

Semantic Context Path Labeling for Semantic Exploration of User Reviews

Salah Ait-Mokhtar and Caroline Brun and Yves Hoppenot and Ágnes Sándor

Naver Labs Europe
6, Chemin de Maupertuis
38240 Meylan, France

firstname.lastname@naverlabs.com

Abstract

In this paper we present a prototype demonstrator showcasing a novel method to perform semantic exploration of user reviews. The system enables effective navigation in a rich contextual semantic schema with a large number of hierarchically structured classes indicating relevant information. In order to identify instances of the structured classes in the reviews, we defined a new Information Extraction task called Semantic Context Path (SCP) labeling, which simultaneously assigns types and semantic roles to entity mentions. Reviews can thus rapidly be explored based on the fine-grained and structured semantic classes. As a proof-of-concept, we have implemented this system for reviews on Points-of-Interest, in English and Korean.

1 Introduction

In this paper, we demonstrate a system that proposes a novel approach for extracting rich and diverse information from informative texts, and allowing effective semantic navigation among the extracted categories. Our system is specifically designed for exploring large quantities of user reviews, which contain a wealth of useful information for potential new users, difficult to exploit in existing review platforms.

The users of our system can navigate in a rich semantic schema representing a large number of hierarchically structured information categories relevant in the reviews, where top classes have multiple layers of subclasses and attribute classes. The semantic navigation can start from the top classes, or directly from the attribute classes.

Mentions of the structured classes in the reviews are automatically assigned labels that we call Semantic Context Paths (SCPs), which denote their types and semantic roles. By semantic roles, we mean classes for which the mentions instantiate attributes. For example, in the

sentence "*See a good movie for \$5!*", the mention \$5 is assigned the SCP label *ShowAndExhibition.Payment.PriceValue*, which means that it is the price of a movie show.

We implemented a supervised SCP tagging system for POI reviews: (1) we created an SCP semantic schema for the domain; (2) we labeled datasets both for English and Korean using an annotation tool that we developed for this task; and (3) we developed and evaluated models for this IE task.

We demonstrate the novel exploration method based on structured semantic classes through a dedicated exploration interface. It is not targeted at end-users at this stage: our purpose is to showcase the new approach in an intuitive way.

In the following sections, we first expose the advantages of exploring user reviews through semantic context paths compared to the methods used in existing social media review sites (Section 2), we then describe the SCP labeling task (Section 3) and finally present our SCP tagging system on POI reviews. We show that performing sequence labeling relying on a state-of-the-art contextual language model extracts with high accuracy great numbers of SCPs from user reviews in multiple languages (Section 4).

2 Semantic Exploration of User Reviews

Popular POI review sites like Google maps or Tripadvisor provide their users with multiple methods for obtaining information on POIs. In this section, we briefly describe the navigation methods offered by existing major user review sites on POIs, and compare them with navigation based on SCPs.

The most widespread way to explore reviews is based on pre-selected frequent terms, as shown in Figure 1.

The user can click on the terms, and thus rapidly access the reviews containing them. However, the main limitation is that the frequent terms are not grouped semantically and thus similar information

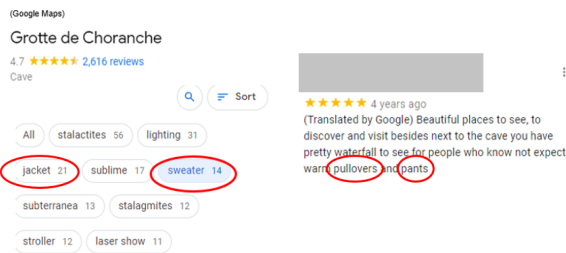


Figure 1: Pre-selected terms in Google Maps

is handled separately (c.f. *jacket* and *sweater*). In addition, important information can be conveyed by clausal expressions like *dress accordingly*, and therefore be missed by term detection algorithms. Finally, in many cases, when talking about relevant topics, like prices, booking, or the availability of food on-site, linguistic expression is so heterogeneous that a method based on term frequency will not detect them.

Keyword search is another way to explore reviews, as shown in Figure 2.

