# Recent Challenges and Innovations in Named Entity Recognition: A Survey of the Subfield

**Neil Kamdar**
School of Information, University of California, Berkeley
INFO 159, Natural Language Processing, Spring 2023
nvk518@berkeley.edu

## Abstract

Named Entity Recognition (NER) is a subfield in natural language processing (NLP) that aims to identify and classify proper nouns including people, locations, and organizations in unstructured text, and its applications are tremendous in a variety of different areas of NLP. In recent years, there have been significant improvements in NER, driven by the availability of large annotated datasets, multi-lingual data, and innovations in approaches. By evaluating advancements in each of the approaches, this survey provides an overview of the latest advancements in NER in 2022 and 2023, including different types of models, evaluation metrics, and challenges. It also discusses the current state-of-the-art approaches and highlights some future directions for research.[1]

## 1 Introduction

Named Entity Recognition (NER) is a method of extracting information, such as people, places, and organizations from text. NER finds itself in many applications, including information retrieval, question answering, and sentiment analysis.

Some notable historical literature that has defined NER include Chiu and Nichols (2016) introducing revolutionary Bidirectional LSTM-CNNs for NER, Akbik et al.'s (2018) work introducing contextual string embedding for significant NER performance increases, and Sutton and McCallum's (2010) investigation into Conditional Random Fields (CRFs) for NER. CRFs, once state-of-the-art for NER in the early 2000s, have since been outperformed by deep learning innovations including LSTM-CNNs-CRF (Ma and Hovy, 2016) and self-attention transformers (Vaswani et al., 2017) allowing for transfer learning and contextualized word embeddings for massive performance increases, especially in domains with limited labeled data.

With the constantly evolving subfield of NER, an understanding of the challenges of NER is necessary to conduct a survey. Some of the difficulties faced in NER include the presence of low-resource environments where data is scarce, lack of language support required for NER, understanding of niche subfield knowledge such as clinical medicine or music, and lack of annotated data and rulesets. From these challenges, we can analyze improvements and the current state of how they influence research being conducted.

## 2 Methodology

The survey was conducted through the analysis of research conducted in 2022 and 2023 and published in major conferences and journals, computational outlets, and through Google Scholar to obtain a breadth of knowledge from multiple venues. They were then investigated, compiled, and organized into key areas of NER.

## 3 Training Data

Datasets used for NER have emerged to solve problems of low resources and to increase the presence of training data for modeling. As of late, NER has been primarily tested in well-formatted documents such as Wikipedia and news articles. However, Antypas et al. (2022) seek to diversify this by using a Twitter social media dataset with baselines and annotated entity types.

### 3.1 Data Generation

Annotations have been usually done manually by domain specialists, but Kim et al. (2022) look to reduce laborious annotations and introduce an "ask-to-generate" approach to labeling to generate NER datasets by asking simple questions to a question-answering system. Chen et al. (2022) have worked on a similar task of data augmentation using style transfer to expand the size and diversity of training

[1] Word Count: 2199

data when presented with low resources by changing a text's style-related attributes to generate synthetic data for training models. Data augmentation using paraphrasing is a promising method for classification tasks, but its effect on NER had yet to be systemically explored. Sharma et al. (2022) employ different state-of-the-art strategies to compare methods of generating paraphrases such as encoder-decoder models, back translation, and GPT-3 variants, and conclude by emphasizing the importance of the choice of paraphraser having a significant impact on NER performance. A larger GPT-3 variant came out on top in their evaluation as it increased NER performance in most cases.

