LETTER

Improving Distantly Supervised Relation Extraction by Knowledge Base-Driven Zero Subject Resolution

Eun-kyung KIM[†a)], ***Nonmember and*** **Key-Sun CHOI**[†b)], ***Member***

SUMMARY This paper introduces a technique for automatically generating potential training data from sentences in which entity pairs are not apparently presented in a relation extraction. Most previous works on relation extraction by distant supervision ignored cases in which a relationship may be expressed via null-subjects or anaphora. However, natural language text basically has a network structure that is composed of several sentences. If they are closely related, this is not expressed explicitly in the text, which can make relation extraction difficult. This paper describes a new model that augments a paragraph with a "salient entity" that is determined without parsing. The entity can create additional tuple extraction environments as potential subjects in paragraphs. Including the salient entity as part of the sentential input may allow the proposed method to identify relationships that conventional methods cannot identify. This method also has promising potential applicability to languages for which advanced natural language processing tools are lacking.

key words: *relation extraction, zero subject, distant supervision, Wikipedia*

1. Introduction

As the demand for structured knowledge has increased, considerable interest has emerged for relation extraction (RE) from a large number of documents written in natural language. RE aims to identify and classify semantic relations between pairs of entities extracted from free text. In particular, distant supervision involves extracting relations without human manual annotation by using a knowledge base (KB) [1]–[3]. The distant supervision approach addresses the problem of creating a significant number of training examples by heuristically aligning entities of text for a given KB to learn a relation extractor. For example, suppose that the tuple `sport`(*Los Angeles Dodgers*, *baseball*) is in the KB and assume the sentence "The *Los Angeles Dodgers* is an American professional *baseball* team..." We can then select this sentence as a training instance of the relation `sport`.

Although the distant supervision strategy is a more effective method of automatically labeling training data than directly supervised labeling, it can only extract relations that are limited to a single complete sentence containing two target entities. Because null subject languages can often leave the subject of a sentence unexpressed, it is difficult to obtain the subject and object entities participating in the KB that can be omitted from a single sentence, particularly in null subject languages such as Korean, Japanese, Arabic, and Swedish. As an example, the first two sentences in Korean Wikipedia articles for "Los Angeles Dodgers" are shown below with their English translations, where the omitted words are specified by the symbol ϕ.

- S_1: Los Angeles Dodgers is a professional baseball team based in Los Angeles, California, USA.
(로스앤젤레스 다저스는 미국 캘리포니아 주 로스앤젤레스를 연고지로 하는 프로 야구 팀이다.)
- S_2: ϕ_1 belongs to the National League West of Major League Baseball.
(ϕ_1 메이저 리그 내셔널 리그 서부 지구 소속이다.)

S_1 contains a subject, object, and predicate. Conversely, the subject entity of ϕ_1 is "Los Angeles Dodgers," which is eliminated in the second sentence. In fact, S_2 is clearly a positive training example for the tuple `affiliations`(*Los Angeles Dodgers*, *National League West*) or `sport`(*Los Angeles Dodgers*, *Major League Baseball*), but we cannot label the training instance S_2 according to an existing distant supervision paradigm based on KB projection for one sentence.

There have been studies on tackling these constraints on RE in two or more sentences [4], [5], where the number of possible paths between entities present in other sentences is increased by integrating dependency graphs generated in a single sentence. The dependency graph, which is the key element of these studies, is known to be effective for RE [1]. However, it is difficult to acquire a highly efficient parser for all languages; thus, practical applications cannot extract relationships in various language environments. As another solution, we can apply a pipelined model to first perform a co-reference resolution [6] or zero-anaphora resolution [7] and then perform RE. However, error propagation between processes is a common problem for many natural language processing (NLP) tasks [8]–[11].

In this paper, we focus on creating additional learning data for improved RE by identifying missing entities with KB-driven zero subject resolution. More specifically, we demonstrate a relation classifier that uses the KB in distant supervision to find the salient entity in a paragraph (i.e., any kind of sequence of contiguous sentences from the text) based on centering theory [12] and use it as a supplementary entity for an incomplete sentence (i.e., subjectless sentence fragment) without parsing.

Manuscript received December 14, 2017.
Manuscript revised June 1, 2018.
Manuscript publicized July 11, 2018.

