Efficient and Robust Knowledge Graph Construction

Ningyu Zhang[1], **Tao Gui**[2], **Guoshun Nan**[3],
[1]Zhejiang University & AZFT Joint Lab for Knowledge Engine, China
[2]Institute of Modern Languages and Linguistics, Fudan University, China
[3]Beijing University of Posts and Telecommunications, China
zhangningyu@zju.edu.cn, tgui@fudan.edu.cn, nanguo2021@bupt.edu.cn

Abstract

Knowledge graph construction which aims to extract knowledge from the text corpus, has appealed to the NLP community researchers. Previous decades have witnessed the remarkable progress of knowledge graph construction on the basis of neural models; however, those models often cost massive computation or labeled data resources and suffer from unstable inference accounting for biased or adversarial samples. Recently, numerous approaches have been explored to mitigate the efficiency and robustness issues for knowledge graph construction, such as prompt learning and adversarial training. In this tutorial, we aim to bring interested NLP researchers up to speed on the recent and ongoing techniques for efficient and robust knowledge graph construction. Additionally, our goal is to provide a systematic and up-to-date overview of these methods and reveal new research opportunities to the audience.

1 Introduction

Motivation: Knowledge Graphs (KGs) regard the knowledge as fact triples in the form of `<subject, predicate, object>`, which can benefit a wide range of natural language processing tasks including question answering (Jia et al., 2021; Fei et al., 2022; Zhang et al., 2021a), fact verification (Zhou et al., 2019), data-to-text generation (Li et al., 2021), commonsense reasoning (Bosselut et al., 2019) and so on. Knowledge graph construction tasks including Named Entity Recognition (NER) (Gui et al., 2019), Relation Extraction (RE) (Zeng et al., 2015) target to extract structural information from unstructured texts, have appealed to researchers in NLP community. While those researchers have largely separated approaches from tasks, they have encountered similar issues such as efficiency and robustness.

Intuitively, efficient and robust knowledge graph construction has been widely investigated due to its potential value of making models scenario-adaptable, data-efficient, and particularly convenient for real-world applications with cold-start issues. In this tutorial, we take a holistic view of knowledge graph construction, introducing the commonalities in the issues and solutions regarding efficiency and robustness. We will explore the approaches of named entity recognition and relation extraction with few-shot labeled data, limited computation resources and approaches to improve the model robustness.

Note that our tutorial is related to green deep learning (Xu et al., 2021) that appeals to researchers to focus on carbon emission and energy usage during model training and inference, and relevant to robust NLP (Omar et al., 2022) which focuses on addressing issues in current models' language understanding capabilities with adversarial attacks. Meanwhile, trends within knowledge graph construction have shifted toward low-resource rather than considering massive labeled data and reliable & trustworthy knowledge graph construction. Notably, it is worth considering the knowledge graph construction tasks as a whole to develop methodologies for efficiency and robustness issues. We will discuss these works and suggest avenues in the future.

Tutorial Content: We will start this tutorial by defining tasks of knowledge graph construction, including named entity recognition relation extraction from sentences or documents. Then, we will give introductions to the basic models, open datasets and tools used in knowledge graph construction covering both English and Chinese (Zhang et al., 2022). We plan to focus on methods that enable efficient knowledge graph construction, such as the distant supervision (Wang et al., 2022b) and data augmentation paradigms of creating training data (Liu et al., 2021), model enhancement methods like meta-learning (Yu et al., 2020), transfer learning (Ma et al., 2022a) and prompt learn-

Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing: Tutorial Abstracts, pages 1–7
November 20-21, 2022. ©2022 Association for Computational Linguistics

ing (Chen et al., 2022d,c,b), parameter-efficient approaches (Ma et al., 2022b; Chen et al., 2022a), including adaptor-based tuning. We will then explore research focusing on robust knowledge graph construction for stable learning with adversarial attacks and selection or semantic biases.

During the tutorial, we plan to deliver lessons learned from the diverse communities involved in knowledge graph construction research and will introduce insights from the industry when building a business knowledge graph in low-resource settings. Section 3 has an outline of tutorial content.

