Artificial Intelligence Review (2025) 58:287
https://doi.org/10.1007/s10462-025-11280-0

A survey on cutting-edge relation extraction techniques based on language models

Jose A. Diaz-Garcia[1] · **Julio Amador Diaz Lopez**[2]

Accepted: 30 May 2025 / Published online: 1 July 2025

Abstract
This comprehensive survey examines the latest advancements in Relation Extraction (RE), a pivotal task in natural language processing essential for applications across biomedical, financial, and legal sectors. This study highlights the evolution and current state of RE techniques by analyzing 137 papers presented at the Association for Computational Linguistics (ACL) conferences from 2020 to 2023, focusing on models that leverage language models. Our findings underscore the dominance of BERT-based methods in achieving state-of-the-art results for RE while also noting the promising capabilities of emerging Large Language Models (LLMs) like T5, especially in few-shot relation extraction scenarios where they excel in identifying previously unseen relations.

Keywords Relation extraction · NLP · Language models · Datasets

1 Introduction

With the rapid growth of data generated and stored on the internet, numerous tools have emerged to process and extract valuable insights from vast information. Given that a significant portion of this data exists as unstructured text, the need for techniques capable of handling such data is evident. The branch of artificial intelligence dedicated to processing text is Natural Language Processing (NLP) (Ranjan et al. 2016). NLP provides practitioners and researchers with a comprehensive toolkit for effectively analyzing unstructured text from diverse sources, including social media (Diaz-Garcia et al. 2024; Coppersmith et al. 2018), public health data (Hossain et al. 2023), research papers (de Winter 2024) and digital humanities (Suissa et al. 2022). Recently, NLP has been heavily influenced by neural

✉ Jose A. Diaz-Garcia
jagarcia@decsai.ugr.es

Julio Amador Diaz Lopez
julio@siftyml.com

[1] Department of Computer Science and A.I., University of Granada, Granada, Spain

[2] SiftyML, London, UK

approaches such as transformers (Gillioz et al. 2020) and recurrent neural networks (Bennet et al. 2024). However, traditional non-neural methods, including dependency grammars (Nagy and Kapusta 2021) and rule-based systems (Xu and Cai 2021), continue to play a significant role.

One of the prominent topics within NLP is Relation Extraction (RE). RE is a task focusing on identifying and extracting the intricate relationships between different entities mentioned in the textual content. Essentially, RE automates discovering connections between words or phrases within a text. For example, in the sentence "The Eiffel Tower is located in Paris," a relation extraction system would automatically identify the relationship "located in" between the entities "Eiffel Tower" and "Paris." Though this example may seem trivial, it can be extrapolated to other domains with far-reaching implications. In fields such as biomedicine, finance, and law (Thomas and Sivanesan 2022), the importance of Relation Extraction becomes undeniable, as it enables the automatic discovery of critical connections within vast amounts of unstructured data.

In the biomedical domain for example, RE plays a crucial role in identifying the complex interactions between compounds and proteins (Zhou et al. 2014). This is particularly significant when analyzing signal transduction mechanisms and biochemical pathways. Protein-protein interactions, for instance, can initiate a wide range of biological processes. RE automates the extraction of these interactions and relationships from vast volumes of unstructured biomedical literature. This not only accelerates the research process but also enhances the precision of data extraction, enabling researchers to systematically uncover critical insights that would be difficult to identify manually.

RE is not just essential for identifying relationships within text, but also indispensable for downstream tasks such as question-answering systems (Chen et al. 2019), decision support systems (Agosti et al. 2018), and expert systems. Its utility extends to the humanities, where it enhances knowledge acquisition in graph-based datasets, enabling more sophisticated and accurate information retrieval and analysis (Zhu et al. 2017).

In recent years, the advancement of RE has been significantly bolstered by the integration of language models. These models, particularly those based on deep learning architectures, have demonstrated a remarkable ability to capture the semantic nuances of textual data. Language models enhance RE by enabling more accurate identification of relationships between entities by effectively understanding and representing the underlying meaning of words and phrases. This capability allows RE systems to move beyond surface-level associations, capturing complex and context-dependent relationships, thereby improving the precision and applicability of RE in various domains.

In this review, we aim to consolidate the latest advancements in RE, particularly emphasizing how recent research has leveraged language models. We aim to provide a comprehensive toolkit encompassing tasks, techniques, models, and datasets. By doing so, we intend to offer a resource that will serve as a springboard for future research and development.

Our survey overviews cutting-edge RE methodologies showcased at the Association for Computational Linguistics (ACL) conferences from 2020 to 2023. Numerous surveys on relation extraction have been conducted in the literature, including the notable work by Bassignana and Plank (2022). In their paper, authors research into the definition of relation extraction, categorizing techniques based on their sub-parts, such as Named Entity Recognition (NER), relation identification, and relation classification. Notably, their survey emphasizes scientific relation extraction, exclusively considering papers from ACL confer-

ences from 2017 to 2021. We extend the analysis to satellite conferences, NAACL, AACL, and EACL. Our focus lies in utilizing language models, and we extend the review timeline to 2023, providing an updated perspective on trends in RE research. It is noteworthy to highlight the study conducted by Wadhwa et al. (2023), which provides a comprehensive review of the impact of Large Language Models (LLMs) like GPT or T5 across various benchmark datasets and baselines. While our work closely relates to theirs, our objective is to expand the scope of analysis beyond LLMs. We aim to incorporate a broader range of language models, conduct thorough comparisons, and undertake an extensive and robust review, encompassing a more comprehensive understanding of the field.

Our review aims to answer the following research questions (RQ):

- **RQ1**: What are the challenges of RE that are being solved by systems that leverage language models?
- **RQ2**: What are the most commonly used language models for the RE problem?
- **RQ3**: Which datasets are used as benchmarks for RE using language models?
- **RQ4**: Are new LLMs like GPT or T5 useful in RE versus widespread models such as BERT?

The subsequent sections of the paper are organized as follows: providing a foundational understanding of essential concepts, Sect. 2 offers a concise overview on the techniques studied. Section 3 studies the intricate details of the methods employed in crafting this comprehensive review. Section 4 provides a comprehensive analysis of the datasets used for RE to the best of our knowledge. Our examination of cutting-edge RE models unfolds in Sect. 5, while Sect. 6 scrutinizes the performance disparities between Language Models (LMs) and LLMs. Addressing the research questions that guided our exploration, Sect. 7 articulates responses derived from our review and findings. Finally, in Sect. 8, we briefly summarize the key takeaways and insights presented in this paper.

2 Background

This section focuses on the theoretical principles that form the basis of our survey. We aim to provide the reader with the necessary conceptual foundations for RE and Language Models.

2.1 Relation Extraction

RE is a task in natural language processing that aims to identify and classify the relationships between entities mentioned in the text. This task can be divided into three main components: NER, relation identification, and relation classification.

NER is a critical initial step in relation extraction. It involves identifying and categorizing named entities within a text, such as the names of individuals, organizations, locations, and other specific entities (Sun et al. 2018). For example, in the sentence “Apple Inc. is headquartered in Cupertino, California,” NER identifies “Apple Inc.” as an organization and “Cupertino, California” as a location. This step is foundational, determining the entities between which relationships will be analyzed.

Relation identification follows NER and involves detecting pairs of entities related to each other in the text. This task can be accomplished using various techniques, including rule-based approaches or machine learning models (Ma et al. 2021). For instance, in the sentence "Steve Jobs co-founded Apple Inc.," relation identification would recognize the entity pair "Steve Jobs" and "Apple Inc." and determine their relationship.

Relation classification is the subsequent process where a predefined label is assigned to each identified entity pair, indicating the nature of their relationship. For example, in the sentence "Barack Obama was born in Hawaii," the entity pair "Barack Obama" and "Hawaii" is classified with the relationship label "PlaceOfBirth" (Pouran Ben Veyseh et al. 2020). This task involves categorizing the relationship as one of several predefined types, such as "PlaceOfBirth," "WorksFor," or "LocatedIn," based on the context of the entities involved.

2.2 Approaches and challenges in Relation Extraction

RE poses several challenges that are currently being addressed and require further research. In our study, we have identified key advancements in RE and will specifically focus on the following areas:

- **Document-level Relation Extraction**: In specific scenarios, relationships between entities extend beyond individual sentences, spanning entire documents. Document-level RE tackles this challenge by considering relationships in a broader context. This task is crucial for applications where understanding connections across multiple sentences or document sections is essential for accurate information extraction.
- **Sentence-level Relation Extraction**: Conversely, sentence-level RE focuses on identifying relationships within individual sentences. This level of granularity is relevant for tasks where relationships are localized and can be adequately captured within the confines of a single sentence. This task is essential for applications with short-form content or when the relations of interest are contained within discrete textual units.
- **Multimodal Relation Extraction**: With the increasing prevalence of multimodal data, RE extends beyond textual information to include visual cues. Multimodal RE integrates text and visual data to enhance understanding of relationships. This task is relevant in domains where textual and visual information jointly contribute to the overall context, such as image captions, medical reports, or social media content.
- **Multilingual Relation Extraction**: In multilingual relation extraction, the challenge lies in extracting relations and training systems in domains where multiple languages are present. This task includes scenarios where documents or texts may contain information in different languages, requiring models to understand and extract relations effectively across language barriers. This task is particularly relevant in diverse linguistic contexts or global applications where information is presented in multiple languages.
- **Few-Shot and Low-Resource Relation Extraction**: Few-shot learning addresses scenarios with limited labeled data for training RE models. It aims to develop models that generalize well even with minimal labeled examples. In low-resource settings, such as specialized domains like biology or medicine, the scarcity of annotated data poses a significant challenge. The high annotation costs in these domains hinder the availability

of large labeled datasets, making it challenging to train accurate RE models using traditional supervised learning methods.

– **Distant Relation Extraction**: Distant RE involves extracting relations between entities based on distant supervision. In this approach, heuristics or existing knowledge bases automatically label large amounts of data. However, this introduces challenges related to noisy labels and the potential inaccuracies in the automatically generated annotations. Addressing these challenges is crucial for improving the reliability of models trained on distantly supervised data.
– **Open Relation Extraction**: OpenRE is dedicated to unveiling previously unknown relation types between entities within open-domain corpora. The primary objective of OpenRE lies in the automated establishment and continual evolution of relation schemes for knowledge bases, accomplished through identifying novel relations within unsupervised data. Within the realm of OpenRE methods, we can find two prominent categories: tagging-based methods and clustering-based methods.

2.3 Language models and large language models

LMs and LLMs process sequences of tokens, denoted as $w = \{w_1, w_2, ..., w_N\}$, by assigning a probability $p(w)$ to the sequence and incorporating information from previous tokens. This is captured mathematically as follows:

$$p(w) = \prod_t p(w_t \mid w_{t-1}, w_{t-2}, ..., w_1) \tag{1}$$

Although $p(w)$ has been traditionally calculated statistically, recently $p(w)$ is being calculated through neural language models (Bengio et al. 2001). However, the methods employed to calculate $p(w)$ vary across applications, diverse architectures from multi-layer perceptrons (Mikolov and Zweig 2012) and convolutional neural networks (Dauphin et al. 2017) to recurrent neural networks (Zaremba et al. 2014).