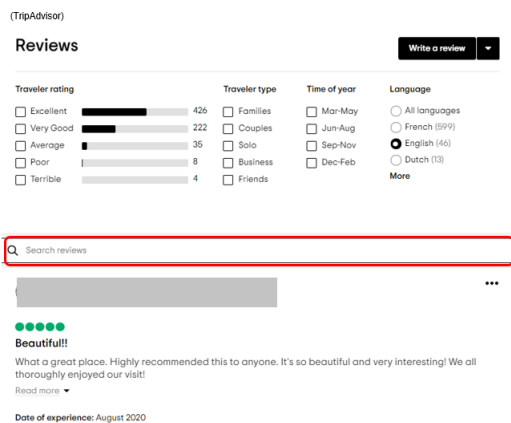


Figure 2: Keyword search in Tripadvisor

Its advantage over the previous method is that the user is not limited by a pre-selected set of terms. However it is slower, since the user needs to enter queries, and getting all the relevant information may require multiple queries. Moreover, some queries will fail either because the information is absent or because the query keywords are not matched, which can be frustrating due to the waste of time.

Finally, a few sites propose pre-selected flat categories for navigation, as illustrated in figure 3.

This method provides fast access to information by clicking on the categories, and improves pre-selected term navigation. However, since the categories are not structured, they can be ambiguous.



Figure 3: Flat categories in Naver Places

E.g. “Price” merges all prices: admission price, price of on-site food, or drink, price of a specific onsite activity, etc.

Taken together, keyword based navigation lacks the generality offered by structured search, however currently proposed structured search does not disambiguate general categories. Compared to these methods, the advantage of navigation based on semantic context paths is that it provides rapid and effective access to precise information on a great diversity of topics.

3 Information Extraction with Semantic Path Labeling

3.1 Presentation of the Task

We introduce Semantic Context Paths (SCPs) as representations of entity mentions enriched with contextual semantic labels. SCP labels encode three types of information: (1) context-free semantic types, as traditionally extracted by Named Entity Recognition (NER) systems, (2) hierarchically represented context-free semantic types, traditionally extracted by fine-grained NER systems, and (3) context-dependent semantic roles specifying the classes for which the mentions instantiate attributes.

As an example, assume we want to extract various kinds of useful information from Point-of-Interest reviews about recreation places (e.g. movie theaters), such as food, prices, visitors, etc.

Figure 4 illustrates how SCP labeling is applied on a real POI comment.

The proposed IE task goes beyond traditional Named Entity Recognition (NER), which merely assigns semantic categories to entity mentions (Akbiik et al., 2018; Pawar et al., 2017), even when the set of categories is fine-grained and hierarchical (Mai et al., 2018; Zhang et al., 2020). In contrast, SCP labels also contain information on the semantic content of the entities or concepts within the

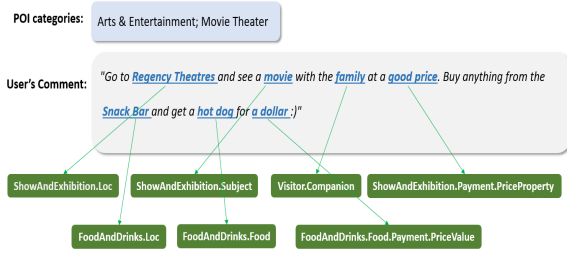


Figure 4: An example of SCP tagging.

mentions, but at the same time also on their semantic roles and types of relations within the text.

In the example in Figure 4, *good price* does not have a simple *PriceProperty* label as it could have in a classical NER system, but *ShowAndExhibition.PriceProperty*, which indicates that it is related to mentions with the same root label, *ShowAndExhibition* – in this example *Regency Theatres* and *movie*, which are respectively tagged as *ShowAndExhibition.Loc*, and *ShowAndExhibition.Subject*.

SCP tagging therefore presents the advantage of extracting rich contextual information, while still being a sequence labeling task.