### 3.2 Data for Multi-language Support

Because manual training data used for NER is expensive, Tedeschi and Navigli (2022) designed a language-agnostic approach to generating fine-grained NER annotations called MultiNERD, which has allowed for the creation of a dataset containing "10 languages, 15 NER categories, and 2 textual genres"[2] using Babel and Wikipedia. Expansion of NER dataset language support into Thai (Buaphet et al., 2022) has proven difficult and discussions of lower-resource languages lacking sufficient resources and the inaccuracies of Wikidata (Lignos et al., 2022) have emerged to examine the importance of making a multilingual NER solution that is functional across a myriad of languages. A similar expansion of the language dataset was made in Marathi, a language of the Indian subcontinent. Joshi (2022) presents L3Cube-MahaCorpus a Marathi dataset with 289 million tokens and 24.8 sentences as well as MahaBERT, MahaAlBERT, and MahaRoBerta language models – a huge step forward for NER and other NLP tasks in Marathi.

Undoubtedly, NER's future presents advancements in additional language support and new data augmentation techniques.

## 4 Methods of Named Entity Recognition

Named Entity Recognition can be carried out through three major approaches: rule-based, machine learning-based, and hybrid-based consisting of a combination of methods. Rule-based NER has fallen significantly in popularity in recent years as many of the papers researched rely on the use of machine learning for solving NER or a combination of both machine learning and linguistic rules.

This paper will go into recent advancements in each of the schools of thought and how they are being used to solve novel problems in NER.

### 4.1 Machine Learning Based

Machine learning methods utilize neural networks to train and test models on datasets. However, this has its shortcomings as massive data is required to build a state-of-the-art NER model. As aforementioned, significant improvements in data augmentation and the availability of data have allowed for novel solutions to NER using machine learning.

#### 4.1.1 Adversarial Learning

Adversarial learning approaches rely on feeding a model with data used to fool it to improve robustness and generalization ability. TransAdv for Zero-Resource Cross-Lingual NER had been proposed by Zhao et al. (2022) that mitigates errors in word-by-word translated data and utilizes "multi-level adversarial learning and multi-model knowledge distillation"[3] to achieve competitive performance compared to state-of-the-art models. Similarly, adversarial learning has proven useful for Relation Extraction (RE) – Qin et al. (2022) model RE with adversarial multi-task learning, where the first training stage recovers given named entities to enhance the main relation extractor, and the adversarial mechanism controls the effect of NER on relation extraction. The result is a state-of-the-art performance on two RE benchmark datasets, which validates the ability for adversarial learning to extract relations.

#### 4.1.2 Few-shot Learning

Few-shot learning is a machine learning approach where data is classified given a model trained on very few labeled examples. While traditional machine learning approaches require large amounts of annotated training data to train accurate models, few-shot learning is often used for low-resource settings without requiring large human-annotated datasets used for supervised deep learning. Sub-optimal performance has been seen when generalizing few-shot learning to unseen target domains, which has motivated Das et al. (2022) who have

[2] https://aclanthology.org/2022.findings-naacl.60.pdf

[3] https://aclanthology.org/2022.findings-emnlp.52.pdf

proposed CONTaiNER, "a novel contrastive learning technique optimizing the inter-token distribution distance for few-shot NER" by using Gaussian-distributed embeddings to differentiate between token categories[4]. The model has proven useful, especially in challenging scenarios by outperforming previous techniques for few-shot NER.

#### 4.1.3 GPT and LMs

With the recent popularization of GPT models within the scientific community, researchers have explored the potential for large language models (LLMs) by evaluating them against common Pre-trained Language Models (PLMs) such as BERT and its derivatives. Hu et al. (2023) work in utilizing ChatGPT for clinical NER in a zero-shot setting, Gutiérrez et al.'s (2022) study of GPT-3 for biomedical NER and relation extraction in a few-shot setting, and González-Gallardo et al.'s (2023) exploration of ChatGPT for entity tagging in historical primary sources has demonstrated the shortcomings of GPT evaluated against PLMs. The latter research highlights difficulties including entity complexity and multilingualism, a sentiment echoed by other attempts to use GPT-3 for NER. González-Gallardo et al. identify issues with using ChatGPT for NER resulting from ChatGPT's confusion with primary sources including first-hand accounts and newspapers due to them being locked behind paywalls, which can "cause biased and out-of-domain responses"[5] The adoption of ChatGPT for NER is a technique that will require future studies and considerations to understand how researchers can leverage the popular model for entity recognition.