Recently, the introduction of self-attention mechanisms (Parikh et al. 2016) has heralded the development of high-capacity models, with transformer architectures (Vaswani et al. 2017) emerging as frontrunners due to their efficient utilization of extensive data, leading to state-of-the-art results. Transformers have revolutionized the field of NLP, demonstrating their potential across various downstream applications, such as question answering (Nassiri and Akhloufi 2023) and text classification (Tezgider et al. 2022) or text generation (Luo et al. 2022b). Within the transformer architecture, there are different types, with two of the most widespread being the encoder-only and the encoder–decoder models (Cai et al. 2022).

Encoder-only models are designed to process input data and generate rich, contextualized representations. These models excel at tasks that involve understanding (Kadlec et al. 2016) and analyzing text, such as NER. The key feature of encoder-only models is their focus on generating deep contextual embeddings of the input sequence, which capture intricate semantic and syntactic information without producing any output sequence.

On the other hand, encoder–decoder models are structured to handle tasks that require both understanding and generation of text. This architecture consists of two distinct components: an encoder and a decoder. The encoder processes the input sequence and generates a set of intermediate representations that encapsulate the input's contextual information.

The decoder then takes these representations and generates the final output sequence. This design is particularly effective for tasks that involve complex text generation, such as machine translation (Cho et al. 2014), where the model translates text from one language to another, and text summarization (Lebanoff et al. 2018). The encoder–decoder structure allows for a dynamic interplay between understanding and generation, making it well-suited for applications that require producing coherent and contextually appropriate output based on the input.

Noteworthy is Bi-directional Encoder Representations from Transformers (BERT) (Devlin et al. 2018). BERT employs a unique approach that involves learning the probability $p(w_i)$ by masking parts of the input sequence and predicting the masked word w_i while considering the surrounding context. Specifically, BERT and other contextual transformer-based models aim to compute the probability of a word given its context by estimating:

$$p(w_i) = p(w_i \mid w_1, \ldots, w_{i-1}, w_{i+1}, \ldots, w_N), \tag{2}$$

Being w_i the target word to be predicted, and $\{w_1, \ldots, w_{i-1}, w_{i+1}, \ldots, w_N\}$ represents the surrounding context words. This approach allows BERT to capture the contextual dependencies within a given sequence effectively.

The computation of these probabilities is made possible by training the transformer architecture on large and diverse datasets. Specifically, models like BERT are trained on extensive corpora, including the BookCorpus (Zhu et al. 2015) and a comprehensive crawl of English Wikipedia. This broad training allows the models to learn from vast amounts of text and diverse text combinations, enabling them to capture a wide range of linguistic patterns and contextual nuances. Notably, BERT has demonstrated remarkable prowess in tasks like RE (Roy and Pan 2021) and knowledge base completion (Yao et al. 2019a; Lan et al. 2021), showcasing its ability to capture intricate relationships within the text.

Variations of BERT, such as RoBERTa, have further refined the architecture by leveraging large-scale pre-training and optimized training procedures, enhancing both efficiency and performance. RoBERTa Liu et al. (2019) achieves state-of-the-art results through techniques like dynamic masking and larger batch sizes.

In the domain of larger computational capacities, is where we encounter LLMs, which represent a significant advancement in NLP. One of the most innovative models in this field is the Generative Pre-trained Transformer (GPT), developed by Radford et al. (2018), alongside other influential models such as Text-to-Text Transfer Transformer (T5) by Raffel et al. (2020), Jiang et al. (2023) or LLaMA (Touvron et al. 2023).

At this point, it is important to distinguish between LMs and LLMs. The primary difference lies in their scale. LMs typically have fewer parameters, whereas LLMs contain billions or even trillions of parameters, leading to far greater computational capacity (Kumar 2024). Due to this, traditional LMs are limited in handling complex tasks, processing large datasets, or capturing nuanced language patterns. In contrast, LLMs can process extensive datasets while maintaining strong contextual awareness. For instance, while a typical LM may have a context window of 1024 tokens, models like GPT-3 can handle up to 8192 tokens, and GPT-4 supports even larger context windows.

These models have offered novel methods for processing and interpreting language, extending beyond traditional approaches. GPT, for example, distinguishes itself through its ability to generate coherent and contextually rich text, revolutionizing text generation tasks

(Kumar et al. 2024). On the other hand, T5 adopts a different approach by treating various NLP challenges as a unified text-to-text problem (Mastropaolo et al. 2021), showcasing its versatility in handling multiple tasks within a single framework.

Furthermore, LMs and LLMs differ not only in size but also in their training processes and applications. LLMs require significantly greater computational resources and can be applied to a wider range of NLP tasks, including those that involve deeper contextual reasoning and long-term dependencies. Each of these models, with its own unique approach, collectively represents the evolution of modern NLP, where theoretical advancements translate into practical applications, enabling the extraction of complex relations and semantic nuances from textual data.

3 Methodology

To ensure that we focus on cutting-edge applications of RE in natural language processing, we have limited our survey to the most recent papers presented at the ACL conferences. These conferences are widely recognized as one of the most important and prestigious conferences in the field of NLP. Specifically, we focus on ACL, NAACL, AACL, and EACL. This review was conducted from September 2023 to January 2024. During this period, the most recent conferences were ACL 2023, AACL 2023, and EACL 2023, marking the endpoint of our survey. We started in 2020, the first year with the significant incorporation of language models like BERT, which was introduced at NAACL 2019. It is important to note that ACL is the most prominent conference in our survey, as AACL, NAACL, and EACL are held biennially.

To narrow down our selection, we searched for papers containing the phrase "Relation Extraction" in either the title or abstract. This was our only search query, designed to be as broad as possible to capture a high volume of papers within the RE domain. This search resulted in a set of papers screened based on their relevance to our research question and quality. Specifically, we only included papers that presented novel approaches or significant advancements in RE techniques using state-of-the-art language models. We exclude papers that are not directly related to the traditional problem of RE and venture into novel domains, such as temporal RE (Mathur et al. 2021; Cui et al. 2021; Tan et al. 2023) or papers with a particular focus on NER (Ye et al. 2022). We also omitted papers that lack a foundation in language models or, at the very least, word embeddings (Zhang et al. 2021a; Yu et al. 2022). Our final exclusion criterion involves papers suggesting novel encoding techniques that, while intriguing, lack widespread adoption within the research community. For instance, the approach presented by Vakil and Amiri (2022), which introduces an exciting graph encoding solution, falls into this category. According to our inclusion criteria, the final set comprised 65 papers. For a better understanding, Fig. 1 illustrates the complete process of our inclusion and exclusion criteria.

Given that our research focuses on cutting-edge techniques for RE and aims to derive key insights about the datasets used to validate these models, our review comprised a total of 137 papers. Of these, 92 papers were presented at ACL conferences from 2020 to 2023. As shown in Fig. 1, 11 of these papers do not propose new techniques but instead introduce new datasets. These dataset-focused papers were reviewed alongside an additional 45 papers, resulting in a total of 56 papers dedicated to dataset proposals. The selection of these 45

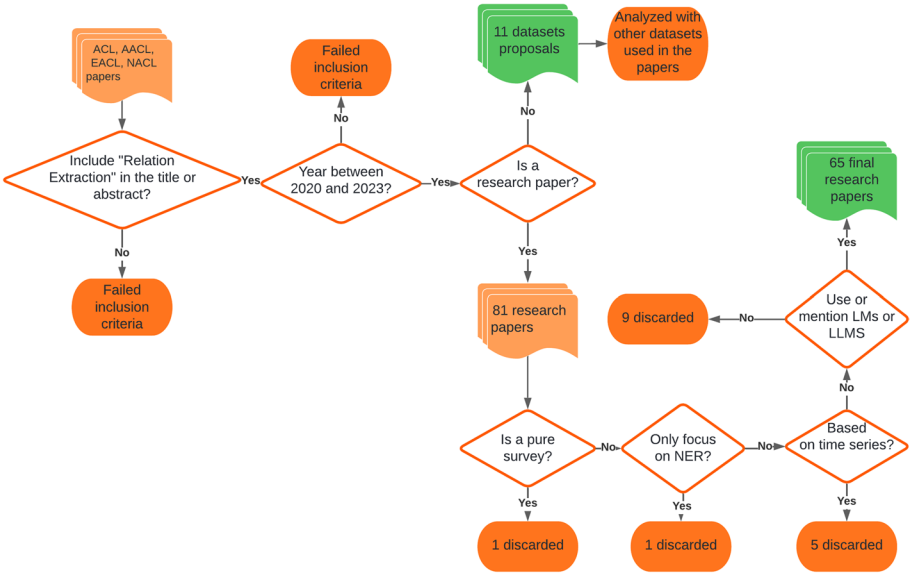

Fig. 1 Graphic explanation of our inclusion/exclusion criteria

Table 1 Survey summary

Category	Count
Total	137
Dataset papers (total)	56
Papers (ACL conferences)	92
Dataset papers (ACL conferences)	11
Research papers (ACL conferences)	81
Passed inclusion criteria	65
Failed inclusion criteria	16

additional papers was based on their datasets being utilized in at least one of the 65 papers included in our final review of ACL conferences. The final set of 65 papers was classified by task, sub-task, and LMs used for RE. A summary of our survey statistics is provided in Table 1.

4 Datasets for RE

We have compiled the most comprehensive datasets used for RE to date. We begin by providing brief descriptions of each dataset. Furthermore, in Sect. 7.3, we describe in more detail the most frequently used datasets for RE. We have categorized datasets into two distinct periods. The first comprises cutting-edge datasets introduced during our review period from 2019 to 2023, while the second includes traditional datasets proposed before this time frame.

4.1 Datasets proposed before the survey period

- **DocRED**: A comprehensive document-level relation extraction dataset, DocRED, is constructed using Wikipedia data. This dataset encompasses two subsets, one manually annotated and the other annotated through distant supervision. DocRED includes 96 diverse relations, including notable ones like *educated_at* and *spouse*. The dataset comprises 155,535 relation mentions (Yao et al. 2019b).
- **SciERC**: A dataset designed for constructing scientific knowledge graphs, covering multi-task annotation for entities, relations, and coreference resolution (Luan et al. 2018).
- **TACRED**: A large-scale dataset annotated via Amazon Mechanical Turk, focusing on common relations between people, organizations, and locations, emphasizing slot filling (Zhang et al. 2017).
- **NYT-10/FB-NYT**: A dataset addressing relation modeling without explicitly labeled text, exploring relationships mentioned in news articles (Riedel et al. 2010).
- **BioCreative V CDR Task Corpus**: A chemical-disease relation extraction resource contributing to biomedical research (Li et al. 2016).
- **ChemProt**: A dataset that extracts various chemical-protein interactions within the biomedical domain (Kim Kjærulff et al. 2012).
- **GDA**: The corpus on gene-disease associations comprises 30,192 titles and abstracts extracted from PubMed articles. These entries have been annotated to identify genes, diseases, and gene-disease associations through distant supervision. A subset of 1000 examples from this corpus forms the test set. The relationships within the corpus are rooted in the genetic association database (Bravo et al. 2015; Becker et al. 2004).
- **FewRel**: A large-scale dataset for few-shot relation classification featuring 100 relations with training, validation, and testing splits (Han et al. 2018).
- **Wiki-ZSL**: A zero-shot learning benchmark for relation extraction utilizing Wikipedia articles, including seen and unseen relations (Levy et al. 2017).
- **Wiki80**: A dataset with 80 relation types designed for evaluating neural relation extraction models (Han et al. 2019).
- **ACE05**: A multilingual training corpus used in the 2005 Automatic Content Extraction (ACE) technology evaluation (Walker et al. 2006).
- **SemEval-2010 Task8**: A benchmark dataset for relation extraction, focusing on multi-way classification of semantic relations between pairs of nominals (Hendrickx et al. 2019).
- **SKE**: A dataset tailored for relation extraction from the Chinese social network Baidu.[1]
- **ERE**: A dataset facilitating comparison of events and relations across different annotation standards (Aguilar et al. 2014).
- **WebNGL**: Crafted using micro-planning data-to-text corpora extracted from existing Knowledge Bases, namely DBpedia, encompassing a total of 21,855 pairs of data and text (Gardent et al. 2017).
- **ADE**: A benchmark corpus for the automated extraction of drug-related adverse effects from medical case reports. The documents undergo a double-annotation process across multiple rounds to ensure uniform and reliable annotations (Gurulingappa et al. 2012).

