Journal of Engineering Faculty, 1(2): 75-81, 2023

Journal of Engineering Faculty

| cumfad.cumhuriyet.edu.tr | Founded: 2023 Available online, ISSN Publisher: Sivas Cumhuriyet Üniversitesi

A Systematic Literature Review on Custom NER with Evidence-Based Software Engineering

Hakan Kekül[1,a], Abdulkadir Şeker[2,b*]

[1]*Sivas Cumhuriyet University, Faculty of Technology, Software Engineering, Sivas, Türkiye*

[2]*Sivas Cumhuriyet University, Faculty of Engineering, Computer Engineering, Sivas, Türkiye*

**Corresponding author*

Review Article

History

Received: 24/11/2023
Accepted: 08/12/2023

ABSTRACT

As the significance of data continues to grows, so does the importance of data analysis methods. Various models are currently being applied, and new models are being proposed all the time. In the context of this study, we conducted a detailed review of Named-Entity Recognition, a data analysis model. We applied the Evidence-Based Software Engineering method, which has been used successfully for many years, as the analysis method. The study analyzed 38 articles selected from a collection of 114 different research articles identified by this method. A detailed presentation of the analyzed data is provided. The study aimed to identify the most effective among the methods using NER. The analysis indicates that BERT was the most successful method in NER studies. It has been found that the "News" domain contains the highest number of NER datasets. The study also provides detailed information on other methods and domains identified. As an original and comprehensive guide, this study serves as an excellent resource for those interested in the field.

Keywords: Evidence-based-software-engineering, Named-entity recognition, NER, Custom NER, EBSE

[a] *hakankekul@cumhuriyet.edu.tr* *0000-0001-6269-8713* | [b] *aseker@cumhuriyet.edu.tr* *0000-0002-4552-2676*

How to Cite: Kekül H, Şeker A (2023) A Systematic Literature Review on Custom NER with Evidence-Based Software Engineering, Journal of Engineering Faculty, 1(2): 75-81

Introduction

Text data is an important data type whose popularity has increased rapidly with the advancement of information technology [1]. Scientific text data can be obtained from articles, letters, e-mails, social media posts, articles, books and many other sources. These texts are important for many commercial and industrial applications and for interpersonal communication. Textual data analysis and processing focuses on an area known as text mining or natural language processing (NLP). Various extracts from the text are used to understand and manipulate this textual data. Features convert text data into numeric or symbolic values, and enable machine learning algorithms to understand this data. The main features of textual data are word frequency, word distribution, sentiment analysis, and semantic extraction.

Named entity recognition (NER) is one of these attributes. NER represents a very important component in the field of NLP. NER is an automatic process that identifies specific entities or names in texts [2]. These entities often include specific information such as names of people, place names, organization names, dates, currency units and other proper names. NER is used in text mining and information extraction applications and enables automatic classification of such data. For example, given the sentence "Selçuk Bayraktar is the CEO of Baykar company" in a text document, the proper names identified by NER can be determined as "Selçuk Bayraktar" and "Baykar". Automatic identification of such information allows for better understanding and processing of text data. Applications of NER include text mining, automatic text classification, information extraction, language translation and many more.

"Custom Named Entity Recognition" which has a special place in the attributes of text data, as the name suggests, refers to NER systems that are customized for specific needs and specific applications [3]. Standard NER systems are trained to recognize common names, place names and dates, but for some applications there may be a need to identify specific and unique entities. For example, it may be necessary to identify a company's specific product names or a special industrial terminology in the content of a document. Custom NER is a NER system that is trained and configured to respond to such specialized needs. This helps professionals working in a specific business or industry to better understand text data and process it more effectively. Custom NER systems offer great flexibility with the ability to define special terms or entities. Particularly used in legal, medical, financial, information security and other specialized fields, Custom NER systems are tailored to the requirements of text data in these specific application areas.

In this paper, we will focus more on the attributes of text data and in particular the importance of Custom NER in text mining and customized analysis applications.

Related Work

NER involves identifying specific entities, such as individuals, locations, and organizations, within a given text [2]. These entities are pre-defined semantic types and play a crucial role in various natural language applications like question answering, text summarization, and machine translation. NER serves as a fundamental component for these applications by pinpointing and categorizing relevant mentions in the text [4]. Many different libraries and natural language processing tools, such as SpaCy in Python, Apache OpenNLP, and TensorFlow, enable the creation of Named Entity Recognition (NER) systems with pre-trained models that can be imported, used, and customized based on specific requirements [5].

