A REVIEW OF RELATION EXTRACTION AND JOINT NER-RE MODELS: FROM PIPELINE TO MULTI-TASK ARCHITECTURES

Louna Alyasein

PhD student, Varna Free University "Chernorizets Hrabar", Bulgaria

Veselina Spasova

Associate Professor, PhD

Nikola Vaptcarov Naval Academy, Bulgaria

Varna Free University "Chernorizets Hrabar", Bulgaria

Abstract: *Named entity recognition (NER) and relation extraction (RE) convert unstructured text into structured representations that support knowledge graph construction, search, and data analysis. This narrative review examines supervised, semi-supervised, and distantly supervised RE and compares pipeline and joint approaches to extracting entities and relations from raw text. It distinguishes how models are trained, how predictions are produced, and what information is supplied at inference time. The review covers multi-head selection, cascade and token-pair tagging, span-based extraction, and recent approaches based on prompting, retrieval, and schema-conditioned generation. It also examines how document-level context, incomplete annotations, and evaluation criteria affect reported results. Joint training can exploit dependencies between entity and relation predictions, while well-designed pipelines remain competitive. Meaningful comparisons therefore require an explicit description of the input, output, supervision, and scoring protocol.*

Keywords: *Natural Language Processing, Named Entity Recognition, Relation Extraction, Joint Models, Multi-Task Learning, Distant Supervision, BERT, Large Language Models*

1. Introduction

The growth of unstructured text across scientific literature, news, and digital records has driven sustained interest in automatically extracting structured knowledge from natural language (Jurafsky and Martin, 2023). Two tasks lie at the core of this effort: Named Entity Recognition (NER), which identifies and classifies entities such as persons, organizations, and locations, and Relation Extraction (RE), which identifies the semantic relations that connect pairs of entities (Goldberg, 2017). Together, these tasks convert raw sentences into structured triples of the form (subject, relation, object), enabling applications such as knowledge graph construction, semantic search, and automated question answering (Young et al., 2018).

For example, given the sentence "Ahmed works at the University of Baghdad," NER identifies "Ahmed" as a Person and "University of Baghdad" as an Organization, while RE determines the relation works_at that connects them, producing the structured triple (Ahmed, works_at, University of Baghdad). This simple example illustrates the central goal of information extraction research: converting free text, which is easy for humans to read but difficult for machines to query, into structured records that can be indexed, searched, and reasoned over automatically.

In a conventional pipeline, the relation component operates on entities identified by an earlier stage. A missed mention can prevent the system from extracting the corresponding relation. Joint approaches exploit dependencies between entities and relations through shared representations, structured prediction, or both (Li and Ji, 2014; Bekoulis et al., 2018). Pipelines can also incorporate entity information effectively and can outperform joint models under matched evaluation conditions (Zhong and Chen, 2021). Large language models (LLMs) extend the design space through in-context learning and structured generation (Wan et al., 2023; Xu et al., 2024).

This review contributes a comparative framework that connects supervision, training organization, inference, and task definition. It uses this framework to explain when pipeline, joint, and generative approaches can be compared directly, and how the comparison changes when systems move from sentence-level classification to document-level extraction. The discussion emphasizes overlapping triples, annotation quality, and generalization to new relation types and domains.

2. Background and Related Work

2.1 Review Methodology

This narrative review retains foundational papers that introduced influential architectures, learning methods, or benchmarks, including benchmark descriptions published before 2014. The literature update through 20 November 2026 prioritizes work published in 2025 and 2026. Targeted searches combined terms for relation extraction and joint entity and relation extraction with terms for document-level extraction, zero-shot learning, distant supervision, and structured information extraction. The update used publisher pages and ACL Anthology records to verify bibliographic details and publication status. Sources were selected for their relevance to the comparative framework, coverage of different task settings, or evidence about benchmark limitations. The resulting synthesis compares design choices and evaluation conditions across representative studies.