3.2 Formal Definition of the Task

A **semantic schema** S is a tuple $\langle C, I, R \rangle$ where C is a finite set of semantic classes, $I \subseteq C$ is the subset of instantiable **classes**, and $R \subset C \times C$ is the set of **attribute relations**. R defines which classes have attributes and their types: $(c_i, c_j) \in R$ means that class c_i has an attribute of type c_j . A class $c \in C$ is a **top class** iff $\nexists c' \in C$ such that $(c', c) \in R$ (i.e. c is not an attribute of another class), and it is **terminal** iff $\nexists c' \in C$ such that $(c, c') \in R$ (i.e. it does not have attributes).

A **Semantic Context Path** (SCP) in S is any non-empty and finite sequence of classes $p = c_1.c_2.c_3\dots c_n$ where c_1 is a top class, and $\forall c_i, c_{i+1} \in p : (c_i, c_{i+1}) \in R$. An SCP $p = c_1.c_2.c_3\dots c_n$ is **acyclic** iff $\forall i, j \in \{1, \dots, n\} : i \neq j \implies c_i \neq c_j$. An SCP $p = c_1.c_2.c_3\dots c_n$ is **instantiable** iff $c_n \in I$. We use P^S to denote the set of all acyclic, instantiable paths of S .

We can now define the task of **Semantic Context Path tagging** as follows: given a semantic schema S and a tokenized text document $d = t_1t_2\dots t_n$, SCP tagging assigns a (possibly empty) subset of P^S to every token in d .

4 System Description

The semantic navigation system has two main components: the IE component, which identifies mentions and labels them with SCPs off-line, and a navigation component, which allows exploring the reviews based on the structured classes. We describe these two components in the following subsections.

4.1 Information Extraction Component

4.1.1 Model Description

We model the SCP labeling task as a multi-label classification problem at the level of tokens: review texts are tokenized, and each token can have one or multiple SCP labels if it is part of a relevant mention, or has the special label O otherwise. The set of possible labels is therefore the set P^S of all acyclic, instantiable paths of a semantic schema S , plus the special label O . We implement the model as a neural network that jointly learns to predict the final path labels and the individual classes that make up the paths. Indeed, since there are many paths that share common individual classes (i.e. they have identical prefixes or suffixes), learning to predict individual classes inside the paths can alleviate the problem of data sparseness with regard to (whole) path labels. Figure 5 below depicts the model architecture.

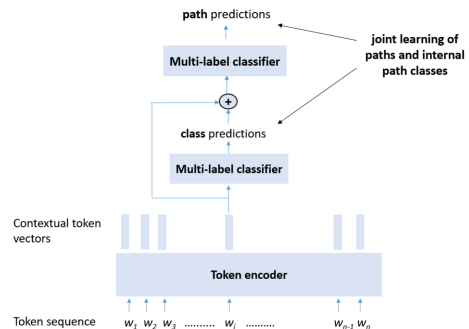


Figure 5: Model architecture

First, the model assigns contextualized word embeddings to each token w_i in the input text using a pre-trained, transformer-based language model, more exactly a RoBERTa model (Liu et al., 2019):

$$x_i = \text{RoBERTa}(w_i; w_1w_2\dots w_n) \quad (1)$$

The contextualized token representations are then fed to a fully connected hidden layer:

$$h_i = \text{relu}(W^h x_i + b^h) \quad (2)$$

From these hidden token representations h_i , we first compute the probabilities of all individual (flat) classes of the semantic schema:

$$p(y^{(c)}|h_i, \theta) = \text{sigmoid}(W^c x_i + b^c) \quad (3)$$

We then concatenate the class probability vector with the hidden token representation h_i (in eq. 2) and feed the result to a linear layer with a sigmoid activation in order to compute the probabilities of the path (SCP) labels:

$$p(y|h_i, \theta) = \text{sigmoid}(W(h_i \oplus p(y^{(c)}|h_i, \theta)) + b) \quad (4)$$

We train the model using the sum of the binary cross-entropy losses of the individual class predictions and whole path predictions.

