

# BERT-based Spatial Information Extraction

Hyeong Jin Shin\*, Jeong Yeon Park\*\*, Dae Bum Yuk\*\*, Jae Sung Lee\*\*

Computer Science Major, School of Electrical Engineering and Computer Science,  
Graduate School of Chungbuk National University, Korea

hjshin1985@gmail.com\*, {parkjeongyeon, daebum1994, jasonlee}@cbnu.ac.kr\*\*

## Abstract

Spatial information extraction is essential to understand geographical information in text. This task is largely divided to two subtasks: spatial element extraction and spatial relation extraction. In this paper, we utilize BERT (Devlin et al., 2018), which is very effective for many natural language processing applications. We propose a BERT-based spatial information extraction model, which uses BERT for spatial element extraction and R-BERT (Wu and He, 2019) for spatial relation extraction. The model was evaluated with the SemEval 2015 dataset. The result showed a 15.4% point increase in spatial element extraction and an 8.2% point increase in spatial relation extraction in comparison to the baseline model (Nichols and Botros, 2015).

## 1 Introduction

Extracting spatial relations from text is a type of relation extraction, focusing on the static and dynamic spatial relations in the text. It is essential for natural language understanding systems, such as robot navigation systems and question-answering systems, to understand geographical relations or to track moving objects.

For example, in the sentence, “Tom is on the box,” we find a static relation in which *Tom* is the trajector, *box* is the landmark, and *on* denotes their static spatial relation. In the following sentence, “He steps down from the box to the ground,” we also find a dynamic spatial relation, in which *He* (*Tom*) is the mover, *steps down* is the trigger, *box* is

the source, and *ground* is the destination. Using simple inference based on the extracted relations, we can infer a new relation: “Tom is on the ground now.”

The task is largely divided into two subtasks<sup>1</sup>: spatial element extraction and spatial relation extraction. Finding candidate elements for spatial relations roles, such as the trajector, landmark, and trigger, defines spatial element extraction. Finding or verifying relations among the role candidates defines spatial relation extraction.

Many natural language processing techniques and machine learning methods have been applied to spatial information extraction. For example, a conditional random field (CRF) model (Lafferty et al., 2001) is used for spatial element extraction, and support vector machine (SVM) (Suykens and Vandewalle, 1999; Roberts and Haragagiu, 2012) and convolutional neural net (CNN) (Mazalov et al., 2015) models are used for spatial relation extraction. Various language resources, such as GloVe (Pennington et al., 2014), WordNet (Salaberri et al., 2015), and PropBank (Salaberri et al., 2015) are also used for spatial information extraction.

In this paper, we propose a BERT-based spatial information extraction model that utilizes BERT (Devlin et al., 2018) extensively for both spatial element extraction and spatial relation extraction. Recently, many context-aware language models have been developed, including not only BERT, but also ELMO (Peters et al., 2018), XLNet (Yang et al., 2019), and GPT (Radford et al., 2018), among others. We chose BERT simply because many downstream applications of the BERT system have been developed for named entity

---

<sup>1</sup> The subtasks are defined in more detail for evaluation in SemEval-2015 Task 8 (Pustejovsky et al., 2015). Spatial element extraction and spatial relation extraction correspond to 1.b and 1.d tasks, respectively, in the definition. However,

in this paper, spatial element extraction task is extended to extract not only spatial elements such as paths, places, motions, spatial entities, for instance, but also spatial signals and motion signals.

recognition and semantic role labeling, which can be easily applied to spatial information extraction.

In Section 2, we briefly summarize related works. In Section 3, we describe our proposed model, which consists of three modules, specifically a spatial element extraction model, a triple candidate generator, and a spatial relation extraction model. Section 4 presents the experimental results of our model. Finally, section 5 concludes the paper.

## 2 Related Works

An early method of spatial information extraction was introduced as spatial role labeling (SpRL) by Kordjamshidi et al. (2011). SemEval-2012 introduced a spatial role labeling task mainly focusing on static spatial relations. SemEval-2013 expanded static spatial relations to capture fine-grained semantics and to include dynamic spatial relations.