2.2 Named Entity Recognition and Relation Extraction

Named Entity Recognition identifies spans of text referring to real-world entities and assigns them a category, typically Person, Organization, or Location, though domain-specific schemes exist for medical, legal, or technical text (Goldberg, 2017). Relation Extraction builds on this output (or, in joint architectures, on a shared representation) to determine the semantic relation that connects a pair of entities, expressed as a structured triple.

Table 1

NER vs. RE

Task	Function	Example
NER	Identifies and classifies entities	Ahmed: Person; University of Baghdad: Organization
RE	Identifies the relation between entities	(Ahmed, works_at, University of Baghdad)

2.3 Related Work

Early relation extraction research relied heavily on hand-crafted lexical and syntactic features combined with classical machine learning classifiers (Goldberg, 2017). The introduction of neural architectures — convolutional and recurrent networks, and later attention mechanisms — substantially improved performance by allowing models to learn feature representations directly from data (Young et al., 2018). A major turning point was the introduction of pretrained Transformer-based language models such as BERT (Devlin et al., 2019), which provided contextual word representations that could be fine-tuned for a wide range of downstream NLP tasks, including relation extraction, with comparatively little task-specific data.

Li and Ji (2014) proposed incremental joint extraction of entity mentions and relations. Bekoulis et al. (2018) subsequently combined neural sequence labeling with multi-head selection, enabling each entity to participate in several labeled relations. PURE (Zhong and Chen, 2021) provides a contrasting design within the same end-to-end extraction task. Its two independently trained encoders use separate contextual representations for entities and relations, with entity information entering the relation model early.

Nasar et al. (2021) provide a broader survey of the NER and RE literature, tracing the field's progression from rule-based systems to deep learning approaches and cataloguing the datasets and evaluation practices used across the community. A more recent strand of joint extraction research has explored tagging-based formulations: Wei et al. (2020) proposed CasRel, a cascade binary tagging framework that first identifies head entities and then, for each head entity, applies relation-specific taggers to identify corresponding tail entities, explicitly

designed to handle overlapping relational triples. Wang et al. (2020) proposed TPLinker, which reformulates joint extraction as a token-pair linking problem using a handshaking tagging scheme, avoiding the exposure bias that arises when earlier stages of a multi-step tagging pipeline are trained on ground-truth conditions but must operate on their own predictions at inference time. Eberts and Ulges (2020) proposed SpERT, a span-based joint entity and relation extraction model built on a pretrained Transformer, which classifies entity spans and relations between them directly rather than relying on sequential tagging. Unified Information Extraction (UIE) uses structured generation to express multiple extraction tasks in a common output representation (Lu et al., 2022).

Nasar et al. (2021) survey both NER and RE, while Diaz-Garcia and Diaz Lopez (2025) focus on language-model-based RE. Xu et al. (2024) place generative approaches within the broader information extraction literature. Building on these surveys, the present review connects how systems obtain supervision, organize learning, and expose intermediate predictions to the evaluation conditions under which their performance is reported.

2.4 Benchmark Datasets

Table 2 covers benchmarks for sentence-level relation classification, extraction from raw text, and document-level RE. In standard relation classification on TACRED or SemEval-2010 Task 8, the input specifies the target pair. In a joint extraction setting such as ACE05, entity detection is part of the task. DocRED (Yao et al., 2019) introduces relations between entities that can have multiple mentions across sentences. Re-DocRED (Tan et al., 2022) and Re²-DocRED (Heng et al., 2026) address omissions in document-level annotations. Benchmark names should be accompanied by a description of the experimental setting, since an annotated corpus can support evaluation with either supplied or predicted entities. Re-TACRED (Stoica et al., 2021) provides revised sentence-level annotations, and FewRel (Han et al., 2018) evaluates classification with a small support set for the target relations.