4.1.2 Experiments

Dataset. The dataset is a collection of POI users’ comments¹ in both English and Korean, randomly sampled from three main categories: Food (food places), Art & Entertainment, and Outdoor & Recreation. Every comment has a main category among these, and a set of one or more sub-categories (e.g. Park for Outdoor & Recreation, or Movie Theater for Art & Entertainment.) Since there was no rich, off-the-shelf hierarchical semantic schema to represent useful information types for POIs, we defined one from scratch, based on a development subset of the data. The final fine-grained schema comprises 42 semantic classes, their combination leading to 185 virtually instantiable semantic paths. For example, *Visit.Accessibility.Mobility.Device* is an SCP associated with the mention *wheelchair* in *"Most parts are fine for wheelchairs."*; *ShowAndExhibition.Subject.Name* is a SCP associated with the mention *Last Judgment* in *"The pride of the Gdansk museum is a triptych of the Last Judgement by Hans Memling (1433-1494)"*, etc. Annotation guidelines have been defined together with the semantic schema.

We (the authors) annotated manually the English version of the dataset according to the semantic schema and the guidelines. Following the same

¹The data was acquired legally through an agreement with a company providing POI-based services. The dataset contains a set of user comments with corresponding POI identifiers, geolocation and categories. It does not contain any user information. No crowd-workers have been involved in the annotation process.

Dataset	English		Korean	
	Train	Test	Train	Test
# docs	3000	1500	2498	498
# tokens	49484	24641	55242	11126
# mentions	5861	3026	5111	1023
# labels	5914	3068	5302	1051

Table 1: Dataset Statistics: number of documents, tokens, annotated mentions and labels. There are approximately 2 annotated mentions per document on average.

schema and annotation guidelines, the Korean version of the dataset was annotated by a native Korean speaker specialized in data annotation (hired under a Naver employment contract). The distribution of SCP labels is unbalanced for both languages, with a long tail of infrequent labels: 50% of the observed path labels in the training set for Korean had a maximum of 10 occurrences. Table 1 reports some statistics regarding both versions of the annotated dataset.

In order to assess the reliability of the annotations, we calculated Krippendorff’s alpha (Hayes and Krippendorff, 2007) on the English annotated dataset. On a token-basis, we obtained a relatively high agreement level, $\alpha = 0.838$, which indicates a good dataset reliability. This dataset is, to our knowledge, one of the finest-grained annotated datasets for information extraction.

Hyperparameters. We used the *xlm-roberta-large* pretrained, transformer-based model from HuggingFace transformers library (Wolf et al., 2020) to produce the token contextualized embeddings (equation 1). Experiments with multiple mono-lingual pre-trained language models either showed similar performance (*roberta-large* for English) or lower performance (*bert-large* for English, and *kobert*, *kobart*, *koelectra* and *KR-BERT-char16424* for Korean.) We used an initial learning rate of $1e^{-5}$ with a scheduler, a batch size of 16, and a maximum of 100 epochs, with an early stopping strategy.

Evaluation. We evaluate the performance of the models as fully-automatic information extraction systems. The input contains raw text (user comments), and the system needs to identify mentions (and their boundaries) and tag them with SCP labels. We measured performance using the traditional precision, recall, and f1 scores in their re-

laxed variants. That is, the boundaries of predicted mentions do not need to match fully the ground-truth mention spans: they are considered correct if they have at least one common token with ground-truth spans, and as long as the predicted SCP label is entirely correct. This relaxed span evaluation is motivated by the target application: in the UI for exploring POI reviews (see section 4.2), when the user navigates in the hierarchical semantic classes and selects one class, mentions of that class are automatically displayed and highlighted within their context (review snippets). Thus, the user can read the immediate context, including any token missing from the identified mention. Performance results are shown in table 2.

Model	English		Korean	
	f1-avg (std)	f1-best	f1-avg (std)	f1-best
En	76.87 (0.35)	77.40	–	64.43
Ko	–	69.05	74.58 (0.45)	75.39
En+Ko	–	78.16	76.14 (0.85)	77.66

Table 2: Performance of the information extraction models on English and Korean test sets. Models *En* and *Ko* were trained on the English and Korean training datasets respectively, while model *En+Ko* was trained on the union of the two training datasets, and yields the best performance for both languages. Average f1 score from 10 runs is reported with standard deviation. f1-best refers to the score of the best run. In cross- or multi-lingual settings, only the model from the best run for one of the two languages was tested on the other, hence the missing f1-avg figures.

