Sebastian Krause
German Research Center for Artificial Intelligence
Alt-Moabit 91c, Berlin, Germany
{kathrin.eichler,feiyu,uszkoreit,leonhard.hennig,sebastian.krause}@dfki.de

Abstract

The task of relation extraction is to recognize and extract relations between entities or concepts in texts. Dependency parse trees have become a popular source for discovering extraction patterns, which encode the grammatical relations among the phrases that jointly express relation instances. State-of-the-art weakly supervised approaches to relation extraction typically extract thousands of unique patterns only potentially expressing the target relation. Among these patterns, some are semantically equivalent, but differ in their morphological, lexical-semantic or syntactic form. Some express a relation that entails the target relation. We propose a new approach to structuring extraction patterns by utilizing entailment graphs, hierarchical structures representing entailment relations, and present a novel resource of gold-standard entailment graphs based on a set of patterns automatically acquired using distant supervision. We describe the methodology used for creating the dataset and present statistics of the resource as well as an analysis of inference types underlying the entailment decisions.

Keywords: Entailment Graphs, Relation Extraction, Textual Entailment

1. Introduction

The task of relation extraction (RE) is to recognize and extract relations between entities or concepts in texts. Dependency parse trees have become a popular source for discovering extraction patterns, which encode the grammatical relations among the phrases that jointly express relation instances. In rule-based RE methods, the patterns are directly applied to extract relation mentions from parsed sentences of free texts (e.g., Yangarber et al. (2000), Alfonseca et al. (2012)). Other methods treat RE as a classification or sequence-labeling problem, but even for those techniques parse tree patterns have proven useful as key classification features (e.g., Zelenko et al. (2003), Bunescu and Mooney (2005)). In order to circumvent manual annotation work needed for supervised learning, recent work in RE concentrates on weakly supervised learning, for example based on techniques of distant supervision (Mintz et al., 2009; Krause et al., 2012). These utilize extensive volumes of pre-existing knowledge for partially labeling large volumes of data, resulting in large numbers of unique candidate patterns acquired from *suspected* mentions of relation instances. Among these patterns, some are semantically equivalent, but differ in their morphological, lexical-semantic or syntactic form. Some express a relation that entails the target relation or that is entailed by the target relation. Others are semantically unrelated to the target relation. The basic assumption made in this work is that patterns are truly reliable if they express a relation that semantically entails the target relation (Romano et al., 2006). This includes all patterns that express the target relation explicitly or a semantically equivalent relation. As an example, the pattern "org|BUYER bought org|ACQUIRED" can be considered to be semantically equivalent to the pattern "org|BUYER purchased org|ACQUIRED", whereas "per|SPOUSE divorced per|SPOUSE" is not semantically equivalent to "per|SPOUSE married per|SPOUSE", but entails a marriage relation.

We propose a new approach to structuring extraction patterns by utilizing entailment graphs, hierarchical structures representing entailment relations, and present a novel resource of entailment graphs based on dependency-structure based extraction patterns. We automatically acquire a set of patterns for three semantic relations using distant supervision, and create gold-standard entailment graphs representing the semantic relationships holding among the patterns. We describe the methodology used for creating the dataset as well as statistics about and an analysis of the dataset. The dataset is intended to be used as a resource for the relation extraction task as well as for evaluating automatically generated entailment graphs and systems for recognizing textual entailment.

2. Related Work

While relation extraction would clearly benefit from considering semantic relationships between patterns, there has been only a limited amount of prior work in structuring patterns. Matrix factorization approaches cluster semantically similar patterns based on argument co-occurrence information (e.g., Riedel et al. (2013)). Other approaches focus on the tree structure of the patterns, and compute similarity metrics based on graph matching techniques or tree edit distance (e.g., Thomas et al. (2011), Liu et al. (2013)). We propose a new approach to structuring extraction patterns by utilizing entailment graphs.