Tutorial slides will be available at https://github.com/NLP-Tutorials/AACL-IJCNLP2022-KGC-Tutorial.

Relevance to AACL: Knowledge graphs benefit many crucial NLP tasks, and knowledge graph construction tasks such as relation extraction and named entity recognition are core tasks in information extraction. A 2018 NAACL tutorial, "Scalable Construction and Reasoning of Massive Knowledge Bases" (Ren et al., 2018), introduced a summary of recent KB, and IE works. More recently, an ACL tutorial, "Multi-modal Information Extraction from Text, Semi-structured, and Tabular Data on the Web" (Dong et al., 2020), provided an overview of information extraction (IE) from Web data with two vital dimensions: the thrust to develop scalable approaches and the diversity in data modality. However, previous tutorials mainly focus on models with rich resources of labeled data and computation, and recent years have witnessed the fast development of efficient and robust knowledge graph construction. On the other hand, the NLP community has paid much attention to robust NLP, such as an EMNLP 2021 tutorial, "Robustness and Adversarial Examples in Natural Language Processing" (Chang et al., 2021). Different from this tutorial in general NLP, we target a small, focused domain of knowledge graph construction and introduce the detailed latest work in limited 3 hours.

2 Type of this Tutorial

This tutorial contains **cutting-edge** approaches in general knowledge graph construction approaches regarding efficiency and robustness issues. However, our coverage of this tutorial will contain **introductory material** of knowledge graph construction for widespread audiences of the NLP community. Besides, we will introduce methods of Chinese knowledge graph construction for Asia audiences.

3 Outline

1. (1 hour) Introduction and Applications

- Named Entity Recognition (NER)
  - Flat NER (Li et al., 2020)
  - Nested NER (Straková et al., 2019)
  - Joint Flat and Nested NER (Wang and Lu, 2020)
- Relation Extraction
  - Supervised Relation Extraction (Lin et al., 2016; Song et al., 2018; Nan et al., 2021)
  - Distance-supervised Relation extraction (Zeng et al., 2015)
  - Open Relation Extraction (Wu et al., 2019; Zhao et al., 2021)
- Knowledge Graph Construction
  - Introduction (Bosselut et al., 2019)
  - Industry Examples
  - Resource Applications and Toolkits
  - Importance of the Efficiency and Robustness

2. (1 hour) Efficient KG Construction

- Data Efficiency
  - Data Augmentation (Chaudhary et al., 2019)
  - Model Enhancement (Chen et al., 2022d)
  - Hybrid Approaches (Hu et al., 2021)
- Model Efficiency
  - Parameter-efficient Learning (Zhou et al., 2021)
  - Efficient Architecture (Zhu, 2021)
- Inference Efficiency
  - Generative Inference (Yan et al., 2021)
  - Non-autoregressive Decoding (Sui et al., 2021)

3. (1 hour) Robust KG Construction

- Robustness Problem Discovery
  - Model Behavior Probing (Cao et al., 2021)
  - Robustness Evaluation (Wang et al., 2021)

- Data Correction
  - Data Denoising (Ma et al., 2021)
  - Data Bias Removal (Mehrabi et al., 2020)
- Robust Model Learning
  - Adversarial Training (Li and Qiu, 2021; Liu et al., 2022)
  - Robust Architecture Design (Zheng et al., 2022; Wang et al., 2022a)
  - Causal Inference (Zhang et al., 2021b)

4 Prerequisites

Anyone with a background in natural language processing can access this tutorial. Moreover, a basic understanding of neural networks, preferably with some knowledge of information extraction, knowledge graph, and pre-trained language models, is helpful.