SemEval-2015 was the first shared task conference to evaluate implementation systems for the SpaceEval annotation scheme, which is the current spatial information annotation scheme (Pustejovsky et al., 2015). Many spatial information extraction systems have been developed based on the SpaceEval annotation scheme.

Nichols and Botros (2015) proposed the SpRL-CWW model, which uses a CRF model (Lafferty et al., 2001) for spatial element extraction and an SVM model (Suykens and Vandewalle, 1999) for spatial relation extraction. It uses many input features for element extraction, such as word embedding using GloVe (Pennington et al., 2014), named entities, part of speech tags and dependency parse labels. SVM is used to filter out correct triples from all possible combinations of triples. D’Souza and Ng (2015) proposed the UTD-SpRL model based on SVM, which includes more than 100 different features generated by a greedy feature selection technique and uses the joint detection of a relation’s arguments. The X-Space model proposed by Salaberry et al. (2015) uses node information, such as the place, position, location and so forth, included in WordNet for spatial element extraction. It also uses argument information in PropBank for spatial relation classification.

A multimodal approach that uses image and text information simultaneously in a multimodal spatial role labeling (mSpRL) shared task was also

presented in CLEF 2017 (Kordjamshidi et al., 2017), but the result was not satisfactory (Zablocki et al., 2017).

Mazalov et al. (2015) extracted spatial roles and their relations by adapting a convolutional neural network based system developed for semantic role labeling. The pre-existing system was successfully adapted to spatial information extraction.

Dan et al. (2020) proposed spatial BERT to predict the spatial relation between two entities given an image involving them. The spatial BERT was composed of a spatial model, implemented with a feed forward network, and a language model, which were implemented with BERT. The language model is used as complementary features to predict unseen (untrained) relations in images. Despite the fact that BERT is used as the language model in this approach, spatial relation extraction is limited to relation detection for the given subject and object entities in the image. Our approach also uses BERT but deals instead with the entire process of relation extraction from raw text; we extract spatial elements from raw text, determine their corresponding spatial roles, and find spatial relations from the spatial roles.

## 3 Spatial Information Extraction Model

We divide the spatial information extraction task into two subtasks: spatial element extraction and spatial relation extraction, according to the ISOspace annotation scheme (ISO, 2014; Pustejovsky et al., 2015). For the integrated system, we pipelined the two subtasks via a triple candidate generator.

Figure 1 shows the overall architecture of our system. A sentence is inputted to the element extractor, which is jointed with one of link’s role modules. The element extractor outputs the spatial elements and spatial roles jointly. The spatial roles are combined to triples as spatial relation candidates by the triple candidate generator. The triple candidates are classified as either valid relations or invalid relations by the relation extractor. Each module is described in the following sections.

For the general architecture of spatial relation extraction, two restrictions are imposed in this work. First, only three arguments are allowed. According to ISOspace, we have certain arguments for each relation, as shown Table 1. To maintain the static architecture of relation extraction, we set the number of arguments of each relation to three.

Therefore, we keep all arguments for QSLink (Qualitative Spatial Link), OLink (Orientation information Link), and MeLink (Measurement Link), but for MvLnk (Movement Link), we choose three arguments out of seven: *mover*, *goal*, and *motion*.

Second, only one prime spatial role is determined in the element extraction stage. In multiple relations in a sentence, an entity may be related to multiple relations in multiple roles. In this case, it is necessary to choose only one role in a typical case. For example, in the sentence in Figure 2, ‘vase’ has two roles for each relation: *trajectory* and *landmark*. Because sharing these two different roles most frequently occurs, we decided to include these roles as one role label, *traLand*. The triple candidate generator interprets this role label as two roles separately, *trajectory* and *landmark*, for triple candidate generation.

