# LMN at SemEval-2022 Task 11: A Transformer-based System for English Named Entity Recognition

**Ngoc Minh Lai**
Van Ho Middle School
nl834771@gmail.com

## Abstract

Processing complex and ambiguous named entities is a challenging research problem, but it has not received sufficient attention from the natural language processing community. In this short paper, we present our participation in the English track of SemEval-2022 Task 11: Multilingual Complex Named Entity Recognition. Inspired by the recent advances in pretrained Transformer language models, we propose a simple yet effective Transformer-based baseline for the task. Despite its simplicity, our proposed approach shows competitive results in the leaderboard as we ranked 12 over 30 teams. Our system achieved a macro F1 score of 72.50% on the held-out test set. We have also explored a data augmentation approach using entity linking. While the approach does not improve the final performance, we also discuss it in this paper.

## 1 Introduction

Recognizing complex named entities (NEs) is a challenging research problem, but it has not received sufficient attention from the natural language processing community (Meng et al., 2021a; Fetahu et al., 2021). Complex NEs can be complex noun phrases (e.g., *National Baseball Hall of Fame and Museum*), gerunds (e.g., *Saving Private Ryan*), infinitives (e.g., *To Build a Fire*), or even full clauses (e.g., *I Capture The Castle*). This ambiguity makes it difficult to recognize them based on their context (Aguilar et al., 2017; Luken et al., 2018; Hanselowski et al., 2018).

In this paper, we describe our participation in the English track of SemEval-2022 Task 11: Multilingual Complex Named Entity Recognition (Malmasi et al., 2022a,b). Inspired by the recent success of Transformer-based pre-trained language models in many NLP tasks (Devlin et al., 2019; Joshi et al., 2019; Lai et al., 2019; Joshi et al., 2020; Tran et al., 2020; Yu et al., 2020; Wen et al., 2021; Lai et al., 2021; Monaikul et al., 2021), we propose a simple but effective Transformer-based baseline for the task. Despite its simplicity, our proposed approach shows promising results: the official ranking indicated that our system achieved a macro $F_1$ score of 72.50% on the test set and ranked 12th out of 30 teams. We have also explored a data augmentation approach using entity linking. While the approach does not improve the final performance, we also discuss it in this paper.

In the following sections, we first describe the related work in Section 2 and the proposed method in Section 3. We then describe the experiments and their results in Section 4. Finally, Section 5 concludes this work and discusses potential future research directions.

## 2 Related Work

Many previous named entity recognition (NER) methods are based on the sequence labeling approach (Collobert et al., 2011; Ma and Hovy, 2016; Lample et al., 2016; Chiu and Nichols, 2016; Lee et al., 2019; Yang et al., 2018; Yang and Zhang, 2018; Lai et al., 2020a; Li et al., 2020). For example, Collobert et al. (2011) introduced a neural architecture that uses convolutional neural networks (CNNs) to encode tokens combined with a CRF layer for the classification. Many other studies used recurrent neural networks (RNNs) instead of CNNs to encode the input and a CRF for the prediction (Ma and Hovy, 2016; Lample et al., 2016). With the recent rise of pre-trained language models, recent NER models typically make use of context-dependent embeddings such as ELMo (Peters et al., 2018) or BERT (Devlin et al., 2019).

While neural-based models have achieved impressive results on popular benchmark datasets like CoNLL03 and OntoNotes (Tjong Kim Sang, 2002; Tjong Kim Sang and De Meulder, 2003; Pradhan et al., 2012), these models typically do not perform well on complex/unseen entities (Augenstein et al., 2017). Complex named entities (e.g., titles



*Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 1438 - 1443
July 14-15, 2022 ©2022 Association for Computational Linguistics