[1] http://ai.baidu.com/broad/download?dataset=sked.

- **GIDS**: A dataset with 50,000 examples addressing relations related to institutions, places of birth, and attended locations (Jat et al. 2018).
- **Twitter 15**: A dataset containing text and image pairs from Twitter posts utilized for multimodal NER and relation extraction tasks (Lu et al. 2018).
- **Twitter 17**: A dataset comprising text and image pairs from Twitter posts, used for multimodal NER and relation extraction tasks (Zhang et al. 2018).
- **MSCorpus**: A dataset concentrating on materials science synthesis procedures, defining a labeling schema specialized for this domain, including various relations (Mysore et al. 2019).
- **CoNLL**: A bilingual (German and English) dataset specifically designed for NER (Tjong Kim Sang et al. 2003).
- **EventStoryLine**: A dataset designed for narrative understanding and event extraction (Mostafazadeh et al. 2016).
- **Causal-TimeBank**: A dataset for relation extraction based on time series focusing on causal relations within the TimeBank corpus (Mirza and Tonelli 2014).
- **Matres**: A dataset for temporal relation extraction based on time series (Ning et al. 2018).
- **RE-QA**: A dataset with ten folds containing 84 relation types in the train, 12 on the dev, and 24 on the test splits. There is no overlap among the relation types of these splits per fold. Each fold contains 840k sentences in the train, 6k in the dev, and 12k in the test split. Half the sentences are negative examples, and gold questions are provided for each relation type (Levy et al. 2017).
- **KBP37**: A sentence-level relation extraction dataset with 21,046 examples collected from 2010 and 2013 KBP documents and the July 2013 dump of Wikipedia (Zhang and Wang 2015).

4.2 Datasets proposed during the survey period

- **DialogRE**: A dataset derived from dialogues in the Friends TV series, comprising 10,168 subject-relation-object tuples, covering diverse relations such as *'positive_impression'* and *'siblings'*. The dataset utilizes BERT-base-uncased with special tokens to demarcate subject and object boundaries within dialogues (Yu et al. 2020).
- **BioRED**: A biomedical relation extraction dataset featuring multiple entity types (e.g., gene, protein, disease, chemical) and relation pairs (e.g., *'gene-disease'*, *'chemical-chemical*) at the document level, tagged on a set of 600 PubMed abstracts, comprising 4178 relations and 13,351 entities (Luo et al. 2022a).
- **Re-DocRED**: An updated version of the DocRED dataset with re-annotation of 4053 documents to address and rectify issues with false negatives due to incomplete annotation in the original dataset (Tan et al. 2022).
- **BioRelEx**: A dataset comprising over 2000 sentences from biological journals with comprehensive annotations of entities such as proteins, genes, and chemicals, and detailed information about their binding interactions (Khachatrian et al. 2019).
- **FOBIE**: A dataset focusing on biological relations between elements, including relations such as *'trade-off'*, *'argument-modifier'*, and *'not-a-trade-off'*. It is automatically

annotated using a rule-based system and employs word embedding encoding rather than larger language models like BERT (Kruiper et al. 2020).

- **FUNSD**: A dataset for form understanding in noisy scanned documents, comprising 199 fully annotated scanned forms addressing tasks including entity labeling and relation extraction (Jaume et al. 2019).
- **SIMILER**: A dataset consisting of 1.1 million annotated Wikipedia sentences representing 36 relations and 14 languages, created specifically for entity and relation extraction tasks (Seagnti et al. 2021).
- **HISTRED**: A historical document-level relation extraction dataset constructed from Yeonhaengnok, a collection of historical records with bilingual annotations for various named entities and relations (Yang et al. 2023).
- **MultiTACRED**: A multilingual version of the TACRED dataset covering 41 person- and organization-oriented relation types across 12 target languages (Hennig et al. 2023).
- **TACRED-Revisited/TACREV**: A label-corrected version of the TACRED dataset, addressing challenging cases by carefully validating the most difficult 5000 examples from the original dataset (Alt et al. 2020a).
- **RE-TACRED**: A comprehensive analysis and re-annotation of the entire TACRED dataset (Alt et al. 2020b).
- **REDfm**: A filtered and multilingual relation extraction dataset divided into two subsets: SREDfm, which includes over 40 million triplet instances across 18 languages, and REDfm, a smaller human-reviewed dataset focusing on seven languages (Huguet Cabot et al. 2023).
- **NMRE**: A multimodal NER dataset containing text, image, and entity annotations, with 9201 text-image pairs and 15,485 entity pairs across 23 relation categories (Zheng et al. 2021).
- **RISeC**: A dataset focused on procedural text within the cooking domain, offering explicit labels for various relations, including temporal relations and manner descriptions (Jiang et al. 2020).
- **EFGC**: A dataset specializing in cooking with co-reference relations segmented by tools, foods, or actions (Yamakata et al. 2020).
- **WIKIDISTANT**: A dataset derived from Wikipedia, composed of 454 target relations with 1,050,246 elements in training, 29,145 in validation, and 28,897 for testing (Han et al. 2020).
- **DrugCombination-Dataset**: A dataset designed for relation extraction involving diverse drug combinations, comprising 1,634 biomedical abstracts and expert-annotated to extract information regarding drug combination efficacy (Tiktinsky et al. 2022).
- **DWIE**: The Deutsche Welle corpus for Information Extraction (DWIE) contains 501,095 tokens from news articles, illustrating 317,204 relations across 65 types, with annotations for named entities, co-reference resolution, and relation extraction (Zaporojets et al. 2021).

5 Cutting-edge RE techniques based on language models

We employed a two-fold approach to facilitate a thorough analysis of the field's evolution. First, we conducted a task-based analysis, offering a comprehensive examination of the techniques used to address each specific task in relation extraction. This structured approach provides a detailed understanding of how various methods have been applied and evolved across different RE tasks. Second, we organized our survey into chronological tables by year, conference, task, sub-task, and model type. This segmentation enables us to explore the progression of techniques over time, identify emerging trends, and derive valuable insights from the evolving landscape of RE methods.

5.1 Task analysis of cutting-edge RE

In this section, we will discuss how new systems and models are addressing the tasks and challenges of RE that were introduced in Sect. 2.1. We will particularly emphasize the systems and models that make use of LMs and LLMs.

5.1.1 Sentence-level Relation Extraction

In Veyseh et al. (2020), Veyseh et al. used an Ordered-Neuron Long-Short Term Memory Networks (ON-LSTM) (Shen et al. 2018) to infer model-based importance scores for every word within sentences. This paper uses dependency trees to calculate importance scores for sets of words based on syntax. These syntax-based scores are then adjusted to align with the model-based importance scores, creating a balanced integration of syntax and semantics. This approach achieved state-of-the-art performance levels on three distinguished RE benchmark datasets: ACE 2005, related to news; SPOUSE with only one entity type and relation, related to whether two entities regarding people are married; and SciERC, related to scientific abstracts. They used BERT to obtain pre-trained word embeddings for the sentences. Specifically, they use the hidden vectors in the last layer of the BERT-base model (with 768 dimensions) to obtain the pre-trained word embeddings for the sentences. They kept BERT's parameters unchanged in the experiments and compared their proposed model, using BERT embeddings, against others that utilized word2vec embeddings (Mikolov et al. 2013). In this study, word embeddings achieved state-of-the-art results, so it was expected that a trend would emerge in 2020, leveraging their potential. In Yu et al. (2020), Yu et al. introduced DialogRE, a dataset derived from dialogues in the Friends TV series. The dataset contains 10,168 subject-relation-object tuples, encompassing a wide range of relations such as "positive_impression" and "siblings". The paper utilize BERT-base-uncased, employing unique tokens to demarcate subject and object boundaries within the dialogues. Notably, this work pioneers RE in dialogue contexts, conducting experiments against traditional CNN, LSTM, and BiLSTM-based models, often coupled with the GloVe encoder (Pennington et al. 2014).

In sentence-level relation extraction, authors are also developing **unsupervised techniques**. In Yuan and Eldardiry (2021), authors explore unsupervised relation extraction, a technique that does not require labeled data to identify relations between entities in text. The paper proposes a new approach that overcomes the limitations of existing variational autoencoder-based methods. The proposed method enhances training stability and outper-

forms state-of-the-art approaches, such as EType+ (Tran et al. 2020), a simple yet effective model that leverages the combination of entity types within each sentence, on the NYT dataset. The authors use a variational autoencoder to learn a low-dimensional representation of sentences that captures their relation information. They also introduce a novel regularization term, encouraging the model to learn more informative representations. The proposed approach is evaluated on the NYT dataset and compared to four state-of-the-art methods. The results show that the proposed method achieves higher F1 scores than the baselines. HiURE (Liu et al. 2022b) advances unsupervised relation extraction by introducing a contrastive learning framework, which is designed to improve the model's ability to distinguish between similar and dissimilar relational patterns. Contrastive learning is a technique that learns by comparing positive and negative examples, encouraging the model to bring similar instances closer in representation space while pushing dissimilar ones apart. In HiURE, this approach is enhanced with hierarchical exemplars, which represent different levels of relational features in a sentence, allowing the model to capture more nuanced relational information. The framework employs exemplar-wise contrastive learning through an iterative process of integrating HiNCE (Hierarchical Noise Contrastive Estimation) and propagation clustering. This method differentiates itself from other unsupervised RE models by addressing the issue of gradual drift, where the model's representations may shift over time, leading to performance degradation. HiURE avoids this issue by incorporating a hierarchical structure into its learning process, ensuring more stable and effective extraction of relationships.

One of the most debated topics in RE is the choice between **pipeline frameworks** and **joint methods**. As discussed in Sect. 2.1, RE involves three sub-techniques. Pipeline frameworks tackle individual tasks in RE, such as entity recognition, relation identification, and relation classification, separately, while joint methods integrate these tasks into a unified model. This distinction raises questions about which approach is more effective for improving performance and efficiency in relation extraction systems. While most approaches treat each task as a separate problem, Wang et al. (2021) propose a joint system for entity detection and relation classification. The key idea is to view entity detection as a particular case of relation classification. The authors introduce a new input space, a two-dimensional table with each entry corresponding to a word pair in sentences. The joint model assigns labels to each cell from a unified label space. After filling the table, the authors apply a Biaffine layer to predict the final relation labels. The proposed UNIRE model is based on BERT-base-uncased, ALBERT, and SciBERT as encoding depending on the dataset and achieves state-of-the-art performance on the ACE04 and ACE05 datasets, outperforming several baseline models (Zhong and Chen 2021).

