Evaluating Joint Named Entity Recognition and Relation Extraction on a Novel Portuguese News Dataset

Derli Aparecida Machado[1], **Lucas Marcondes Pavelski**[2], **Rita Cristina Galarraga Berardi**[1]

[1]Programa de Pós-Graduação em Computação Aplicada – PPGCA
Av. Sete de Setembro, 3165 Rebouças, Curitiba, Brasil.

[2]Mestrado Profissional em Poder Legislativo – MPPL
Câmara dos Deputados – Praça dos Três Poderes, Brasília, Brasil.

derli@alunos.utfpr.edu.br, lucas.pavelski@camara.leg.br,
rita.berardi@utfpr.edu.br

Abstract. *Information Extraction (IE) techniques transform unstructured data into structured knowledge, which is useful for constructing knowledge graphs and downstream tasks. Named Entity Recognition (NER) and Relation Extraction (RE) are fundamental IE tasks for identifying entities and semantic relations in texts. This work presents a Portuguese annotated news dataset for NER and RE across multiple domains. The objective is to make the dataset publicly available and evaluate a state-of-the-art technique (Partition Filter Network) for these tasks. Results show consistent performance in joint extraction and highlight the impact of domains, mainly on RE, which requires contextual understanding.*

Resumo. *Técnicas de Extração de Informação (EI) transformam dados não estruturados em conhecimento estruturado, útil para construção de grafos de conhecimento e tarefas subsequentes. Reconhecimento de Entidades Nomeadas (NER) e Extração de Relações (ER) são tarefas fundamentais de EI para identificar entidades e relações semânticas em textos. Este trabalho apresenta um novo conjunto de dados de notícias em português anotado para NER e ER em múltiplos domínios. O objetivo é disponibilizar a base e avaliar uma técnica do estado da arte (Partition Filter Network) para essas tarefas. Resultados mostram desempenho consistente na extração conjunta e evidenciam o impacto dos domínios, principalmente em RE, que requer entendimento contextualizado.*

1. Introduction

Knowledge Graphs (KGs) are essential structures for organizing, retrieving, and integrating information across a wide range of applications, including search engines, recommendation systems, and content management platforms [Hogan et al. 2021]. By structuring information extracted from different contexts, KGs enable more efficient knowledge retrieval and support data-driven applications that require precise and context-aware information.

One main problem of building KGs is structuring information from text data. Information Extraction (IE) aims to transform unstructured textual content into structured

representations suitable for storage and processing [Jurafsky and Martin 2026]. Among the most common IE tasks are Named Entity Recognition (NER), responsible for identifying and classifying textual entities, and Relation Extraction (RE), which identifies semantic relationships between entities. Together, these tasks enable the extraction and organization of structured information from textual data.

News portals represent a continuous source of information about people, organizations, locations, and events [Opdahl et al. 2023]. Due to their diversity of topics and writing styles, news articles constitute a relevant scenario for evaluating IE techniques. Additionally, news texts frequently contain complex semantic relations and domain-specific vocabulary associated with areas such as politics, sports, economics, and entertainment, which directly influences extraction performance.

Although IE has been extensively studied, much of the literature focuses on improving extraction architectures or constructing new datasets and knowledge resources [Popovski et al. 2020, De et al. 2025, Blouin et al. 2024, Nunes et al. 2024]. In the Portuguese language context, resources for joint Named Entity Recognition and Relation Extraction in the news domain remain limited, particularly for datasets covering multiple domains.

More recently, joint extraction approaches have gained attention as an alternative to traditional pipeline architectures for NER and RE. Instead of treating NER and RE as independent tasks, joint models aim to capture dependencies between entities and relations while reducing error propagation between stages [Li et al. 2022]. Transformer-based architectures have played an important role in the advancement of these approaches, particularly in multilingual and domain-adapted scenarios [Utino and Marcacini 2025].