Spatial relation	QSLink	<b>trajectory, landmark, trigger</b>
	OLink	<b>trajectory, landmark, trigger</b>
	MvLnk	<b>mover, goal, motion, midpoint, landmark, source, path</b>
	MeLink	<b>trajectory, landmark, measure</b>

Table 1: Spatial roles of each link

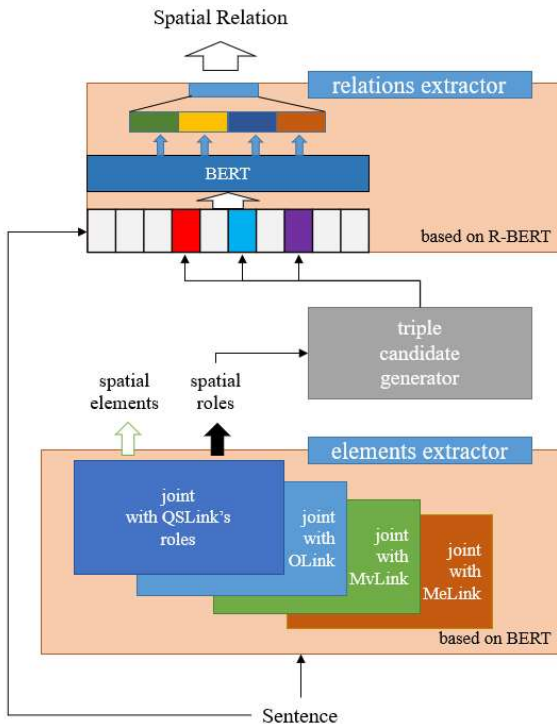


Figure 1: Pipeline model for spatial information extraction

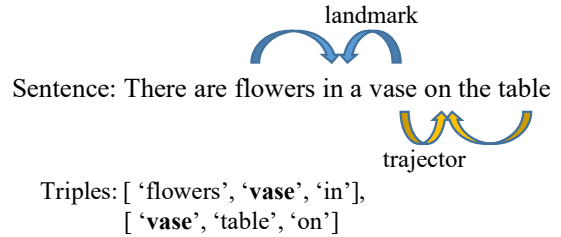


Figure 2: Example of sharing *trajectory* and *landmark* roles: *vase* has two roles, *trajectory* and *landmark*

### 3.1 Spatial Element Extraction

Spatial element extraction is a problem of sequence labeling, which can be easily solved with BERT. The structure of the model is shown in Figure 3. A sentence is segmented into word pieces and they are inputted to BERT to extract the spatial elements and spatial roles jointly. In previous methods, many features are extracted through preprocessing for learning by CRFs (Nichols and Botros, 2015). However, the BERT-based spatial element extraction module does not require any preprocessing for feature extraction; rather, it requires only raw text as input for fine tuning.

Multi-layer perceptron (MLP), used for a classifier on top of BERT, performs fully connected layer computation and produces IOB-based tags for annotation. Because BERT is based on word pieces (Wu et al., 2016), the outputs are also word pieces. For sequence labeling, we labeled only the first word piece. For example, we can assume that ‘flower’ is divided into two word pieces ‘flow’ and ‘##er’, with only ‘flow’ then annotated as a normal tag, such as a Spatial Entity tag, whereas ‘##er’ is annotated as an Other tag.

We use a joint model for spatial element extraction and spatial role extraction. Two classifiers are located on top of the BERT system and share the same parameters for BERT fine tuning. We noted an improvement in the joint model over the single model during a preliminary test on Korean data (Kim and Lee, 2016).

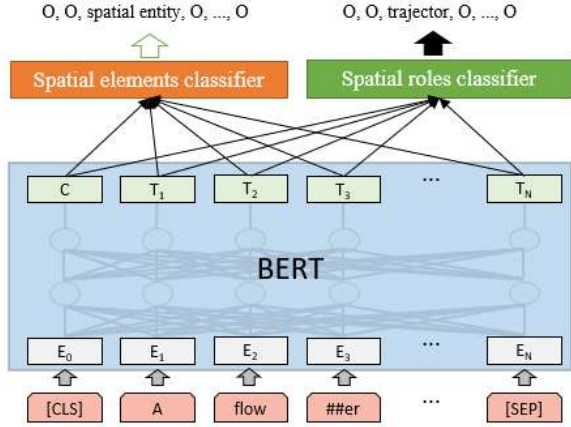


Figure 3: Joint model for spatial elements and spatial role extraction

### 3.2 Triple Candidate Generator

Because the spatial role extractor produces only entity tags, we do not know which entity is related to which entity, especially when there are multiple relations. Moreover, we do not know the relation type to which they belong. The triple candidate generator produces all possible combinations of given spatial roles for the spatial relation extractor to determine which combination and type should be chosen.