In addition to general NER, there are also domain-specific custom NER studies aimed at solving problems within narrower scopes [6]. Medical literature, published by researchers in the field, contains an enormous amount of information. One of the study proposes a hybrid-based approach for identifying entities from medical literature, leveraging a newly created dictionary for application routes, dosage forms, and symptoms, trained on annotated entities using a blank SpaCy machine learning model [7]. Another paper introduces a chatbot designed to increase crime awareness, utilizing classification and generative models. The chatbot includes spam detection, a system for registering complaints, and a custom named entity recognition model to extract structured information. The primary objective is to provide efficient and user-friendly complaint registration methods using natural language processing, ultimately promoting social good [8]. In a paper used custom NER, it introduces a novel weighted distributional semantic model for unsupervised Named Entity Recognition (NER) in the agricultural domain, addressing challenges such as the absence of annotated data, domain-specific vocabulary, entity ambiguity, and contextual variation [9]. One of the studies that focuses on Turkish text conducts a thorough examination of Turkish named entity recognition, evaluating the performance of state-of-the-art models across diverse datasets to assess their generalization capabilities. The results, backed by statistical tests, reveal that Transformer-based language models consistently achieve the highest weighted F1 scores, ranging from 80.8% in tweets to 96.1% in news articles [10]. Another study investigates the effectiveness of various word embeddings for the cyber security Named Entity Recognition (NER) task, comparing general-purpose pre-trained word embeddings (both non-contextual and contextual) with task-adapted embeddings fine-tuned on a specific supervised dataset. The findings suggest that, for cyber security NER, fastText outperforms GloVe and BERT in utilizing pre-trained embeddings [11].

Methodology

There are resources that an expert of any field should refer to in order to solve the problems encountered during his/her professional life. The most effective and fastest of these resources is the scientific literature. However, it is time-consuming and difficult to find the most effective solutions to a particular problem within the scientific literature. The Evidence-Based Medicine (EBM) approach, which is specially proposed in the field of Medicine, where this problem is experienced, has been very successful [12]. According to this approach, a standard is provided to summarize the literature accurately and efficiently for the solution of a problem. After the success of the EBM approach in the field of Medicine, Evidence-based software engineering (EBSE), which aims to reach the fastest and most accurate solutions to the problems that software engineering specialists may encounter throughout their professional lives, has been proposed in the literature. This approach is based on the principles of evidence-based medicine, which have been successfully applied for many years. According to this perspective, a software engineer cannot solve all the problems of his/her professional life with his/her graduation knowledge alone. Therefore, it is important to have access to successful and proven methods for similar problems. This access is possible through the scientific publication of proven methods. However, it is not practical for every expert to review all the publications in detail to analyze a specific problem [13]. Therefore, it is inevitable that the literature in a particular field should be effectively summarized and presented in an organized way. Moreover, there is a requirement for this process to be subject to a standardized procedure in each study. In order to address this challenge, the Evidence-Based Medicine (EBM) approach has been proposed [12].

The successful application of the evidence-based medicine model has led to the questioning of the applicability of this method in different fields. Experimental studies have shown that the method gives successful results in different fields [14]. After the first studies that recommended the use of this method in the field of software engineering [15], different studies have been carried out on how to apply the method to software engineering [16].

For this purpose, Kitchenham et al. were the first to propose the use of this method in a literature review in the field of software engineering [15]. However, this study reflects the point of view of a medical doctor. For this reason, the researchers published a different study in which they added the perspective of a software engineer to the method [16].

The proposed method consists of the following six steps[15].

1. Defining an answerable question.
2. Finding results that provide the best evidence.
3. Critically evaluate the evidence.
4. Integrating critical evaluation with software engineering expertise.
5. Evaluating the process
6. Making inferences for evidence-based software engineering.

In this study, the methodology has been developed and applied by taking these steps into account.

Defining Answerable Questions – Research Questions

In the literature searches carried out during our study, we have tried to answer the following research questions.

RQ 1: What are the most successful models in NER studies?

RQ 2: In which domains are there custom NER studies?

Finding the Best Evidence – Search Strategy

The search process was carried out with the idea that using research questions as search terms may be a more accurate approach. The databases used in the search process are IEEE Xplore (http://ieeexplore.ieee.org), ACM Digital Library (www.acm.org/dl), ISI Web of Science (www.isinet.com /products/citation/wos), Science Direct (https://www. sciencedirect.com/), Google Scholar (http://scholar.google. com). The articles found were evaluated according to the inclusion and exclusion criteria.