This work presents a new Portuguese-language dataset constructed from news articles and annotated for joint entity and relation extraction. Furthermore, the performance of a joint extraction architecture based on the Partition Filter Network is evaluated across different news domains, analyzing how domain characteristics influence NER and RE performance. The results demonstrate the feasibility of applying joint extraction techniques to Portuguese news datasets and provide insights into the impact of domain variability on IE tasks. Therefore, the contributions of this work are: a new Portuguese annotated news dataset, and the evaluation of state-of-the-art techniques for joint NER and RE, exploring configurations and performance in different text domains.

This paper is organized as follows. Section 2 discusses related work, while Section 3 provides a brief background on the concepts used in the paper. Section 4 details the proposed dataset, the joint extraction model, and the experiments. Section 5 presents the results, and Section 6 concludes the paper.

2. Related Work

In general, Information Extraction tasks such as NER and RE have been addressed using pipeline-based approaches [Nasar et al. 2021], where each task is performed independently. In this setting, entities are first identified and used as input for relation extraction models. Although simple, pipeline architectures are susceptible to error propagation, since errors in entity recognition directly affect relation extraction performance [Celian et al. 2025].

Recently, joint extraction approaches have gained attention in the literature as a strategy to perform NER and RE tasks while reducing inter-task error propagation and effectively capturing dependencies between entities and relations [Giorgi et al. 2019]. According to recent surveys, advances in this area have been strongly associated with the adoption of Transformer-based architectures [Celian et al. 2025].

One of the main challenges in joint extraction is balancing shared contextual representations while avoiding negative transfer between tasks. To address this issue, the Partition Filter Network (PFN) [Yan et al. 2021] partitions representations into task-specific and shared components, enabling the model to jointly learn NER and RE while preserving task specialization. Motivated by these advantages, this work adopts PFN as the base architecture for jointly modeling NER and RE tasks.

Another challenge in Information Extraction concerns domain adaptation. Although Transformer-based models achieve strong results on general-domain benchmarks, their performance often degrades in specialized or heterogeneous domains. Studies demonstrate this limitation by evaluating state-of-the-art models on technical documents from the oil and gas industry, reporting a substantial performance reduction compared to general-domain datasets [Ticianeli et al. 2024].

Joint extraction in Portuguese is constrained by a scarcity of annotated data, as most existing resources focus solely on NER, such as HAREM [Santos et al. 2006] (general text) and LeNER-Br [Luz de Araujo et al. 2018] (legal domain). For joint NER and RE, datasets are rare and domain-specific, such as Bete [Lucas et al. 2023] (diabetes healthcare). Other Portuguese RE studies are restricted to narrow domains like finance [Reyes et al. 2021] and labor law [Quinta de Castro and Da Silva 2025], or evaluate joint models on English corpora like NYT [Li et al. 2022]. Consequently, a Portuguese news dataset featuring joint NER and RE annotations is still lacking. We address this gap by introducing and evaluating a new news dataset designed specifically for joint extraction.

In this context, the present work evaluates a joint extraction approach based on PFNs on a dataset of Portuguese news from multiple domains, including both highly specialized domains and general news portals. Our analysis investigates how domain heterogeneity impacts the performance of NER and RE, particularly in relation extraction tasks, which require a more specialized contextual understanding.

3. Background

NER is a fundamental task in Information Extraction and plays a central role in Natural Language Processing (NLP) applications. According to [Jurafsky and Martin 2026], the objective of NER is to identify textual mentions of real-world entities and classify them into predefined categories, such as Person (PER), Location (LOC), and Organization (ORG). Because entity identification directly impacts downstream extraction tasks, NER serves as the foundation for more complex applications, including Question Answering, Sentiment Analysis, Knowledge Base Population, and Relation Extraction (RE). As highlighted by [Nasar et al. 2021] and [Al-Moslmi et al. 2020], entity recognition is a critical step in transforming unstructured textual data into structured representations that can be linked to external knowledge sources.