[†]The authors are with School of Computing, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea.
a) E-mail: kekeeo@kaist.ac.kr
b) E-mail: kschoi@kaist.edu

DOI: 10.1587/transinf.2017EDL8270

2. Proposed Framework

2.1 Zero Subject Resolution at the Paragraph Level

Normally, a paragraph that consists of a group of sentences deals with a coherent topic, so any reference can be omitted as long as the context provides the subject to which it is referring. In this paper, we attempt to deal with the null-subject problem to process RE beyond the sentence level. Because the subjectivity of an entity can be determined by how it is expressed in a paragraph, the "**salience**" can be used to identify the subject entity among the entities in the paragraph.

There are certain cues for identifying salience. Unsurprisingly, salient entities tend to be mentioned in the title or first sentence and are then mentioned frequently throughout. However, being included in the title (or first sentence) is neither necessary nor a sufficient condition for salience. Based on these observations, we believe that a KB-based projection of a paragraph that already contains a variety of evidence for an entity is better than developing simple heuristics. The key premise is that the distribution of entities present in a paragraph can be projected into a directed-weighted graph composed of nodes (i.e., KB entities) and edges (i.e., relationships between entities). Salient entities are nodes in a graph that are considered to have salience with regard to some criteria. In general, a node's salience is measured in various ways depending on its application. In this paper, we define salient entities as those with a major effect on the cohesion that occurs in a graph. This assumption is not arbitrary; some of these regularities have been recognized in centering theory. The proposed approach automatically abstracts a paragraph into a set of tuples derived from entities within the paragraph. This is a representation that reflects the referential information and salience of entities with a predefined KB.

To illustrate the problem, Fig. 1 (a) shows a single paragraph with an incomplete sentence (i.e., the second sentence); entities are marked with parenthesized boundaries. When entities of a given paragraph are identified, an entity degree matrix is created (Fig. 1 (c)) by retrieving corresponding entity pairs that exist in the KB tuple (Fig. 1 (b)). For example, the gray tile portion of the entity degree matrix in Fig. 1 (c) shows that the "Los Angeles" entity is referenced once for the head entity position and twice for the tail position. Because salience is related to the reference of entities in a paragraph, "Los Angeles Dodgers," which is the most frequently appearing head entity at five times in this paragraph, is selected as a salient entity for assigning the latent subject entity.

S1: [Los Angeles Dodgers] is a professional [baseball] team based in [Los Angeles], [California], [USA].
S2: Ø1 belongs to the [National League West] of [Major League Baseball].

(a) Paragraph (Entity-marked with '[ ]')

```
home(Los Angeles Dodgers, California)
home(Los Angeles Dodgers, Los Angeles)
league(Los Angeles Dodgers, National League West)
affiliations(Los Angeles Dodgers,Major League Baseball)
sport(Los Angeles Dodgers, Baseball)
isPartOf(Los Angeles, California)
country(California, USA)
largestCity(California, Los Angeles)
country(National League West, USA)
sport(National League West, Baseball)
team(National League West, Los Angeles Dodgers)
country(Major League Baseball, USA)
```

(b) Tuples in KB between given Entities

Head \ Tail	Los Angeles Dodgers	baseball	Los Angeles	California	USA	National League West	Major League Baseball
Los Angeles Dodgers	-	1	1	1		1	1
baseball		-					
Los Angeles			-	1			
California			1	-	1		
USA					-		
National League West	1	1			1	-	
Major League Baseball					1		-

(c) Entity Degree Matrix

Fig. 1 Example graph with given relation tuples between a pair of entities in the paragraph.

Accordingly, a pair of entities that appear together in a single sentence or head a salient entity and appear within a paragraph is considered a potential relation instance. In the case of the second sentence in Fig. 1, the concealed subject ϕ_1 becomes "Los Angeles Dodgers" and provides an opportunity for acquiring possible labeled instances via heuristic alignments, such as `affiliations`(*Los Angeles Dodgers*, *National League West*) or `sport`(*Los Angeles Dodgers*, *Major League Baseball*). Creating numerous training examples is useful, even for relations between two entities that are not explicitly included in one sentence. In this stage, we may derive a relatively large weight for the context feature associated with `affiliations` such as "…belongs to…" that are ignored in conventional distant supervised training steps.