Table 2

Common Benchmark Datasets for RE and Joint NER-RE

Dataset	Level	Typical task	Entity input and annotations
SemEval-2010 Task 8 (Hendrickx et al., 2010)	Sentence	Relation classification	Supplied nominal pair
TACRED (Zhang et al., 2017) and Re-TACRED (Stoica et al., 2021)	Sentence	Relation classification	Supplied entity pair with original or revised labels

Dataset	Level	Typical task	Entity input and annotations
ACE05 (Walker et al., 2006)	Sentence in common joint setups	Entity and relation extraction	Entities predicted from text in the end-to-end setup
NYT (Riedel et al., 2010)	Sentence or entity-pair bag	Distantly supervised RE	Knowledge-base alignment supplies weak labels
FewRel (Han et al., 2018)	Sentence	Few-shot relation classification	Supplied pair and labeled support examples
DocRED (Yao et al., 2019)	Document	Document-level RE	Entity clusters supplied in standard RE setup
Re-DocRED (Tan et al., 2022)	Document	Document-level RE	Revised DocRED annotations with additional relation triples
Re²-DocRED (Heng et al., 2026)	Document	Document-level joint extraction evaluation	Further audited relation annotations for document-level extraction

Sections 3 and 4 review supervision and joint modeling. Sections 5 and 6 use these perspectives to compare representative systems and their trade-offs. Sections 7–10 discuss interpretation, applications, recent LLM-based methods, and research directions. Sections 11 and 12 present the review limitations and conclusions.

3. Relation Extraction Methods

3.1 Supervised Relation Extraction

Supervised RE uses manually labeled examples of relations, often with entity mentions or candidate pairs. Earlier models relied on lexical and syntactic features, while neural approaches learn contextual representations and can be fine-tuned from pretrained encoders such as BERT (Goldberg, 2017; Young et al., 2018; Devlin et al., 2019). How much task-specific data a model requires depends on the domain, the quality of pretraining, and the relation inventory. In end-to-end extraction, training data must also indicate the entity spans or other representations required by the output protocol.

Supervised methods can achieve high accuracy when high-quality labeled data are available; the primary limitation is that manual annotation is expensive and time-consuming, particularly for specialized or low-resource domains, and models trained on one domain often generalize poorly to another, as relation semantics and typical sentence structures can vary substantially across domains such as biomedical, legal, and news text.

3.2 Semi-Supervised Relation Extraction

Semi-supervised approaches combine a small labeled dataset with a much larger unlabeled corpus, using techniques such as self-training, bootstrapping, co-training, and pseudo-labeling. An initial model is trained on the labeled examples, used to label the unlabeled data, and then retrained on its own high-confidence predictions. This iterative process reduces

annotation cost but introduces the risk of error propagation and confirmation bias, whereby incorrect predictions are reinforced over successive training iterations, gradually degrading model quality if left unchecked.

In practice, semi-supervised methods are most useful when a reasonably reliable initial model can be trained on a modest labeled seed set, and when confidence thresholds for adding new pseudo-labeled examples are set conservatively enough to limit the accumulation of errors across training iterations.

3.3 Distant Supervision

Distant supervision derives labels by aligning an external knowledge base with text. The alignment assumes that mentions of a pair recorded in the knowledge base provide evidence for its relation, which can introduce false positives and omit relations that are absent from the knowledge base (Riedel et al., 2010). For example, “Steve Jobs visited Apple headquarters” contains both entities without expressing that Jobs founded Apple.

Multi-instance learning groups sentences mentioning the same entity pair and allows a subset of them to license a relation label. Selective attention can reduce the influence of sentences with weak evidence within each group (Lin et al., 2016). At document level, Menotti et al. (2026) identify contradictions and violations of ontology constraints in distantly supervised DocRED data. Their results illustrate how structural regularization can complement instance weighting by addressing inconsistencies among triples.

3.4 Evaluation Metrics

Precision, recall, and F1 are standard metrics. Precision is TP / (TP + FP), recall is TP / (TP + FN), and F1 is their harmonic mean. In relation classification, a true positive typically requires the correct relation label for a supplied, ordered entity pair. For end-to-end extraction, it also requires the predicted arguments to match reference entities under the stated span or entity-cluster criterion. Reports should specify exact or partial matching, whether argument types must match, the treatment of the no-relation class, and micro- or macro-averaging. Document-level evaluations should specify the treatment of coreference clusters and whether evidence sentences are scored. Incomplete annotations can cause valid predictions to be counted as false positives, as shown by the re-annotation work on DocRED (Yao et al., 2019; Tan et al., 2022; Heng et al., 2026).