RE complements NER by identifying semantic associations between recognized

entities in a given context. According to [Jurafsky and Martin 2026], the goal of RE is to detect and classify the semantic relationship that holds between pairs of entities. For example, Figure 1 shows a sentence with recognized entities of type `TITLE`, `PERSON`, and `DATETIME`, and an `acted_in` relation connecting the actor (subject) to the movie (object). Formally, the task aims to extract triples of the form (e_i, r, e_j), where e_i and e_j represent entities and r denotes the semantic relation between them. Since RE strongly depends on the quality and contextual interpretation of entity representations, both tasks are closely connected and often benefit from joint modeling approaches.

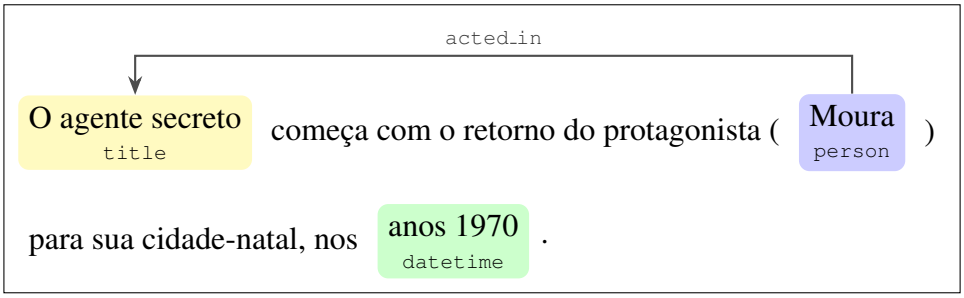


Figure 1. Example of NER and RE annotations in a typical news article sentence.

Bidirectional Encoder Representations from Transformers (BERT), introduced by [Devlin et al. 2018], represent a significant advancement in Natural Language Processing. Unlike traditional language models that process text unidirectionally, BERT is based on the Transformer encoder architecture [Vaswani et al. 2017] and employs bidirectional contextual representations through self-attention mechanisms. This allows the model to simultaneously consider both left and right contexts when encoding words. In Information Extraction tasks, contextual representations are especially important because entity recognition and relation extraction frequently depend on semantic context distributed across the sentence. In this work, we adopt BERTimbau [Souza et al. 2020] as the language representation model due to its adaptation to the Portuguese language.

Class imbalance is another important challenge in IE tasks, particularly in RE datasets where a few majority classes dominate the training process. In our dataset, relations such as `also_known_as` and `is` are substantially more frequent than minority relations such as `coach_of` and `created_by`. To address this issue, [Lin et al. 2020] proposed Focal Loss, an extension of the standard Cross-Entropy Loss designed to reduce the influence of easily classified samples during optimization. The objective of Focal Loss is to introduce a modulating factor that down-weights "easy" examples, forcing the model to focus more strongly on difficult and minority class instances.

The Partition Filter Network (PFN), introduced by [Yan et al. 2021], is a neural architecture designed for the joint extraction of entities and relations. Traditional approaches often treat NER and RE as sequential pipeline tasks, which may lead to error propagation, or as parallel tasks with limited information sharing. In contrast, PFN explicitly models the bidirectional interaction between entity recognition and relation extraction through a partitioning and filtering mechanism.

The core idea of PFN is to divide hidden representations into three components: an entity-specific partition, a relation-specific partition, and a shared partition containing information relevant to both tasks. A partition filter unit then controls the information

flow between these components, allowing the model to preserve task-specific representations while simultaneously learning shared contextual dependencies. This strategy helps mitigate negative transfer, a phenomenon in which irrelevant information from one task negatively affects the performance of another. Given the strong dependency between NER and RE in news texts, PFN was selected in this work because of its ability to jointly model both tasks while reducing error propagation and improving contextual interaction between entities and relations.

The combination of Transformer-based contextual representations, imbalance-aware optimization techniques, and joint extraction architectures provides an effective framework for IE in heterogeneous Portuguese-language news domains. In this context, this work investigates how these techniques contribute to improving the robustness of NER and RE performance across different journalistic domains.

4. Methodology

This section details the proposed Portuguese News Dataset, the model architecture, the techniques applied to enhance performance, and the experimental setup.