5 Reading list

- "Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion", (Dong et al., 2014)
- "Fonduer: Knowledge Base Construction from Richly Formatted Data", (Wu et al., 2018)
- "A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios", (Hedderich et al., 2021)
- "Few-Shot Named Entity Recognition: An Empirical Baseline Study", (Huang et al., 2021)
- "Knowledge Extraction in Low-Resource Scenarios: Survey and Perspective", (Deng et al., 2022)
- "Uncertainty-Aware Label Refinement for Sequence Labeling ", (Gui et al., 2020)
- "Reasoning with latent structure refinement for document-level relation extraction", (Nan et al., 2020)
- "KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction", (Chen et al., 2022d)

6 Presenters

Ningyu Zhang is an associate professor at Zhejiang University, leading the group about KG and NLP technologies. He is also a researcher at Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies (AZFT), Co-PI of the Alibaba Open Business Knowledge Graph[1], which is devoted to benefiting e-commerce applications and to discovering socioeconomic values. He is a member of ACL, a member of the Youth Working Committee of the Chinese Information Processing Society of China, and the member of the Language and Knowledge Computing Professional Committee of the Chinese Information Processing Society of China. He has published many papers in top international academic conferences and journals such as ICLR, ACL, ENNLP, NAACL, and IEEE/ACM Transactions on Audio Speech and Language. He has served as PCs for NeurIPS, ICLR, KDD, ICML, AAAI, IJCAI and reviewer for ARR, TKDE, and TKDD. He has received the best paper award from the China Conference on Knowledge Graph and Semantic Computing and the best paper nominations from the International Joint Conference on Knowledge Graphs. He has won first place in the TREC Precision Medicine 2020 sponsored by the National Institute of Standards and Technology (NIST) and fourth place in the international semantic evaluation competition (SemEval 2021 Task4) sponsored by ACL. He has given multiple talks on information extraction and knowledge graph.

Email: zhangningyu@zju.edu.cn

Homepage: https://person.zju.edu.cn/en/ningyu

Tao Gui is an associate professor at the Institute of Modern Languages and Linguistics of Fudan University. He is the key member of the FudanNLP group[2]. He is a member of ACL, a member of the Youth Working Committee of the Chinese Information Processing Society of China, and the member of the Language and Knowledge Computing Professional Committee of the Chinese Information Processing Society of China. He has published more than 40 papers in top international academic conferences and journals such as ACL, ENNLP, AAAI, IJCAI, SIGIR, and so on. He has served as area chair or PCs for SIGIR, AAAI, IJCAI, TPAMI, and ARR. He has received the Outstanding Doctoral Dissertation Award of the Chinese Informa-

[1] https://kg.alibaba.com/

[2] https://nlp.fudan.edu.cn

tion Processing Society of China, the area chair favorite Award of COLING 2018, the outstanding Paper Award of NLPCC 2019, and a scholar of young talent promoting projects of CAST.

Email: tgui@fudan.edu.cn

Homepage: https://guitaowufeng.github.io

Guoshun Nan is a tenure-track professor in the School of Cyber Science and Engineering, Beijing University of Posts and Telecommunications (BUPT). He is a key member of the National Engineering Research Center of Mobile Network Security and a member of the Wireless Technology Innovation Institute of BUPT. Before starting his academic career, he also worked in Hewlett-Packard Company (China) for more than four years as an engineer. He is a member of ACL. He has a broad interest in information extraction, model robustness, multimodal retrieval, cyber security and the next generation of wireless networks. He has published more than ten papers in top-tier conferences such as ACL, CVPR, EMNLP, SIGIR, IJCAI, CKIM and Sigcomm. He served as a reviewer for ACL, EMNLP, AAAI, IJCAI, Neurocomputing and IEEE Transaction on Image Processing.

Email: nanguo2021@bupt.edu.cn

References

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4762–4779, Florence, Italy. Association for Computational Linguistics.

Boxi Cao, Hongyu Lin, Xianpei Han, Le Sun, Lingyong Yan, Meng Liao, Tong Xue, and Jin Xu. 2021. Knowledgeable or educated guess? revisiting language models as knowledge bases. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 1860–1874. Association for Computational Linguistics.

Kai-Wei Chang, He He, Robin Jia, and Sameer Singh. 2021. Robustness and adversarial examples in natural language processing. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts*, pages 22–26, Punta Cana, Dominican Republic & Online. Association for Computational Linguistics.