2.2 Sentence Vector for RE

We succeeded in culling omitted entity candidates extracted from the paragraph, as described in the previous sub-section, but it is difficult to resolve an omitted entity's exact position in a sentence without parsing. Therefore, placing the omitted entity at an arbitrary position in the training phase may contaminate the context of the existing sentence. We tackle this issue by presenting a new way of constructing quadruple subsequences based on each entity making up a relation to learning context representations. To do this, we convert each sentence into a word-level matrix where each row is a sentence vector extracted from our model. Figure 2 is an overview of the proposed architecture. A sentence with two annotated entities is first divided into four subsequences based on the two marked entities to focus on each entity part and independently learn each subsequence's context representation: the left context subsequence of e_1 (L_{e_1}), the right context subsequence of e_1 (R_{e_1}), the left context subsequence of e_2 (L_{e_2}), and the right context subsequence of e_2 (R_{e_2}). Each subsequence may contain sentences in which related paired entities are not explicitly mentioned directly; this is completely excluded in conventional distant supervised RE methods. That is, a sentence augmented to

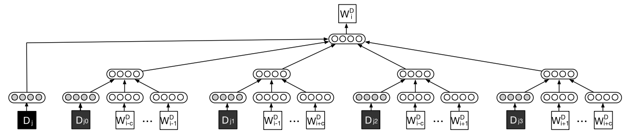

Fig. 2 Model architecture. The context of a single sentence is divided on the basis of the entity participating in the relation type.

include a pair of related entities by the salient entity can participate in developing context information with two null context subsequences such as (*null*, *null*, L_{e_2}, R_{e_2}) without contamination of the existing sentence.

The model computes distributed representations for the context subsequences by using a distributed memory model of paragraph vectors (PV-DM, i.e., Document Vectors [13]). In this work, we use PV-DM for sentence encoding, but other sentence embedding models may also be used. PV-DM is an extension of Word2Vec [14] for learning document embedding and was first applied to training with the entire corpus completely unsupervised. The main difference in Word2Vec is the use of an additional paragraph token from the previous sentence in the document in the context window. This paragraph token maps to a vector space when a different matrix from that employed to map the word is used. More specifically, given the set of a sentence $p = \{s_1, s_2, \ldots, s_n\}$ and the sequence of words $y(s_i) = \{w_{i_1}, w_{i_2}, \ldots, w_{i_m}\}$ sampled from the sentence $s_i \in p$, PV-DM learns the ρ dimensional embeddings of the sentence $s_i \in p$, and each word w_j sampled from $y(f_i)$, i.e., $s_i \in R^\rho$ and $\overrightarrow{w_j} \in R^\rho$, respectively. The model works by considering the word $w_j \in y(s_i)$ that occurs in the context of the sentence s_i and tries to maximize the following log-likelihood:

$$\sum_{j=i_1}^{i_m} \log \Pr(w_j|s_i). \tag{1}$$

The probability is defined as

$$\Pr(w_j|s) = \frac{\exp(\overrightarrow{s} \cdot \overrightarrow{w_j})}{\Sigma_{w \in V} \exp(\overrightarrow{s} \cdot \overrightarrow{w})}, \tag{2}$$

where V is the vocabulary of all words across all sentences in p.

2.3 Distantly Supervised RE

We define our RE task as follows: Given a sentence s' that is a complementary form of an incomplete (e.g. subject-less) sentence s with marked entities e_1 and e_2 and a set of relations $R = \{r_1, \ldots, r_n\}$, we formulate the task of identifying the semantic relation as the following standard classification problem:

$$f : (P, E) \rightarrow R, \tag{3}$$

where P is the set of all paragraphs, the paragraph $p \in P$ is the set of contiguous sentences $\{s'_1, s'_2, \ldots, s'_m\}$, and E is the set of entity pairs. Our training objective is to learn a representation of the sentences in a paragraph such that a regression layer can predict the correct label. Basically, RE has three main feature types: contextual (or lexical), semantic, and syntactic information. In this paper, we only focus on contextual information that can be obtained from the sentence itself to analyze the effect of the RE architecture that learns the embedding of a word sequence after compensating for sentences with zero subject entity resolution. In other words, we construct sentence feature vectors from the words themselves in s' at the training stage. We use the following values for all experiments: initial learning rate ($\alpha = 0.025$) and minimum learning rate ($\alpha_{min} = 0.002$). The learning rate decreases linearly in each epoch from the initial rate to the minimum rate. We use the following parameters during training: (1) we optimize the embedding vector size to 400, (2) we optimize the left/right fixed context window size to five, (3) we use the min count (β) to 1, and (4) we run the experiment with the 20 training epochs. We set β to 1 in our model to ensure that we treat all tokens in the context as meaningful and use them for training. All PV-DM training is done with the Gensim library. A multi-class logistic regression classifier is employed in the next step given the sentence embedding inferred from the vectorized model. Once the model is trained, each paragraph in the test dataset is inferred directly from the model.