4.1. Portuguese News Dataset

One of the primary contributions of this work is a novel dataset for NER and RE tasks in Portuguese. The main source of news is the Globo news website and its corresponding subdomains [1]. The 194 articles comprising the dataset were sampled from multiple domains, such as sports, politics, and entertainment.

The annotation process was performed using the Doccano platform [Nakayama et al. 2018]. Four experts, including the author, labeled the corpus based on well-established definitions. To ensure data quality, the annotation adopted a cross-validation approach: each document was initially annotated by one expert and subsequently reviewed by another. Discrepancies were resolved through consensus to ensure the reliability of the ground truth without relying solely on a single perspective. The final dataset comprises $12,521$ entities across 17 classes and $1,813$ positive semantic relations over 15 categories.

Table 2 summarizes the distribution of the entity types. Categories such as `person` ($2,131$ entities) and `title` ($1,993$ entities) are the most frequent classes, while classes like `fac` and `commodity` represent minority samples. The relation classes are presented in Table 1. There are 15 relation categories in total. `is` (820 instances) and `also_known_as` (382 instances) represent the majority classes, whereas relations like `coach_of` and `formerly_played_for` have only 14 and 34 instances, respectively [2]

4.2. Negative Sampling and Data Augmentation

Articles (up to $8,280$ words, 625.21 ± 740.49 on average) are split into sentences to accommodate the BERT context limit and because most relations are intrasentential. Consequently, the 194 news articles are divided into $3,576$ annotated sentences.

[1] Available at `https://somos.globo.com/marcas/`.

[2] The full description, workflow and annotated dataset are available for download at `https://github.com/derliaparecida/NewsER` under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license.

Table 1. Distribution of relation types (total of 2,379).

Relation Type	Frequency	Relation Type	Frequency
is	820	member_of	57
also_known_as	382	part_of	54
participated_in	255	owner_of	43
located_in	202	created_by	41
acted_in	128	manager_of	36
works_for	121	formerly_played_for	34
is_related_to	109	coach_of	14
plays_for	83		

Table 2. Distribution of entity types (total of 12,521).

Entity Type	Frequency	Entity Type	Frequency
person	2131	datetime	368
title	1993	event	337
org	1580	agricultural_product	321
gpe	1487	sports	254
food	818	policy_or_law	242
product_or_service	763	food_item	242
gov_body	646	commodity	194
loc	511	fac	163
norp	471		

To help distinguish non-existent relations [Zhang et al. 2017], we generated a directed Cartesian product of all entities in each sentence. Unannotated pairs $\{(e_i, e_j)|i \neq j\}$ were labeled as $(e_i, \texttt{no_relation}, e_j)$, yielding $43,615$ negative relations.

We also used data augmentation [Wei and Zou 2019] by replacing head and tail entities in existing triples with candidates from type-constrained gazetteers. This preserves semantic consistency and directionality while introducing lexical variety. Consequently, the training set expanded from $3,576$ to $5,359$ instances (adding $1,862$ synthetic samples), ensuring minority classes reached at least 300 samples each to improve model learning.

4.3. Experimental Setup

We adopt the PFN [Yan et al. 2021] model with the following refinements:

BERTimbau as the sentence encoder: since it is specialized for the Portuguese language [Souza et al. 2020];

Focal Loss: replaces the standard cross-entropy with Focal Loss (using $\gamma = 2.0$), which down-weights easy-to-classify negative samples and focuses the gradient updates on "hard" minority instances [Lin et al. 2020];

Class Weighting: assigns higher penalty weights to underrepresented classes during the loss calculation to ensure more equitable feature learning [King and Zeng 2001].

The dataset was partitioned into training, validation, and testing sets using an 80/10/10 split. The experiments were conducted in four incremental phases to evaluate the impact of each technique:

1. **Baseline:** PFN with BERTimbau and standard settings.
2. **PFN-A:** Integration of data augmentation.
3. **PFN-A-F:** Integration of augmentation and focal loss.
4. **PFN-A-F-W:** Full configuration including augmentation, Focal Loss, and Class Weighting.