Additionally, because LMs typically go years without updates, Onoe et al. (2022) propose a framework of analysis of LMs' abilities to infer about new entities by creating an automatic data collection pipeline sourced from Wikipedia and concluding greater access to textual information decreases perplexity. The emphasize the difficulties LMs have in making inferences about entities they have not been trained on. This is significant in proving the importance of updating models in a dynamic world and for future research.

An "EACL Outstanding Paper" award-winning study conducted by Epure and Hennequin (2023) investigates NER on conversational music recommendation queries. The researchers incorporated a human subject study to understand human linguistic capabilities and compare them with the BERT, RoBERTa, and MPNet PLMs. Since music entities come from creative thought, many entities are without proper capitalization and contain misspellings. Although this contributes a significant amount of noise, the results, interestingly, showed challenges for both the humans and models, where humans had higher precision, but the model had a higher recall due to pre-training increasing entity exposure.

#### 4.1.4 Other Neural Networks

The GLARA approach is a machine learning-based method for NER suggested by Zhao et al. (2021) that utilizes graph neural networks to generate new labeling rules from unlabeled data. By investigating semantic relations between candidate rules, GLARA creates augmented rules for the generation of weak labels used when training the supervised NER model and improving the accuracy of NER.

Many aforementioned problems in language modeling have been worked towards using Deep Learning. Gated graph neural networks have been key in understanding challenging issues using rulesets. Haisa and Altenbek (2022) employ a deep-learning approach to improving NER for the Kazakh language using a gated graph neural network (GGNN) on a graph structure consisting of entities as nodes and inclusion relations as edges, outperforming previous methods' F1 scores by 88.04%. De Cao et al. (2022) propose the mGENRE sequence-to-sequence modeling system for Multilingual Entity Linking that predicts the name of the target entity when resolving language-specific mentions in a Knowledge Base that is multilingual.

### 4.2 Rule-based

Linguistic rule-based approaches rely on the researchers being experts in understanding language grammar. The creation and maintenance of rules is time-intensive to manually create and the dominance of machine learning models for NER have caused rule-based NER approaches to decline in popularity. Simultaneously, new innovations in data augmentation have made it easier to solve NER using machine learning. While researching papers from major conferences, the decrease in methods relying solely on linguistic rules was obvious as many instead focused on machine learning approaches. However, there were few studies that

[4] https://aclanthology.org/2022.acl-long.439.pdf

[5] https://arxiv.org/pdf/2303.17322.pdf

still employed rule-based methods for niche domains that outperformed popular machine learning models.

An interesting use case for linguistic rule-based NER is information extraction for mechanical-electrical-plumbing (MEP) from the web. Leveraging the advantages of rule-based approaches is essential, given the scarcity of MEP data available and the inherent complexity of MEP information that makes classical deep learning challenging. The method incorporates a "suffix-based matching algorithm on text segments" for NER and a matching algorithm on dependency paths using a dependency tree for Relation Extraction (Wu et al., 2022). The results illustrated the rule-based approach outperformed deep learning models by 37% to 49%, showing the viability of rule-sets in a machine learning-dominated field.

VALET is a rule-based Python framework for information extraction and NER "offering direct support for mixing heterogeneous information—lexical, orthographic, syntactic, corpus-analytic—in a succinct syntax that supports context-free idioms" (Freitag et al., 2022). The researchers emphasize the importance of rule-based information extraction in the early stages of model development and for annotating examples fed into machine learning extraction models.

### 4.3 Hybrid-based

Hybrid approaches rely on the use of both Rule-based and machine learning-based methods. Span-based approaches typically have a two-stage task consisting of extraction of a span and classification.