3.5 Comparison of RE Methods

Table 3

Comparison of Relation Extraction Methods

Criterion	Supervised	Semi-Supervised	Distant Supervision
Label source	Manual	Manual + automatic	Knowledge base
Setup cost	High	Medium	Relatively low
Label quality	High	Medium	Can be low
Scalability	Limited	Good	Very high
Main limitation	Labeling cost	Error propagation	Noisy labels

No single method is universally superior; modern systems increasingly combine manually labeled data, semi-labeled data, and distantly supervised data with Transformer-based encoders and noise-reduction mechanisms, choosing the mix that best matches the data available and the accuracy requirements of the target application.

4. Joint Models for NER and Relation Extraction

A traditional pipeline architecture executes NER and RE sequentially: entities are first identified, and the recognized entities are then passed to a separate RE module. This design is simple and modular, and each component can be developed and trained independently, potentially on different datasets and with different modeling choices. However, it is highly vulnerable to error propagation — if NER fails to correctly identify an entity, RE cannot extract the corresponding relation, regardless of how well the RE model itself performs (Li and Ji, 2014).

Joint systems couple entity and relation prediction through shared representations or a structured objective. In a common multi-task design, a shared encoder produces contextual token representations, which an entity component uses to identify spans and a relation component uses to score candidate pairs. The two components can share information during training even when relation prediction at inference depends on entities selected by an earlier decoding step (Bekoulis et al., 2018; Eberts and Ulges, 2020). The extent of error propagation therefore depends on how candidate entities enter the relation component.

A common strategy for training such models is multi-task learning, in which shared layers capture general linguistic information — word meaning, syntax, and context — while task-specific layers (e.g., a CRF or softmax layer for NER; an attention or bilinear layer for RE) specialize in each task. The overall objective combines both task losses in a weighted joint loss function, $L_{total} = \lambda_{NER} \cdot L_{NER} + \lambda_{RE} \cdot L_{RE}$, where the weights λ_{NER} and λ_{RE} balance the relative importance of each task; poorly tuned weights can cause one task to

dominate training at the expense of the other. In practice, these weights are often tuned empirically on a validation set, or adjusted dynamically during training based on each task's relative learning progress.

Bekoulis et al. (2018) formulate relation extraction as multi-head selection. Each entity can have several labeled links, allowing a sentence to express multiple triples that share an argument. In “Mira joined North University and directs its robotics laboratory,” one possible representation contains (Mira, works_at, North University) and (Mira, directs, robotics laboratory). The relations share Mira, while the second object is a separate span. This kind of relation overlap differs from nested mentions, which occur when one annotated entity span lies within another.

End-to-end extraction describes the input and output of the complete system. A system that receives raw text and returns entities and relations can use a separately trained pipeline, a joint model, or a generative decoder. Joint training describes how parameters and objectives are shared. These two properties can be evaluated independently (Zhong and Chen, 2021; Eberts and Ulges, 2020).

Beyond multi-head selection, tagging- and span-based formulations offer alternative ways of building joint, end-to-end extraction models. CasRel (Wei et al., 2020) treats relation extraction as identifying, for each head entity, the tail entities associated with it under each relation type, using relation-specific taggers — an approach well suited to sentences containing overlapping relational triples. TPLinker (Wang et al., 2020) instead links pairs of tokens directly through a handshaking tagging scheme, avoiding the exposure bias that can arise when a model is trained under ground-truth conditions but must rely on its own earlier predictions at inference time. SpERT (Eberts and Ulges, 2020) takes a span-based approach, built on a pretrained Transformer encoder, classifying candidate entity spans and the relations between them directly rather than through sequential tagging. Together with multi-head selection, these approaches illustrate that there is no single dominant formulation for joint extraction; rather, the choice of tagging scheme or span-based formulation represents a further design dimension, alongside the shared-representation and multi-task-learning choices discussed above.