During the experimental phase, all configurations were trained for 30 epochs to compare performance. After identifying the best-performing configuration (PFN-A-F-W), a final training run of 100 epochs was conducted to ensure full convergence. The remaining parameters follow the recommendations of the PFN proposal [Yan et al. 2021]: AdamW optimizer, learning rate of 2×10^{-5}, batch size of 20, PFN hidden size of 300, dropout and drop-connect of 0.1.

We evaluate the model performance using Precision, Recall, and F1-score. We report micro-averaged scores to assess the overall per-sample performance. Due to the high class imbalance, we also show macro-averaged scores for entity types, relation types, and news article domains to assess performance across minority classes.

5. Results

Table 3 shows the micro F1-score of the PFN model in different configurations for the NER and RE tasks on the validation dataset using 30 epochs. The baseline PFN model already demonstrates good NER performance (F1-score of 0.820), but RE is considerably harder (F1-score of 0.506). By adding Data Augmentation, Focal Loss, and Class Weighting, there was a mild improvement (F1-score of 0.822 and 0.579 for NER and RE, respectively). Therefore, PFN-A/F/W is considered the best configuration.

Table 3. Micro F1-scores for configurations with data augmentation, focal loss, and class weights, using 30 epochs on validation dataset.

Model	Aug.	Focal	Weights	NER F1	RE F1
Baseline PFN	Off	Off	Off	0.820	0.506
PFN-A	On	Off	Off	0.814	0.544
PFN-A/F	On	On	Off	0.810	0.572
PFN-A/F/W	On	On	On	**0.822**	**0.579**

Table 4 shows the evaluation of the PFN-A/F/W configuration trained for 100 epochs on the test dataset. Overall, the NER prediction achieves over 0.83 for precision, recall, and F1-score. The prediction of RE triples has a recall of 0.577, a precision of 0.636, and an F1-Score of 0.605. Domains like `valor` and `valoroneinterno` show good performance (NER F1-Score above 0.80 and RE F1-Score above 0.60), while `g1` and `oglobo` have lower F1-Scores, mainly on RE (below 0.45). This is primarily due to `valor` and `valoroneinterno` having highly focused articles on themes like economy and finance, while `g1` and `oglobo` contain general factual news. Domains such as `techtudo`, `cbn`, `umsoplaneta`, `gshow`, `casavogue`, `crescer`, `receitas`, `gq` are not included in Table 4 because there are not enough NER or RE test samples.

Macro precision, recall, and F1-scores for NER entities are presented in Table 5. Entity types `agricultural_product`, `gov_body`, `gpe`, `person`, `sports`, and

Table 4. Micro Precision, Recall and F1-scores for each domain on the test dataset using PFN-A/F/W with 100 epochs.

Domain	NER				RE			
	Precision	Recall	F1-Score	#	Precision	Recall	F1-Score	#
valor	0.890	0.8725	0.881	102	0.800	0.546	0.649	22
marieclaire	0.800	0.800	0.800	5	1.000	1.000	1.000	1
valor-investe	0.917	1.000	0.957	11	1.000	0.833	0.909	6
ge	0.822	0.887	0.853	141	0.500	0.467	0.483	30
memoriaglobo	1.000	1.000	1.000	3	1.000	1.000	1.000	1
glamour	0.674	0.756	0.713	41	0.333	0.273	0.300	11
epocanegocios	0.886	0.886	0.886	35	0.750	0.750	0.750	4
pegn	0.739	0.773	0.756	22	0.500	0.250	0.333	4
globorural	0.896	0.885	0.890	80	0.571	1.000	0.727	4
valoroneinterno	0.844	0.862	0.853	138	0.762	0.640	0.696	25
quem	0.800	0.897	0.846	58	0.600	0.286	0.387	21
g1	0.812	0.876	0.843	217	0.563	0.360	0.439	25
extra	0.833	0.862	0.848	58	0.714	0.833	0.769	6
oglobo	0.768	0.808	0.787	229	0.353	0.353	0.353	17
monet	0.900	0.900	0.900	10	0.667	0.400	0.500	5
All	0.834	0.858	0.846	1214	0.636	0.577	0.605	182

`policy_or_law` show good performance (F1-Score above 0.88), possibly because these are associated with proper nouns. However, entities like `datetime`, `event`, and `food` have relatively poor results (F1-Score below 0.7), since there are many more ways to express these entity types.