For example, in the sentence shown in Figure 4, we have two trajectors, ‘bike’ and ‘puppy’; two landmarks, ‘warehouse’ and ‘gate’; and two triggers, ‘by’ and ‘in front of’. The triple candidate generator produces all combinations of trajector, landmark, and trigger. In this case, it produces 8 ( $2*2*2$ ) triple candidates. Generally, for a set of trajector  $T$ , a set of landmark  $L$ , and a set of trigger  $G$ , we have a number of Cartesian product triple candidates:  $|T|*|L|*|G|$ .

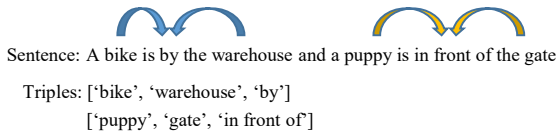


Figure 4: Example for multiple triplets in a sentence, represented with triple  $[trigger, trajector, landmark]$

### 3.3 Spatial Relation Extraction

Spatial relation extraction is a task to identify the relation between given entities, in our case, triple entities. A similar task has been done by using BERT in semantic role labeling (SRL), in which a relationship is classified for two given semantic

role arguments (Wu and He, 2019). This model showed the best performance in SRL, and we refer to this model as R-BERT in this paper. We adopted R-BERT for spatial relation extraction, but we modified two aspects of the model. We extended two arguments to three arguments, and we include null argument for the case of a movement link.

Figure 5 shows the structure of the modified model. A sentence, marked with a triple candidate, is inputted to BERT. The BERT outputs of each of the roles in a triple and CLS token are averaged and go through the fully connected network. The four outputs are concatenated and then go through fully connected network again. The *softmax* of the output is the final result to determine the validity of the triple relation.

For the argument span, the input format is changed with the start index and end index along with the words [words, start index, end index]. In the case of two arguments, we utilize formulae (1) to (5) for the spans of  $i$  and  $j$ , and  $k$  and  $m$ :

$$H'_0 = W_0(\tanh(H_0)) + b_0 \quad (1)$$

$$H'_1 = W_1[\tanh(\frac{1}{j-i+1}\sum_{t=i}^j H_t)] + b_1 \quad (2)$$

$$H'_2 = W_2[\tanh(\frac{1}{m-k+1}\sum_{t=k}^m H_t)] + b_2 \quad (3)$$

$$h'' = W_3[\text{concat}(H'_0 + H'_1 + H'_2)] + b_3 \quad (4)$$

$$p = \text{softmax}(h'') \quad (5)$$

We extended the model to operate for three arguments of spatial information extraction. We added one additional *tanh* output for a trigger with span  $q$  and  $r$ , as shown in formula (6). We also modified formula (4) to formula (7) to include the trigger:

$$H'_3 = W_4[\tanh(\frac{1}{r-q+1}\sum_{t=r}^q H_t)] + b_4 \quad (6)$$

$$h'' = W_3[\text{concat}(H'_0 + H'_1 + H'_2 + H'_3)] + b_3 \quad (7)$$

For the null argument in the case of a movement link, we utilize the last character in the sentence. For example, the sentence ‘‘John leaves from school.’’ contains a *mover*, ‘‘John,’’ and a *motion*, ‘‘leaves,’’ but it does not have a *goal*. In this case, we represent the *goal* as the null argument. Therefore, we have a three-argument span: [‘John’, 0, 0], [‘.’, 4, 4], and [‘leaves’, 1, 1].