Aditi Chaudhary, Jiateng Xie, Zaid Sheikh, Graham Neubig, and Jaime G. Carbonell. 2019. A little annotation does a lot of good: A study in bootstrapping low-resource named entity recognizers. 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 2019, Hong Kong, China, November 3-7, 2019*, pages 5163–5173. Association for Computational Linguistics.

Xiang Chen, Lei Li, Shumin Deng, Chuanqi Tan, Changliang Xu, Fei Huang, Luo Si, Huajun Chen, and Ningyu Zhang. 2022a. Lightner: A lightweight tuning paradigm for low-resource ner via pluggable prompting. In *Proceedings of the 29th International Conference on Computational Linguistics*. International Committee on Computational Linguistics.

Xiang Chen, Lei Li, Ningyu Zhang, Xiaozhuan Liang, Shumin Deng, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2022b. Decoupling knowledge from memorization: Retrieval-augmented prompt learning. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Xiang Chen, Lei Li, Ningyu Zhang, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2022c. Relation extraction as open-book examination: Retrieval-enhanced prompt tuning. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, pages 2443–2448. ACM.

Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2022d. Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, pages 2778–2788. ACM.

Shumin Deng, Ningyu Zhang, Hui Chen, Feiyu Xiong, Jeff Z. Pan, and Huajun Chen. 2022. Knowledge extraction in low-resource scenarios: Survey and perspective. *CoRR*, abs/2202.08063.

Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014*, pages 601–610. ACM.

Xin Luna Dong, Hannaneh Hajishirzi, Colin Lockard, and Prashant Shiralkar. 2020. Multi-modal information extraction from text, semi-structured, and tabular data on the web. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 23–26, Online. Association for Computational Linguistics.

Zichu Fei, Qi Zhang, Tao Gui, Di Liang, Sirui Wang, Wei Wu, and Xuan-Jing Huang. 2022. Cqg: A simple and effective controlled generation framework for multi-hop question generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6896–6906.

Tao Gui, Jiacheng Ye, Qi Zhang, Zhengyan Li, Zichu Fei, Yeyun Gong, and Xuan-Jing Huang. 2020. Uncertainty-aware label refinement for sequence labeling. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2316–2326.

Tao Gui, Yicheng Zou, Qi Zhang, Minlong Peng, Jinlan Fu, Zhongyu Wei, and Xuanjing Huang. 2019. A lexicon-based graph neural network for Chinese NER. 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 1040–1050, Hong Kong, China. Association for Computational Linguistics.

Michael A. Hedderich, Lukas Lange, Heike Adel, Jannik Strötgen, and Dietrich Klakow. 2021. A survey on recent approaches for natural language processing in low-resource scenarios. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 2545–2568. Association for Computational Linguistics.

Xuming Hu, Chenwei Zhang, Yawen Yang, Xiaohe Li, Li Lin, Lijie Wen, and Philip S. Yu. 2021. Gradient imitation reinforcement learning for low resource relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 2737–2746. Association for Computational Linguistics.

Jiaxin Huang, Chunyuan Li, Krishan Subudhi, Damien Jose, Shobana Balakrishnan, Weizhu Chen, Baolin Peng, Jianfeng Gao, and Jiawei Han. 2021. Few-shot named entity recognition: An empirical baseline study. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 10408–10423. Association for Computational Linguistics.

Zhen Jia, Soumajit Pramanik, Rishiraj Saha Roy, and Gerhard Weikum. 2021. Complex temporal question answering on knowledge graphs. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*, pages 792–802. ACM.

Junyi Li, Tianyi Tang, Wayne Xin Zhao, Zhicheng Wei, Nicholas Jing Yuan, and Ji-Rong Wen. 2021. Few-shot knowledge graph-to-text generation with pretrained language models. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1558–1568, Online. Association for Computational Linguistics.

Linyang Li and Xipeng Qiu. 2021. Token-aware virtual adversarial training in natural language understanding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(9):8410–8418.

Xiaonan Li, Hang Yan, Xipeng Qiu, and Xuanjing Huang. 2020. FLAT: chinese NER using flat-lattice transformer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 6836–6842. Association for Computational Linguistics.

Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2016. Neural relation extraction with selective attention over instances. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics.

