

CLEF HIPE-2026: Evaluating Accurate and Efficient Person–Place Relation Extraction from Multilingual Historical Texts^{*}

Juri Opitz¹, Corina Raclé¹, Emanuela Boros³, Andrianos Michail¹,
Matteo Romanello², Maud Ehrmann³, and Simon Clematide¹

¹ University of Zurich, Switzerland

² Swiss Art Research Infrastructure (SARI), University of Zurich

³ École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

hipe-2026@googlegroups.com

<https://hipe-eval.github.io/HIPE-2026>

Abstract. HIPE-2026 is a CLEF evaluation lab dedicated to person–place relation extraction from noisy, multilingual historical texts. Building on the HIPE-2020 and HIPE-2022 campaigns, it extends the series toward semantic relation extraction by targeting the task of identifying person–place associations in multiple languages and time periods. Systems are asked to classify relations of two types—**at** (“Has the person ever been at this place?”) and **isAt** (“Is the person located at this place around publication time?”)—requiring reasoning over temporal and geographical cues. The lab introduces a three-fold evaluation profile that jointly assesses accuracy, computational efficiency, and domain generalization. By linking relation extraction to large-scale historical data processing, HIPE-2026 aims to support downstream applications in knowledge-graph construction, historical biography reconstruction, and spatial analysis in digital humanities.

Keywords: Relation Extraction · Multilingual NLP · Digital Humanities · Historical Texts · Shared Task

1 Introduction

Historical documents bear a wealth of information that can help reconstructing past events, biographies, and the development of social networks. However, the digitization of historical documents typically results in data that are noisy, multilingual, and weakly structured. Therefore, robust approaches to mining historical data are necessary to support a wide range of information needs among scholars in the social sciences and history [10,21,16,13,4,25,31].

HIPE-2026 is a CLEF Evaluation Lab dedicated to the extraction of person–place relations in multilingual historical documents. A person–place relation corresponds to a semantic link between an individual and a location as evidenced

^{*} The version of record is available at https://doi.org/10.1007/978-3-030-99739-7_44.

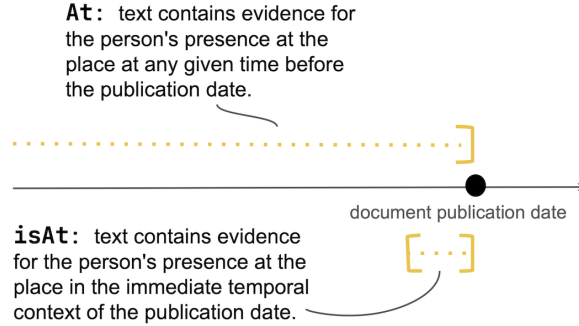


Fig. 1. Temporal scope of **at** and **isAt** relations relative to publication time.

in a document. Such relations may indicate where a person is said to be at a given moment, where they lived or worked, or places connected to notable moments in their life (e.g., birthplaces, residences, visits, travel destinations). Together, these relations can help answer the question of *Who was where when?*, and support the reconstruction of individuals' geographical and temporal trajectories.

These implicit or explicit, spatio-temporal relations cannot be detected through simple document co-occurrence of entity mentions. Rather, it requires temporal reasoning, geographical inference, and interpretation of noisy historical texts—often with sparse or indirect contextual cues—to detect and qualify person–place relations with appropriate degrees of certainty.

The objective of HIPE-2026 is to advance the automatic detection of such relations, enabling the reconstruction of individuals' movements in space and time and the tracing of life trajectories in support of digital humanities scholarship. The task is designed to be approachable by both generative AI systems (LLMs) and more traditional classification models. HIPE-2026 builds on HIPE-2020 and HIPE-2022, which targeted named-entity recognition and linking in historical corpora [5,6] and advances the series toward relation extraction (RE).

2 Task Description

Participants are tasked with determining the relationship of each person–place pair mentioned in a historical document. Each pair consists of two entities, one person and one location, each of which appears in the text through one or more mentions. For each pair, systems must assess whether the text provides evidence of the person's presence at the location, taking into account the document's temporal horizon relative to its publication date (see Figure 1 for an illustration).