3. Experiments

3.1 Dataset

We evaluated the performance of our proposed method by training and testing Spanish, Polish, Korean, and Greek versions of Wikipedia as the textual corpus. Specifically, we used a snapshot from July 2017†. We used DBpedia [15] as background knowledge for supervision; this is a large KB of entities and relationships. Because DBpedia provides tuple downloads in multiple languages††, it is advantageous for building an efficient RE model with several languages.

3.2 Implementation Details

Distantly supervised RE can be viewed as a two-step process: (A) detecting entities of interest and (B) determining the relationship between the possible set of entities. We explain this in detail in the following sub-sections.

3.2.1 Entity Detection

We processed the Wikipedia text by using the following steps. First, paragraphs consisting of two or more consecutive sentences and that were separated by blank lines or different section names were extracted from a Wikipedia article. Second, entity detection of a given paragraph was performed by using WikiLinks, which links one page to another to provide inter-entity access. Because a Wikipedia article is collaboratively co-edited by a large number of people worldwide, it can be considered as a kind of crowd-sourcing-based entity detecting system. Each DBpedia entity is directly derived from the URL of the source Wikipedia article. For example, a typical sentence of the Wikipedia text may appear similar to "*[[Steve Jobs]] was [[Businessmen]] in [[United States]],*" and the brackets are used to

†https://dumps.wikimedia.org/

††http://wiki.dbpedia.org/downloads-2016-10

Table 1 Statistics of dataset. **Standard** denotes the labeling results of two entities in a sentence according to the existing distant supervision paradigm. $\mathcal{A}$**(title)** interprets the title of the Wikipedia document as the head entity because the title is the protagonist in the document, whereas $\mathcal{A}$**(Salient)** represents the extension of the salient entity in the paragraph as the subject entity.

Languages	**Data**	**#(Sentences)**
Greek	Standard	52,687
	$\mathcal{A}$(Title)	125,705
	$\mathcal{A}$(Salient)	187,607
Korean	Standard	163,658
	$\mathcal{A}$(Title)	414,825
	$\mathcal{A}$(Salient)	590,644
Polish	Standard	390,661
	$\mathcal{A}$(Title)	1,199,276
	$\mathcal{A}$(Salient)	1,573,486
Spanish	Standard	1,231,759
	$\mathcal{A}$(Title)	2,791,537
	$\mathcal{A}$(Salient)	3,631,466

represent the WikiLink. In this snippet, there are three DBpedia entities: "dbr[†]:Steve Jobs," "dbr:Businessmen," and "dbr:United States." We intend to detect these as entities. In practice, an alternative entity detection system may be required because Wikipedia links relatively little text; however, that endeavor is beyond this paper's scope.

3.2.2 Relation Classification

Once entities are detected, salient entities are selected from each paragraph by calculating the out-degree centrality based on the entity degree matrix derived from the DBpedia tuple. Then, as described earlier, the training data are automatically generated from a sentence containing a pair of related entities exists in a single sentence range, or a sentence containing a single tail entity that can be related to a salient entity in a paragraph. These labeled training data were leveraged for sentence embedding to generate a feature vector.

The statistics of the training data constructed under the distant supervision assumption are shown in Table 1. We made one conventional labeling scheme and two extensions (one is proposed salient-based extension and the other is an extension with the title). Experimental results show that the amount of training data required for RE is significantly increased in all four languages. The number of relation types used for in the case of the Korean experimental data, totaling 228 DBpedia relations are used. The number of relationship types for the other languages were 239, 131, and 271 in Greek, Polish, and Spanish, respectively. Details about target relations used as distant supervision are described in Appendix.

To further demonstrate the effectiveness of the salient entity method, we performed an additional experiment with two added heuristic extension methods for Korean data, as presented in Table 2. As presented in Table 2, because the method of sentence augmentation by adding the omitted entity to the sentences had a higher score than **Standard**,

[†]Prefix used in this paper: dbr='http://dbpedia.org/resource/'.

Table 2 Experimental results before and after extensions were labeled based on the distant supervision paradigm with various baselines for entity discovery.

	Precision	Recall	F1 Score
Standard	0.44	0.40	0.38
$\mathcal{A}$(Title)	0.47	0.42	0.38
$\mathcal{A}$(Max)	0.52	0.48	0.45
$\mathcal{A}$(First)	0.51	0.46	0.43
$\mathcal{A}$(Salient)	**0.58**	**0.54**	**0.52**

the proposed sentence augmentation method increased the positive learning instances for RE. In particular, the use of salience to achieve the best coherent paragraph-driven-graphs yielded a better RE performance than other heuristic methods.