Kun Liu, Yao Fu, Chuanqi Tan, Mosha Chen, Ningyu Zhang, Songfang Huang, and Sheng Gao. 2021. Noisy-labeled NER with confidence estimation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 3437–3445. Association for Computational Linguistics.

Qin Liu, Rui Zheng, Bao Rong, Jingyi Liu, Zhihua Liu, Zhanzhan Cheng, Liang Qiao, Tao Gui, Qi Zhang, and Xuan-Jing Huang. 2022. Flooding-x: Improving bert's resistance to adversarial attacks via loss-restricted fine-tuning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5634–5644.

Ruotian Ma, Tao Gui, Linyang Li, Qi Zhang, Xuanjing Huang, and Yaqian Zhou. 2021. SENT: sentence-level distant relation extraction via negative training. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 6201–6213. Association for Computational Linguistics.

Ruotian Ma, Yiding Tan, Xin Zhou, Xuanting Chen, Di Liang, Sirui Wang, Wei Wu, and Tao Gui. 2022a. Searching for optimal subword tokenization in cross-domain ner. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 4289–4295. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Ruotian Ma, Xin Zhou, Tao Gui, Yiding Tan, Linyang Li, Qi Zhang, and Xuanjing Huang. 2022b. Template-free prompt tuning for few-shot NER. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5721–5732, Seattle, United States. Association for Computational Linguistics.

Ninareh Mehrabi, Thamme Gowda, Fred Morstatter, Nanyun Peng, and Aram Galstyan. 2020. Man is to person as woman is to location: Measuring gender bias in named entity recognition. In *HT '20: 31st ACM Conference on Hypertext and Social Media, Virtual Event, USA, July 13-15, 2020*, pages 231–232. ACM.

Guoshun Nan, Zhijiang Guo, Ivan Sekulic, and Wei Lu. 2020. Reasoning with latent structure refinement for document-level relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 1546–1557. Association for Computational Linguistics.

Guoshun Nan, Jiaqi Zeng, Rui Qiao, Zhijiang Guo, and Wei Lu. 2021. Uncovering main causalities for long-tailed information extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9683–9695.

Marwan Omar, Soohyeon Choi, Daehun Nyang, and David Mohaisen. 2022. Robust natural language processing: Recent advances, challenges, and future directions. *IEEE Access*, 10:86038–86056.

Xiang Ren, Nanyun Peng, and William Yang Wang. 2018. Scalable construction and reasoning of massive knowledge bases. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pages 10–16, New Orleans, Louisiana. Association for Computational Linguistics.

Linfeng Song, Yue Zhang, Zhiguo Wang, and Daniel Gildea. 2018. N-ary relation extraction using graph-state lstm. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2226–2235.

Jana Straková, Milan Straka, and Jan Hajic. 2019. Neural architectures for nested NER through linearization. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 5326–5331. Association for Computational Linguistics.

Dianbo Sui, Chenhao Wang, Yubo Chen, Kang Liu, Jun Zhao, and Wei Bi. 2021. Set generation networks for end-to-end knowledge base population. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 9650–9660. Association for Computational Linguistics.

Jue Wang and Wei Lu. 2020. Two are better than one: Joint entity and relation extraction with table-sequence encoders. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 1706–1721. Association for Computational Linguistics.

Xiao Wang, Shihan Dou, Limao Xiong, Yicheng Zou, Qi Zhang, Tao Gui, Liang Qiao, Zhanzhan Cheng, and Xuan-Jing Huang. 2022a. Miner: Improving out-of-vocabulary named entity recognition from an information theoretic perspective. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5590–5600.

Xiao Wang, Qin Liu, Tao Gui, Qi Zhang, Yicheng Zou, Xin Zhou, Jiacheng Ye, Yongxin Zhang, Rui Zheng, Zexiong Pang, Qinzhuo Wu, Zhengyan Li, Chong Zhang, Ruotian Ma, Zichu Fei, Ruijian Cai, Jun Zhao, Xingwu Hu, Zhiheng Yan, Yiding Tan, Yuan Hu, Qiyuan Bian, Zhihua Liu, Shan Qin, Bolin Zhu, Xiaoyu Xing, Jinlan Fu, Yue Zhang, Minlong Peng, Xiaoqing Zheng, Yaqian Zhou, Zhongyu Wei, Xipeng Qiu, and Xuanjing Huang. 2021. TextFlint: Unified multilingual robustness evaluation toolkit for natural language processing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations*, pages 347–355, Online. Association for Computational Linguistics.