Zhong and Chen (2021) introduce a pipeline framework for RE that capitalizes on the results obtained from language models to create a straightforward and highly accurate approach. This approach entails employing distinct models for entity recognition and relation extraction, with the latter model taking entity pairs as its input. The models are trained independently, utilizing cross-sentence context to enhance overall performance. For each task, the authors harnessed the power of pre-trained language models, specifically BERT, ALBERT, and SciBERT, constructing two separate encoders for entity recognition and relation extraction. Both tasks saw significant performance improvements by training entity and relation models independently and having the relation model use the entity model for input features.

The authors in Yan et al. (2022) compared the performance of pipeline and joint approaches to Entity and Relation Extraction (ERE) tasks. The research focused on two specific datasets, ACE2005 and SciERC, which are widely used for evaluating information extraction methods. The authors used eight different methods, including purely sequential methods and four combined models, to see how well they could extract entities and relationships from the text. Their experiments used two pre-trained language models: BERT-base-uncased for the ACE2005 dataset and Seibert-sci-vocab-uncased for the SciERC dataset. The study revealed that while pipeline approaches can yield competitive results, the best-performing joint approach consistently outperformed the best pipeline model across various metrics. This finding underscores the potential advantages of joint models, particularly in leveraging the interdependencies between tasks. The ERE tasks performed in this study operate primarily at the sentence-level, extracting entities and relations from individual sentences within the documents.

Jie et al. (2022) introduced an approach for solving math word problems, which can be considered a sentence-level task. Their method employs explainable deductive reasoning steps to construct target expressions iteratively. Each step involves a primitive operation over two quantities, defining their relation. By framing the task as a complex RE problem, their model significantly outperforms existing baselines as BERT-Tree (Liang et al. 2021) and MultiE&D (Shen and Jin 2020) in accuracy and explainability. The primary objective is to identify relations between numbers, where mathematical expressions represent these relations. The study uses three English and one Chinese dataset, outperforming state-of-the-art methods established by previous works (Koncel-Kedziorski et al. 2016; Wang et al. 2017; Amini et al. 2019; Patel et al. 2021). The authors leverage BERT as the encoder for their proposed model. Specifically, they utilize Chinese BERT and Chinese RoBERTa for the Math23k dataset and BERT and RoBERTa (Liu et al. 2019) for the English datasets. The experimentation involves also multilingual BERT and XLM-RoBERTa (Conneau et al. 2020). The pre-trained models are initialized from HuggingFace's Transformers, with the best-performing model identified as related to RoBERTa.

In Wang et al. (2022b), the authors examine the issue of entity bias in sentence-level relation extraction. The paper's main contribution is the introduction of CORE (Counterfactual Analysis Relation Extraction): a model-agnostic technique explicitly designed to address this problem. The authors showcase that CORE effectively enhances the efficacy and generalization of prevalent RE models by distilling and mitigating biases entity-awarely. The approach involves formulating existing RE models as causal graphs and applying counterfactual analysis to post-adjust biased predictions during inference. Language models, particularly RoBERTa, are significant because of the flexibility with which CORE can be seamlessly integrated into popular RE models to boost their performance and mitigate biases without necessitating retraining. In contrast to approaches incorporating graphs or additional information, (Liang et al. 2022) introduced an innovative method for RE that exclusively extracts multi-granularity hierarchical features from the original input sentences. The model employs hierarchical mention-aware segment attention and global semantic attention to capture features at various levels. These features are then aggregated for relation prediction. While emphasizing the effectiveness of BERT and SpanBERT (Joshi et al. 2020) as encoders, the authors also incorporate a BiLSTM model.

The authors in Zhou and Chen (2022) leveraged RoBERTa as the foundational framework for enhancing the baseline model in sentence-level relation extraction. The paper pres-

ents a dual-fold contribution, marked by two advancements. Firstly, the authors introduce an improved baseline specifically tailored for sentence-level relation extraction, strategically tackling two prominent challenges prevalent in existing models: effective entity representation and handling noisy or ill-defined labels. Secondly, the study establishes and underscores the efficacy of pre-trained language models, particularly RoBERTa, in sentence-level relation extraction. This efficacy is substantiated by achieving state-of-the-art results on the TACRED dataset and its refined iterations. Notably, the proposed model stands out as a solution that adeptly addresses two critical issues impacting the performance of contemporary RE models: entity representation and dealing with noisy or ill-defined labels.

In 2022, LLMs such as GPT (Radford et al. 2018) became mainstream. Yang and Song (2022) introduced a prompt-based fine-tuning approach to enhance relation extraction. The foundation of their method involved utilizing RoBERTa large as the base model, subjected to fine-tuning on the training sets of respective datasets through the prompt-based fine-tuning methodology. Throughout the fine-tuning process, the authors employed prompts featuring a masked relation label, tasking the model with predicting the obscured label based on the input sentence. A carefully crafted prompt learning curriculum was also implemented to facilitate the model's adaptation to a multi-task setting, introducing tasks of increasing difficulty. The experimental findings underscore the efficacy of the proposed model. Notably, the model attained state-of-the-art results on benchmark datasets, namely TACRED and SemEval 2010. These outcomes affirm the method's prowess in advancing the state-of-the-art in RE tasks.

5.1.2 Document-level Relation Extraction

Document-level RE's most discussed topics are evidence selection, joint and graph models, and new training strategies.

Evidence selection refers to identifying and selecting relevant sentences or text from a document that provides crucial information to determine relationships between entities. In Huang et al. (2021a), the authors present E2GRE, a model designed for document-level relation extraction. E2GRE tackles the challenge of handling relations spanning multiple sentences by jointly extracting relations and their supporting evidence sentences using BERT as an input encoder. BERT plays a central role in the E2GRE framework by contextualizing document text and entity mentions. The last four layers of BERT are used for RE and evidence prediction. Notably, E2GRE enhances BERT's attention mechanism to focus on relevant context, using attention probabilities as supplementary features for evidence prediction. While E2GRE achieves state-of-the-art results in evidence prediction on the DocRED dataset, it needs to attain the same level of accuracy in relation classification. The model uses RoBERTa (Liu et al. 2019) as an input encoder, broadening its capabilities in document-level relation extraction.

In Nan et al. (2020), the authors introduce an approach to document-level relation extraction. This model sets itself apart from existing methods in the field by employing a refinement strategy that progressively aggregates pertinent information and evidence to facilitate multi-hop reasoning. This distinctive strategy, Latent Structure Refinement (LSR), contrasts traditional models reliant on fixed syntactic trees. Instead, LSR dynamically learns the document's structure and utilizes this acquired knowledge to inform predictions. The model's underlying encoding mechanisms draw from both GloVe and BERT, allowing it to achieve

state-of-the-art results when applied to datasets such as CDR (Li et al. 2016), GDA (Han and Sun 2016), and DocRED (Yao et al. 2019b), especially when employing BERT encoding. Similarly to the preceding paper, Huang et al. (2021b) propose a system for document-level RE that focuses on selecting evidence sentences rather than performing RE directly. The authors present a method for heuristically selecting informative sentences from a document that can be used as evidence for identifying the relationship between a given entity pair. The selected evidence sentences can then be input into a BiLSTM-based RE model. The authors show that their method for evidence sentence selection outperforms graph neural network-based methods, as GAIN-GloVe (Weng et al. 2020), for document-level RE on benchmark datasets, including DocRED, GDA, and CDR.

Xu et al. (2022) introduces a paper that complements traditional document-level RE tasks. The framework, named SIEF, can be integrated with established approaches like BERT. SIEF enhances the performance of DocRE by introducing importance scores to sentences, directing the model's attention to evidence-rich sentences. This approach derive consistent and robust predictions. The methodology involves calculating sentence importance scores and implementing a sentence-focusing loss to promote model robustness. The study underscores the synergistic relationship between SIEF and modern language models, such as BERT. Document-level relation extraction has also been explored from a **joint modeling perspective**. Eberts and Ulges (2021) present a comprehensive approach to entity-level RE that covers all stages of the RE process. Their proposed joint model extracts mentions, clusters them into entities, and classifies relations jointly. The joint model can leverage shared parameters and training steps across the different sub-tasks, which improves efficiency and performance compared to a pipeline approach where each sub-task is trained separately. The authors also use a multi-instance learning approach that aggregates information from multiple mentions of the same entity to improve performance. The model incorporates BERT to encode contextual information. The proposed joint model achieves state-of-the-art results on the challenging DocRED dataset, demonstrating the power of joint modeling for entity-level RE.

Zhang et al. (2023) proposed a system, TaG, which adopted a joint approach based on **graph-based** methodologies, incorporating techniques such as table filling, graph construction, and information propagation. To capture rich contextual information, the authors employed BERT and RoBERTa as encoders over the tables, with particular success achieved through leveraging RoBERTalarge. In Xu and Choi (2022) authors introduced COREF, a Graph Compatibility (GC) approach rooted in the bidirectional interaction between coreference resolution and RE. The GC method capitalizes on task-specific traits to actively influence the decisions of distinct tasks, establishing explicit task interactions that avoid the isolated decoding of each task. The authors employed SpanBERT to capture intricate contextual dependencies for encoding,

Document-level relation extraction involves identifying and classifying relationships between entities within entire documents, which presents unique challenges compared to sentence-level extraction. Traditional training and evaluation techniques often fall short because they may need to adequately capture the complexities of long-range dependencies and contextual nuances inherent in document-level data. Standard methods might focus on sentence-by-sentence analysis, overlooking the importance of the document's overall structure and coherence. Consequently, **new training techniques** are required to effectively manage the large-scale and multi-sentence contexts in which entities are situated.

Xiao et al. in (2022), introduced SAIS, an approach employing language models, such as BERT, RoBERTa, and SciBERT, to encode extensive contextual dependencies. This strategic use of language models addresses the complexities of capturing extended contexts and nuanced interactions among entities in document-level relation extraction. The primary contribution of the SAIS method lies in its explicit guidance for models to adeptly capture pertinent contexts and entity types during document-level relation extraction. This emphasis results in enhanced extraction quality and more interpretable predictions from the model.

Chen et al. (2023) introduce a novel perspective by evaluating models based on language understanding and reasoning capabilities. Employing the Mean Average Precision (MAP) metric and subjecting models to RE-specific attacks, the authors shed light on significant disparities in decision rules between state-of-the-art models and human approaches. Notably, these models often depend on spurious patterns while overlooking crucial evidence words, adversely affecting their robustness and generalization in real-world scenarios. Although the paper does not propose new models for RE and does not explicitly use language models, it aligns with our research goal of understanding how language models capture information for relation extraction. Instead, the authors compare and evaluate various baselines, including BERT and RoBERTa, providing valuable insights into their comparative effectiveness.

In 2023, Guo et al. (2023) introduced PEMSCL, to enhance relation prediction accuracy by integrating discriminability and robustness. The approach employs a pairwise moving-threshold loss with entropy minimization, incorporates adapted supervised contrastive learning, and introduces a novel negative sampling strategy. For encoding, the authors used ATLOP (Zhou et al. 2021). Ma et al. (2023) introduced a method called DREEAM. This approach uses evidence information to guide the attention modules to emphasize important evidence. One notable aspect is that the model is self-trained to learn relationships between entities using automatically generated evidence from large amounts of data without requiring explicit evidence annotations. This approach's encoder for evidence retrieval is RoBERTa, which significantly contributes to the system's overall effectiveness. DREEAM currently stands as the most accurate system for DocumentRE.