3.3 Evaluation

There is no gold-standard annotated dataset with distant supervision, so evaluations typically use the held-out strategy. This was done automatically by withholding half of the DBpedia relationship knowledge during training and comparing the newly discovered relationship instances against the withheld data. The ratio of training and testing in our experiment was 8 : 2. We conducted several baselines to validate the approach in the experiments. We obtained the precision, recall, and F1 scores for each relation type in the experiment. Then, we obtained the sum of the weighted averages for each performance measure from each relationship type rather than simply the harmonic mean. Table 3 shows the results in terms of the precision (P), recall (R), and F1 score (F). The proposed approach outperformed the baseline for the four experimental languages (Greek, Korean, Polish, and Spanish) by 9%, 4%, 2%, and 5%, respectively, according to the F1 score. We conducted the paired t-test to confirm the significance of the salience-based zero subject resolution quality improvements over title-based subject resolution. For Greek, Korean, Polish and Spanish, p values for $\mathcal{A}$(Salient) against $\mathcal{A}$(Title) are 8.11E-10, 4.51E-11, 1.78E-9, and 1.24E-5. When p values are less than 0.05, the results are statistically significant. Although there is much room for improvement in precision and recall, our results indicated that our approach can be useful for extracting a large number of relationship types by using labeled data in a KB without using advanced NLP tools such as parsers.

To comprehensively evaluate the performance of the proposed approach, we compared our algorithm with several variants of RE model with respect to the augmented sentence of the zero entity as follows. In this experiment, sentences used for training in augmented form from a salient entity of the paragraph added the salient entity to the beginning of the original sentence.

- **TF-IDF** consists of the vector in which every sentence represents the calculation of the term frequency-inverse document frequency value for each word. We fed this vector to the logistic regression classifier to classify the relationship.

Table 3 Experimental results for the test data. **Ext:SalientEntity** represents the proposed approach, which showed a statistically significant improvement over the **Ext:TitleEntity** system according to the t-test.

	Standard			$\mathcal{A}$(Title)			$\mathcal{A}$(Salient)			
Languages	P	R	F	P	R	F	P	R	F	*p*-value
Greek	0.47	0.40	0.42	0.47	0.40	0.42	0.53	0.49	**0.51**	8.11×10^{-10}
Korean	0.51	0.46	0.48	0.51	0.47	0.48	0.58	0.54	**0.52**	4.51×10^{-11}
Polish	0.58	0.55	0.57	0.59	0.55	0.57	0.61	0.58	**0.59**	1.78×10^{-9}
Spanish	0.51	0.45	0.47	0.49	0.42	0.45	0.56	0.48	**0.52**	1.24×10^{-5}

- **AvgWord2Vec** averages the word embeddings (Word2Vec) in a sentence to generate sentence representations as features for the logistic regression classifier.
- **Sen2Vec** is an unsupervised sentence vector framework that is used to learn distributed representations for texts; here, the basic PV-DM was used to obtain sentence representations. We tested this feature with the logistic regression classifier.
- **SubSeq2Vec** is the proposed model, in this paper, created by integrating subsequence embeddings with the logistic regression classifier. The superscripts at the end indicate the type of integration: Q for integrating to the four subsequences with the proposed method and T for integrating to the three subsequences, such as 1) a word sequence in front of the head entity, 2) a word sequence between the head and tail entities, and 3) a word sequence behind the tail entity.
- **CNN, LSTM** represent the results of performing RE with the deep neural network frameworks convolutional neural network (CNN) and long short-term memory (LSTM), respectively. Both models used pre-trained word embedding from Korean Wikipedia (including only sentences for which the omitted subject had not been resolved). The parameter settings of CNN and RNN were the same as those in [16] and [17], respectively.

The proposed model exhibited a higher classification performance than several classification models. In particular, **SubSeq2Vec** showed a 35%–52% improvement in the F1 score compared **Sen2Vec**. The results indicate that the integrating architecture of the four subsequence embeddings effectively captured the semantic meanings and their inherent relations and obtained more accurate sentence representations for RE. Although deep neural network-based RE models record highly effective performances in recent years, it can be seen that the missing information in the pre-trained word embedding is an obstacle in completing an effective RE. We have found that we can improve relationship extraction performance by adding new information that is not explicitly revealed in the sentence range. In other words, if the hidden subject or object is identified in the text and the cross-reference problem is solved, we expect to be able to generate more robust word embedding referring to the deep neural network-based RE.