5. A Unified Framework for Comparing Extraction Systems

Table 4 applies the framework to representative methods. The supervision column describes the data or signal used for the target task. The training and inference columns distinguish when components learn together from how their predictions depend on each other at inference. For example, CasRel (Wei et al., 2020) learns linked extraction steps while still

conditioning tail prediction on a selected head. GLiREL (Boylan et al., 2025) illustrates a separate distinction: zero-shot prediction for new relation labels can still rely on a trained model and supplied entity candidates.

Table 4

Supervision, Inference, and Task Definition in Representative Systems

System	Supervision and training	Inference	Task input and output
PURE (Zhong and Chen, 2021)	Labeled data, separate encoders	Entity detection then pair classification	Text to entities and relations
Multi-head selection (Bekoulis et al., 2018)	Labeled data, joint objective	Entity tags and multiple labeled links	Text to entities and relations
CasRel (Wei et al., 2020)	Labeled data, linked tagging objectives	Head detection followed by relation-specific tail tagging	Text to triples
TPLinker (Wang et al., 2020)	Labeled data, joint tagging	Token pairs decoded as spans and links	Text to triples
SpERT (Eberts and Ulges, 2020)	Labeled data, shared encoder	Span classification then pair scoring	Text to entities and relations
GLiREL (Boylan et al., 2025)	Trained classifier, zero-shot transfer to new labels	Scores supplied entity pairs and candidate labels	Text, entities, labels to relations
DocIE approach (Popovic et al., 2025)	Pretrained LLM with synthetic demonstrations	Retrieves examples and prompts extraction	Document to entities and relations

These distinctions also apply to prompting. A single LLM can be used in a pipeline with separate prompts for entities and relations or asked to generate both outputs in one response. The term joint extraction describes the requested output in the latter setting. Claims about joint training require information about the model's training objectives (Xu et al., 2024; Lu et al., 2022; Popovic et al., 2025).

6. Comparative Analysis

Table 5

Practical Trade-offs Across Extraction Designs

Criterion	Modular pipeline	Joint discriminative model	Prompted generation
Intermediate outputs	Entity predictions can be inspected directly	Candidate spans and relation scores depend on model design	Outputs may include text spans or evidence if requested
Main error mechanisms	Missed mentions remove candidate pairs	Missed spans, task interference, or inconsistent links	Unsupported facts, schema violations, or invalid output
Schema changes	May require updating one or both components	May require training or adaptation	May be specified through prompts, examples, or retrieved schemas

Criterion	Modular pipeline	Joint discriminative model	Prompted generation
Evaluation priority	Report results with supplied and predicted entities	Specify span, type, and relation matching	Report validation failures and inference cost alongside F1

A modular pipeline supports component-level error analysis and replacement, although changes in the entity module may alter the inputs received by the relation module. Joint training can share useful information across tasks, but its benefit depends on the training objective, candidate selection, and domain. PURE outperformed earlier joint models with the same pretrained encoders on ACE04, ACE05, and SciERC in its reported experiments (Zhong and Chen, 2021). This result makes a strong pipeline an appropriate baseline for studies attributing gains to joint modeling.

7. Discussion

Recent work broadens the settings in which entity and relation extraction can be applied. Pretrained encoders support discriminative extraction with limited adaptation, while prompting and retrieval can support rapidly changing schemas. Comparisons become most informative when models receive equivalent inputs, predict the same output structures, and are scored on the same benchmark version under the same matching rules (Zhong and Chen, 2021; Xu et al., 2024; Diaz-Garcia and Diaz Lopez, 2025; Heng et al., 2026).

CasRel and TPLinker explicitly address overlapping triples, while SpERT scores relations between candidate spans and multi-head selection allows an entity to take part in multiple labeled links (Bekoulis et al., 2018; Wei et al., 2020; Wang et al., 2020; Eberts and Ulges, 2020). Evaluation can distinguish triples sharing one argument from triples sharing both arguments. This distinction helps determine whether a model handles multiple labels for the same entity pair. Nested mention recognition presents an additional challenge that should be scored at the entity level. ARETO (Liao et al., 2026) further examines overlap in which triples share an entity and a relation label, drawing attention to distinctions that aggregate triple F1 can obscure.