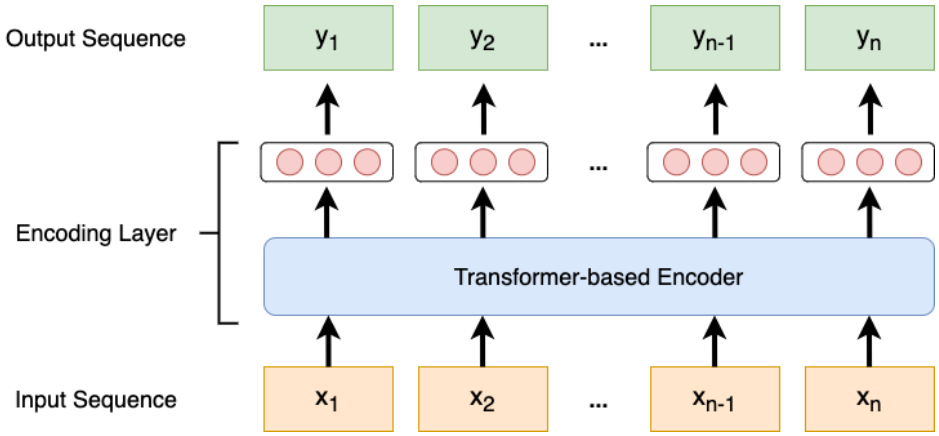


Figure 1: Overview of our Transformer-based model.

| Tag | Description |
|---|---|
| {B,I}-PER | A named entity of a *person* |
| {B,I}-LOC | A named entity of a *location* |
| {B,I}-GRP | A named entity of a *group* |
| {B,I}-CORP | A named entity of a *corporation* |
| {B,I}-PROD | A named entity of a *product* |
| {B,I}-CW | A named entity of a *creative work* |
| O | Not a named entity |

Table 1: The label set.

of creative works) are typically not simple nouns and are harder to recognize. The challenges of NER for recognizing complex entities and in low-context situations was recently outlined by Meng et al. (2021b). Other work has extended this to multilingual and code-mixed settings (Fetahu et al., 2021).

## 3 Method

### 3.1 Baseline model

Similar to many previous studies (Lample et al., 2016; Chiu and Nichols, 2016), we formulate the task as a sequence labeling problem. Given an input sequence consisting of $n$ tokens $(x_1, ..., x_n)$, the goal is to predict a sequence of labels $(y_1, ..., y_n)$, where $y_i$ is the label corresponding to token $x_i$. Table 1 describes the label set. We follow the BIO format: B denotes the beginning of a named entity, I denotes the continuation of a named entity, and O corresponds to tokens that are not part of any named entity.

Figure 1 shows a high-level overview of our Transformer-based model. Our model first forms a contextualized representation for each input token using a Transformer encoder (Devlin et al., 2019). Let $\mathbf{H} = (\mathbf{h}_1, ..., \mathbf{h}_n)$ be the output of the encoder where $\mathbf{h}_i \in \mathbb{R}^d$. After that, we can predict the final label $y_i$ for each input token $x_i$:

$$\begin{aligned} \mathbf{y}_i &= \text{softmax}(\text{FFNN}_\theta(\mathbf{h}_i)) \\ y_i &= \arg\max_j \mathbf{y}_{ij} \end{aligned} \quad (1)$$

where $\text{FFNN}_\theta$ is a trainable feedforward network. $\mathbf{y}_i$ is the predicted probability distribution over the label set for the token $x_i$. The model is fine-tuned end-to-end via minimizing the typical cross-entropy loss.

Unlike many previous studies (Lample et al., 2016; Chiu and Nichols, 2016), our model does not have a CRF layer (Lafferty et al., 2001). A recent paper suggested that when using a pretrained Transformer language model for sequence labeling, adding a CRF layer may not improve the performance substantially (Chen et al., 2019).

### 3.2 Data Augmentation

To increase the size of the training set, we have also experimented with a simple data augmentation approach (Figure 2). For example, consider the sentence "*The main contractor was Ssangyong Engineering and Construction.*", which is an example in the training set of the English track of Multi-CoNER. In this case, "*Ssangyong Engineering and Construction*" is a named mention referring to a Korean corporation. To create a new training example, we can replace the named mention with a different entity that is also a corporation.

More specifically, in this example, we first use an entity linker[1] to link the named mention to its corresponding entity in Wikidata, a large-scale knowledge graph. From the found Wikidata page, we