An interesting side result is that since the IE component is built on top of a multilingual pre-trained language model (*xlm-roberta-large*), it is applicable to other languages in a zero-shot setting. Although we have not performed a quantitative evaluation on other languages than English and Korean, due to the lack of labeled data, a preliminary qualitative evaluation shows promising results. An example of SCP labeling of Arabic is provided in Figure 6.

4.2 Exploration of User Reviews

We have designed an interface on top of Naver map² to demonstrate the application of SCP tagging for exploring rich and detailed information conveyed in user reviews. Currently, in the demo, we cover two languages, English and Korean, but since the system is operational for a great number of other languages, these could be added as well.

²<https://m.map.naver.com/>

كانت تجربة جميله ولم يسعفتني الوقت للاطلاع على كامل المتحف الزيارة كانت مجانية بسبب انها اخر ساعة يوجد بالمتحف لوحات جميلة جدا ولكن التصوير ممنوع	
المتحف	ShowAndExhibition.Loc
مجانى	ShowAndExhibition.Payment.PriceValue
اخر ساعة	Visit.Time.TimeValue
تحف	ShowAndExhibition.Loc
لوحات	ShowAndExhibition.Subject
التصوير	Visit.Regulation.RegulationTopic
التصوير	RecreationAndSport
ممنوع	Visit.Regulation.RegulationValue

Figure 6: An example of SCP labeling in Arabic.

When a user opens the map, she first selects the language, and the POI types (one or several) that she would like to explore. The current system covers 3 POI types: Arts & Entertainment, Outdoors & Recreation, and Food.

Hovering over a POI in the map makes its name appear along with a circular chart displaying the different topics (i.e. the top classes of the semantic schema) that are covered by the reviews, as well as the number of available reviews. The width of the particular categories is determined by the number of subcategories covered by the reviews (c.f. Figure 7).

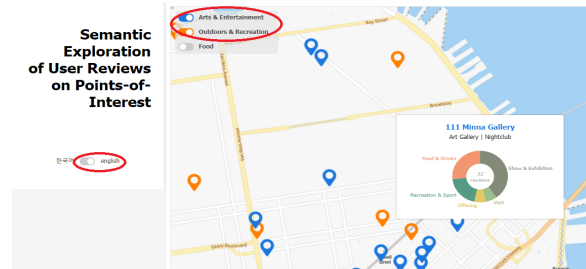


Figure 7: Visualizing on the map the different topics covered by the reviews of a given POI.

In order to explore the reviews in more detail, the user can click on the circular chart, which will expand into a sunburst displaying all the subcategories that have mentions in the reviews. As we indicated in the introduction, the demo system is not targeted at end users. We have chosen the sunburst interface since it allows a comprehensive and straightforward visualization of a great number of SCPs: the innermost circle contains the main categories, and each subsequent layer displays the subcategories in the same segment. Thus, at a glimpse, the chart offers an overview of all the categories and SCPs covered by a review. Moving the mouse over any category opens a callout with the snippets