The *textual entailment* paradigm captures the semantic relationship holding between two textual expressions T (text) and H (hypothesis): T entails H if the meaning of H can be inferred from the meaning of T (Dagan and Glickman, 2004). Textual entailment is defined as a semantic relation between exactly two text expressions. With entailment being a transitive relation, entailment relations holding among a set of expressions can be represented in a hierarchical structure, referred to as *entailment graphs* (Berant et al., 2010). Entailment graphs have been built for various types of expressions, including propositional templates (Berant et al., 2010), typed predicates (Berant et al., 2011; Berant et al., 2012), open IE propositions (Levy et al., 2014), and text fragments (Kotlerman et al., 2015). A sample graph from Berant et al. (2010) is depicted in Figure 1, where $\rightarrow$ denotes a unidirectional and $\leftrightarrow$ a bidirectional entailment (i.e., paraphrase) relation.

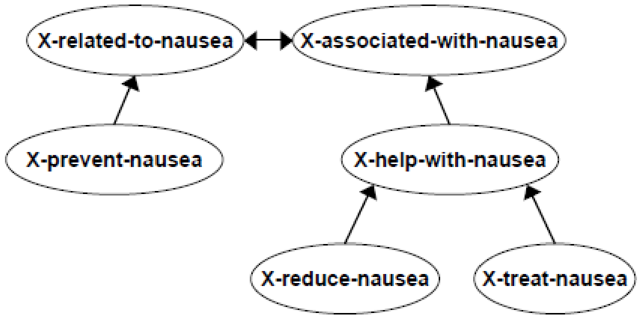


Figure 1: Entailment graph for propositional templates

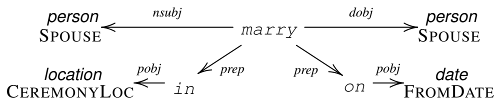
In order to evaluate automatically generated entailment graphs, Bentivogli and Magnini (2014) created a gold-standard entailment graph dataset based on text fragments representing complaints extracted from Italian customer interactions. Starting with a manual annotation of so-called modifiers, i.e., tokens that can be removed without affecting the fragment's comprehension, they automatically derive entailment relations holding between a fragment and its subfragments (fragments from which modifiers were removed). All other entailment decisions required for building the entailment graphs are acquired by manually annotating T/H pairs of different fragments or subfragments. The number of T/H pairs to be annotated is minimized by manually clustering fragments into topics and by skipping unnecessary comparisons based on previous annotator decisions. In the final step, transitive closure edges are added to the graph and a consistency check is performed to ensure the transitivity of the resulting graph. Their final dataset contains 19 textual entailment graphs, one for each topic cluster, with - altogether - 760 nodes and 2316 entailment edges.

Kotlerman et al. (2015) use the same procedure to construct a dataset of entailment graphs based on text fragments extracted from English customer interactions. Their dataset consists of 29 entailment graphs, with 756 nodes and 7862 edges. In our work, we adapt the procedure proposed by Bentivogli and Magnini (2014) to the construction of entailment graphs for relation extraction patterns instead of text expressions. To our knowledge, the corpus we present is the first corpus of entailment graphs of this kind.

3. Relation extraction patterns

We acquire patterns using the automatic pattern discovery system Web-DARE, which is based on distant supervision. As defined by most distant supervision systems (e.g., Mintz et al. (2009), Alfonseca et al. (2012)), Web-DARE regards a sentence as a candidate of a relation mention if it contains the (main) entities of a relation instance of a fact knowledge base. Web-DARE utilizes facts from Freebase (Bollacker et al., 2008) for annotating relation mentions in candidate sentences and learns pattern candidates from sentence parses generated using MaltParser (Nivre et al., 2007). Unlike most other relation extraction systems, Web-DARE can deal with n-ary relations, not only binary relations. Furthermore, just as in the Snowball system (Agichtein, 2006), Web-DARE rules assign the semantic role labels to the relation arguments. The following example rule of Web-DARE for the relation *marriage* contains four arguments, two married persons plus the wedding location and the starting date of the marriage. The notation person|SPOUSE represents a placeholder for an entity mention of type person, which is assigned the role label SPOUSE at extraction time.

Example 1



The method is applied to 39 relations from the domains *Awards*, *Business* and *People* modeled in Freebase. About 2.8M instances of these relations were retrieved from Freebase as seed knowledge, from which about 200,000 were turned into search queries, resulting in almost 20M downloaded web pages. 3M sentences matched by seed facts were utilized to learn more than 1.5M pattern candidates for the relation extraction task.