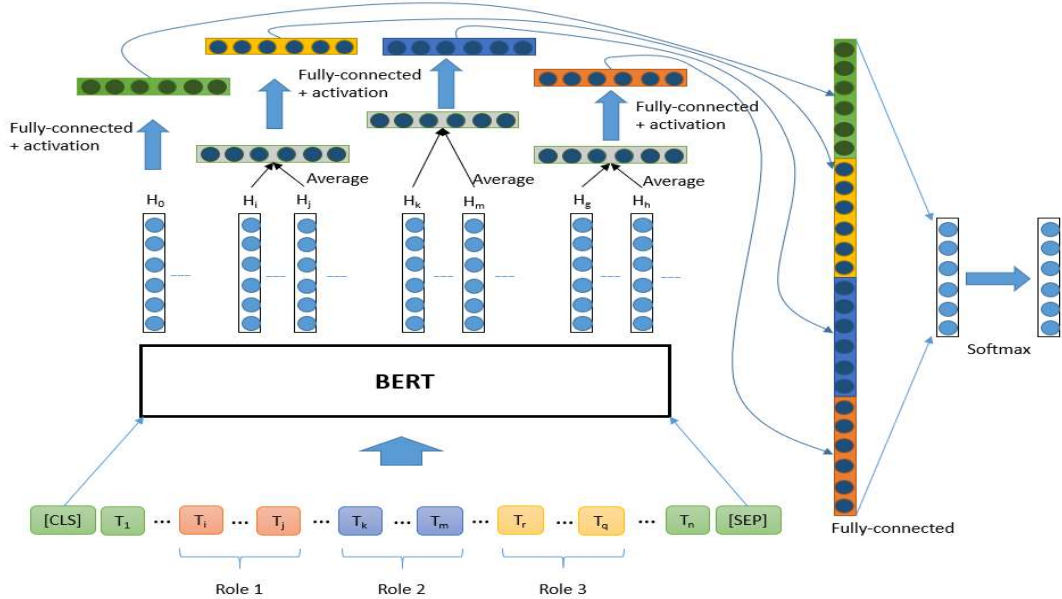


Figure 5: Spatial relation extraction model using R-BERT

## 4 Experiments

### 4.1 Settings and dataset

We use the pre-trained cased BERT<sub>BASE</sub> version model for fine tuning available at the GitHub site<sup>2</sup>. For both the spatial element model and the relation role extraction model, the hyper-parameters in Table 2 are used.

An experiment was conducted with a dataset of SemEval-2015 task 8: SpaceEval. Table 3 shows the statistics of the dataset.

Because non-motion events and MeLink are usually not evaluated in spatial information extraction tasks, they were excluded from our experiment. We also added invalid triplets generated by the triple candidate generator for training. These triplets accounted for approximately 40% of all data and were used as negative data in the training data.

Batch size	1
Learning rate	2e-5
Hidden size	768
Dropout rate	0.1

Table 2: Hyper-parameters for BERT element extraction

place	path	spatial entity	measure
2,175	561	1,746	201
motion	motion signal	spatial signal	non-motion
940	674	939	418

QSLink	OLink	MvLink	MeLink
1,005	251	783	74

Table 3: Number of spatial elements and relations in the SpaceEval dataset

### 4.2 Results

Because the SpRL-CWW model (Nichols and Botros, 2015) was best in SpaceEval, it was used as the baseline model in this evaluation. In our model, spatial elements and spatial roles are jointly trained and extracted. Because the spatial roles depend on the link relation type, we have four types of joint models here: QSLink, OLink, MvLink, and MeLink. The evaluation results for these models are shown in Table 4. Overall, the performance for element extraction was better than that for role extraction. Moreover, the Joint-with-QSLink model showed the worst performance, whereas the Joint-with-MeLink model showed the best performance.

<sup>2</sup> bert-base-cased model : [github.com/google-research/bert](https://github.com/google-research/bert)

Joint models	elements			roles		
	prec	rec	F1	prec	rec	F1
Joint w/ QSLink	96.1	84.5	90.0	93.1	78.6	85.2
Joint w/ OLink	94.8	86.8	90.6	86.7	84.0	85.3
Joint w/ MvLink	93.2	88.2	90.6	96.8	91.0	93.8
Joint w/ MeLink	95.2	88.4	91.7	97.3	97.3	97.3
average	94.8	87.0	90.7	93.5	87.7	90.4

Table 4: Precision, recall and F1 scores of both spatial element and role extraction

Table 5 shows a performance comparison for all elements of the baseline model and our model. The performance was compared with our Joint-with-QSLink model, which showed the worst performance. The results show that the performance of every spatial element increased significantly in comparison with the baseline model. Additionally, the micro-average of our model was 15.4% point better than that of the baseline model. This demonstrates that the BERT-based deep-learning model for element extraction can more effectively improve the performance compared to traditional machine-learning models, such as CRF models, which use features extracted through various preprocessors.