Distant supervision presents two distinct challenges when used for document-level extraction. Missing relations reduce the completeness of the target set, while incompatible relation labels can produce contradictions across triples. Re-annotation studies address the former, and ontology-based consistency constraints address the latter (Tan et al., 2022; Heng et al., 2026; Menotti et al., 2026). Both strands motivate reporting the source and revision of the annotations as part of evaluations.

Schema flexibility introduces another design choice. A system may classify from a fixed relation inventory, score labels newly supplied at inference, or return relations in a structured output described by the user. These settings differ in how relation meanings are specified and how outputs can be validated (Diaz-Garcia and Diaz Lopez, 2025; Boylan et al., 2025; Ho et al., 2026).

8. Applications and Open Challenges

Entity and relation extraction supports knowledge graph population, semantic search, and question answering (Young et al., 2018; Xu et al., 2024). Once text-derived facts have been normalized and integrated into a graph, link prediction can propose additional connections. KG-BERT (Yao et al., 2019) belongs to this downstream knowledge graph completion setting. For applications with audit requirements, extracted relations can be linked to their supporting passages so the source of each assertion remains traceable.

Errors can arise from boundary mismatches, invalid argument types, incorrect relation direction, missing cross-sentence links, or unsupported facts. These errors require different remedies. Span or mention-cluster errors motivate improvements to entity recognition and coreference, while inconsistent relation types motivate schema constraints. Missing reference triples warrant annotation review (Yao et al., 2019; Tan et al., 2022; Heng et al., 2026; Menotti et al., 2026). System comparisons should identify which error types are counted and how invalid or unparseable outputs are handled.

9. The Role of Large Language Models

In-context learning conditions an LLM on instructions and, where available, task examples without updating its parameters at inference time. GPT-RE (Wan et al., 2023) illustrates this approach to relation classification. GLiREL (Boylan et al., 2025) offers a discriminative alternative that scores candidate relation labels for supplied entities, including labels unseen during task-specific training. Li et al. (2025) study a hybrid strategy in which generation supports retrieval for zero-shot RE. These examples illustrate distinct uses of language models and require comparison under the same assumptions about candidate entities and label availability.

At document level, Zhang et al. (2025) filter candidate relation labels for individual entity pairs before extracting relation facts. Popovic et al. (2025) retrieve synthetic demonstrations for document-level entity and relation extraction and report that this setting remains challenging for LLMs. SchemaRAG (Ho et al., 2026) retrieves relevant schema elements before extraction and evaluates the effect on accuracy, latency, and token use. As an

adjacent example of schema-conditioned multi-task information extraction, GLiNER2 (Zaratiana et al., 2025) combines entity recognition, text classification, and hierarchical extraction in a compact encoder framework. These studies motivate evaluation that reports format validity, schema compliance, relation accuracy, and inference cost separately.

10. Future Research Directions

Future comparisons should evaluate matched pipeline, joint, and prompted systems on document-level inputs with both supplied and predicted entities. Reporting results for each setting would help quantify how much error originates in relation prediction and how much comes from entity detection or clustering. A second direction is to combine distantly supervised data with explicit type and relation constraints, then validate results against an independently annotated sample (Heng et al., 2026; Menotti et al., 2026). For schema-conditioned LLMs, a useful comparison would vary the schema size, example retrieval, and evidence requirements while recording exact-match F1, unsupported triples, invalid outputs, and computational cost (Popovic et al., 2025; Ho et al., 2026).

11. Limitations of this Review

The review uses a narrative selection of foundational and recent studies, primarily from English-language venues. Its coverage reflects the scope of the targeted searches and the relevance of publications to the proposed framework. Reported results come from studies with varying annotation schemes, datasets, and evaluation protocols. Direct comparisons are therefore limited to results reported under sufficiently similar conditions, and the review does not pool benchmark scores.

12. Conclusion

The review relates supervision, training organization, inference, and task definition across approaches to relation extraction. Supervised learning depends on the availability of task annotations, semi-supervised methods can extend a limited labeled set, and distant supervision provides scale while introducing omissions and inconsistencies. Joint training can capture dependencies between entity and relation predictions, while two-stage systems can also perform end-to-end extraction from raw text and remain competitive under controlled comparisons.