Moro et al. (2013) employ various filtering strategies to identify invalid patterns, e.g., frequency filtering and semantic filtering, where the latter selects patterns containing at least one relation relevant word sense. We observe that although semantic filtering can provide clues to evaluating the usefulness of relation extraction patterns, it cannot capture whether the meaning of a given pattern expresses that the target relation really holds. This includes patterns that express the target relation R explicitly, patterns that express a relation that is semantically equivalent to R, and patterns that express a relation that entails R. For example, patterns P1 to P3 below are all semantically related to the target relation marriage. However, only patterns P1 and P2 indicate that two persons were involved in a marriage relation. P1 expresses the relation explicitly, P2 entails the relation.

P1: person|SPOUSE was married to person|SPOUSE

P2: person|SPOUSE is widow of person|SPOUSE

P3: person|SPOUSE is in love with person|SPOUSE

As being aware of these semantic relations holding among patterns can be of help in the pattern selection process, we capture these relations in the form of entailment graphs.

4. Annotation procedure

The goal of the annotation procedure is to identify all entailment relations holding among different relation extraction patterns expressing or indicating the same relation. We define a pattern T to entail a pattern H if the meaning expressed by H can be inferred from the meaning expressed by T. For creating entailment graphs, we identify entailment relations between patterns via manual annotation based on the guidelines described in Sections 4.1. and 4.2.. As with a large number of expressions, the task of comparing each possible T/H pair becomes unfeasible, we reduce the manual annotation workload by restricting the possible pairs based on logical considerations. One such consideration is that we can reduce the complexity of the task, and thus the number of inconsistencies in annotation, by dividing the annotation into two steps:

1. Identification of semantically equivalent patterns
2. Annotation of unidirectional entailment relations

Second, we reduce the manual annotation work by removing entailment pairs, for which entailment is not possible based on the set of arguments contained in the pattern. The rationale behind this is that a pattern T cannot entail a pattern H if H contains more arguments (i.e., is more specific) than T. The annotation guidelines are summarized in the following.

4.1. Identification of semantically equivalent patterns

The goal of this first step is to construct sets of semantically equivalent patterns, referred to as *equivalence classes*. Semantically equivalent patterns correspond to patterns, for which textual entailment holds in both directions, i.e., patterns expressing the same meaning. For example, the pattern "organization|BUYER bought organization|ACQUIRED" is semantically equivalent to the pattern "organization|BUYER purchased organization|ACQUIRED" (in this case, due to the synonymy of "buy" and "purchase").

The input to this first annotation step are *argument clusters*, i.e., clusters of patterns grouped automatically based on the number and type of arguments identified in the pattern. For example, for the *marriage* relation, we added all patterns with the argument combination {person|SPOUSE, person|SPOUSE} to one cluster, all patterns with the combination {person|SPOUSE, person|SPOUSE, date|FROMDATE} to another, and so on. The underlying assumption is that patterns can only be semantically equivalent if their arguments are identical.

For each pattern, the annotator first determines whether it entails the base relation, e.g., for the *marriage* relation, if from the semantics of the pattern we can tell that the pattern links two people that are or were[1] married. Patterns expressing the base relation are associated to equivalence classes. For grouping patterns into the same equivalence class, we distinguish the following types of equivalence:

Identity Patterns containing the same set of words, possibly differing in word order, e.g., "In date|FROMDATE person|SPOUSE marries person|SPOUSE" ↔ "person|SPOUSE marries person|SPOUSE in date|FROMDATE"

Preposition variations Differing prepositions assigning the same meaning in the given context, e.g., "person|SPOUSE married to person|SPOUSE in date|FROMDATE" ↔ "person|SPOUSE married to person|SPOUSE from date|FROMDATE"

Tense variations e.g., "person|SPOUSE marries person|SPOUSE" ↔ "person|SPOUSE married person|SPOUSE"

Morphological variations such as derivation, e.g., "person|SPOUSE marries person|SPOUSE" ↔ "person|SPOUSE marriage to person|SPOUSE"

Synonymy e.g., "person|SPOUSE marries person|SPOUSE" ↔ "person|SPOUSE weds person|SPOUSE"