Critically Evaluating the Evidence – Inclusion and Exclusion Criteria

Within the scope of the study, care was taken to include articles that generally complied with our quality assessment rules. The quality assessment criteria defined for this purpose are described in the next section. In addition, it was sought as another criterion that the datasets obtained with the articles should be shared publicly. As a result of the research, a total of 114 articles were identified in different databases. As a result of the analyses carried out separately by the authors, the necessary selection and elimination processes were carried out. The first steps of this analysis eliminate survey studies and duplicate studies. Then, inclusion and exclusion criteria were applied on the remaining articles. As a result of this process, a total of 38 studies were selected.

Integrating Critical Assessment with Software Engineering Expertise - Quality Assessment Rules - QARs

This step is the selection of the final articles to be included in the study. Five quality assessment questions were used to evaluate the articles. According to these questions, the score of each article for the relevant question was determined in the range of 0 - 1. The quality assessment questions are as follows. We accept papers that get score 3 and above.

QAR1: Is the Sample Source Code for the development of the NER dataset cited?

The answer to this question is scored as 0 or 1.

QAR2: Is there multilingual support?

The answer to this question is scored as 0 for one language, 0.5 for two languages and 1 for more than two languages.

QAR3: Does it support multiple tasks?

The answer to this question is scored as 0 or 1.

QAR4: What is the annual citation average?

This value is calculated as the ratio of the number of citations a publication has received since its year of publication to the number of years it has been in publication, and articles are normalized between 0-1.

QAR5: Are study codes and datasets shared?

The answer to this question is scored as 0 or 1.

Evaluation of the Process - Data extraction strategy and synthesis of data

At this stage of the study, it was evaluated how the selected articles answered the research questions. For this purpose, an information extraction form was created. In this form, the title of the article, the year of publication, the best model and the fields covered were marked. As a result, information on the extent to which the analysed publication contributed to the research questions was collected.

Results

In this section, we discuss the answers to our research questions. We provide a detailed analysis of the answers to our research question in the scientific papers that we included in our literature review. We also provide an overview of the Custom NER field. In this respect, we provide guidance to researchers who want to work in this field. We present detailed analyses of 38 articles selected within the scope of the study. The evaluation results and details of the 38 articles selected according to our Quality Assessment Rules are clearly presented in table 3.

What are the most successful models in studies in the field of NER?

This section aims to answer the question asked in RQ 1: What are the most successful models in NER studies? Four different models stand out among the 38 research articles examined. Additionally, best models could not be determined for some proposed NER models. The information resulting from the analysis is presented in table 1. According to the table, the most used and successful model in the field of NER is BERT. It is also understood that LSTM, Transformers and Word Embedding methods are used predominantly. It is seen that researchers make analyzes with different models. However, it is seen that BERT models produce better results than other models in terms of NER.

Table 1. Best Models

Ref	Best Model
[17] [29] [31] [41] [43]	LSTM
[18] [24] [28] [36]	Tranformers
[19] [20] [21] [22] [23] [25] [26] [27] [32] [45] [46] [47] [48] [51]	BERT
[30] [34] [42]	Word Embedding
[33] [35] [37] [38] [39] [40] [44] [49] [50] [52] [53]	Other / Undetermined

Table 2. Domains

Ref	Domain
[17] [23] [24] [25] [27] [29] [32] [33] [34] [35] [36] [40] [44] [50]	News
[18] [19] [41] [45] [46] [47] [48]	Health
[20] [30] [31]	Biomedical
[21] [51]	AI Scientific Papers
[22]	Social Media
[26] [28] [38] [49] [53]	Wikipedia
[43]	Trafic
[52]	Web Sites
[37] [39] [42]	Mix/ Undetermined