In Zhao et al. (2023d), a continual document-level relation extraction model is introduced to address the challenges of distinguishing analogous relations and preventing overfitting to stored samples. The model employs a learning framework that integrates a contrastive learning objective with a linear classifier training one. It generates memory-insensitive relation prototypes by combining static and dynamic representations, which helps maintain robustness against noise from stored samples. The model uses memory augmentation to create more training samples for replay, enhancing its adaptability in continual learning scenarios. Focal knowledge distillation is also used to assign greater weights to similar relations, ensuring that the model concentrates on the fine distinctions between them. For encoding, the authors use BERT, leveraging its robust contextual embeddings.

Zhang et al. (2021a) discussed the challenges of extracting information from visually rich documents (VRDs) in their paper (Zhang et al. 2021b). The complexity arises from the need to incorporate both visual and textual features in the extraction process. The authors propose an approach that adapts the affine model used in dependency parsing to the entity RE task and conduct experiments on the FUNSD dataset and a real-world customs dataset to compare different representations of the semantic entity, different VRD encoders, and different relation decoders. They also employ two training strategies, multi-task learning with entity labeling and data augmentation, to further improve their model's performance.

The proposed model outperforms previous baselines, such as LayoutLM (Xu et al. 2020), which leverages BERT for 2D position embedding encoding instead of BiLSTM, on the FUNSD dataset. Additionally, it achieves high performance on the real-world customs dataset. The authors also analyze the performance of different language models, such as BERT and LayoutLM, on their language mixed data and find that encoding layout information into language models significantly enhances the model's ability to understand the spatial relationships and contextual relevance of entities, leading to improved accuracy in relation extraction tasks.

Finally, Yuan et al. (2023) present SENDIR, a model specifically tailored to tackle the complexities of document-level reasoning in event-event relation extraction. This model introduces an approach by learning sparse event representations, enabling effective discrimination between intra- and inter-sentential reasoning. The paper addresses the challenges associated with comprehending lengthy texts. Experimental validation of the model was conducted across three datasets related to event detection: EventStoryLine, Causal-TimeBank, and MATRES. The SENDIR model incorporates language models to enhance its capabilities, employing a BERT-base-uncased in conjunction with a BiLSTM.

5.1.3 Multilingual and multimodal Relation Extraction

Multilingual RE has benefited from several developments in recent years. In Seagnti et al. (2021), the authors introduce a new dataset, SMILER, comprising 1.1 million annotated sentences spanning 14 languages. To tackle this dataset, they propose the HERBERTa model, a BERT-based pipeline integrating two independent BERT models for sequence classification and entity tagging. The authors employ the multilingual BERT model, M-BERT, pre-trained on monolingual corpora in 104 languages and fine-tuned BERT models tailored for specific languages. Notably, the HERBERTa model achieves performance close to state-of-the-art in multilingual relation extraction.

Yang et al. in (2023) introduce a new dataset for historical RE based on the Yeonhaengnok, a collection of historical records written initially in Hanja and later translated into Korean. The dataset contains 5816 data instances and includes ten types of named entities and 20 relations. The authors propose a bilingual RE model to extract relations from the dataset that leverages Korean and Hanja contexts. The model leverages pre-trained language models, including KLUE, a BERT-based model designed explicitly for Korean natural language processing tasks, and AnchiBERT (Tian et al. 2021a), tailored for Hanja. By leveraging both Korean and Hanja contexts, the proposed model outperforms other monolingual models on the HistRED dataset, demonstrating the effectiveness of employing multiple language contexts in RE tasks. Authors in Hennig et al. (2023) introduced MultiTACRED. This extension of the TACRED dataset spans 12 different languages and aims to explore the intricacies of multilingual relation extraction further. Unlike proposing novel models, the paper focuses on meticulously evaluating monolingual, cross-lingual, and multilingual models. This assessment dig into the performance across all 12 languages the dataset covers. Additionally, the study scrutinizes the effectiveness of pre-trained mono- and multilingual encoders, particularly highlighting the role of BERT-based in tackling challenges posed by multilingual natural language processing tasks and cross-lingual transfer learning scenarios. The research in Huguet Cabot et al. (2023) also contributes to multilingual RE by introducing REDfm, a dataset designed for multilingual models. This dataset

encompasses various relation types across multiple languages, offering higher annotation coverage and more evenly distributed classes than current datasets. Additionally, the authors propose a multilingual RE model named mREBEL. As an extension of the REBEL model, mREBEL adopts a seq2seq architecture to extract triplets, encompassing entity types across various languages. The authors conducted pre-training for mREBEL using mBART-50 (Liu et al. 2020) and fine-tuned it on the RED FM dataset. The results demonstrate that mREBEL surpasses the performance of existing models on the REDfm dataset.

In **multimodal RE**, recent innovations have focused on integrating visual and textual information. In Zheng et al. (2023a), authors addressed multimodal entity and RE through an approach that integrates visual and textual information. The model uses techniques such as back-translation and multimodal divergence estimation to exploit a unified transformer. By using convolutional neural networks and BERT-base-uncased as encoders for visual and textual content, the study uses Twitter-2015, Twitter-2017, and MNRE datasets designed explicitly for multimodal tasks, achieving state-of-the-art results. In the same line, (Wu et al. 2023) proposed a graph-based framework for multimodal RE. The framework involves representing the input image and text with visual and textual scene graphs, which are then fused into a unified cross-modal graph. The graph is then refined using the graph information bottleneck principle to filter out less informative features. The model uses latent multimodal topic features to enhance the feature contexts, and then the decoder uses these enriched features to predict the relation label. The model is pre-trained using CLIP (vit-base-patch32), which performed better than BERT on the MRE (Zheng et al. 2021) dataset, comprising 9201 text-image pairs. Hu et al. present in (2023) an innovative method that introduces a strategy for synthesizing object-level, image-level, and sentence-level information. This approach enhances the capacity for reasoning across various modalities, facilitating improved comprehension of similar and disparate modalities. The proposed method underwent testing on the MNRE dataset, achieving state-of-the-art results. BERT-base-uncased served as the encoder for textual information.

5.1.4 Few-shot and low-resource Relation Extraction

The need to enhance model performance in scenarios where labeled data is scarce has led to the development of specialized techniques such as few-shot, low-resource, and zero-shot learning. These approaches are essential for making RE systems more adaptable and practical, especially in real-world applications where obtaining large, annotated datasets is often challenging or impractical. Few-shot and low-resource RE methods address these limitations by enabling models to learn effectively from minimal data, reducing the dependency on extensive manual annotation. Additionally, the incorporation of external knowledge, innovative training strategies, and advanced semantic matching techniques has enhanced these methods, resulting in more precise and broadly applicable models.

In Yang et al. (2021a), the authors address the challenge of inaccurate relation classification in **low-resource data** domains driven by limited samples and a knowledge deficit. They introduced the ConceptFERE scheme, an advancement in few-shot RE algorithms. This scheme incorporates a concept-sentence attention module that selects the most relevant concept for each entity by calculating semantic similarity between sentences and ideas. A self-attention-based fusion module also bridges the gap between concept and sentence embeddings from different semantic spaces. The paper adopts BERT as the foundational

model for relation classification. This BERT model undergoes pre-training on a large text corpus and fine-tuning on the FewRel (Han et al. 2018) dataset. The authors initialize the BERT parameters with BERT-base-uncased. Teru (2023) tackles the low-resource scenario in RE by addressing the challenge of acquiring high-quality labeled data. The study explores the impact of data augmentation as a strategy to enhance data quality and improve model performance. The author leverages pre-trained translation models to generate diverse data augmentations for a given sentence in German and Russian. Then (Teru 2023) uses lexically constrained decoding strategies to obtain paraphrased sentences while retaining the head and the tail entities. The proposed REMIx system uses a BERT-base-uncased as the encoder. The paper's main contribution is demonstrating the effectiveness of data augmentation and consistency training for semi-supervised relation extraction, achieving competitive results on four benchmark datasets.

Building on previous advances, the creation of realistic benchmarks and domain-specific datasets, such as those focusing on biomedical and document-level RE, has been crucial in testing the applicability of these approaches in diverse, **low-resource environments**. Popovic et al. (2022) introduced a novel benchmark named FREDo, which is explicitly designed for Few-Shot Document-Level RE (FSDLRE). Unlike existing benchmarks that primarily target sentence-level tasks; the authors claim FREDo provides a more realistic testing ground. The authors proposed two approaches to address FSDLRE tasks that incorporates a modification to the state-of-the-art few-shot RE approach, MNAV. While results demonstrate the superiority of the proposed methods over the baseline built on BERT, they also reveal the ongoing challenges associated with realistic cross-domain tasks, such as domain adaption or the high proportion of negative samples. The study underscores the significance of realistic benchmarks and emphasizes the necessity for substantial advancements in few-shot RE approaches to make them applicable in real-world, low-resource scenarios. Contributing to the domain of low-resource domains, Tiktinsky et al. (2022) introduce an expert-annotated dataset, DrugCombination-Dataset, tailored for the intricate task of N-ary RE involving drug combinations from the scientific literature. To offer more robust results, they conducted a comparative study, pitting various baselines against prominent scientific language models, namely SciBERT (Beltagy et al. 2019), BlueBERT (Peng et al. 2019), PubmedBERT (Gu et al. 2021), and BioBERT (Lee et al. 2020). The results underscore the effectiveness of domain-adaptive pre-training, with PubmedBERT demonstrating the most robust performance among the considered models.

In Xu et al. (2023b), Xu et al. introduced a novel approach called NBR (Natural Language Inference-based Biomedical Relation extraction). NBR addresses the challenges of annotation scarcity and instances without pre-defined labels by converting biomedical RE into a natural language inference formulation, providing indirect supervision, and improving generalization in low-resource settings. The approach also incorporates a ranking-based loss to calibrate abstinent instances and selectively predict uncertain cases. Experimental results demonstrate that NBR outperforms existing state-of-the-art methods on three widely used biomedical RE benchmarks: ChemProt, DDI, and GAD. The paper leverages BioLinkBERTlarge, a BERT variation designed explicitly for biomedical text, to enhance the performance of the proposed NBR approach. By fine-tuning BioLinkBERTlarge on the NBR task, the paper demonstrates the effectiveness of leveraging domain-specific pre-training for improved biomedical RE in low-resource environments. The research in Zhou et al. (2023) proposes a method that pre-trains and fine-tunes the RE model using consistent

objectives based on contrastive learning. Contrastive learning often results in one relation forming multiple clusters in the representation space. The authors introduce a multi-center contrastive loss that facilitates the formation of various clusters for a single relation during pre-training, aligning it more effectively with the subsequent fine-tuning process. Notably, PubmedBERT and BERT-base-uncased are the chosen encoders for pre-training and fine-tuning. Finishing with low-resource scenarios, (Xu et al. 2023a) introduces S2ynRE, a system designed to address the challenge of low-resource RE by harnessing the power of LLMs such as GPT-2 and GPT−3.5. These LLMs generate synthetic data and adapt to the target domain, automatically producing substantial amounts of coherent and realistic training data. The proposed algorithm employs a two-stage self-training approach, iteratively and alternately learning from synthetic and golden data. The effectiveness of the system is demonstrated across five diverse datasets. The paper uses BERT for baseline comparison and encoding, showcasing that combining LLMs surpasses the current state-of-the-art in relation extraction.