Paraphrase Multi-word expressions used interchangeably within the given context, e.g., "person|SPOUSE marries person|SPOUSE" ↔ "person|SPOUSE established marriage with person|SPOUSE"

Passivization e.g., "Organization|BUYER bought Organization|ACQUIRED" ↔ "Organization|ACQUIRED was bought by Organization|BUYER"

Argument labelling variation Semantically equivalent patterns, in which the arguments of the predicates are aligned by different syntactic functions, e.g., "Organization|BUYER bought Organization|ACQUIRED" ↔ "sold Organization|ACQUIRED to Organization|BUYER"

4.2. Annotation of unidirectional entailment

For annotating unidirectional entailment relations *within* each argument cluster, we select one representative of each equivalence class and generate all possible pairings of representatives. Based on the logical considerations described in Section 4., for annotating entailment relations holding *across* patterns from different argument clusters, we generate all pairs, for which the set of arguments of H is a subset of the set of arguments of T, as T can only entail H if it is at least as specific as H.

For each generated pair, a human annotator decides, whether entailment holds or not. During annotation, we encountered cases, in which two patterns contradicted each other, e.g., "person|SPOUSE person|SPOUSE divorced in date|TODATE" ⊥ "person|SPOUSE was married to person|SPOUSE until death in date|TODATE. However, these cases were rare and we did not annotate them separately, as the entailment graphs we construct only capture binary decisions (entailment, non-entailment). Following the manual annotation, we created entailment graphs based on the annotated entailment decisions, checked the transitivity of each graph and identified and removed inconsistencies.

4.3. Inference types

In order to investigate the types of inference underlying the entailment decisions, we carried out an additional annotation step. For this, we randomly selected a T/H pair of patterns from each pair of equivalence classes linked by an entailment relation, and analysed the nature of entailment, distinguishing among the following inference types:

Additional modifier (MOD) The inferring pattern contains additional information expressed in the form of a modifier, e.g., "person|SPOUSE was married to person|SPOUSE until death" → "person|SPOUSE was married to person|SPOUSE" (additional information: until death).

[1] Ignoring tense aspects is also the approach taken in the RTE challenges (Dagan et al., 2005) Note that we do not consider tense variations to be equivalent if they cause a meaning change, e.g., "person|SPOUSE marries person|SPOUSE" ≠ "person|SPOUSE was going to marry person|SPOUSE." Here, the second pattern, unlike the first, presumes that the marriage has not taken place yet.

Ontological (ONTO) The inference is drawn based on ontological knowledge, such as hyperonymy. For example, ”person|WINNER’s novel won prize|AWARD” → ”person|WINNER’s book won prize|AWARD” (a novel is a kind of book). When annotating ontological relations, we used the WordNet ontology (Fellbaum, 1998) as our reference.

Reasoning (REAS) The inference is drawn by reasoning, e.g., based on general world knowledge, temporal knowledge, or logical inference. For example, the decision that ”person|SPOUSE wife of person|SPOUSE” → ”person|SPOUSE married person|SPOUSE” is based on inferencing that ”wife” refers to the female role of a married couple. Note that in some cases, inference could be drawn based on a combination of different ontological relations. For example, according to WordNet, ”wife” is a hyponym of ”spouse”, which is linked to ”marriage” via a member holonym relation, which in turn is derivationally related to ”marry”. In other cases, reasoning goes beyond lexical inference, as in ”was nominated for prize|AWARD lost to person|WINNER” → ”person|WINNER won prize|AWARD.”

5. Corpus setup

We created an entailment graph corpus based on relation extraction patterns for three semantic relations: *marriage*, *acquisition*, and *award honor*. For each relation, we applied Moro et al. (2013)’s semantic filtering with the least restrictive configuration and selected about 500 of the most frequently occurring Web-DARE patterns[2]. Based on these patterns, we created entailment graphs according to the procedure described in Section 4.