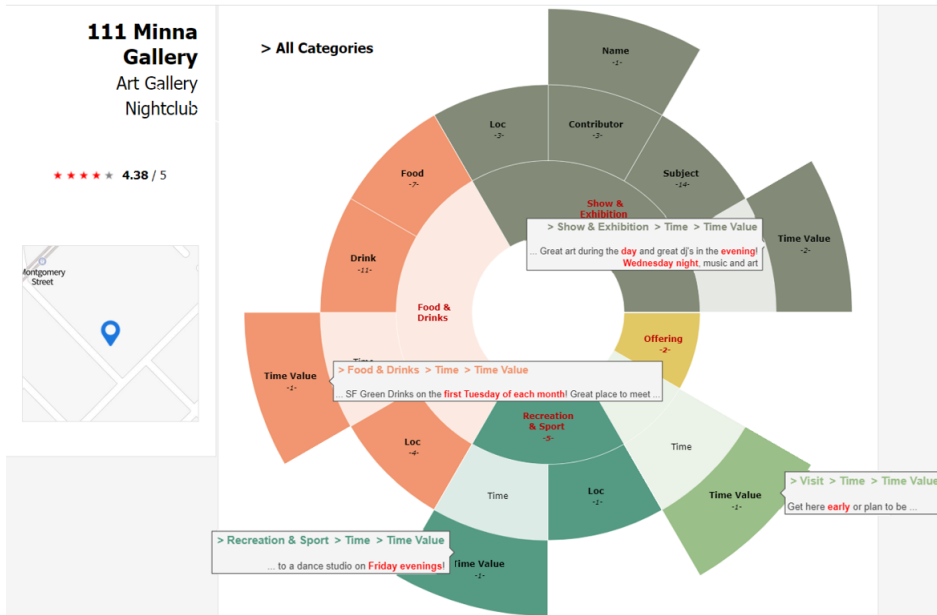


Figure 8: Visualizing snippets containing class mentions. In this example, besides a tip on the best time for a general visit, we find information on time when you can dance (Recreation&Sport.Time), when you can get specific drinks (Food&Drinks.Time), as well as on when you can listen to music (Show&Recreation.Time). In the interface, snippets of one category can be visualized at a time by hovering over it; here we present the snippets of several categories on the same figure for a concise presentation.

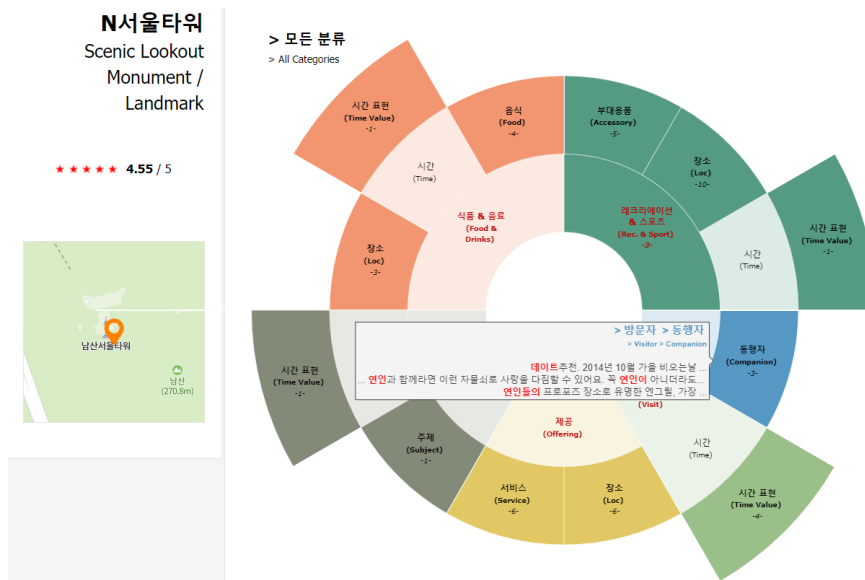


Figure 9: Categories extracted from Korean reviews on the Seoul Tower.

of the relevant user reviews (Figure 8).

If review snippets are not sufficient to understand the information conveyed in the review, the user can click on the callouts to access the full reviews (Figure 10).

Figure 9 shows the categories extracted from the Korean reviews on the Seoul Tower, and the snippets about recommended visitor companions.

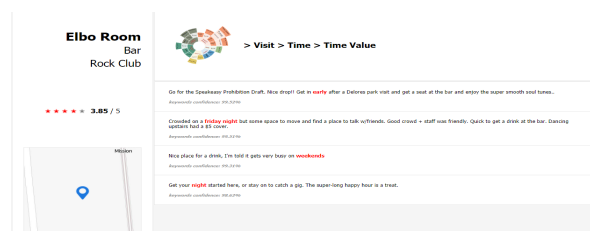


Figure 10: Full review texts for mentions of show time.

5 Conclusion

We have presented a new method for the semantic exploration of user reviews.