Table 3. Selected Research Articles

NER Name	Year	Num. of Cite. (Year)	QAR1	QAR2	QAR3	QAR4	QAR5	TOTAL
CoNLL-2003[17]	2003	203,6190476	1	1	1	0,380952381	1	4,38
BC5CDR[18]	2016	534,5	1	0	1	1	1	4,00
GENIA[19]	2003	67,52380952	1	0	1	0,126330794	1	3,13
BLUE[20]	2019	151,5	1	0	1	0,28344247	1	3,28
SciERC[21]	2018	74	1	0	1	0,138447147	1	3,14
WNUT 2017[22]	2017	51	1	0	1	0,095416277	1	3,10
ONTONOTES 5.0[23]	2013	31,72727273	1	1	1	0,059358789	1	4,06
CoNLL 2002[24]	2003	230,6190476	1	1	1	0,43146688	1	4,43
ACE 2005 [25]	2006	0,944444444	0	1	1	0,001766968	1	3,00
Few-NERD[26]	2021	42	1	0	1	0,07857811	1	3,08
ACE 2004[27]	2005	3,578947368	0	1	1	0,006695879	1	3,01
IPM NEL[28]	2014	46,3	1	0	1	0,086623012	1	3,09
DWIE[29]	2020	11,25	1	0	1	0,021047708	1	3,02
BROAD TWITTER CORPUS[30]	2016	16	1	0	1	0,029934518	1	3,03
CrossNER[31]	2020	21	1	0	1	0,039289055	1	3,04
BioRED[32]	2022	19	1	0	1	0,03554724	1	3,04
BB (Bacteria Biotope)[33]	2019	8,2	1	0	1	0,015341441	1	3,02
CONLL-2000[34]	2000	47,08333333	1	0	1	0,088088556	1	3,09
GUM [35]	2017	41,14285714	1	0	1	0,076974475	1	3,08
GeoWebNews[36]	2019	8,8	1	0	1	0,016463985	1	3,02
MEDIA[37]	2004	3,35	1	1	1	0,00626754	0	3,01
WikiNEuRal[38]	2021	15	1	1	1	0,028063611	1	4,03
AMALGUM[39]	2020	3,75	1	0	1	0,007015903	1	3,01
BC4CHEMD[40]	2015	33,11111111	1	0	1	0,061947822	1	3,06
I2B2 [41]	2020	1	1	0	1	0,001870907	1	3,00
CORD-r[42]	2023	0	1	0	1	0	1	3,00
FUNSD-r[42]	2023	0	1	0	1	0	1	3,00
FindVehicle[43]	2023	0	1	0	1	0	1	3,00
NAAMAPADAM[44]	2023	2	1	1	0	0,003741815	1	3,00
SHARE/CLEF 2014[45]	2014	8	1	0	1	0,014967259	1	3,01
BC7 NLM-CHEM[46]	2022	6	1	0	1	0,011225444	1	3,01
RARE DISEASES [47]	2021	3,333333333	1	0	1	0,006236358	1	3,01
THYME-2016[48]	2016	33,375	1	0	1	0,062441534	1	3,06
BIOGRAPHICAL[49]	2022	2,5	1	0	1	0,004677268	1	3,00
FLUE [50]	2022	17,5	1	0	1	0,032740879	1	3,03
LPSC [51]	2022	0	1	0	1	0	1	3,00
TASTESET[52]	2022	2	1	0	1	0,003741815	1	3,00
STEM-ECR[53]	2020	8,5	1	0	1	0,015902713	1	3,02

In which domains are there Custom NER studies?

In this section, we focus on answering RQ 2 by trying to determine which domains are used to create the NER dataset. A detailed analysis of the domains covered by the 38 research articles examined within the scope of the research is given in Table 2. As can be seen from here, eight different domains have been identified. Only 3 studies showed that it was not a specific domain but a combination of different domains. Therefore, these articles are assigned to the mix category. If we look at custom domains, it can be seen that the domain with the most NERs is the NEWs category. Health and Wikipedia domains follow. It can be seen that different NER data sets have been created in the biomedical domain. Other domains are Scientific Papers, Social Media, Traffic and Websites.

Discussion

The increase in NER data sets in recent years and attracting the attention of researchers has increased the studies in this field. This study serves as a guide for experts who will conduct research in this field. In this context, a systematic literature analysis of NER studies in different fields has emerged with an evidence-based software engineering approach. Within the scope of the study, 38 research articles were identified. Our study focused on finding answers to two different research questions. As a result of our research, we tried to identify the most successful models used in the NER field. As a result of this research, it was seen that BERT, LSTM, Transformers and Word Embedding methods were used predominantly. It is understood that the most successful and most used model among these models is BERT. The second focus of our study, the question of which domains the NER datasets are created for, is the focus of our second research question. As a result of the analysis, eight different domains were identified. Among these, it can be seen that the domain with the most NER data sets is News. It has been observed that many NER data sets have been created in the health domain, which comes in second place.