The hybrid-based methodology has been carried out using a combination of rule-based systems and machine learning. Reich et al. (2022) have employed this approach when leveraging expert-guided rule systems to improve the generalization of NER and use GAT to improve the suggested model's performance on adversarial data. Entity Disambiguation (ED) defines the problem of reducing the ambiguity of classification and Ayoola et al. (2022) use a cutting-edge approach with a structured knowledge base and an ED model that links entities by reasoning over a rule-based, symbolic knowledge base using a machine learning model for relation extraction – the "first time an end-to-end differentiable symbolic KB has been used for ED".[6]

#### 4.3.1 Span-based

Recently, nested NER, where one token may belong to multiple mentions, has been a topic of development. Span-based approaches, with their two-stage tasks, have emerged as popular options for solving nested NER through the extraction of possible spans and classification of their respective category. Wan et al. (2022) discuss the problem of nested NER and propose the utilization of retrieval-based span-level graphs with a neural network backbone to improve the span representation and incorporate similar neighbor entities into the span representation. This has led to overall improvements in F1 score when tested on three benchmarks and recall increases for low-frequency entities. Nested NER has the issues of errors being propagated, span boundaries being ignored, and difficulties in long entity recognition. Huang et al. (2022) discuss solutions with a span selection framework called Extract-Select for tackling the problems. It has a three-fold approach which has led to state-of-the-art evaluation metrics when compared to benchmark datasets: it uses a trained extractor to select results pertaining to the given entity category, a hybrid selection mechanism in the extractor, and an extractor and discriminator for evaluation trained using Generational Adversarial Training (GAT)[7].

## 5 Conclusion

Named Entity Recognition is a key subfield of NLP that involves extracting key entities from text. With this comes a myriad of challenges ranging from entity ambiguity to data sparsity. However, recent advancements in data augmentation and within multiple approaches to NER including machine learning-based, rule-based, and hybrid-based have sought to increase access and accuracy of NER in a variety of fields, such as music and mechanical-electrical-plumbing. They look to work with other innovations, including those in LLMs, to understand future directions for research.

## References


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–


[6]https://aclanthology.org/2022.naacl-main.210.pdf

[7]https://aclanthology.org/2022.findings-acl.154.pdf

1649, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Dimosthenis Antypas, Asahi Ushio, Jose Camacho-Collados, Vitor Silva, Leonardo Neves, and Francesco Barbieri. 2022. Twitter topic classification. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3386–3400, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Tom Ayoola, Joseph Fisher, and Andrea Pierleoni. 2022. Improving entity disambiguation by reasoning over a knowledge base. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2899–2912, Seattle, United States. Association for Computational Linguistics.

Weerayut Buaphet, Can Udomcharoenchaikit, Peerat Limkonchotiwat, Attapol Rutherford, and Sarana Nutanong. 2022. Thai nested named entity recognition corpus. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1473–1486, Dublin, Ireland. Association for Computational Linguistics.

Shuguang Chen, Leonardo Neves, and Thamar Solorio. 2022. Style transfer as data augmentation: A case study on named entity recognition. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1827–1841, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

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.

Sarkar Snigdha Sarathi Das, Arzoo Katiyar, Rebecca Passonneau, and Rui Zhang. 2022. CONTaiNER: Few-shot named entity recognition via contrastive learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6338–6353, Dublin, Ireland. Association for Computational Linguistics.

Nicola De Cao, Ledell Wu, Kashyap Popat, Mikel Artetxe, Naman Goyal, Mikhail Plekhanov, Luke Zettlemoyer, Nicola Cancedda, Sebastian Riedel, and Fabio Petroni. 2022. Multilingual autoregressive entity linking. *Transactions of the Association for Computational Linguistics*, 10:274–290.

Elena Epure and Romain Hennequin. 2023. A human subject study of named entity recognition in conversational music recommendation queries. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1281–1296, Dubrovnik, Croatia. Association for Computational Linguistics.

Dayne Freitag, John Cadigan, Robert Sasseen, and Paul Kalmar. 2022. Valet: Rule-based information extraction for rapid deployment. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 524–533, Marseille, France. European Language Resources Association.