The system underlying the demonstrator relies on an IE component which identifies relevant mentions in the reviews and labels them with Semantic Context Paths, denoting their types and semantic roles. We have implemented an SCP tagger that extracts information from user reviews on Points-of-Interest. We have defined a dedicated semantic schema, created datasets, and developed sequence labeling models for the task. The IE component was quantitatively evaluated on English and Korean, and showed promising qualitative results on other languages.

We have designed a review exploration interface exploiting the output of the SCP tagger. A sunburst chart in the interface allows navigation among the classes, and rapid access to relevant information in the reviews. Future work includes integrating a wider range of languages to the review exploration interface.

References

- David Ahn. 2006. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pages 1–8.
- Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. [Contextual string embeddings for sequence labeling](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1638–1649, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019a. [Matching the blanks: Distributional similarity for relation learning](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2895–2905, Florence, Italy. Association for Computational Linguistics.
- Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019b. [Matching the Blanks: Distributional Similarity for Relation Learning](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2895–2905, Florence, Italy. Association for Computational Linguistics.
- Robert Bossy, Louise Deléger, Estelle Chaix, Mouhamadou Ba, and Claire Nédellec. 2019. [Bacteria Biotope at BioNLP Open Shared Tasks 2019](#). In *Proceedings of The 5th Workshop on BioNLP Open Shared Tasks*, pages 121–131, Hong Kong, China. Association for Computational Linguistics.
- H. Cunningham, K. Humphreys, R. Gaizauskas, and Y. Wilks. 1997. Software infrastructure for natural language processing. In *Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP-97)*.
- 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*.
- Michael Glass and Alfio Gliozzo. 2018. [Discovering implicit knowledge with unary relations](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1585–1594, Melbourne, Australia. Association for Computational Linguistics.
- Andrew F Hayes and Klaus Krippendorff. 2007. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1:77–89.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. [RoBERTa: A Robustly Optimized BERT Pretraining Approach](#). *arXiv:1907.11692 [cs]*. ArXiv: 1907.11692.
- Khai Mai, Thai-Hoang Pham, Minh Trung Nguyen, Tuan Duc Nguyen, Danushka Bollegala, Ryohei Sasano, and Satoshi Sekine. 2018. [An empirical study on fine-grained named entity recognition](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 711–722, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Arvind Neelakantan and Ming-Wei Chang. 2015. Inferring missing entity type instances for knowledge base completion: New dataset and methods. *arXiv preprint arXiv:1504.06658*.
- Sachin Pawar, Girish K. Palshikar, and Pushpak Bhat-tacharyya. 2017. [Relation extraction : A survey](#).
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. 2012. [Brat: A web-based tool for nlp-assisted text annotation](#). In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL ’12*, pages 102–107, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pi-eric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame,

- Quentin Lhoest, and Alexander Rush. 2020. [Transformers: State-of-the-Art Natural Language Processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.
- Shanchan Wu and Yifan He. 2019. [Enriching pre-trained language model with entity information for relation classification](#). *CoRR*, abs/1905.08284.
- Wuti Xiong, Fei Li, Ming Cheng, Hong Yu, and Donghong Ji. 2019. [Bacteria Biotope Relation Extraction via Lexical Chains and Dependency Graphs](#). In *Proceedings of The 5th Workshop on BioNLP Open Shared Tasks*, pages 158–167, Hong Kong, China. Association for Computational Linguistics.
- Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. 2020. [LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6442–6454, Online. Association for Computational Linguistics.
- Qi Zhang, Chao Liu, Ying Chi, Xuansong Xie, and Xi-anheng Hua. 2019. [A Multi-Task Learning Framework for Extracting Bacteria Biotope Information](#). In *Proceedings of The 5th Workshop on BioNLP Open Shared Tasks*, pages 105–109, Hong Kong, China. Association for Computational Linguistics.
- Tao Zhang, Congying Xia, Chun-Ta Lu, and Philip Yu. 2020. [MZET: Memory augmented zero-shot fine-grained named entity typing](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 77–87, Barcelona, Spain (Online). International Committee on Computational Linguistics.