In the realm of **few-shot learning**, Brody et al. in (2021) tackle the challenge of performance variability across relation types in few-shot relation classification models. While these models have demonstrated impressive results, their reliance on entity-type information makes distinguishing between relations involving similar entity types challenging. To address this limitation, the authors propose modifying the training routine to enhance the models' ability to differentiate such ties. The suggested enhancements augment the training data with relations involving similar entity types. Through evaluations on the FewRel 2.0 dataset, the paper demonstrates that these modifications substantially improve performance on unseen relations, achieving up to a 24% enhancement under a P@50 problem (precision with 50 examples). The study utilizes BERT as a pre-trained language model to initialize the few-shot relation classification models.

The creation of relation prototypes is explored in Han et al. (2021). The authors propose few-shot RE (FSRE), an approach based on hybrid prototypical networks and relation-prototype contrastive learning. This method leverages entity and relation information to improve the model's generalization ability. FSRE achieves state-of-the-art results on two datasets, FewRel 1.0 and 2.0, and outperforms existing models by a significant margin. The authors also use BERT as the encoder to obtain contextualized embeddings of query and support instances. In Liu et al. (2022a), Liu et al. proposed an approach to low-shot RE that unifies zero-shot and few-shot RE tasks. The method, Multi-Choice Matching Networks (MCMN) with triplet-paraphrase pre-training, achieves state-of-the-art performance on three low-shot RE tasks. The datasets used in the experiments are FewRel and TACRED, widely used benchmarks for low-shot RE. The paper assimilates BERT by using it as a backbone model for MCMN and pre-training it with triplet-paraphrase. The experimental results show that MCMN with triplet-paraphrase pre-training outperforms previous methods in all three low-shot RE tasks and achieves state-of-the-art performance.

Qin and Joty in (2022) investigate a concept called continuous few-shot Relation Learning (CFRL). CFRL is relevant to real-life situations where there is usually enough data available for an established task, but only a small amount of labeled data for new tasks that come up. Acquiring large labeled datasets for every new relation is expensive and sometimes impractical for quick deployment. CFRL aims to quickly adapt to new environments or tasks by exploiting previously acquired knowledge. The proposed model is based on embedding space regularization and data augmentation. The authors assimilate BERT and

other language models by using them as feature extractors for the input sentences. They show that their method generalizes to new few-shot tasks and avoids catastrophic forgetting of previous tasks. The experiments are conducted on three datasets, and the results demonstrate that the proposed method outperforms state-of-the-art approaches. Notably, it surpasses EMR (Wang et al. 2019), which was the leading method in continual RE through lifelong learning, across all three datasets. The authors also conduct ablation studies to show the effectiveness of each component of their method.

Moving to the complex area of **zero-shot learning**, the authors of Zhao et al. (2023a) present a fine-grained semantic matching method tailored for zero-shot relation extraction, explicitly capturing the matching patterns inherent in relational data. Additionally, the paper introduces a context distillation method designed to mitigate the negative impact of irrelevant components on context matching. The effectiveness of the proposed method is evaluated on two datasets: Wiki-ZSL and FewRel. For the encoder model, BERT-base-uncased is employed as the input instance encoder. Comparative assessments are conducted against several state-of-the-art methods, including REPrompt, a competitive seq2seq-based ZeroRE approach that utilizes GPT-2 to generate pseudo data for new relations during fine-tuning. The results demonstrate that the proposed method surpasses the performance of all others in the comparison. In Gururaja et al. (2023), the authors tackle challenges in few-shot RE across domains. The study invistegate the influence of linguistic representations, specifically syntactic and semantic graphs derived from abstract meaning representations and dependency parses, on the performance of RE models in few-shot transfer scenarios. Employing BERT-base-uncased, the authors extract embeddings for each entity's constituent tokens and max-pool them into embeddings. These embeddings initialize the feature vectors of nodes in the linguistic graph. The graph-aware model integrates these BERT-derived graph-based features, enhancing RE performance in few-shot scenarios. Validation involves two datasets: one focusing on cooking relations, especially between food and elaborations in recipes, and the other related to materials.

In Najafi and Fyshe (2023), Najafi and Fyshe explore the domain of zero-shot relation extraction. The authors propose an approach that elude the need for manually crafted gold question templates by generating questions for unseen relations. The critical advancement lies in the introduction of the OffMML-G training objective. This objective fine-tunes question-and-answer generators specifically tailored for ZeroShot-RE. The result is the generation of semantically relevant questions for the answer module based on the given head entity and relationship keywords. The T5 model serves as the answer generator, having undergone pre-training and fine-tuning on five distinct question-answering datasets. Hu et al. (2021b) presents an approach named Gradient Imitation Reinforcement Learning (GIRL) tailored for Low Resource Relation Extraction. GIRL's primary challenge is extracting semantic relations between entities from text when confronted with a scarcity of labeled data. To address this limitation, GIRL uses gradient imitation to create pseudo labels with fewer biases and errors. It also improves the relation labeling generator within a reinforcement learning framework. The results demonstrate GIRL's superiority over several state-of-the-art methods on benchmark datasets, SemEval and TACRED, achieving competitive performance with significantly less labeled data. The authors employed the BERT default tokenizer with a max-length of 128 for data preprocessing to encode contextualized entity-level representations for the relation labeling generator.

Chen and Li in (2021) address zero-shot RE by leveraging text descriptions of both seen and unseen relations to learn attribute vectors as semantic representations. Their strategy enables accurate predictions of unseen ties. The significant influence of BERT on ZS-BERT underscores the power of leveraging sentence-BERT's (Reimers and Gurevych 2019) contextual representation learning capabilities for encoding input sentences and relation descriptions. Notably, ZS-BERT ranks as the fifth most accurate system for zero-shot relation classification to date, underscoring the enduring impact of BERT as a powerful tool in relation extraction.

Finally, while not directly centered on relation extraction, (Arcan et al. 2022) exploit the capabilities of RE for constructing a chatbot using a pertinent corpus of language data. The application was tailored to industrial heritage in the 18th and 19th centuries, focusing on the industrial history of canals and mills in Ireland. The authors curated a corpus from relevant newspaper reports and Wikipedia articles, employed the Saffron toolkit to extract pertinent terms and relations within the corpus, and utilized the extracted knowledge to query the British Library Digital Collection and the Project Gutenberg library. Although the paper does not explicitly mention BERT, it is the first in our survey to reference the T5 model. The authors suggest that leveraging the capabilities of T5 in future enhancements could lead to further improvements in the system's performance. By utilizing T5's architecture, designed for a wide range of NLP tasks, the model can better capture complex relationships and improve its ability to generalize across different relation types, ultimately addressing the limitations observed in current few-shot models.

5.1.5 Distant supervision Relation Extraction

Distant supervision for RE provides a scalable and efficient way to generate large amounts of labeled training data, which is often scarce and expensive to obtain manually. In traditional supervised learning, high-quality labeled data is required for training models. However, manually annotating such data for RE tasks is labor-intensive, time-consuming, and costly, particularly for large and diverse datasets. Distant supervision automates this process by aligning existing structured data from knowledge bases with unstructured text, generating training examples without human intervention. Distant supervision techniques often generate training data by automatically aligning a knowledge base with unstructured text, which can lead to inaccuracies. These inaccuracies arise in the form of false positives (incorrectly labeled data) and false negatives (missing relations) due to the incompleteness of the knowledge base. Noisy data can degrade model performance by introducing misleading patterns and errors during training, resulting in poor generalization and lower accuracy. Filtering or mitigating noisy labels helps models capture true relational patterns, enhancing robustness and improving relation extraction. This is crucial for RE tasks, as accurate relationship identification directly impacts downstream applications. Based on this motivation, most papers involving distant supervision for relation extraction focus on the necessity of **techniques to address noisy data** and **mitigate the impact of negative examples**.

In Hao et al. (2021), the authors addressed the problem of false negatives in distant supervision for relation extraction. The proposed two-stage approach leverages the memory mechanism of deep neural networks to filter out noisy samples. It utilizes adversarial training to align unlabeled data with training data and generate pseudo labels for additional supervision. The approach achieves state-of-the-art results on two benchmark datasets,

NYT10 and GIDS, outperforming several baseline models, including BERT. Sui et al. in (2021) addressed the issue of label noise in distantly supervised relation extraction, which often results in inaccurate or incomplete outcomes. To mitigate this challenge, the authors introduce a denoising method based on multiple-instance learning. This method involves using multiple platforms to identify reliable sentences collectively. It is implemented within a federated learning framework, ensuring data decentralization and privacy protection by separating model training from direct access to raw data. In this study, BERT and RoBERTa are used as the encoders.

In the same line, Xie et al. (2021) focus on the problem of missing relations caused by the incompleteness of the knowledge base (false negatives) rather than reducing wrongly labeled relations (false positives). To address this issue, they propose a pipeline approach called RERE using BERT for encoding, which first performs sentence classification with relational labels and then extracts the subjects/objects. Furthermore, the proposed method, RERE, consistently outperforms existing approaches over the NYT dataset, even when learned with many false positive samples. However, the authors also experimented with a Chinese dataset (ACE05), using RoBERTa for encoding, but achieved poor performance. Sun et al. in (2023) introduces a framework, UG-DRE, which employs uncertainty estimation technology to guide the denoising of pseudo labels in distant supervision data. The proposal incorporates an instance-level uncertainty estimation method, gauging the reliability of pseudo labels with overlapping relations. Dynamic uncertainty thresholds are introduced for different types of ties to filter high-uncertainty pseudo labels. The authors employ BERT-base-uncased to capture contextual representations of documents and entities, facilitating the generation of pseudo labels and the denoising process. Results illustrate that using the discussed stages of uncertainty estimation, dynamic uncertainty threshold creation, and the iterative e-labeling strategy, the UG-DRE framework effectively mitigates noise in distant supervision data, enhancing performance in document-level RE tasks.

In addition to research addressing noisy and inaccurate data, we also found a line of work focused on **enhancing distant supervision** for relation extraction by incorporating external knowledge, **fine-grained entity types**, and **innovative training** strategies.

In Christopoulou et al. (2021), the authors introduced a multi-task probabilistic approach for distantly supervised RE, employing a variational autoencoder. The paper's primary contribution lies in integrating Knowledge Base priors into the variational autoencoder framework to enhance sentence space alignment and improve interpretability. To assess the proposed approach's effectiveness, the authors evaluated two benchmark datasets, NYT10 and WikiDistant. As an encoder, the authors leveraged a BiLSTM model in conjunction with 50-dimensional embeddings provided by GloVe. In Dai et al. (2021a), the authors propose to leverage a Universal Graph (UG) that contains external knowledge to provide additional evidence for relation extraction. They introduce two training strategies: Path Type Adaptive Pretraining, which enhances the model's adaptability to various UG paths by pre-training on diverse path types, and Complexity Ranking Guided Attention, which enables the model to learn from both simple and complex UG paths by ranking their relevance and guiding attention accordingly. The paper evaluates the proposed framework on two datasets: a biomedical dataset (crafted by the authors using data from the Unified Medical Language System and Medline) and the NYT10 dataset. The results show that the proposed methods outperform several baseline approaches, including the distantly supervised method SENT+KG introduced in Dai et al. (2019).