Conclusion

Nowadays, when data has reached an incredible size and continues to increase, it is very difficult to analyze data of this size with human power or statistical methods. For this reason, the use of artificial intelligence algorithms in data analysis is common. Many different methods are used to transform unstructured data into a structure that computer systems can understand. NER models are one of these methods. NER models, especially designed specifically for certain areas, have increased their use in recent years. However, new models constantly produced in different domains cause a researcher to spend significant time deciding which model to choose. This study presents an analysis that will reduce research time and provide access to the right resources. With our study, it has been revealed in which domains data sets were created in the field of NER and which models achieved the highest success values with these data sets. As a result of the analysis, some suggestions were revealed for future studies. The most important of these is that not many NER studies have been found, especially in the field of computer science. For this reason, we recommend that NER be produced in the field of computer science in future studies. In addition, researching how Artificial Intelligence robots, which have become widespread recently, will contribute to the creation of automatic NER is one of the interesting topics.

References

[1] I. H. Sarker, "Deep Learning: A Comprehensive" Overview on Techniques, Taxonomy, Applications and Research Directions," SN Comput Sci, vol. 2, no. 6, pp. 1–20, Nov. 2021, doi: 10.1007/S42979-021-00815-1/FIGURES/6.

[2] Z. Nasar, S. W. Jaffry, and M. K. Malik, "Named Entity Recognition and Relation Extraction," ACM Computing Surveys (CSUR), vol. 54, no. 1, Feb. 2021, doi: 10.1145/3445965.

[3] B. Jehangir, S. Radhakrishnan, and R. Agarwal, "A survey on Named Entity Recognition — datasets, tools, and methodologies," Natural Language Processing Journal, vol. 3, p. 100017, Jun. 2023, doi: 10.1016/J.NLP.2023.100017.

[4] J. Li, A. Sun, J. Han, and C. Li, "A Survey on Deep Learning for Named Entity Recognition," IEEE Trans Knowl Data Eng, vol. 34, no. 1, pp. 50–70, Jan. 2022, doi: 10.1109/TKDE.2020.2981314.

[5] H. Shelar, G. Kaur, N. Heda, and P. Agrawal, "Named Entity Recognition Approaches and Their Comparison for Custom NER Model," 2020, doi: 10.1080/0194262X.2020.1759479.

[6] L. Stepanyan, "Automated Custom Named Entity Recognition and Disambiguation," International Journal of Signal Processing, vol. 5, 2020, Accessed: Nov. 12, 2023. [Online]. Available: http://iaras.org/iaras/journals/ijsp

[7] R. Ramachandran and K. Arutchelvan, "Named entity recognition on bio-medical literature documents using hybrid based approach," J Ambient Intell Humaniz Comput, vol. 1, pp. 1–10, Mar. 2021, doi: 10.1007/S12652-021-03078-Z/FIGURES/5.

[8] S. Surana, J. Chekkala, and P. Bihani, "Chatbot based Crime Registration and Crime Awareness System using a custom Named Entity Recognition Model for Extracting Information from Complaints," International Research Journal of Engineering and Technology, 2021, Accessed: Nov. 12, 2023. [Online]. Available: www.irjet.net

[9] G. Veena, V. Kanjirangat, and D. Gupta, "AGRONER: An unsupervised agriculture named entity recognition using weighted distributional semantic model," Expert Syst Appl, vol. 229, p. 120440, Nov. 2023, doi: 10.1016/J.ESWA.2023.120440.

[10] O. Ozcelik and C. Toraman, "Named entity recognition in Turkish: A comparative study with detailed error analysis," 2022, doi: 10.1016/j.ipm.2022.103065.

[11] S. Srivastava, B. Paul, and D. Gupta, "Study of Word Embeddings for Enhanced Cyber Security Named Entity Recognition,"

Procedia Comput Sci, vol. 218, pp. 449–460, Jan. 2023, doi: 10.1016/J.PROCS.2023.01.027.

[12]D. L. Sackett, W. M. C. Rosenberg, J. A. M. Gray, R. B. Haynes, and W. S. Richardson, "Evidence based medicine: what it is and what it isn't. 1996.," Clinical orthopaedics and related research, vol. 455, no. 7023. British Medical Journal Publishing Group, pp. 3–5, 2007. doi: 10.1136/bmj.312.7023.71.