Spatial Elements	baseline	ours	$\Delta$ (ours - base)
Place	74.7	86.8	+12.1
Path	61.7	94.9	+33.2
Spatial entity	80.8	89.9	+9.1
Motion	76.9	94.3	+17.4
Motion signal	78.6	90.7	+12.1
Spatial signal	70.9	85.9	+15.0
Measure	79.1	98.3	+19.2
Non-motion	56.4	89.4	+33.0
Micro-average	74.6	90.0	+15.4

Table 5: Performance comparison of spatial element extraction (F1 score)

Table 6 shows a comparison of the spatial relation extraction performance outcomes of selected models. We measured their average performances for only three link types (QSLink, OLink, and MvLink) because a researcher investigating other previous models chose them for evaluation. (Even though the relation type was limited to static

relations in SemEval-2012 and SemEval-2013, we included the two best research results, those of Roberts and Haragagiu (2012) and Mazalov et al. (2015), for comparison.) Our model outperformed all of the other models compared here. This proves that our relation classification model based on R-BERT is very effective for spatial relation extraction.

Models and techniques	prec.	recall	F1	$\Delta$ (F1)
Baseline (Nichols and Botros, 2015) CRF/SVM	56.0	51.0	53.0	0.0
Salaberri et al. (2015) WordNet	54.0	51.0	53.0	0.0
Roberts and Haragagiu (2012) † SVM	60.3	53.4	56.6	+3.6
Mazalov et al. (2015) ‡ CNN	60.5	56.2	58.3	+5.3
Proposed BERT	62.7	59.8	61.2	+8.2

Table 6: Spatial relation extraction performance for comparison

†: SemEval 2012 dataset, relation with general type

‡: SemEval-2013 dataset, average of two datasets

### 4.3 Ablation study

In order to observe the effects of the proposed features, in this case the *traLand* tag, and the joint model of the spatial elements and roles, we conducted an ablation test. Table 7 shows the result. The simple spatial role extraction model (w/o a joint model in the table) performs very poorly at 25.2% F1. Our analysis shows that this occurs because the data for some roles are very sparse in the SemEval dataset, degrading the performance. This problem is mitigated in the joint model with the use of the element data.

The model without *traLand* tag is tested in two ways: replacing the tag with a *trajactor* tag and doing so with a *landmark* tag. The two models are degraded by 22.5% point F1 and by 14.8% point F1, respectively.

These results show that the two proposed features had a positive impact on the performance.

In particular, the joint model feature greatly improved the performance.

models	prec	recall	F1	$\Delta$ (F1)
proposed	62.7	59.8	61.2	0
w/o joint	25.7	24.6	25.2	-36.0
w/o <i>traLand</i> (repl w/ <i>trajector</i> )	26.5	71.6	38.7	-22.5
w/o <i>traLand</i> (repl w/ <i>landmark</i> )	43.5	49.7	46.4	-14.8

Table 7. Ablation test of models without using the joint training feature and a dual-role tag (*traLand*)

## 5 Conclusion

Spatial information extraction is necessary for many applications, such as robot navigation and question-answering systems, to understand geographical information in text. This task is processed largely with two subtasks: spatial element extraction and spatial relation extraction.

In this paper, we proposed a BERT-based spatial information extraction model that uses BERT (Devlin et al., 2018) for spatial element extraction and R-BERT (Wu and He, 2019) for spatial relation extraction. The two modules are connected with a pipeline through a triple candidate generator.

Spatial elements are extracted jointly with spatial roles that are input for spatial relation extraction. The joint model contributes to increase the performance of spatial role extraction in some cases, which is more useful for relation extraction. R-BERT, which was originally used for semantic role labeling, was modified here to handle three arguments and a null argument for spatial relation extraction.

Our model was evaluated with the SemEval 2015 dataset. The result showed a 15.4% point improvement in spatial element extraction and an 8.2% point improvement in spatial relation extraction in comparison to the baseline model (Nichols and Botros, 2015). This proves that our BERT-based model is very effective for spatial information extraction.