[1] https://github.com/laituan245/EL-Dockers

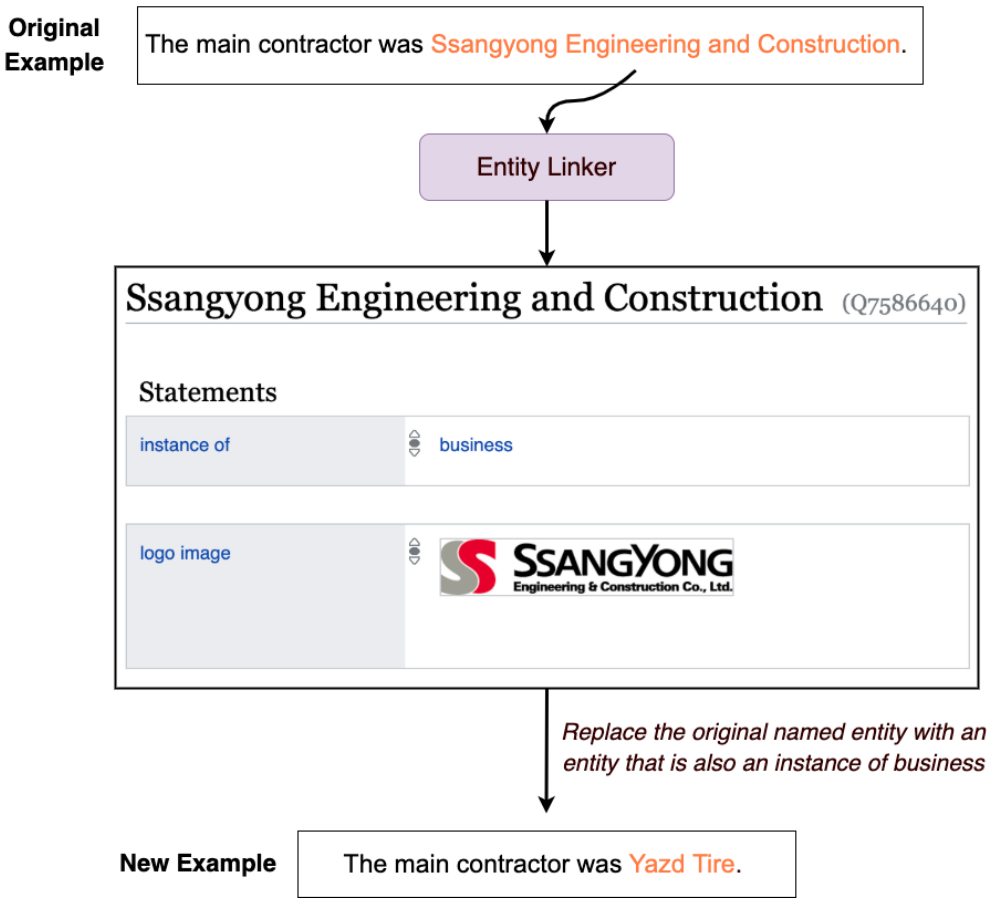


Figure 2: Our data augmentation approach.

can extract all types of information about the entity. We can utilize these types of information to find a new entity that is different but highly similar to the original entity. For simplicity, in this work, we simply try to find a new entity of the same Wikidata type as the original entity. At the end, we will have a new example (e.g., "*The main contractor was Yazd Tire.*").

## 4 Results

### 4.1 Data and Experimental Setup

The learning rate is set to be 2e-5, and the batch size is 32. We experimented with different numbers of training epochs, 10 and 20. We use Huggingface's Transformer library (Wolf et al., 2020) to experiment with various Transformer language models:

- **BERT**. Devlin et al. (2019) introduced a language representation model named BERT, which is pre-trained using two tasks: masked language modeling (MLM) and next sentence prediction (NSP). We used the large version of BERT (i.e., bert-large-uncased) in this work.
- **RoBERTa**. Liu et al. (2019) proposed an improved recipe for training BERT models. The modifications include: (1) training the model longer, with bigger batches, over more data; (2) removing the NSP objective; (3) training on longer sequences; and (4) dynamically changing the masking pattern applied to the training data. We used the large version of RoBERTa (i.e., roberta-large) in this work.
- **ALBERT**. Lan et al. (2020) introduced ALBERT, a BERT-based model with two parameter reduction techniques: factorized embedding parameterization and cross-layer parameter sharing. We used the xxlarge version of ALBERT (i.e., albert-xxlarge-v2) in this work.