Table 4 Precision, recall, and F1 scores of the test data for various RE models in a Korean data experimental environment.

	Precision	Recall	F1 Score
TF-IDF	**0.65**	0.39	0.38
AvgWord2Vec	0.52	0.45	0.43
Sen2Vec	0.44	0.36	0.35
SubSeq2Vec^T	0.51	0.42	0.40
SubSeq2Vec^Q	0.58	**0.54**	**0.52**
CNN	0.57	0.31	0.39
LSTM	0.57	0.36	0.42

4. Conclusion

We focused on the distant supervision paradigm for RE from passages not containing both entities expected to participate in a relation. We showed that a distant supervision-based model that does not require labeling can be used to represent the context of sentences and the surrounding relationship mentions to enable relation classification at the paragraph level. Experiments were conducted with four different languages—Spanish, Korean, Greek, and Polish—to demonstrate the model's effectiveness for practical application without the use of natural language processing tools in various language data.

Acknowledgments

This work was supported by the Industrial Strategic Technology Development program (10072064, Development of Novel Artificial Intelligence Technologies to Assist Imaging Diagnosis of Pulmonary, Hepatic, and Cardiac Diseases and Their Integration into Commercial Clinical PACS Platforms) funded by the Ministry of Trade Industry and Energy (MI, Korea) and Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIP) (2013-0-00109, WiseKB: Big data based self-evolving knowledge base and reasoning platform).

References

[1] M. Mintz, S. Bills, R. Snow, and D. Jurafsky, "Distant supervision for relation extraction without labeled data," Proc. 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, ACL '09, Stroudsburg, PA, USA, pp.1003–1011, Association for Computational Linguistics, 2009.

[2] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D.S. Weld, "Knowledge-based weak supervision for information extraction of

overlapping relations," Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Volume 1, HLT '11, Stroudsburg, PA, USA, pp.541–550, 2011.
[3] S. Riedel, L. Yao, A. McCallum, and B.M. Marlin, "Relation extraction with matrix factorization and universal schemas," Joint Human Language Technology Conference/Annual Meeting of the North American Chapter of the Association for Computational Linguistics, pp.74–84, 2013.
[4] N. Peng, H. Poon, C. Quirk, K. Toutanova, and W.T. Yih, "Cross-sentence N-ary relation extraction with graph LSTMs," Trans. Association for Computational Linguistics, vol.5, pp.101–115, 2017.
[5] C. Quirk and H. Poon, "Distant supervision for relation extraction beyond the sentence boundary," Proc. 15th Conference of the European Chapter of the Association for Computational Linguistics, ed. M. Lapata, P. Blunsom, and A. Koller, pp.1171–1182, Association for Computational Linguistics, 2017.
[6] K. Clark and C.D. Manning, "Entity-centric coreference resolution with model stacking," Proc. 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp.1405–1415, The Association for Computer Linguistics, 2015.
[7] R. Mitkov, "Anaphora resolution: The state of the art," Tech. Rep., University of Wolverhampton, Wolverhampton, 1999.
[8] C. Quirk and S. Corston-Oliver, "The impact of parse quality on syntactically-informed statistical machine translation," Proc. 2006 Conference on Empirical Methods in Natural Language Processing, ed. D. Jurafsky and Gaussier, pp.62–69, ACL, 2006.
[9] B. Yang and C. Cardie, "Joint inference for fine-grained opinion extraction," ACL, vol.1, pp.1640–1649, 2013.
[10] D. Han, P. Martínez-Gómez, Y. Miyao, K. Sudoh, and M. Nagata, "Effects of parsing errors on pre-reordering performance for Chinese-to-Japanese smt," PACLIC, National Chengchi University, Taiwan, 2013.
[11] D. Zeng, K. Liu, Y. Chen, and J. Zhao, "Distant supervision for relation extraction via piecewise convolutional neural networks," Proc. 2015 Conference on Empirical Methods in Natural Language Processing, ed. L. Márquez, C. Callison-Burch, J. Su, D. Pighin, and Y. Marton, pp.1753–1762, The Association for Computational Linguistics, 2015.
[12] M. Walker, A. Joshi, and E. Prince, eds., Centering Theory in Discourse, Clarendon Press, Oxford, 1998.
[13] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," Proc. 31st International Conference on Machine Learning (ICML-14), pp.1188–1196, 2014.
[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
[15] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "DBpedia – A crystallization point for the web of data," Web Semantics: Science, Services and Agents on the World Wide Web, vol.7, no.3, pp.154–165, 2009.
[16] Y. Kim, "Convolutional neural networks for sentence classification," Proc. 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, pp.1746–1751, 2014.
[17] F. Li, M. Zhang, G. Fu, and D. Ji, "A neural joint model for entity and relation extraction from biomedical text," BMC Bioinformatics, vol.18, no.1, pp.198:1–198:11, 2017.