## Acknowledgments

This research was supported by the AURI (Korea Association of University, Research institute and Industry) grant funded by the Korea Government (MSS: Ministry of SMEs and Startups). (No. S2929950, HRD program for 2020)

## References

- Soham Dan, Hangfeng He, and Dan Roth. 2020. [Understanding Spatial Relations through Multiple Modalities](#). *Proceedings of The 12th Language Resources and Evaluation Conference (LREC2020)*. Pages 2368-2372.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. [Bert: Pre-training of deep bidirectional transformers for language understanding](#). *arXiv preprint arXiv: 1810.04805*.
- Jennifer D’Souza and Vincent Ng. 2015. [UTD: Ensemble-based spatial relation extraction](#). *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*.
- ISO (2014), 24617-7:2014(E). [Language resource management – Semantic annotation framework – Part 7: Spatial information \(ISOSpace\)](#)
- Bogyum Kim and Jae Sung Lee. 2016. [Extracting Spatial Entities and Relations in Korean Text](#). *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. pages 2389-2396.
- Parisa Kordjamshidi, Martijn Van Otterlo, and Marie-Francine Moens. 2011. [Spatial role labeling: Towards extraction of spatial relations from natural language](#). *ACM Transactions on Speech and Language Processing (TSLP)* 8(3): 1-36.
- Parisa Kordjamshidi, Taher Rahgooy, Marie-Francine Moens, James Pustejovsky, Umar Manzoor, and Kirk Roberts. 2017. [CLEF 2017: Multimodal spatial role labeling task working notes](#). *CLEF (Working Notes)*.
- John Lafferty, Andrew McCallum, and Fernando C.N. Pereira. 2001. [Conditional random fields: Probabilistic models for segmenting and labeling sequence data](#). *Proceeding of the 18th International Conference on Machine Learning 2001 (ICML 2001)*, pages 282-289.
- Alexey Mazalov, Bruno Martins, and David Matos. 2015. [Spatial Role Labeling with Convolutional Neural Networks](#). *Proceedings of the 9th Workshop on Geographic Information Retrieval*.
- Eric Nichols and Fadi Botros. 2015. [SpRL-CWW: Spatial relation classification with independent multi-class models](#). *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*.
- Jeffery Pennington, Richard Socher, and Christopher D. Manning. 2014. [GloVe: Global vectors for word representation](#). *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. pages 1532-1543. <https://portal.acm.org/citation.cfm?id=655813>.

- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep contextualized word representation](#). *arXiv preprint arXiv: 1802.05365*.
- James Pustejovsky, Parisa Kordjamshidi, Marie-Francine Moens, Aaron Levine, Seth Dworman, and Zachary Yocum. 2015. [Semeval-2015 task 8: Spaceeval](#). *Proceedings of the 9th International Workshop on Semantic Evaluation (semeval 2015)*. pages 884-894.
- Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. [Improving Language Understanding by Generative Pre-Training](#). URL [https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf)
- Kirk Roberts and Sanda M. Haragagiu. 2012. [UTD-SpRL: A joint approach to spatial role labeling](#). *\*SEM 2012: The First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*. pages 419-424.
- Haritz Salaberri, Olatz Arregi, and Beñat Zapirain. 2015. [IXAGroupEHUSpaceEval: \(X-Space\) A WordNet-based approach towards the automatic recognition of spatial information following the ISO-Space annotation scheme](#). *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*.
- J.A.K. Suykens and J. Vandewalle. 1999. [Least squares support vector machine classifiers](#). *Neural processing letters* 9(3):293-300.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). *Advances in neural information processing systems*. pages 5998-6008.
- Shanchan Wu and Yifan He. 2019. [Enriching pre-trained language model with entity information for relation classification](#). *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. pages 2361-2364.
- Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. [Google’s neural machine translation system: Bridging the gap between human and machine translation](#). *arXiv preprint arXiv: 1609.08144*.
- Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. [Xlnet: Generalized autoregressive pretraining for language understanding](#). *Advances in neural information processing systems*. pages 5753-5763.
- Éloi Zablocki, Patrick Bordes, Laure Soulier, Benjamin Piwowarski, and Patrick Gallinari. 2017. [LIP6@ CLEF2017: Multi-modal spatial role labeling using word embeddings](#). *CLEF (Working Notes)*.