### 4.2 Results on the Development Set

Table 3 shows the overall results on the development set of the English track of MultiCoNER. We see that ALBERT-xxlarge trained with 20 epochs outperforms all other baseline models on the development set. As such, we use this model to generate predictions for the test set. The model achieved a

| Original Example | Generated Example |
|---|---|
| the guardian described the album 's release as one of the 50 key events ... | metro described the album 's release as one of the 50 key events ... |
| the game uses a battery packed random-access memory in order to save progress . | the game uses a battery packed delay line memory in order to save progress . |
| in the end the best placed rider was wilfried cretskens who finished 61st . | in the end the best placed rider was harald andersson who finished 61st . |
| it was broadcast on the channel animal planet , with episodes having aired between 2001 and 2003 . | it was broadcast on the channel true4u , with episodes having aired between 2001 and 2003 . |

Table 2: Some of the newly generated examples.

| | Prec. | Recall | F1 |
|---|---|---|---|
| RoBERTa-large (10 epochs) | 85.63 | 87.82 | 86.68 |
| BERT-large (10 epochs) | 86.02 | 88.34 | 87.14 |
| ALBERT-xxlarge (10 epochs) | 86.81 | 88.7 | 87.7 |
| ALBERT-xxlarge (20 epochs) | 86.47 | 89.49 | 87.91 |

Table 3: Overall results on the development set. Macro scores (%) are shown.

macro F1 score of 72.50% on the held-out test set. Note that the baseline models shown in Table 3 are trained using only the original training set (without any data augmentation).

### 4.3 Analysis of the Data Augmentation Approach

For each example in the training set, we used the data augmentation approach (Section 3.2) to generate a new example. Table 2 shows some of the newly generated examples.

We used all of the original and newly generated examples to train a new RoBERTa-large model (the number of epochs is 10). The model performs worst than the RoBERTa-large model trained with only the original examples. Nevertheless, we still believe the approach has a lot of potential, and we leave further exploration to future work.

## 5 Conclusion

In future work, we plan to conduct a thorough error analysis and apply visualization techniques to understand our models better (Murugesan et al., 2019). In addition, as pretrained Transformer models are typically computationally expensive and have many parameters, we are also interested in reducing the computational complexity of our baseline models using compression techniques (Sanh et al., 2019; Lai et al., 2020b; Sun et al., 2020).

## References


Gustavo Aguilar, Suraj Maharjan, Adrian Pastor López-Monroy, and Thamar Solorio. 2017. A multi-task approach for named entity recognition in social media data. In *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pages 148–153, Copenhagen, Denmark. Association for Computational Linguistics.

Isabelle Augenstein, Leon Derczynski, and Kalina Bontcheva. 2017. Generalisation in named entity recognition: A quantitative analysis. *Comput. Speech Lang.*, 44:61–83.

Qian Chen, Zhu Zhuo, and Wen Wang. 2019. Bert for joint intent classification and slot filling. *ArXiv*, abs/1902.10909.

Jason P.C. Chiu and Eric Nichols. 2016. Named entity recognition with bidirectional LSTM-CNNs. *Transactions of the Association for Computational Linguistics*, 4:357–370.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa. 2011. Natural language processing (almost) from scratch. *J. Mach. Learn. Res.*, 12:2493–2537.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Besnik Fetahu, Anjie Fang, Oleg Rokhlenko, and Shervin Malmasi. 2021. Gazetteer Enhanced Named Entity Recognition for Code-Mixed Web Queries. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1677–1681.

Andreas Hanselowski, Hao Zhang, Zile Li, Daniil Sorokin, Benjamin Schiller, Claudia Schulz, and Iryna Gurevych. 2018. UKP-athene: Multi-sentence textual entailment for claim verification. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 103–108, Brussels, Belgium. Association for Computational Linguistics.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. SpanBERT: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8:64–77.

Mandar Joshi, Omer Levy, Luke Zettlemoyer, and Daniel Weld. 2019. BERT for coreference resolution: Baselines and analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5803–5808, Hong Kong, China. Association for Computational Linguistics.

John D. Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *ICML*.

Tuan Lai, Heng Ji, ChengXiang Zhai, and Quan Hung Tran. 2021. Joint biomedical entity and relation extraction with knowledge-enhanced collective inference. *arXiv preprint arXiv:2105.13456*.

Tuan Lai, Quan Hung Tran, Trung Bui, and Daisuke Kihara. 2019. A gated self-attention memory network for answer selection. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5953–5959, Hong Kong, China. Association for Computational Linguistics.