[13]D. Berry, W. T.-I. T. on Software, and undefined 2003, "Comments on" Formal methods application: an empirical tale of software development"," ieeexplore.ieee.org, Accessed: Mar. 10, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1205183/

[14]M. Jorgensen, … T. D.-11th I. I., and undefined 2005, "Teaching evidence-based software engineering to university students," ieeexplore.ieee.org.

[15]B. A. Kitchenham, T. Dybå, and M. Jørgensen, "Evidence-based software engineering," in Proceedings - International Conference on Software Engineering, 2004, pp. 273–281. doi: 10.1109/icse.2004.1317449.

[16]T. Dybå, B. A. Kitchenham, and M. Jorgensen, "Evidence-based software engineering for practitioners," IEEE Softw, vol. 22, no. 1, pp. 58–65, Jan. 2005, doi: 10.1109/MS.2005.6.

[17]E. F. Tjong, K. Sang, and F. De Meulder, "Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition," 2003, Accessed: Nov. 22, 2023. [Online]. Available: http://lcg-www.uia.ac.be/conll2003/ner/

[18]J. Li et al., "BioCreative V CDR task corpus: a resource for chemical disease relation extraction," Database, vol. 2016, p. 68, 2016, doi: 10.1093/database/baw068.

[19]J.-D. Kim, Y. Tateisi, J.-D. Kim, T. Ohta, Y. Tateisi, and J. Tsujii, "GENIA corpus-A semantically annotated corpus for bio-textmining," BIOINFORMATICS, vol. 19, pp. 180–182, 2003, doi: 10.1093/bioinformatics/btg1023.

[20]Y. Peng, S. Yan, and Z. Lu, "Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets", Accessed: Nov. 22, 2023. [Online]. Available: https://biocreative.

[21]Y. Luan, L. He, M. Ostendorf, and H. Hajishirzi, "Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction", Accessed: Nov. 22, 2023. [Online]. Available: http://labs.semanticscholar.

[22]L. Derczynski, E. Nichols, and M. Van Erp, "Results of the WNUT2017 Shared Task on Novel and Emerging Entity Recognition," pp. 140–147, 2017, Accessed: Nov. 22, 2023. [Online]. Available: https://stackexchange.com

[23]R. Weischedel et al., "Ontonotes release 5.0 ldc2013t19," Linguistic Data Consortium, Philadelphia, PA, vol. 23, p. 170, 2013.

[24]E. F. Tjong Kim Sang, "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), 2002. [Online]. Available: https://aclanthology.org/W02-2024

[25]C. Walker, S. Strassel, J. Medero, and K. Maeda, "Ace 2005 multilingual training corpus-linguistic data consortium," URL: https://catalog. ldc. upenn. edu/LDC2006T06, 2005.

[26]N. Ding et al., "Few-NERD:A Few-shot Named Entity Recognition Dataset," in ACL-IJCNLP, 2021.

[27]A. Mitchell, S. Strassel, S. Huang, and R. Zakhary, "Ace 2004 multilingual training corpus," Linguistic Data Consortium, Philadelphia, vol. 1, p. 1, 2005.

[28]L. Derczynski et al., "Analysis of Named Entity Recognition and Linking for Tweets," Oct. 2014, doi: 10.1016/j.ipm.2014.10.006.

[29]"ACE 2004 Multilingual Training Corpus - Linguistic Data Consortium." Accessed: Nov. 22, 2023. [Online]. Available: https://catalog.ldc.upenn.edu/LDC2005T09

[30]L. Derczynski, K. Bontcheva, and I. Roberts, "Broad Twitter Corpus: A Diverse Named Entity Recognition Resource," pp. 1169–1179.

[31]Z. Liu et al., "CrossNER: Evaluating Cross-Domain Named Entity Recognition," 2020.

[32]L. Luo, P.-T. Lai, C.-H. Wei, C. N. Arighi, and Z. Lu, "BioRED: a rich biomedical relation extraction dataset," Brief Bioinform, vol. 2022, no. 5, pp. 1–12, doi: 10.1093/bib/bbac282.

[33]R. Bossy, L. Deléger, E. Chaix, M. Ba, and C. Nédellec, "Bacteria biotope at BioNLP open shared tasks 2019," in Proceedings of the 5th workshop on BioNLP open shared tasks, 2019, pp. 121–131.