For each document, systems must classify each candidate pair according to the following relation types:

- at** The text provides evidence that the person was present at the place at any point in time *prior to the publication date*. This relation is right-bounded by the publication date but may hold at any time in the past. It is labeled as:

- **true**: explicit evidence supports the relation;
- **probable**: the relation can be inferred from contextual clues and is a likely assumption;
- **false**: no evidence, or contradictory evidence, is present.

isAt captures whether the text provides implicit or explicit evidence that the person was at the location in the *immediate temporal context of the publication date*. The decision is binary (+/-): + means that there is implicit or explicit evidence the person was at the location shortly before the publication date, while - indicates the absence of such evidence.

Relation between at and isAt. The **isAt** relation can be viewed as a temporal refinement of **at**: it captures whether there is evidence (implicit or explicit) for the **at** relation in the closer temporal scope of the article’s date. Conceptually, assigning **isAt**=+ presupposes that an **at** relation is true, or at least, probable; Predictions where **at**=**false** and **isAt**=+ are therefore epistemically inconsistent, but practically permitted. In sum, if **at** is **true** or **probable**, **isAt** would further specify whether the person’s presence falls within the document’s temporal horizon.

Abductive interpretation and evidential reasoning. The distinction between explicit, probable, and absent evidence is grounded in an abductive view of language understanding. Hobbs’ framework of “Interpretation as Abduction” [11] argues that discourse interpretation consists in identifying the minimal set of assumptions which, together with background knowledge, explain why an utterance would be a rational and coherent thing to say. Under this view, meaning is not limited to what is explicitly stated, but includes what must be assumed to provide a coherent explanation of the discourse. Applied to person-place relations in historical texts, this perspective motivates the **probable** label: relations may be supported by indirect cues such as event participation, institutional roles, or narrative coherence, even when no explicit locative statement is present. Conversely, the **false** label reflects cases where no such abductive explanation is warranted based on the text alone. The **isAt** refinement further constrains abductive inference temporally, requiring that assumptions about presence remain compatible with the document’s publication horizon.

In addition, we allow systems to optionally provide free-text explanations or indicate other relevant background knowledge that informed their decisions. Evaluating such explanations is not a primary objective of the task, but participants are encouraged to analyze and report pieces of information that could help us better understand prediction rationales.

Example-based Illustration of the Task

To illustrate the task, we present an excerpt from an article of the American newspaper *Tabor City Tribune* on June 6, 1960. The article reports on a military scout meeting and involves multiple person-place candidate relation pairs, as shown in Table 1.

Loris Scouts Win Camporee Top Honor WINNERS — Members of Boy Scout Troop 843. »»ho won top honors at an Horry IMstrict Scout Camporee at Clear Pond March 25-27 are shown above hiking down a road. The lads won after engaging in compassing, rope throwing, Morse code, hiking with compass and individual cooking. The troop is sponsored by the First Baptist Church. Boy Scout Troop 843 of Loris won top honors at the Horry District Scout Camporee March 25-27 at Clear Pond between Conway and Myrtle Beach. Competing against some of the finest units in the county, the boys won top honors after participating in Compassing, rope throwing, Morse Code, hiking with compass and individual cooking. Friday night the troops! competing joined in a camp fire meeting and heard an address by Col. Gruenwald, commanding officer of the Myrtle Beach Air Force Base. The Loris troop, sponsored by the First Baptist church, was the only troop to have all its Scout leaders present: Francis Ragan, George Rent/ and George Lav.

A clear case is the relation between the person entity *Col. Gruenwald* and the place entity *Myrtle Beach Air Force Base*. The article explicitly identifies the colonel as the “commanding officer of the” air force base, thereby supporting a true label for the **at** relation, and a positive one for **isAt**. In contrast, although the place *Myrtle Beach* is also mentioned, the text does not state that Col. Gruenwald was physically present in the city; his connection is institutional rather than locational. This motivates the probable label for the **at** relation and a negative one for **isAt**, highlighting the distinction between affiliation and physical presence.

The relation between Col. Gruenwald and Clear Pond (as well as the Horry District Scout Camporee) provides another instructive example. The article reports that he delivered an address at a campfire meeting during the camporee, which took place at Clear Pond. This constitutes sufficient evidence to annotate both locations with a TRUE **at** relation and a positive **isAt** relation, even though his presence is inferred indirectly through the event.

Several negative annotations demonstrate the importance of avoiding inference beyond the text. Conway and Loris are mentioned as geographic reference points, but there is no indication that Col. Gruenwald visited either place. Consequently, these relations are marked as false, despite their proximity to the actual event location.

Table 1. Annotated person-place candidate relations for the example article.