Tuan Manh Lai, Trung Bui, Doo Soon Kim, and Quan Hung Tran. 2020a. A joint learning approach based on self-distillation for keyphrase extraction from scientific documents. *arXiv preprint arXiv:2010.11980*.

Tuan Manh Lai, Quan Hung Tran, Trung Bui, and Daisuke Kihara. 2020b. A simple but effective bert model for dialog state tracking on resource-limited systems. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8034–8038. IEEE.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California. Association for Computational Linguistics.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. Albert: A lite bert for self-supervised learning of language representations. *ArXiv*, abs/1909.11942.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Yangming Li, Han Li, Kaisheng Yao, and Xiaolong Li. 2020. Handling rare entities for neural sequence labeling. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6441–6451, Online. Association for Computational Linguistics.

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*, abs/1907.11692.

Jackson Luken, Nanjiang Jiang, and Marie-Catherine de Marneffe. 2018. QED: A fact verification system for the FEVER shared task. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 156–160, Brussels, Belgium. Association for Computational Linguistics.

Xuezhe Ma and Eduard Hovy. 2016. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1064–1074, Berlin, Germany. Association for Computational Linguistics.

Shervin Malmasi, Anjie Fang, Besnik Fetahu, Sudipta Kar, and Oleg Rokhlenko. 2022a. MultiCoNER: a Large-scale Multilingual dataset for Complex Named Entity Recognition.

Shervin Malmasi, Anjie Fang, Besnik Fetahu, Sudipta Kar, and Oleg Rokhlenko. 2022b. SemEval-2022 Task 11: Multilingual Complex Named Entity Recognition (MultiCoNER). In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Association for Computational Linguistics.

Tao Meng, Anjie Fang, Oleg Rokhlenko, and Shervin Malmasi. 2021a. GEMNET: Effective gated gazetteer representations for recognizing complex entities in low-context input. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1499–1512, Online. Association for Computational Linguistics.

Tao Meng, Anjie Fang, Oleg Rokhlenko, and Shervin Malmasi. 2021b. GEMNET: Effective gated gazetteer representations for recognizing complex entities in low-context input. In *Proceedings of the 2021 Conference of the North American Chapter of*

*the Association for Computational Linguistics: Human Language Technologies*, pages 1499–1512.

Natawut Monaikul, Giuseppe Castellucci, Simone Filice, and Oleg Rokhlenko. 2021. Continual learning for named entity recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(15):13570–13577.

Sugeerth Murugesan, Sana Malik, Fan Du, Eunyee Koh, and Tuan Manh Lai. 2019. Deepcompare: Visual and interactive comparison of deep learning model performance. *IEEE computer graphics and applications*, 39(5):47–59.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *NAACL*.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes. In *Joint Conference on EMNLP and CoNLL - Shared Task*, pages 1–40, Jeju Island, Korea. Association for Computational Linguistics.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *ArXiv*, abs/1910.01108.

Zhiqing Sun, Hongkun Yu, Xiaodan Song, Renjie Liu, Yiming Yang, and Denny Zhou. 2020. MobileBERT: a compact task-agnostic BERT for resource-limited devices. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2158–2170, Online. Association for Computational Linguistics.

Erik F. Tjong Kim Sang. 2002. Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition. In *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.

Quan Hung Tran, Nhan Dam, Tuan Lai, Franck Dernoncourt, Trung Le, Nham Le, and Dinh Phung. 2020. Explain by evidence: An explainable memory-based neural network for question answering. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5205–5210, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Haoyang Wen, Ying Lin, Tuan Lai, Xiaoman Pan, Sha Li, Xudong Lin, Ben Zhou, Manling Li, Haoyu Wang, Hongming Zhang, et al. 2021. Resin: A dockerized schema-guided cross-document cross-lingual cross-media information extraction and event tracking system. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*, pages 133–143.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric 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.

Jie Yang, Shuailong Liang, and Yue Zhang. 2018. Design challenges and misconceptions in neural sequence labeling. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3879–3889, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Jie Yang and Yue Zhang. 2018. NCRF++: An open-source neural sequence labeling toolkit. In *Proceedings of ACL 2018, System Demonstrations*, pages 74–79, Melbourne, Australia. Association for Computational Linguistics.

Juntao Yu, Bernd Bohnet, and Massimo Poesio. 2020. Named entity recognition as dependency parsing. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6470–6476, Online. Association for Computational Linguistics.