Likewise, macro precision, recall, and F1-score test values for RE using PFN-A/F/W with 100 epochs are presented in Table 6. The best performing relation type extractions are `also_known_as` and `is` (F1-Score above 0.67). These relations possibly connect semantically similar terms, which favors BERT inference. Relations like `located_in` and `participated_in` have relatively poor performance (F1-Score below 0.47), which could be explained by the fact that these require understanding of a broader context, like information outside a single sentence. For example, in the sentence of Figure 2, the model must identify `Libertadores` as an event where `Danilo` plays for the team `Bahia` to extract the triple (`Danilo`, `participated_in`, `Libertadores`). Such cases are considerably more challenging than relations that are commonly associated with explicit lexical patterns.

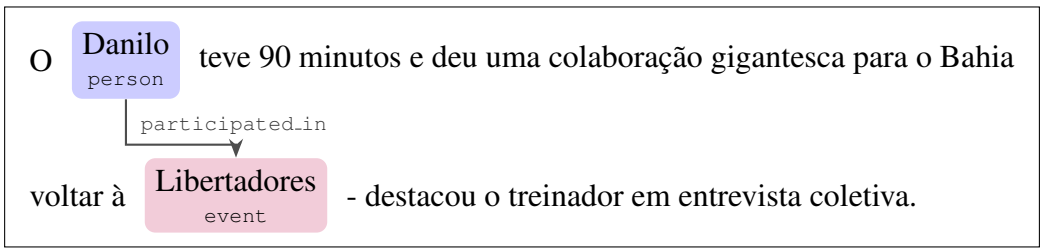


Figure 2. Example of sentence with `participated_in` relation.

Table 5. Macro NER Precision, Recall and F1-scores the test dataset using PFN-A/F/W with 100 epochs.

Entity type	Precision	Recall	F1-Score	#
agricultural_product	0.909	0.909	0.909	33.0
commodity	0.778	0.824	0.800	17.0
datetime	0.643	0.600	0.621	30.0
event	0.639	0.677	0.657	34.0
fac	0.733	0.786	0.759	14.0
food	0.673	0.636	0.654	55.0
food_item	0.938	0.652	0.769	23.0
gov_body	0.902	0.902	0.902	51.0
gpe	0.925	0.948	0.936	155.0
loc	0.764	0.793	0.778	53.0
norp	0.853	0.892	0.872	65.0
org	0.857	0.881	0.869	143.0
person	0.843	0.921	0.881	216.0
policy_or_law	0.909	1.000	0.952	20.0
product_or_service	0.693	0.722	0.708	72.0
sports	0.857	0.923	0.889	26.0
title	0.863	0.879	0.871	207.0
All (average)	0.834	0.858	0.845	1214.0

Overall, the results demonstrated that PFN can be used for NER and RE jointly prediction on Portuguese news. More precisely, NER performance is mostly good, even when there are many entities types. The performance on RE is possibly impacted by the number of classes, domain variability and the low amount of data in some relations.

6. Conclusion

This paper proposes a new annotated dataset of 194 multi-domain news articles in the Portuguese language for NER and RE tasks. It also proposes an enhanced PFN model training procedure to deal with class imbalance and low data samples. The results in terms of micro and macro F1-Scores are competitive for both NER (around 0.8) and RE (around 0.6) tasks.

When compared to the original PFN study [Yan et al. 2021], it is essential to distinguish the benchmarks used. While the original model achieved RE F1-scores above 0.90 on datasets like NYT and WebNLG, those results are obtained under partial match criteria and with a much lower number of entity types and relation types. For a more direct comparison, the datasets ACE04 and ACE05 employ exact-match evaluation and consist of news texts, reporting F1-scores from 0.83 to 0.85 for NER and 0.51 to 0.55 for RE. In this context, our baseline results on the proposed dataset are competitive, especially considering the complexity of our corpus. The dataset has 17 entity categories and 15 semantic relation types, compared to the 7 entities and 6 relations found in the ACE benchmarks.