Person	Place	at	isAt
Col. Gruenwald	Myrtle Beach Air Force Base	true	+
Col. Gruenwald	Clear Pond	true	+
Col. Gruenwald	Myrtle Beach	probable	−
Col. Gruenwald	Loris	false	−
Col. Gruenwald	Conway	false	−
Col. Gruenwald	Horry District Scout Camporee	true	+
Francis Ragan	Myrtle Beach Air Force Base	false	−
Francis Ragan	Conway	false	−
Francis Ragan	Myrtle Beach	false	−
George Lav	Myrtle Beach Air Force Base	false	−
George Lav	Conway	false	−
George Rent	Myrtle Beach Air Force Base	false	−
George Rent	Myrtle Beach	false	−
George Rent	Loris	true	−
George Rent	Conway	false	−
George Rent	Horry District Scout Camporee	true	+

3 Data

HIPE-2026 data consists of two sets of data:

Test Set A: historical newspaper articles in French, German, English, and Luxembourgish spanning roughly 200 years (19th–20th centuries), drawn from the HIPE-2022 data.

Surprise Test Set B: French literary texts (16th–18th century), used to assess domain generalization; only the **at** relation is evaluated.

Pilot Study. A total of 119 person–place pairs were annotated by three independent annotators, drawn from the English and French development sets of HIPE-2022. The inter-annotator agreement (Cohen’s kappa) ranged from 0.7 to 0.9 for the **at** relation, and from 0.4 to 0.9 for **isAt**, indicating moderate to high consistency. Large language models also showed promising alignment with human judgments: GPT-4o reached up to 0.8 agreement with the gold standard for **at**, while **isAt** results were lower and more variable (0.2–0.7). The study further highlighted the high inference cost of current models and the need for scalable methods that can handle the multiplicative growth of candidate entity pairs in historical documents.

4 Evaluation

We consider three evaluation profiles.

The **accuracy profile** rewards high-performing systems and encourages the exploration of frontier models, advanced prompting strategies, or agent-based approaches. System responses are evaluated per test set using *macro-averaged Recall* (also known as *balanced accuracy*). Macro Recall is defined as follows:

1. For each label ℓ in the label set L , compute its recall:

$$\text{Recall}(\ell) = \frac{\text{\#examples with label } \ell \text{ correctly predicted}}{\text{\#examples whose gold label is } \ell}.$$

2. The final score is the arithmetic mean of the per-label recalls:

$$\text{MacroRecall} = \frac{1}{|L|} \sum_{\ell \in L} \text{Recall}(\ell).$$

This metric is principled and interpretable, ensuring that all labels contribute equally regardless of class imbalance [15,22]. For Test Set A, which includes three labels for **at** and two labels for **isAt**, macro Recall is computed separately for each relation and then averaged to obtain the final system ranking.

The **accuracy-efficiency profile** promotes lightweight, scalable methods, including smaller LLMs or task-specific classifiers. This profile balances predictive performance with efficiency and resource usage. This is motivated by the increasing cost of running very large models and, in our case, by the scale of digitized material together with the quadratic nature of the task (person–location

pairs). Participant teams will be surveyed regarding parameter count and model size; together with accuracy, these factors are integrated into a robust ranking metric that computes a balanced score reflecting system efficiency.

Finally, the **generalization profile** tests system accuracy on the Surprise Test Set B, based on the **at**-relation only and using *macro-Recall*.

Annotated data, baselines and scoring tools are released via GitHub⁴ under a CC-BY 4.0 license, and participation guidelines are published on Zenodo⁵.

5 Related Work

Relation Extraction (RE) Benchmarks and Their Limitations. Early Open Information Extraction (Open IE) approaches extracted unrestricted relations without predefined schemas [9]. Subsequent benchmarks introduced controlled settings with fixed relation inventories, notably TACRED [30] for sentence-level and DocRED [29] for document-level RE. These English-only datasets, covering several dozen relation types, have advanced the field but suffer from annotation incompleteness. Re-annotation efforts such as Re-TACRED [2,23] and Re-DocRED [26] reduce false negatives, yet the benchmarks remain confined to clean, modern English. In contrast, HIPE-2026 focuses on a single relation type—person—place—across multiple languages, time periods, and OCR-derived historical sources that involve orthographic variation and temporal reasoning. Previous HIPE tasks (2020, 2022) addressed multilingual named entity recognition and linking in historical newspapers [7,8] but not relation extraction. A recent survey highlights progress in multilingual RE through cross-lingual transfer and annotation projection while underscoring the lack of benchmarks addressing multilinguality, noise, and domain shift [1]. HIPE-2026 fills this gap with a historically and linguistically diverse evaluation setting.