Zifeng Wang, Rui Wen, Xi Chen, Shao-Lun Huang, Ningyu Zhang, and Yefeng Zheng. 2022b. Finding influential instances for distantly supervised relation extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*.

Ruidong Wu, Yuan Yao, Xu Han, Ruobing Xie, Zhiyuan Liu, Fen Lin, Leyu Lin, and Maosong Sun. 2019. Open relation extraction: Relational knowledge transfer from supervised data to unsupervised data. 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 219–228, Hong Kong, China. Association for Computational Linguistics.

Sen Wu, Luke Hsiao, Xiao Cheng, Braden Hancock, Theodoros Rekatsinas, Philip Alexander Levis, and Christopher Ré. 2018. Fonduer: Knowledge base construction from richly formatted data. In *Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018*, pages 1301–1316. ACM.

Jingjing Xu, Wangchunshu Zhou, Zhiyi Fu, Hao Zhou, and Lei Li. 2021. A survey on green deep learning. *arXiv preprint arXiv:2111.05193*.

Hang Yan, Tao Gui, Junqi Dai, Qipeng Guo, Zheng Zhang, and Xipeng Qiu. 2021. A unified generative framework for various NER subtasks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 5808–5822. Association for Computational Linguistics.

Haiyang Yu, Ningyu Zhang, Shumin Deng, Hongbin Ye, Wei Zhang, and Huajun Chen. 2020. Bridging text and knowledge with multi-prototype embedding for few-shot relational triple extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6399–6410, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1753–1762, Lisbon, Portugal. Association for Computational Linguistics.

Ningyu Zhang, Qianghuai Jia, Shumin Deng, Xiang Chen, Hongbin Ye, Hui Chen, Huaixiao Tou, Gang Huang, Zhao Wang, Nengwei Hua, and Huajun Chen. 2021a. Alicg: Fine-grained and evolvable conceptual graph construction for semantic search at alibaba. In *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021*, pages 3895–3905. ACM.

Ningyu Zhang, Xin Xu, Liankuan Tao, Haiyang Yu, Hongbin Ye, Shuofei Qiao, Xin Xie, Xiang Chen, Zhoubo Li, Lei Li, et al. 2022. Deepke: A deep learning based knowledge extraction toolkit for knowledge base population. *arXiv preprint arXiv:2201.03335*.

Wenkai Zhang, Hongyu Lin, Xianpei Han, and Le Sun. 2021b. De-biasing distantly supervised named entity recognition via causal intervention. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 4803–4813. Association for Computational Linguistics.

Jun Zhao, Tao Gui, Qi Zhang, and Yaqian Zhou. 2021. A relation-oriented clustering method for open relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9707–9718.

Rui Zheng, Bao Rong, Yuhao Zhou, Di Liang, Sirui Wang, Wei Wu, Tao Gui, Qi Zhang, and Xuanjing Huang. 2022. Robust lottery tickets for pre-trained language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 2211–2224. Association for Computational Linguistics.

Jie Zhou, Xu Han, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2019. GEAR: Graph-based evidence aggregating and reasoning for fact verification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 892–901, Florence, Italy. Association for Computational Linguistics.

Xin Zhou, Ruotian Ma, Tao Gui, Yiding Tan, Qi Zhang, and Xuanjing Huang. 2021. Plug-tagger: A pluggable sequence labeling framework using language models. *arXiv preprint arXiv:2110.07331*.

Wei Zhu. 2021. Autorc: Improving BERT based relation classification models via architecture search. In *Proceedings of the ACL-IJCNLP 2021 Student Research Workshop, ACL 2021, Online, JUli 5-10, 2021*, pages 33–43. Association for Computational Linguistics.