The proposed dataset and training scheme provide many possibilities for future work, such as using multi-domain approaches for NER and RE tasks [Bassignana et al. 2023]. Future studies could also compare the PFN model with

Table 6. Macro RE Precision, Recall and F1-scores the test dataset using PFN-A/F/W with 100 epochs.

Relation type	Precision	Recall	F1-Score	#
acted_in	0.667	0.500	0.571	4
also_known_as	0.724	0.840	0.778	25
coach_of	0.000	0.000	0.000	1
created_by	0.500	0.500	0.500	2
is	0.726	0.627	0.673	59
is_related_to	0.700	0.538	0.609	13
located_in	0.478	0.458	0.468	24
member_of	0.500	0.667	0.571	6
owner_of	0.500	0.500	0.500	2
part_of	0.000	0.000	0.000	1
participated_in	0.526	0.323	0.400	31
plays_for	0.444	0.667	0.533	6
works_for	0.875	0.875	0.875	8
All (average)	0.633	0.577	0.597	182

advanced semi-supervised learning techniques using large language models to overcome the low number of samples [Wang et al. 2026]. Also, the dataset can be extended to encompass other tasks related to the creation of Knowledge Graphs, such as entity linking and Knowledge Graph augmentation [Jurafsky and Martin 2026].

References

Al-Moslmi, T., Gallofre Ocana, M., L. Opdahl, A., and Veres, C. (2020). Named Entity Extraction for Knowledge Graphs: A Literature Overview. *IEEE Access*, 8:32862–32881.

Bassignana, E., Ginter, F., Pyysalo, S., van der Goot, R., and Plank, B. (2023). Multi-CrossRE A Multi-Lingual Multi-Domain Dataset for Relation Extraction. In Alumäe, T. and Fishel, M., editors, *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 80–85, Tórshavn, Faroe Islands. University of Tartu Library.

Blouin, B., Armand, C., and Henriot, C. (2024). A dataset for named entity recognition and entity linking in Chinese historical newspapers. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 385–394, Torino, Italia. ELRA and ICCL.

Celian, R., Gandon, Fabien, Catherine, F., Franck, M., and Hanna, A. A. (2025). A systematic review of relation extraction task since the emergence of Transformers. arXiv:2511.03610 [cs.CL].

De, S., Sanyal, D. K., and Mukherjee, I. (2025). Fine-tuned encoder models with data augmentation beat ChatGPT in agricultural named entity recognition and relation extraction. *Expert Systems with Applications*, 277:127126.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Version Number: 2.

Giorgi, J., Wang, X., Sahar, N., Shin, W. Y., Bader, G. D., and Wang, B. (2019). End-to-end Named Entity Recognition and Relation Extraction using Pre-trained Language Models. arXiv:1912.13415 [cs].

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., et al. (2021). Knowledge graphs. *ACM Computing Surveys (Csur)*, 54(4):1–37.

Jurafsky, D. and Martin, J. H. (2026). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. 3rd edition.

King, G. and Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, 9(2):137–163.

Li, Z., Ma, H., Lv, Y., and Shen, H. (2022). Joint Extraction of Entities and Relations in the News Domain. In Chen, W., Yao, L., Cai, T., Pan, S., Shen, T., and Li, X., editors, *Advanced Data Mining and Applications*, volume 13725, pages 81–92. Springer Nature Switzerland, Cham. Series Title: Lecture Notes in Computer Science.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2020). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327.

Lucas, P. et al. (2023). Bete: A brazilian portuguese dataset for named entity recognition and relation extraction in the diabetes healthcare domain. In *Proceedings of the 12th Brazilian Conference on Intelligent Systems (BRACIS 2023)*, pages 256–267. Springer / Sociedade Brasileira de Computação (SBC).

Luz de Araujo, P. H., de Campos, T. E., de Oliveira, R. R. R., Stauffer, M., Couto, S., and Bermejo, P. (2018). LeNER-Br: a dataset for named entity recognition in brazilian legal text. In *Computational Processing of the Portuguese Language (PROPOR)*, pages 313–321. Springer.