Relation Extraction for Biographical and Historical Domains. Biographical RE extends general RE toward structured knowledge bases. The *Biographical* dataset aligns Wikipedia with Pantheon and Wikidata via distant supervision to derive ten relations [17]. OpenIE has been used to extract RDF triples from Wikipedia biographies [24]. The *Guided Distant Supervision (GDS)* variant adapts this approach to German, denoises labels through external constraints, and explores cross-lingual transfer [18]. Although these studies demonstrate the potential of distant supervision for domain-specific RE, historical data remain largely unexplored, a gap that HIPE-2026 aims to close.

Multilingual, Noisy, and Domain-Shift Relation Extraction. Multilingual RE methods rely on cross-lingual transfer, annotation projection, or zero-shot learning (see [1,31]). However, few benchmarks incorporate noise robustness, domain shift, or low-resource historical languages, with emerging datasets addressing

⁴ <https://github.com/hipe-eval/HIPE-2026-data>

⁵ <https://zenodo.org/records/17800136>

explicitly historical materials [28,20,19]. This motivates shared tasks like HIPE-2026, which emphasize relation mining in noisy, multilingual, historical contexts.

Denoising Techniques. Distant supervision enables large-scale RE but introduces substantial noise. Recent denoising methods include *hierarchical contrastive learning* (HiCLRE) for filtering noisy instances [12] and *NLI*-based validation (DSRE + NLI) [32]. These techniques are relevant for OCR-derived noisy data.

Efficiency in Relation Extraction. Beyond accuracy, current evaluations increasingly emphasize computational efficiency. Shared tasks such as *SustainNLP 2020* measured energy consumption [27], and *EfficientQA* enforced memory limits [14]. Efficiency-focused modeling continues to gain traction [3]. By combining accuracy and efficiency-focused evaluation in a historical multilingual RE task, HIPE-2026 establishes a new benchmark for robust and sustainable NLP.

6 Conclusion

HIPE-2026 extends the HIPE evaluation series into new research directions. It promotes the development of robust methods for extracting person–place relations from multilingual historical text, enabling applications in temporal and spatial analysis across the humanities and social sciences, as well as in the analysis and interpretation of literary texts.

To formalize and isolate this challenge, the task is framed as a relation classification problem conditioned on a person–place pair within its document context. This formulation also facilitates the use of generative AI systems, which are expected to perform well on such contextual reasoning tasks even when applied to historical, digitized texts that diverge from clean modern conditions.

Acknowledgements. HIPE-2026 is funded through the project *Impresso – Media Monitoring of the Past II. Beyond Borders: Connecting Historical Newspapers and Radio*; SNSF 213585 and Luxembourg National Research Fund 17498891.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Ali, M., Speck, R., Zahera, H.M., Saleem, M., Moussallem, D., Ngonga Ngomo, A.C.: Multilingual relation extraction: A survey. *IEEE Access* **13**, 151907–151933 (2025)
2. Alt, C., Gabryszak, A., Hennig, L.: TACRED revisited: A thorough evaluation of the TACRED relation extraction task. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J. (eds.) *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. pp. 1558–1569. Association for Computational Linguistics, Online (Jul 2020). <https://doi.org/10.18653/v1/2020.acl-main.142>, <https://aclanthology.org/2020.acl-main.142/>