The annotation was conducted by three annotators. In step 1 (equivalence class identification), two annotators worked in parallel. Inter-annotator agreement was 0.88 for *marriage*, 0.83 for *acquisition*, and 0.88 for *award honor*, corresponding to almost perfect agreement. The main source of disagreement between annotators were terms or expressions with ambiguous semantics. For example, annotators disagreed as to whether ”person|SPOUSE eloped with person|SPOUSE” entails ”person|SPOUSE married person|SPOUSE” (”elope” often, but not always, refers to marrying secretly) or as to whether ”person|WINNER author wins prize|AWARD” and ”person|WINNER writer wins prize|AWARD” are semantically equivalent (”author” and ”writer” belong to the same synset in WordNet, but are slightly different in meaning.).

A subgraph of the entailment graph for *marriage* is shown in Figure 2[3]. Table 1 shows the statistics of our corpus: 1. the number of patterns entailing the target relation, 2. the number of equivalence classes (ECs), 3. the number of patterns contained in the largest equivalence class, and 4./5. the number of uni- and bidirectional entailment relations.

[2]As our patterns were generated in a fully automatic way, some of them suffered from incorrect parsing. We annotated these patterns if the part relevant for deciding on entailment was semantically interpretable based on the given dependency tree.

[3]Please note that for reasons of clarity and simplicity, the figure only shows the textual representation of the patterns, not the dependency tree structures.

Unidirectional relations correspond to entailment relations holding across equivalence classes, bidirectional relations to entailment relations holding across patterns belonging to the same equivalence class. Table 2 shows the distribution of inference types per relation.

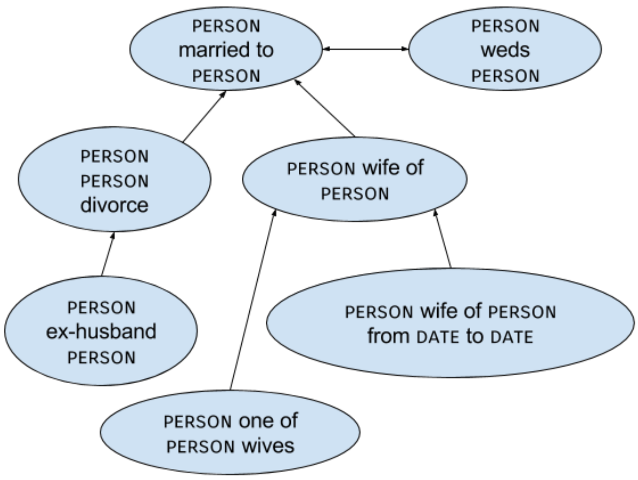


Figure 2: Sample subgraph for marriage relation

relation	# of patterns	# of ECs	max EC size	edges (uni)	edges (bi)
acquisition	161	77	32	122	1796
marriage	265	117	44	225	3262
award honor	412	224	49	977	4852
overall	838	418	49	571	9910

Table 1: Corpus statistics: number of entailing patterns; number of equivalence classes; size of the largest equivalence class; number of uni-/bidirectional entailment edges.

relation	MOD	ONTO	REAS
acquisition	67%	13%	20%
marriage	68%	5%	27%
award honor	59%	11%	30%

Table 2: Distribution of entailment types

6. Conclusion

We presented a new linguistic resource, a corpus of textual entailment graphs based on relation extraction patterns. The graphs in our corpus differ from the ones created by Berant et al. (2010) in that they contain n-ary relations, with $n > 2$. In comparison to the entailment graph corpora by Bentivogli and Magnini (2014) and Kotlerman et al. (2015), the new graphs are more generic and have stronger expressiveness since they are not based on textual expressions, but on dependency structures containing semantic arguments of the target relations. Our corpus can be utilized in several ways: First, as a gold-standard for evaluating both automatically created entailment graphs and textual entailment systems, in particular systems making use of syntactic information. Second, as a resource for fine-grained modelling of semantic context of the target relation. For example, the graph structure can be utilized to identify relation mentions expressing information that is more specific than the target relation, e.g., instances of prize winners that are authors or those of marriages that ended in a divorce. Our resource is publicly available under `http://sargraph.dfki.de/download.html`. Future work includes the extension of the corpus with additional relations and to further automatize the annotation procedure. In particular, automatically identifying patterns that share syntactic structure would allow us to further reduce the number of pairs to be manually annotated.