Nakayama, H., Kubo, T., Kamura, J., Taniguchi, Y., and Liang, X. (2018). doccano: Text Annotation Tool for Human.

Nasar, Z., Jaffry, S. W., and Malik, M. K. (2021). Named Entity Recognition and Relation Extraction: State-of-the-Art. *ACM Comput. Surv.*, 54(1):20:1–20:39.

Nunes, S., Mario Jorge, A., Amorim, E., Sousa, H., Leal, A., Silvano, P. M., Cantante, I., and Campos, R. (2024). Text2Story lusa: A dataset for narrative analysis in European Portuguese news articles. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15773–15782, Torino, Italia. ELRA and ICCL.

Opdahl, A. L., Al-Moslmi, T., Dang-Nguyen, D.-T., Gallofré Ocaña, M., Tessem, B., and Veres, C. (2023). Semantic Knowledge Graphs for the News: A Review. *ACM Computing Surveys*, 55(7):1–38.

Popovski, G., Seljak, B. K., and Eftimov, T. (2020). A Survey of Named-Entity Recognition Methods for Food Information Extraction. *IEEE Access*, 8:31586–31594.

Quinta de Castro, P. V. and Da Silva, N. F. F. (2025). Labor Lex: A New Portuguese Corpus and Pipeline for Information Extraction in Brazilian Legal Texts. In Aletras, N., Chalkidis, I., Barrett, L., Goanță, C., Preoțiuc-Pietro, D., and Spanakis, G., editors, *Proceedings of the Natural Legal Language Processing Workshop 2025*, pages 184–206, Suzhou, China. Association for Computational Linguistics.

Reyes, D., Trajano, D., Manssour, I., Vieira, R., and Bordini, R. (2021). Entity Relation Extraction from News Articles in Portuguese for Competitive Intelligence Based on BERT. In *Anais da X Brazilian Conference on Intelligent Systems*, Porto Alegre, RS, Brasil. SBC.

Santos, D., Cardoso, N., Seco, N., Vilela, R., Rocha, P., Afonso, S., and Barreiro, A. (2006). HAREM: An advanced NER evaluation contest for Portuguese. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, pages 1986–1989, Genoa, Italy. European Language Resources Association (ELRA).

Souza, F., Nogueira, R., and Lotufo, R. (2020). BERTimbau: pretrained BERT models for Brazilian Portuguese. In *9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear)*.

Ticianeli, G. H. G., Candido Junior, A., Guilherme, I. R., Penteado, B. E., Perrout, S. R., Morelli, L. H., Paiola, P. H., and Garcia, G. L. (2024). Comparison of Relation Extraction Models for Generating Knowledge Graphs in the Oil Drilling Domain. In *Anais do XXI Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2024)*, pages 707–718, Brasil. Sociedade Brasileira de Computação - SBC.

Utino, M. Y. R. and Marcacini, R. M. (2025). BERT vs. LLM2Vec: A Comparative Study of Embedding Models for Semantic Information Retrieval. In *Anais do XXII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2025)*, pages 427–438, Brasil. Sociedade Brasileira de Computação - SBC.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is All you Need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Wang, X., Shen, H., Qi, H., and Dai, Y. (2026). Enhancing Low-Resource Joint Entity and Relation Extraction Using Large Language Models within Semi-Supervised Learning. *ACM Transactions on Knowledge Discovery from Data*, 20(3):1–25.

Wei, J. and Zou, K. (2019). EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. 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 6381–6387, Hong Kong, China. Association for Computational Linguistics.

Yan, Z., Zhang, C., Fu, J., Zhang, Q., and Wei, Z. (2021). A Partition Filter Network for Joint Entity and Relation Extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 185–197, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Zhang, Y., Zhong, V., Chen, D., Angeli, G., and Manning, C. D. (2017). Position-aware Attention and Supervised Data Improve Slot Filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 35–45, Copenhagen, Denmark. Association for Computational Linguistics.