Carlos-Emiliano González-Gallardo, Emanuela Boros, Nancy Girdhar, Ahmed Hamdi, Jose G. Moreno, and Antoine Doucet. 2023. Yes but.. can chatgpt identify entities in historical documents?

Bernal Jiménez Gutiérrez, Nikolas McNeal, Clay Washington, You Chen, Lang Li, Huan Sun, and Yu Su. 2022. Thinking about gpt-3 in-context learning for biomedical ie? think again.

Gulizada Haisa and Gulila Altenbek. 2022. Deep learning with word embedding improves kazakh named-entity recognition. *Information*, 13(4).

Yan Hu, Iqra Ameer, Xu Zuo, Xueqing Peng, Yujia Zhou, Zehan Li, Yiming Li, Jianfu Li, Xiaoqian Jiang, and Hua Xu. 2023. Zero-shot clinical entity recognition using chatgpt.

Peixin Huang, Xiang Zhao, Minghao Hu, Yang Fang, Xinyi Li, and Weidong Xiao. 2022. Extract-select: A span selection framework for nested named entity recognition with generative adversarial training. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 85–96, Dublin, Ireland. Association for Computational Linguistics.

Raviraj Joshi. 2022. L3cube-mahacorpus and mahabert: Marathi monolingual corpus, marathi BERT language models, and resources. *CoRR*, abs/2202.01159.

Hyunjae Kim, Jaehyo Yoo, Seunghyun Yoon, Jinhyuk Lee, and Jaewoo Kang. 2022. Simple questions generate named entity recognition datasets. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6220–6236, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Constantine Lignos, Nolan Holley, Chester Palen-Michel, and Jonne Sälevä. 2022. Toward more meaningful resources for lower-resourced languages. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 523–532, Dublin, Ireland. 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.

Yasumasa Onoe, Michael Zhang, Eunsol Choi, and Greg Durrett. 2022. Entity cloze by date: What LMs know about unseen entities. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 693–702, Seattle, United States. Association for Computational Linguistics.

Han Qin, Yuanhe Tian, and Yan Song. 2022. Enhancing relation extraction via adversarial multi-task learning. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6190–6199, Marseille, France. European Language Resources Association.

Aaron Reich, Jiaao Chen, Aastha Agrawal, Yanzhe Zhang, and Diyi Yang. 2022. Leveraging expert guided adversarial augmentation for improving generalization in named entity recognition. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1947–1955, Dublin, Ireland. Association for Computational Linguistics.

Saket Sharma, Aviral Joshi, Namrata Mukhija, Yiyun Zhao, Hanoz Bhathena, Prateek Singh, Sashank Santhanam, and Pritam Biswas. 2022. Systematic review of effect of data augmentation using paraphrasing on named entity recognition. In *NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research*.

Charles Sutton and Andrew McCallum. 2010. An introduction to conditional random fields.

Simone Tedeschi and Roberto Navigli. 2022. MultiNERD: A multilingual, multi-genre and fine-grained dataset for named entity recognition (and disambiguation). In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 801–812, Seattle, United States. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

Juncheng Wan, Dongyu Ru, Weinan Zhang, and Yong Yu. 2022. Nested named entity recognition with span-level graphs. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 892–903, Dublin, Ireland. Association for Computational Linguistics.

Lang-Tao Wu, Jia-Rui Lin, Shuo Leng, Jiu-Lin Li, and Zhen-Zhong Hu. 2022. Rule-based information extraction for mechanical-electrical-plumbing-specific semantic web. *Automation in Construction*, 135:104108.

Xinyan Zhao, Haibo Ding, and Zhe Feng. 2021. GLaRA: Graph-based labeling rule augmentation for weakly supervised named entity recognition. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3636–3649, Online. Association for Computational Linguistics.

Yichun Zhao, Jintao Du, Gongshen Liu, and Huijia Zhu. 2022. TransAdv: A translation-based adversarial learning framework for zero-resource cross-lingual named entity recognition. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 742–749, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.