Recent work extends the comparison to document-level extraction, unseen relation types, and schema-conditioned prediction. The suitability of a method depends on whether entity candidates are available, how much the target schema changes, how reliably annotations represent the underlying relations, and what constraints apply to inference. Benchmark

reporting that makes these conditions explicit is necessary for interpreting progress across the different settings reviewed here.

References

Bekoulis, G., Deleu, J., Demeester, T., & Develder, C. Joint Entity Recognition and Relation Extraction as a Multi-Head Selection Problem. – In: *Expert Systems with Applications*. 2018, vol. 114, pp. 34 – 45. ISSN 0957-4174.

Boylan, J., Hokamp, C., & Gholipour Ghalandari, D. GLiREL: Generalist Model for Zero-Shot Relation Extraction. – In: *NAACL*. 2025. https://aclanthology.org/2025.naacl-long.418/

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. – In: *NAACL*. 2019.

Diaz-Garcia, J. A., & Diaz Lopez, J. A. A survey on cutting-edge relation extraction techniques based on language models. – In: *Artificial Intelligence Review*. 2025, vol. 58, article 287. ISSN 0269-2821. https://doi.org/10.1007/s10462-025-11280-0

Eberts, M., & Ulges, A. Span-based Joint Entity and Relation Extraction with Transformer Pre-training (SpERT). – In: *ECAI*. 2020.

Goldberg, Y. Neural Network Methods for Natural Language Processing. San Rafael, CA: Morgan & Claypool, 2017.

Han, X., Zhu, H., Yu, P., Wang, Z., Yao, Y., Liu, Z., & Sun, M. FewRel: A Large-Scale Supervised Few-Shot Relation Classification Dataset with State-of-the-Art Evaluation. – In: *EMNLP*. 2018. https://aclanthology.org/D18-1514/

Hendrickx, I., Kim, S. N., Kozareva, Z., et al. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals. – In: *Proceedings of the 5th International Workshop on Semantic Evaluation, ACL*. 2010.

Heng, C. K., Tong, S. W., & Wong, J. W. S. Re²-DocRED: Revisiting Revisited-DocRED for Joint Entity and Relation Extraction. – In: *EACL*. 2026. https://aclanthology.org/2026.eacl-long.213/

Ho, S. Y. B., Coles, A., Larsson, E., Marshall, E., Bodenstab, N., & Vozila, P. SchemaRAG: Dynamic Large Schema Reduction for LLM-driven Structured Information Extraction. – In: *ACL Industry Track*. 2026. https://aclanthology.org/2026.acl-industry.78/

Jurafsky, D., & Martin, J. H. Speech and Language Processing (3rd ed., draft). Stanford: Stanford University, 2023.

Li, Q., & Ji, H. Incremental Joint Extraction of Entity Mentions and Relations. – In: *ACL*. 2014.

Li, Z., Zhang, F., Peng, T., Liu, H., & Cheng, J. Generation-Augmented Retrieval: Rethinking the Role of Large Language Models in Zero-Shot Relation Extraction. – In: *Findings of EMNLP*. 2025, pp. 17928 – 17941. https://aclanthology.org/2025.findings-emnlp.974/

Liao, J., et al. ARETO: A joint entity and relation extraction model for the triple overlapping problem. – In: *Knowledge-Based Systems*. 2026, article 115178. ISSN 0950-7051. https://doi.org/10.1016/j.knosys.2025.115178

Lin, Y., et al. Neural Relation Extraction with Selective Attention over Instances. – In: *ACL*. 2016.

Lu, Y., et al. Unified Structure Generation for Universal Information Extraction. – In: *ACL*. 2022. https://aclanthology.org/2022.acl-long.395/

Menotti, L., Marchesin, S., & Silvello, G. Ontology-Driven Structural Regularization for Document-Level Relation Extraction. – In: *EMNLP*. 2026, pp. 756 – 773. https://aclanthology.org/2026.emnlp-main.35/

Nasar, Z., Jaffry, S. W., & Malik, M. K. Named Entity Recognition and Relation Extraction: State-of-the-Art. – In: *ACM Computing Surveys*. 2021, vol. 54, no. 1, pp. 1 – 39. ISSN 0360-0300.