In contrast to prior research, Hu et al. (2021a) challenges the assumption of distant supervision by proposing that the relation between entity pairs should be context-independent. This assumption introduces context-agnostic label noises, leading to a phenomenon known as gradual drift. Addressing this challenge, the authors propose MetaSRE, a method that leverages unlabeled data to enhance the accuracy and robustness of RE by generating quality assessments on pseudo labels. MetaSRE employs two networks: the Relation Classification Network (RCN) and the Relation Label Generation Network (RLGN). The RCN is trained on labeled data, while the RLGN generates high-quality pseudo labels from unlabeled data. Subsequently, these pseudo labels train a meta-learner that can adapt to new domains and tasks. Experimental evaluations conducted on two public datasets, SemEval 2010 Task 8 and FewRel, demonstrated that MetaSRE surpassed other state-of-the-art methods in accuracy and robustness. Authors fine-tuned BERT on labeled data, utilizing it to generate features for the RCN. BERT was employed to generate embeddings for the unlabeled data, which the RLGN then used to create pseudo labels. In Shahbazi et al. (2020), Shahbazi presents a system that transforms entity mentions into fine-grained entity types (FGET), enhancing precision and explainability. Notably, it draws on typical relations from the FB-NYT dataset (Riedel et al. 2010), such as "place lived" and "capital". This paper relies on traditional word embeddings like skip-gram, showcasing their enduring efficacy in relation extraction.

Finally, Ma et al. (2021) proposes an approach for sentence-level distant RE using negative training (NT) and presents a framework called SENT that filters noisy data and improves performance for downstream applications. The approach is based on the idea that selecting complementary labels of the given label during training decreases the risk of providing noisy information and prevents the model from overfitting the noisy data. The model trained with NT can separate the noisy data from the training data, significantly protecting the model from noisy information. The paper uses BERT to encode the input sentences and achieve state-of-the-art performance on TACRED and NYT-10 datasets.

5.1.6 Open Relation Extraction

OpenRE has seen significant advancements with approaches designed to address limitations of traditional clustering-based methods and enhance relation discovery in open settings. The problem with clustering-based methods is that they may need help to capture the advanced contextual information stored in vectors, leading to similar relations with different relationships in the same cluster. In Zhao et al. (2021), Zhao et al. leverage clustering methods and propose OpenRE, a method that addresses existing clustering-based approaches' limitations. Specifically, the authors propose a relation-oriented clustering method that leverages labeled data to learn a relation-oriented representation. The technique minimizes the distance between instances with the same relation and gathers instances towards their corresponding relation centroids to form the cluster structure. The authors demonstrate their method's effectiveness on two datasets, FewRel and TACRED, achieving a 29.2% and 15.7% reduction in error rate compared to current state-of-the-art methods. The proposed method also leverages BERT as the encoder.

Zhao et al. (2023b) introduces actively supervised clustering for Open Relation Extraction. This approach involves alternating between clustering learning and relation labeling. The aim is to furnish essential guidance for clustering while minimizing the need for addi-

tional human effort. The paper proposes an active labeling strategy to discover clusters associated with unknown relations dynamically. In their experiments, the authors use TACRED and FewRel. Notably, they also incorporate the FewRel-LT dataset, an extended version of FewRel featuring a more diverse distribution of relations particularly in the long taiL, the paper references using BERT as an encoder. As a follow-up, (Zhao et al. 2023c) introduces a method that tackles the issue of encountering unknown relations within the test set. Using BERT as an encoder to recognize known relations, the authors introduced a new relation, NOTA (none-of-the-above), to capture and understand previously unlearned relations.

5.1.7 Multi-task

Most of the studies related to multi-task RE address the challenges of relation extraction at either the document or sentence-level, particularly in low-resource settings. Most propose techniques to bridge the gaps in data availability and improve performance in these challenging scenarios. We also identified techniques that are effective across different levels of granularity, such as document-level and sentence-level relation extraction.

Tian et al. (2021b) incorporate dependency-type information to improve relation extraction further. By including the syntactic instruction among connected words, the proposed Attentive Graph Neural Network (A-GCN) model can better distinguish the importance of different word dependencies and improve the accuracy of relation extraction. The authors introduce type information into the A-GCN modeling to incorporate this information, effectively enhancing the model's performance. Furthermore, they used BERT as the model's encoder and introduced unique tokens to annotate the entities in the text, further improving the model's performance. Overall, their approach demonstrates the effectiveness of leveraging dependency trees and dependency types for relation extraction and achieves state-of-the-art performance on SemEval-2010 task 8 (Hendrickx et al. 2019) and ACE05 (Walker et al. 2006) datasets.

The traditional feature extraction methods often extract features sequentially or in parallel. This process can result in suboptimal feature representation learning and limited task interaction. In the domain of **joint models**, Yan et al. (2021) introduce an approach to joint entity and RE utilizing a partition filter network. Yan et al.'s partition filter network addresses this issue by generating task-specific features, enabling more effective two-way interaction between tasks. This approach ensures that features extracted later maintain direct contact with those extracted earlier. The authors conducted extensive experiments across six datasets, demonstrating that their method outperformed other baseline approaches as PURE (Zhong and Chen 2021) in some datasets. BERT-base-cased, ALBERTxxlarge-v1, and sciBERT-sci-vocab-uncased were employed as encoders for NER and RE partitions, but the main component in the classification stage is an LSTM. Continuing with joint methodologies, (Zheng et al. 2023b) introduce Jointprop, a comprehensive graph-based framework designed for joint semi-supervised entity and relation extraction, specifically tailored to address the challenges posed by limited labeled data. By leveraging heterogeneous graph-based propagation, Jointprop adeptly captures global structural information among individual tasks, allowing for a more integrated approach to understanding the relationships between entities and their corresponding relations. This framework facilitates the propagation of labels across a unified graph of labeled and unlabeled data and exploits interactions within the unlabeled data to enhance learning. Experimental results underscore the effec-

tiveness of this approach, producing state-of-the-art performance on renowned datasets, including SciERC, ACE05, ConLL, and SemEval, with notable improvements in F1 scores for both entity recognition and relation extraction tasks. The paper discusses using language models to obtain contextualized representations for each token. Although the specific language model used is not mentioned, the paper suggests that BERT could be a potential model for creating these representations. This implies that Jointprop could gain from the detailed contextual embeddings offered by transformer-based models, thus improving its performance in settings with limited resources.

Eyal et al. (2021) confronts the persistent challenge of acquiring a substantial volume of training data for relation extraction. The authors introduce a bootstrapping process for generating training datasets for relation extraction, specifically designed to be swiftly executed even by individuals without expertise in NLP. Their approach leverages search engines across syntactic graphs to procure positive examples, identifying sentences syntactically akin to user-input examples. This method is applied to relations sourced from TACRED and DocRED, demonstrating that the resultant models exhibit competitiveness with those trained on manually annotated data and data obtained through distant supervision. To further enhance the dataset, the authors employ pre-trained language models like BERT and RoBERTa for data augmentation. This strategy involves generating additional relation examples using GPT-2 and manual annotation. In particular, the authors modify the text by substituting the entity pair with unique tokens, feeding the altered text as input to a BERT model. The relation between the two entities is captured by concatenating the final hidden states corresponding to their respective start tokens. This representation undergoes classification through a dedicated head, and the model is fine-tuned for relation classification.

A notable challenge in RE lies in the variation of relation types across different datasets, complicating efforts to amalgamate them for training a unified model. To tackle this issue, Fu and Grishman (2021) propose a method that employs prototypes of relation types to discern their relatedness within a multi-task learning framework. These prototypes are constructed by selecting representative examples of each relation type and using them to augment related types from a distinct dataset. The authors conduct experiments using two datasets featuring similar relation schemas: the ACE05 and ERE dataset (Aguilar et al. 2014). While the paper mentions using an encoder, it does not specify the language model leveraged in the experiments.

5.2 Observations

Regarding models, we identified four main groups: those employing LSTMs, those utilizing CNNs and graph-based approaches, and the largest group leveraging transformers like BERT. It is important to note that many works incorporate language models like BERT at some stage for encoding, with classification often performed using other techniques, such as LSTMs. However, in this section, we categorized each work based on its primary component. The presence or absence of BERT as an encoder or as part of the main RE component is detailed in Table 8.

Additionally, we observed that not all works adhere to the entire three-stage RE pipeline, with some focusing solely on entity detection or classifying pre-tagged entities within a dataset. To present our findings concisely, we have organized all reviewed works into tables categorized by conference and year, offering a comprehensive year-conference-task

perspective on the cutting-edge developments in RE. It is also important to note that many reviewed papers do not explicitly specify how many stages of the relation extraction process they perform. We have inferred this information based on their models, methodologies, and reported results in such cases. Table 2 presents a summary of insights from the ACL conference, while Table 3 covers insights from the ACL satellite conferences. Finally, Table 4 offers a comprehensive overview of the number of papers that utilized each model or addressed each task. This summary allows us to derive key insights into trends over the years.

One of the key insights from the trend analysis is related to the models used. In the first 2 years of conferences, 2020 and 2022, we observe a trend toward using various models, including CNNs and LSTMs, though they are less prevalent. By 2022 and 2023, however, the focus has shifted to transformer-based techniques, highlighting their growing dominance and effectiveness.

When examining the subtasks, it is clear that relation identification is the least frequently addressed. Most datasets are tagged with pre-identified relations and are used primarily for relation classification. This subtask, the primary focus in the literature, remains the most relevant component of relation extraction techniques.

Examining the tasks and challenges within relation extraction reveals some compelling insights. Specifically, document-level relation extraction has gained significant prominence in the past 2 years, mainly due to the rise of transformer-based techniques capable of processing more intricate and extensive contexts. Data show that 54% of papers addressing this issue utilize transformer architecture, supporting this trend.

The trend is even more pronounced in few-shot and low-resource scenarios, where the generalizability of language models offers the best solutions. Nearly 100% of the papers in these areas employ transformer-based approaches, reflecting a broader shift towards these techniques in recent years. This shift is likely tied to the widespread adoption and democratization of language models.

These observations are further supported by our findings in Tables 6 and 7, highlighting the best models' performance in these domains.

6 Large language models vs language models in RE

We reviewed various baseline approaches in the literature and compared the effectiveness of different language models for encoding relations. We surveyed and analyzed existing studies that used various language models, including GPT-3, T5, and traditional models like BERT or RoBERTa. Our aim is two-fold: firstly, to assess the relative performance of different model categories in the challenging task of RE, and secondly, to evaluate whether the emergence of LLMs represents a significant advancement over traditional models in performance. Our study dig into Few-Shot relation extraction, document-level relation extraction, and traditional relation extraction. We have selected the top five models for each benchmark dataset associated with these tasks: FewRel, DocRed, and TACRED. This approach allows us to capture various methodologies and performances across distinct facets of RE challenges. Table 5 presents the results for TACRED. For document-level relation extraction on the DocRED dataset, the outcomes are detailed in Tables 6 and 7 compiles the results for few-shot relation extraction using the FewRel dataset.