7. Acknowledgements

This work was partially supported by the German Federal Ministry of Education and Research (BMBF) through the project All Sides (grant 01IW14002), the German Federal Ministry of Economics and Technology through the project SD4M (grant 01MD15007B), and a Google Focused Research Award for Natural Language Understanding (project LUcKY). We would like to thank Evelyn Nowak, Gergana Popova, Vivien Macketanz, and Oliver Marten for their annotation work, and the anonymous reviewers for their valuable feedback.

8. Bibliographical References

Agichtein, E. (2006). Confidence Estimation Methods for Partially Supervised Information Extraction. In *Proceedings of the Sixth SIAM International Conference on Data Mining*.

Alfonseca, E., Filippova, K., Delort, J.-Y., and Garrido, G. (2012). Pattern Learning for Relation Extraction with a Hierarchical Topic Model. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 54–59.

Bentivogli, L. and Magnini, B. (2014). An Italian Dataset of Textual Entailment Graphs for Text Exploration of Customer Interactions. In *Proceedings of the first Italian Computational Linguistics Conference*, Pisa, Italy.

Berant, J., Dagan, I., and Goldberger, J. (2010). Global learning of focused entailment graphs. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1220–1229.

Berant, J., Dagan, I., and Goldberger, J. (2011). Global Learning of Typed Entailment Rules. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 610–619, Stroudsburg, PA, USA.

Berant, J., Dagan, I., Adler, M., and Goldberger, J. (2012). Efficient Tree-based Approximation for Entailment Graph Learning. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 117–125, Jeju Island, Korea.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., and 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*, pages 1247–1250, New York, NY, USA. ACM.

Bunescu, R. C. and Mooney, R. J. (2005). A Shortest Path Dependency Kernel for Relation Extraction. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 724–731.

Dagan, I. and Glickman, O. (2004). Probabilistic Textual Entailment: Generic Applied Modeling of Language Variability. In *Learning Methods for Text Understanding and Mining*.

Dagan, I., Glickman, O., and Magnini, B. (2005). The PASCAL Recognising Textual Entailment Challenge. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment*.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Bradford Books.

Kotlerman, L., Dagan, I., Magnini, B., and Bentivogli, L. (2015). Textual entailment graphs. *Natural Language Engineering*, 21:699–724, 11.

Krause, S., Li, H., Uszkoreit, H., and Xu, F. (2012). Large-Scale Learning of Relation-Extraction Rules with Distant Supervision from the Web. In *Proceedings of the 11th International Semantic Web Conference, Part I*, pages 263–278.

Levy, O., Dagan, I., and Goldberger, J. (2014). Focused Entailment Graphs for Open IE Propositions. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, pages 87–97, Ann Arbor, Michigan.

Liu, H., Hunter, L., Kešelj, V., and Verspoor, K. (2013). Approximate Subgraph Matching-Based Literature Mining for Biomedical Events and Relations. *PLoS ONE*, 8(4).

Mintz, M., Bills, S., Snow, R., and 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*, pages 1003–1011.

Moro, A., Li, H., Krause, S., Xu, F., Navigli, R., and Uszkoreit, H. (2013). Semantic rule filtering for web-scale relation extraction. In *Proceedings of the 12th International Semantic Web Conference*. Springer.

Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., Marinov, S., and Marsi, E. (2007). Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.

Riedel, S., Yao, L., McCallum, A., and Marlin, B. M. (2013). Relation Extraction with Matrix Factorization and Universal Schemas. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 74–84.

Romano, L., Kouylekov, M., Szpektor, I., Dagan, I., and Lavelli, A. (2006). Investigating a generic paraphrase-based approach for relation extraction. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*.

Thomas, P., Pietschmann, S., Solt, I., Tikk, D., and Leser, U. (2011). Not all links are equal: Exploiting Dependency Types for the Extraction of Protein-Protein Interactions from Text. In *Proceedings of BioNLP 2011 Workshop*, pages 1–9. Association for Computational Linguistics.

Yangarber, R., Grishman, R., and Tapanainen, P. (2000). Automatic Acquisition of Domain Knowledge for Information Extraction. In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 940–946.

Zelenko, D., Aone, C., and Richardella, A. (2003). Kernel methods for relation extraction. *The Journal of Machine Learning Research*, 3:1083–1106.