[34]E. F. Tjong, K. Sang, and S. Buchholz, "Introduction to the CoNLL-2000 Shared Task: Chunking," pp. 127–132, 2000, Accessed: Nov. 23, 2023. [Online]. Available: www.uia.ac.be/conll2000/chunking/

[35]A. Zeldes, "The GUM corpus: creating multilayer resources in the classroom," Lang Resour Eval, vol. 51, no. 3, pp. 581–612, Sep. 2017, doi: 10.1007/s10579-016-9343-x.

[36]M. Gritta, M. T. Pilehvar, and N. Collier, "A Pragmatic Guide to Geoparsing Evaluation," Oct. 2018.

[37]L. Charnay and ; France-Telecom R&d, "The French MEDIA/EVALDA project: the evaluation of the understanding capability of Spoken Language Dialogue Systems," J. Goulian, no. 2.

[38]S. Tedeschi, V. Maiorca, N. Campolungo, F. Cecconi, and R. Navigli, "WikiNEuRal: Combined neural and knowledge-based silver data creation for multilingual NER," in Findings of the Association for Computational Linguistics: EMNLP 2021, 2021, pp. 2521–2533.

[39]L. Gessler, S. Peng, Y. Liu, Y. Zhu, S. Behzad, and A. Zeldes, "AMALGUM -- A Free, Balanced, Multilayer English Web Corpus," Jun. 2020.

[40]M. Krallinger et al., "The CHEMDNER corpus of chemicals and drugs and its annotation principles," J Cheminform, vol. 7, no. S1, p. S2, Dec. 2015, doi: 10.1186/1758-2946-7-S1-S2.

[41]N. Milosevic, G. Kalappa, H. Dadafarin, M. Azimaee, and G. Nenadic, "MASK: A flexible framework to facilitate de-identification of clinical texts," May 2020.

[42]C. Zhang et al., "Reading Order Matters: Information Extraction from Visually-rich Documents by Token Path Prediction," Oct. 2023.

[43]R. Guan et al., "FindVehicle and VehicleFinder: A NER dataset for natural language-based vehicle retrieval and a keyword-based cross-modal vehicle retrieval system," Apr. 2023.

[44]A. Mhaske et al., "Naamapadam: A Large-Scale Named Entity Annotated Data for Indic Languages".

[45]D. L. Mowery et al., “Task 2: ShARe/CLEF eHealth Evaluation Lab 2014”.

[46]T. Almeida, R. Antunes, J. F. Silva, J. R. Almeida, and S. Matos, “Chemical identification and indexing in PubMed full-text articles using deep learning and heuristics,” Database, vol. 2022, 2022, doi: 10.1093/database/baac047.

[47]H. Dong, V. Suárez-Paniagua, H. Zhang, M. Wang, E. Whitfield, and H. Wu, “Rare Disease Identification from Clinical Notes with Ontologies and Weak Supervision”, Accessed: Nov. 21, 2023. [Online]. Available: https://github.com/hanxiao/bert-as-service

[48]S. Bethard, G. Savova, W.-T. Chen, L. Derczynski, J. Pustejovsky, and M. Verhagen, “SemEval-2016 Task 12: Clinical TempEval”, Accessed: Nov. 21, 2023. [Online]. Available: https://github.com/stylerw/thymedata

[49]T. Ranasinghe, S. Jones, C. Orăsan, and R. Mitkov, “Biographical: A Semi-Supervised Relation Extraction Dataset; Biographical: A Semi-Supervised Relation Extraction Dataset,” p. 10, doi: 10.1145/3477495.3531742.

[50]R. Sanjay Shah et al., “WHEN FLUE MEETS FLANG: Benchmarks and Large Pre-trained Language Model for Financial Domain”.

[51]Y. Zhuang, E. Riloff, K. L. Wagstaff, R. Francis, M. Golombek, and L. Tamppari, “Exploiting Unary Relations with Stacked Learning for Relation Extraction,” pp. 126–137.

[52]A. Wróblewska et al., “TASTEset-RECIPE DATASET AND FOOD ENTITIES RECOGNITION BENCHMARK A PREPRINT”, Accessed: Nov. 21, 2023. [Online]. Available: https://www.allrecipes.com/,

[53]J. D’Souza, A. Hoppe, A. Brack, M. Y. Jaradeh, S. Auer, and R. Ewerth, “The STEM-ECR Dataset: Grounding Scientific Entity References in STEM Scholarly Content to Authoritative Encyclopedic and Lexicographic Sources,” pp. 11–16, 2020, doi: 10.25835/0017546.