Popovic, N., Kangen, A., Schopf, T., & Färber, M. DocIE@XLLM25: In-Context Learning for Information Extraction using Fully Synthetic Demonstrations. – In: *XLLM*. 2025, pp. 298 – 309. https://aclanthology.org/2025.xllm-1.26/

Riedel, S., Yao, L., & McCallum, A. Modeling Relations and Their Mentions without Labeled Text. – In: *ECML PKDD*. 2010. https://doi.org/10.1007/978-3-642-15939-8_10

Stoica, G., Platanios, E. A., & Póczos, B. Re-TACRED: Addressing Shortcomings of the TACRED Dataset. – In: *AAAI*. 2021, vol. 35, no. 15, pp. 13843 – 13850. https://doi.org/10.1609/aaai.v35i15.17631

Tan, Q., Xu, L., Bing, L., Ng, H. T., & Aljunied, S. M. Revisiting DocRED: Addressing the False Negative Problem in Relation Extraction. – In: *EMNLP*. 2022. https://aclanthology.org/2022.emnlp-main.580/

Walker, C., Strassel, S., Medero, J., & Maeda, K. ACE 2005 Multilingual Training Corpus. Philadelphia: Linguistic Data Consortium, 2006.

Wan, Z., Cheng, F., Mao, Z., Liu, Q., Song, H., Li, J., & Kurohashi, S. GPT-RE: In-context Learning for Relation Extraction using Large Language Models. – In: *EMNLP*. 2023.

Wang, Y., Yu, B., Zhang, Y., Liu, T., Zhu, H., & Sun, L. TPLinker: Single-stage Joint Extraction of Entities and Relations Through Token Pair Linking. – In: *COLING*. 2020.

Wei, Z., Su, J., Wang, Y., Tian, Y., & Chang, Y. A Novel Cascade Binary Tagging Framework for Relational Triple Extraction (CasRel). – In: *ACL*. 2020.

Xu, D., Chen, W., Peng, W., et al. Large Language Models for Generative Information Extraction: A Survey. – In: *Frontiers of Computer Science*. 2024, vol. 18, no. 6, article 186357. ISSN 2095-2228.

Yao, L., Mao, C., & Luo, Y. KG-BERT: BERT for Knowledge Graph Completion. 2019. arXiv:1909.03193.

Yao, Y., et al. DocRED: A Large-Scale Document-Level Relation Extraction Dataset. – In: *ACL*. 2019. https://aclanthology.org/P19-1074/

Young, T., Hazarika, D., Poria, S., & Cambria, E. Recent Trends in Deep Learning Based Natural Language Processing. – In: *IEEE Computational Intelligence Magazine*. 2018, vol. 13, no. 3, pp. 55 – 75. ISSN 1556-603X.

Zaratiana, U., Pasternak, G., Boyd, O., Hurn-Maloney, G., & Lewis, A. GLiNER2: Schema-Driven Multi-Task Learning for Structured Information Extraction. – In: *EMNLP System Demonstrations*. 2025. https://aclanthology.org/2025.emnlp-demos.10/

Zhang, Y., Zhong, V., Chen, D., Angeli, G., & Manning, C. D. Position-aware Attention and Supervised Data Improve Slot Filling. – In: *EMNLP*. 2017.

Zhang, F., Yu, H., Cheng, J., & Xu, H. Entity Pair-guided Relation Summarization and Retrieval in LLMs for Document-level Relation Extraction. – In: *Findings of NAACL*. 2025, pp. 4022 – 4037. https://aclanthology.org/2025.findings-naacl.224/

Zhong, Z., & Chen, D. A Frustratingly Easy Approach for Entity and Relation Extraction. – In: *NAACL*. 2021. https://aclanthology.org/2021.naacl-main.5/