Appendix: Implementation Details

A.1 Relationships in Greek

administrativeDistrict affiliation album almaMater anthem archipelago architect artist aspectRatio assembly associate associatedRocket author automobilePlatform award bandMember basedOn battle birthPlace bodyStyle broadcastArea builder canton capital capitalCountry category chairman chancellor channel chiefEditor child citizenship city class coach commander company composer computingPlatform constellation country county coverArtist cpu creator crosses currency currentMember dean deathCause deathPlace deputy designer developer director discoverer distributor district division draftTeam editor education employer endPoint engine era ethnicGroup ethnicity europeanAffiliation europeanParliamentGroup field firstAscentPerson firstFlight firstRace firstWin followingEvent format formerBandMember foundationPlace foundedBy founder garrison genre governor ground hasVariant headquarter highestPosition homeArena hometown honours house ideology illustrator industry inflow influenced influencedBy ingredient institution instrument internationalAffiliation isPartOf island keyPerson knownFor language languageFamily largestCity lastRace lastWin layout leader leaderName league leftTributary license lieutenant locatedInArea location magazine mainInterest majorIsland majorShrine manufacturer mayor mergedIntoParty militaryBranch mission monarch mountainRange mouthCountry mouthPlace movement municipality museum musicComposer musicFusionGenre musicalArtist musicalBand nationalAffiliation nationalTeam nationality nextMission nominee notableCommander notableIdea notableWork occupation operatingSystem operator opponent origin outflow owner owningCompany painter parent parentCompany parentMountainPeak parentOrganisation partner party pastMember patron period person philosophicalSchool picture pictureFormat place playerInTeam politicalPartyInLegislature politicalPartyOfLeader position predecessor president previousEvent previousMission previousWork primeMinister producer product profession programmingLanguage province publisher recordLabel recordedIn region regionServed related relative religion residence restingPlace rightTributary riverMouth rocket routeStart royalAnthem saint season secondTeam service shuttle significantBuilding similar sisterCollege sisterNewspaper source sourceConfluence spokenIn spokesperson sportSpecialty spouse stadium starring startPoint state stateOfOrigin style stylisticOrigin subsequentWork subsidiary successor superintendent team thirdTeam timeZone translator type usedInWar usingCountry variantOf veneratedIn vicePresident writer youthWing

A.2 Relationships in Korean

academicDiscipline album alliance almaMater anthem archipelago architect arrondissement artist assembly associatedAct author award basedOn battle beatifiedBy binomial birthPlace board bodyDiscovered bodyStyle builder canonizedBy capital category chairperson channel child cinematography citizenship city class closingFilm club colour commandStructure commander composer computingInput computingMedia computingPlatform contractor country countryOrigin coverArtist creativeDirector creator crew crosses currency currentMember deathCause deathPlace department derivative designCompany designer developer director distributor division doctoralAdvisor doctoralStudent domain editing education endingTheme engine engineer equipment era ethnicity executiveProducer family field firstAscentPerson firstFlight format formerTeam foundedBy gameEngine garrison genre genus governmentType ground guest hasVariant headquarter hometown hubAirport ideology illustrator industry inflow influenced influencedBy instrument isPartOfMilitaryConflict island keyPerson kingdom knownFor language languageFamily largestCity lastAppearance lastFlight launchPad launchSite leader leaderName league leftTributary license literaryGenre locatedInArea location lyrics mainInterest mainOrgan maintainedBy majorIsland manager managerClub manufacturer militaryBranch mission mouthCountry mouthRegion musicBy musicComposer musicFusionGenre musicSubgenre musicalArtist musicalBand musicians namedAfter narrator nationalTeam nationality nextMission nonFictionSubject notableIdea notableWork occupation officialLanguage openingFilm openingTheme operatingSystem operator opponent order origin outflow owner owningOrganisation parent parentCompany parentMountainPeak parentOrganisation partner party pastMember philosophicalSchool phylum place populationPlace position precursor predecessor president previousMission previousWork producer product programmingLanguage publisher recordLabel recordedIn region regionServed related relative religion residence restingPlace rightTributary rocketFunction routeEnd routeStart season selection series service significantBuilding significantProject sisterNewspaper source sourceConfluence sourceConfluenceCountry sourceConfluenceRegion sourceCountry sourceMountain sourcePlace sourceRegion spacecraft species sportGoverningBody spouse starring subsequentWork subsidiary successor targetAirport team tenant timeZone translator type usedInWar usingCountry vein veneratedIn vicePresident voice writer youthClub