3. Boylan, J., Hokamp, C., Ghalandari, D.G.: GLiREL - generalist model for zero-shot relation extraction. In: Chiruzzo, L., Ritter, A., Wang, L. (eds.) *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. pp. 8230–8245. Association for Computational Linguistics, Albuquerque, New Mexico (Apr 2025). <https://doi.org/10.18653/v1/2025.naacl-long.418>, <https://aclanthology.org/2025.naacl-long.418/>
4. Cardoso, S.D., Da Silveira, M., Pruski, C.: Construction and exploitation of an historical knowledge graph to deal with the evolution of ontologies. *Knowledge-Based Systems* **194**, 105508 (2020)
5. Ehrmann, M., Romanello, M., Bircher, S., Clematide, S.: Introducing the clef 2020 hipe shared task: Named entity recognition and linking on historical newspapers. In: *European Conference on Information Retrieval*. pp. 524–532. Springer (2020)
6. Ehrmann, M., Romanello, M., Doucet, A., Clematide, S.: Introducing the hipe 2022 shared task: named entity recognition and linking in multilingual historical documents. In: *European Conference on Information Retrieval*. pp. 347–354. Springer (2022)
7. Ehrmann, M., Romanello, M., Flückiger, A., Clematide, S.: Overview of CLEF HIPE 2020: Named Entity Recognition and Linking on Historical Newspapers. In: Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Joho, H., Lioma, C., Eickhoff, C., Névél, A., Cappellato, L., Ferro, N. (eds.) *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. pp. 288–310. *Lecture Notes in Computer Science*, Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-58219-7_21
8. Ehrmann, M., Romanello, M., Najem-Meyer, S., Doucet, A., Clematide, S.: Overview of HIPE-2022: Named Entity Recognition and Linking in Multilingual Historical Documents. In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 13th International Conference of the CLEF Association, CLEF 2022, Bologna, Italy, September 5–8, 2022, Proceedings*. pp. 423–446. Springer-Verlag, Berlin, Heidelberg (Sep 2022), https://doi.org/10.1007/978-3-031-13643-6_26
9. Etzioni, O., Banko, M., Soderland, S., Weld, D.S.: Open information extraction from the web. *Commun. ACM* **51**(12), 68–74 (Dec 2008), <https://doi.org/10.1145/1409360.1409378>
10. Fokkens, A., Ter Braake, S., Ockeloen, N., Vossen, P., Legêne, S., Schreiber, G., et al.: Biographynet: Methodological issues when nlp supports historical research. In: *LREC*. pp. 3728–3735 (2014)
11. Hobbs, J.R., Stickel, M.E., Appelt, D.E., Martin, P.: Interpretation as abduction. *Artificial intelligence* **63**(1-2), 69–142 (1993)
12. Li, D., Zhang, T., Hu, N., Wang, C., He, X.: HiCLRE: A hierarchical contrastive learning framework for distantly supervised relation extraction. In: Muresan, S., Nakov, P., Villavicencio, A. (eds.) *Findings of the Association for Computational Linguistics: ACL 2022*. pp. 2567–2578. Association for Computational Linguistics, Dublin, Ireland (May 2022). <https://doi.org/10.18653/v1/2022.findings-acl.202>, <https://aclanthology.org/2022.findings-acl.202/>
13. Lucchini, L., Sinatra, R., Emery, C., Panzarasa, P., Servedio, V.D.P., Riccaboni, M., Cattuto, C.: Following the footsteps of giants: Modeling the mobility of historically notable individuals. *EPJ Data Science* **8**(7) (2019). <https://doi.org/10.1140/epjds/s13688-019-0215-7>