A.3 Relationships in Polish

alliance almaMater anthem archipelago architect artist associatedBand associatedMusicalArtist author award beatifiedBy beatifiedPlace binomial binomialAuthority birthPlace bodyStyle canonizedBy canonizedPlace capital ceo chairman chairperson cinematography citizenship city class commander composer computingMedia computingPlatform constellation country county creator crest deathPlace designer director discoverer distributor domain editing editor engine europeanParliamentGroup family field firstAscentPerson format foundedBy founder gameEngine garrison genre genus ground head headquarter hometown ideology industry instrument internationalAffiliation isPartOfMilitaryConflict island keyPerson kingdom language languageFamily leader leaderName league locatedInArea location manager managerClub manufacturer militaryRank mountainRange musicComposer musicalArtist musicalBand narrator nationality nonFictionSubject notableWork occupation officialLanguage operator order origin owner owningOrganisation party patron person phylum picture place portrayer position predecessor previousWork producer product programmingLanguage province publisher recordLabel recordedIn rector related relatedMeanOfTransportation religion residence service silCode similar sourcePlace spokenIn spouse subsequentWork successor team thumbnail timeZone trainer type vicePresident writer youthWing

A.4 Relationships in Spanish

academyAward achievement administrator affiliation aircraftUser album alliance almaMater anthem archipelago architect architecturalStyle artery artist associate associateEditor author authority automobilePlatform award bandMember battle binomial binomialAuthority birthPlace broadcastArea broadcastNetwork builder buildingType campus capital category chairman chairperson chancellor channel chiefEditor child cinematography citizenship city class club colour commander commune company compiler composer computingInput computingMedia computingPlatform constellation constructionMaterial councilArea country county coverArtist creator curator currency currentProduction deathCause deathPlace debutTeam department designer destination developer director discoverer distributor division doctoralAdvisor doctoralStudent domain editing editor education educationSystem emmyAward employer ethnicity event explorer family field firstAscentPerson firstRace firstWin flagBearer format formerBandMember formerTeam foundationPlace foundedBy founder gameArtist gameEngine gender genre genus goldenGlobeAward goyaAward grammyAward guest hasVariant headquarter homeArena homeStadium hometown hubAirport humanDevelopmentIndexRankingCategory ideology illustrator industry influenced influencedBy instrument internationalAffiliation isPartOfMilitaryConflict keyPerson kingdom knownFor language languageFamily languageRegulator largestCity lastAppearance leader leaderParty league license literaryGenre locatedInArea location locationCity locationCountry lyrics mainOrgan majorShrine management manager manufacturer map mayor militaryBranch militaryRank militaryUnit monarch mountainRange mouthPosition movement municipality musicComposer musicalArtist musicalBand musicians nationalOlympicCommittee nationalTeam nationality network nonFictionSubject notableWork occupation officerInCharge officialLanguage officialOpenedBy olympicOathSwornByAthlete openingTheme operatingSystem operator order origin owner owningCompany owningOrganisation parent parentOrganisation parish partner party period person personFunction phylum picture pictureDescription pictureFormat place placeOfBurial position precursor predecessor premierePlace presenter president previousInfrastructure previousWork primeMinister principal producer product productionCompany profession programmeFormat programmingLanguage province provost publisher recordLabel recordedIn rector region regionServed related relation relative religion religiousHeadLabel religiousOrder residence riverMouth routeEnd routeEndLocation routeJunction routeStart routeStartLocation saint schoolPatron season secretaryGeneral series service servingRailwayLine showJudge sisterStation sourceConfluence sourcePlace species spokenIn spokesperson sport sportSpecialty spouse starring state style subsequentInfrastructure subsequentWork subsidiary successor taoiseach team territory thumbnail timeZone torchBearer tradeMark trainer translator type university veneratedIn viceChancellor vicePrincipal voice voiceType webcast writer youthWing