14. Min, S., Boyd-Graber, J., Alberti, C., Chen, D., Choi, E., Collins, M., Guu, K., Hajishirzi, H., Lee, K., Palomaki, J., Raffel, C., Roberts, A., Kwiatkowski, T., et al.: Neurips 2020 efficientqa competition: Systems, analyses and lessons learned. In: Escalante, H.J., Hofmann, K. (eds.) *Proceedings of the NeurIPS 2020 Competition and Demonstration Track*. *Proceedings of Machine Learning Research*, vol. 133, pp. 86–111. PMLR (2021), <https://proceedings.mlr.press/v133/min21a.html>, also arXiv:2101.00133
15. Opitz, J.: A closer look at classification evaluation metrics and a critical reflection of common evaluation practice. *Transactions of the Association for Computational Linguistics* **12**, 820–836 (06 2024). https://doi.org/10.1162/tac1_a_00675, https://doi.org/10.1162/tac1_a_00675
16. Opitz, J., Born, L., Nastase, V., Pultar, Y.: Automatic reconstruction of emperor itineraries from the *regesta imperii*. In: *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*. pp. 39–44 (2019)
17. Plum, A., Ranasinghe, T., Jones, S., Orăsan, C., Mitkov, R.: Biographical: A semi-supervised relation extraction dataset. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 3121–3130. ACM (2022). <https://doi.org/10.1145/3477495.3531742>, <https://dl.acm.org/doi/10.1145/3477495.3531742>
18. Plum, A., Ranasinghe, T., Purschke, C.: Guided distant supervision for multilingual relation extraction data: Adapting to a new language. In: Calzolari, N., Kan, M., Hoste, V., Lenci, A., Sakti, S., Xue, N. (eds.) *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC–COLING 2024)*. pp. 7982–7992. ELRA and ICCL (2024), <https://aclanthology.org/2024.lrec-main.703/>
19. Quaresma, P., Finatto, M.J.B.: Information extraction from historical texts: A case study. In: Finatto, M.J., Vieira, R., Pollak, S., Luz, S. (eds.) *Proceedings of the Workshop on Digital Humanities and Natural Language Processing (DHandNLP 2020)*. *CEUR Workshop Proceedings*, vol. 2607, pp. 49–56. CEUR-WS.org, Évora, Portugal (Mar 2020), <https://ceur-ws.org/Vol-2607/short2.pdf>
20. Rodríguez-Ortega, A., et al.: Relation extraction from noisy ocr historical texts. *Journal of Data Mining and Digital Humanities* (2022)
21. Schich, M., Song, C., Ahn, Y.Y., Mirsky, A., Martino, M., Barabási, A.L., Helbing, D.: A network framework of cultural history. *Science* **345**(6196), 558–562 (2014). <https://doi.org/10.1126/science.1240064>
22. Sebastiani, F.: An Axiomatically Derived Measure for the Evaluation of Classification Algorithms. In: *Proceedings of the 2015 international conference on the theory of information retrieval*. pp. 11–20 (2015)
23. Stoica, G., Platanios, E.A., Póczos, B.: Re-tacred: Addressing shortcomings of the tacred dataset (2021), <https://arxiv.org/abs/2104.08398>
24. Sugimoto, G., Daza, A., Boer, V.d.: Closer reading of RDF generated by NLP on Wikipedia biography: Comparative analysis. In: *Research Conference on Metadata and Semantics Research*. pp. 41–54. Springer (2023)
25. Tamper, M., Kettunen, M., Mäkelä, E., Ruotsalo, T., Leskinen, P., Hyvönen, E.: BiographySampo: A linked open data service for prosopographical biography research. In: *Proceedings of the Digital Humanities Conference (DH 2023)*. Graz, Austria (2023), <https://seco.cs.aalto.fi/projects/biographysampo/>
26. Tan, Q., Xu, L., Bing, L., Ng, H.T., Aljunied, S.M.: Revisiting DocRED - addressing the false negative problem in relation extraction. In: Goldberg, Y., Kozareva, Z., Zhang, Y. (eds.) *Proceedings of the 2022 Conference on Empirical*

- Methods in Natural Language Processing. pp. 8472–8487. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), <https://aclanthology.org/2022.emnlp-main.580/>
27. Wang, A., Wolf, T.: Overview of the SustaiNLP 2020 shared task. In: Moosavi, N.S., Fan, A., Shwartz, V., Glavaš, G., Joty, S., Wang, A., Wolf, T. (eds.) Proceedings of SustaiNLP: Workshop on Simple and Efficient Natural Language Processing. pp. 174–178. Association for Computational Linguistics (2020). <https://doi.org/10.18653/v1/2020.sustainlp-1.24>, <https://aclanthology.org/2020.sustainlp-1.24>
 28. Yang, S., Choi, M., Cho, Y., Choo, J.: HistRED: A historical document-level relation extraction dataset. In: Rogers, A., Boyd-Graber, J., Okazaki, N. (eds.) Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 3207–3224. Association for Computational Linguistics, Toronto, Canada (Jul 2023). <https://doi.org/10.18653/v1/2023.acl-long.180>, <https://aclanthology.org/2023.acl-long.180/>
 29. Yao, Y., Ye, D., Li, P., Han, X., Lin, Y., Liu, Z., Liu, Z., Huang, L., Zhou, J., Sun, M.: DocRED: A large-scale document-level relation extraction dataset. In: Korhonen, A., Traum, D., Màrquez, L. (eds.) Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 764–777. Association for Computational Linguistics, Florence, Italy (Jul 2019), <https://aclanthology.org/P19-1074/>
 30. Zhang, Y., Zhong, V., Chen, D., Angeli, G., Manning, C.D.: Position-aware attention and supervised data improve slot filling. In: Palmer, M., Hwa, R., Riedel, S. (eds.) Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 35–45. Association for Computational Linguistics, Copenhagen, Denmark (Sep 2017), <https://aclanthology.org/D17-1004/>
 31. Zhong, L., Wu, J., Li, Q., Peng, H., Wu, X.: A comprehensive survey on automatic knowledge graph construction. *ACM Computing Surveys* **56**(4), 1–62 (2023)
 32. Zhou, K., Qiao, Q., Li, Y., Li, Q.: Improving distantly supervised relation extraction by natural language inference. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 14047–14055 (2023)