Table 2 Summary of key papers in ACL categorized by model group and task

Paper	Conference	Year	Model group	Task/challenge	Subtask
Shahbazi et al. (2020)	ACL	2020	CNN	Distant supervised RE	Relation classification
Nan et al. (2020)	ACL	2020	Graph based	Document level	Relation classification
Yu et al. (2020)	ACL	2020	Transformer	Sentence level	Relation identification + relation classification
Tian et al. (2021b)	ACL	2021	CNN	Sentence level/multi-task	Relation classification
Ma et al. (2021)	ACL	2021	LSTM	Distant supervised RE	Relation identification + Relation classification
Huang et al. (2021b)	ACL	2021	LSTM	Document level	Relation classification
Xie et al. (2021)	ACL	2021	Transformer	Distant supervised RE	NER + Relation classification
Yang et al. (2021b)	ACL	2021	Transformer	Few shot/low resource	relation classification
Huang et al. (2021a)	ACL	2021	Transformer	Document level	Relation classification
Wang et al. (2021)	ACL	2021	Transformer	Sentence level	NER + Relation identification + relation classification
Qin and Joty (2022)	ACL	2022	Transformer	Few shot/low resource	relation classification
Liu et al. (2022a)	ACL	2022	Transformer	Few shot/low resource	relation classification
Jie et al. (2022)	ACL	2022	Transformer	Sentence level	NER (quantities) + relation classification
Gururaja et al. (2023)	ACL	2023	Graph based	Few shot/low resource	relation classification
Zheng et al. (2023b)	ACL	2023	Graph based	Multi-task	NER + relation classification
Wu et al. (2023)	ACL	2023	Graph based	Multilingual and multimodal Relation Extraction	NER + Relation identification + relation classification
Xu et al. (2023b)	ACL	2023	Transformer	Few shot/low resource	relation classification
Hennig et al. (2023)	ACL	2023	Transformer	Multilingual and multimodal Relation Extraction	Relation classification
Huguet Cabot et al. (2023)	ACL	2023	Transformer	Multilingual and multimodal Relation Extraction	NER + Relation classification
Zhao et al. (2023b)	ACL	2023	Transformer	Open RE	Relation classification
Zhao et al. (2023b)	ACL	2023	Transformer	Document level	Relation classification
Zhao et al. (2023a)	ACL	2023	Transformer	Few shot/low resource	relation identification + relation classification

Table 2 (continued)

Paper	Conference	Year	Model group	Task/challenge	Subtask
Zheng et al. (2023a)	ACL	2023	Transformer	Multilingual and multimodal Relation Extraction	NER + Relation classification
Xu et al. (2023a)	ACL	2023	Transformer	Few shot/low resource	NER + relation identification + relation classification
Zhao et al. (2023c)	ACL	2023	Transformer	Open RE	Relation identification + relation classification
Zhang et al. (2023)	ACL	2023	Transformer	Document level	NER + Relation classification
Zhou et al. (2023)	ACL	2023	Transformer	Few shot/low resource	NER + relation classification
Sun et al. (2023)	ACL	2023	Transformer	Distant supervised RE	Relation classification
Yuan et al. (2023)	ACL	2023	Transformer	Document level	Relation classification
Hu et al. (2023)	ACL	2023	Transformer	Multilingual and multimodal relation extraction	NER + Relation identification + relation classification
Zhao et al. (2023d)	ACL	2023	Transformer	Document level	Relation classification
Yang et al. (2023)	ACL	2023	Transformer	Multilingual and multimodal relation extraction	Relation classification

Examining the aggregated results reveals a clear dominance of BERT-based approaches, with BERT and RoBERTa collectively representing a substantial 75% of the primary outcomes. In contrast, LLMs contribute merely 25% to the overall results. A temporal analysis unveils intriguing trends, particularly in the latest year, 2023. This year has proven pivotal, delivering state-of-the-art results in two out of the three experiments. Notably, BERT and RoBERTa continue to secure the top positions, while other LLMs, such as T5, claim third place, highlighting the sustained prominence of BERT-based models.

In the context of the benchmark dataset DocRED (Table 6), the top-performing models consistently employ RoBERTa large as their encoder, a noteworthy observation with implications for addressing our RQ4. It is striking that there is a conspicuous absence of papers or experiments using language models other than RoBERTa or BERT.

LLMs like RoBERTa and BERT are widely used and effectively extract relationships between different document elements. This is also true for extracting relationships at the sentence-level. The reason for this effectiveness is linked to the Masked Language Model (MLM) strategy. This strategy involves training the model on a large body of text by hiding a word within the context of other words, similar to how the model processes relation classification tasks, which involve identifying relationships between two entities.

Upon deeper scrutiny, we find that RoBERTa's consistent prominence in top positions for document-level relation extraction is linked to intrinsic details of the model:

Table 3 Summary of key papers in EACL, NAACL, and AACL categorized by model group and task

Paper	Conference	Year	Model group	Task/challenge	Subtask
Yuan and Eldardiry (2021)	EACL	2021	CNN	Sentence level	Relation classification
Hao et al. (2021)	EACL	2021	CNN	Distant supervised RE	Relation identification + relation classification
Sui et al. (2021)	EACL	2021	CNN	Distant supervised RE	Relation classification
Zhang et al. (2021b)	EACL	2021	LSTM	Document level	NER + Relation classification
Yan et al. (2021)	EACL	2021	LSTM	Multi-task	NER + Relation classification
Brody et al. (2021)	EACL	2021	Transformer	Few shot/low resource	Relation classification
Eberts and Ulges (2021)	EACL	2021	Transformer	Document level	NER + Relation identification + relation classification
Zhao et al. (2021)	EACL	2021	Transformer	Open RE	Relation classification
Eyal et al. (2021)	EACL	2021	Transformer	Multi-task	Relation classification
Fu and Grishman (2021)	EACL	2021	Transformer	Multi-task	NER + Relation identification + relation classification
Han et al. (2021)	EACL	2021	Transformer	Few shot/low resource	relation classification
Hu et al. (2021a)	EACL	2021	Transformer	Distant supervised RE	Relation classification
Hu et al. (2021b)	EACL	2021	Transformer	Few shot/low resource	Relation classification
Seagnti et al. (2021)	EACL	2021	Transformer	Multilingual and multimodal relation extraction	NER + Relation classification
Christopoulou et al. (2021)	NAACL	2021	LSTM	Distant supervised RE	NER + Relation identification + Relation classification
Zhong and Chen (2021)	NAACL	2021	Transformer	Sentence level	NER + Relation identification + Relation classification
Chen and Li (2021)	NAACL	2021	Transformer	Few shot/low resource	Relation classification
Yang and Song (2022)	AACL	2022	Transformer	Sentence level	Relation classification
Yan et al. (2022)	AACL	2022	Transformer	Sentence level	NER + Relation classification
Arcan et al. (2022)	AACL	2022	Transformer	Few shot/low resource	NER + Relation identification + relation classification
Zhou and Chen (2022)	AACL	2022	Transformer	Sentence level	NER + Relation classification
Xu et al. (2022)	NAACL	2022	Graph based	Document level	Relation classification
Wang et al. (2022b)	NAACL	2022	Graph based	Sentence level	NER + Relation classification
Xu and Choi (2022)	NAACL	2022	Graph based	Document level	NER + Relation classification

Table 3 (continued)

Paper	Conference	Year	Model group	Task/challenge	Subtask
Tiktinsky et al. (2022)	NAACL	2022	Transformer	Few shot/low resource	NER + Relation identification + relation classification
Liang et al. (2022)	NAACL	2022	Transformer	Sentence level	Relation classification
Popovic et al. (2022)	NAACL	2022	Transformer	Few shot/low resource	Relation classification
Liu et al. (2022b)	NAACL	2022	Transformer	Sentence level	Relation classification
Xiao et al. (2022)	NAACL	2022	Transformer	Document level	NER + Relation classification
Ma et al. (2023)	EACL	2023	Transformer	Document level	Relation identification + Relation classification
Teru (2023)	EACL	2023	Transformer	Few shot/low resource	Relation classification
Guo et al. (2023)	EACL	2023	Transformer	Document level	Relation classification
Najafi and Fyshe (2023)	EACL	2023	Transformer	Few shot/low resource	Relation identification + relation classification

- RoBERTa uses a much larger dataset for pre-training, encompassing more data and longer sequences. This extensive training regime contributes to the model's robustness and generalization.
- RoBERTa employs only the MLM objective during pre-training, discarding the Next Sentence Prediction (NSP) objective. Sentences are treated as continuous streams of text without explicit sentence separation, making it particularly suitable for the challenges posed by document-level relation extraction.
- RoBERTa allows training with longer sequences, surpassing the 512-token limit BERT uses. This capability provides a more comprehensive and extended context, making it well-suited for document-level relation extraction tasks.

LLMs excel in few-shot learning scenarios. This proficiency is attributed to their extensive pre-training on massive volumes of textual data, enabling them to grasp diverse linguistic patterns and acquire broad knowledge. The efficacy of this pre-training becomes evident as LLMs demonstrate a remarkable ability to generalize effectively to new tasks with only a limited number of examples, positioning them as ideal candidates for few-shot learning applications. We identify substantial potential for further exploration and research involving LLMs in this domain.

7 Discussion

In our descriptive analysis, we observed a growing interest in leveraging language models to address RE problems. Specifically, transformer-based techniques (Tables 2 and 3) have gained traction, with the number of related publications increasing steadily, reaching its highest point in the most recent year.

Regarding cutting-edge RE research, Fig. 2 illustrates publication trends across the surveyed papers. It is important to note that some conferences are held biennially, which may affect yearly publication counts. To address this, we created a year-conference graph in

Table 4 Summary of models and tasks by conference and year

Conference	Year	CNN	LSTM	Transformer	Graph based	Sentence level	Document level	Multi-task	Distant supervised RE	Few shot low resource	Multilingual and multimodal
ACL	2020	1	0	1	1	1	0	0	1	0	0
ACL	2021	2	2	5	0	1	2	1	3	1	1
ACL	2022	0	0	7	1	1	0	0	0	3	0
ACL	2023	0	0	14	1	3	3	1	1	6	5
EACL	2021	3	2	9	0	2	2	2	3	3	1
NAACL	2021	0	1	2	0	1	0	0	1	1	0
AACL	2022	0	0	5	0	3	0	0	0	2	0
NAACL	2022	0	0	5	3	3	3	0	0	2	0
EACL	2023	0	0	6	0	0	3	0	0	2	0

Fig. 3. While the number of publications in 2023 appears lower than in 2022 -due to the presence of only three and two relevant conferences- respectively, the overall trend remains upward. Although 2022 also featured three conferences and saw a decrease compared to 2021, this decline can be attributed to other fluctuations rather than a decline in research interest. When viewed over multiple years, the trend remains positive, reflecting steady growth in the field. Focusing specifically on the ACL conference, the leading venue in this domain, we observe that the highest number of RE-related papers were included in the most recent edition. This sustained growth underscores the field's dynamism and highlights the increasing research and development efforts in RE.

To provide stronger evidence that the trend indicates a growing importance of the technique, we conducted a correlation analysis between year and number of publications using Pearson correlation analysis (Freedman et al. 2007). We performed two analyses: one examining the relationship between year and publications, and another incorporating conference-year-publications for a more comprehensive evaluation. The cumulative graph by year yields a correlation coefficient of 0.68, while the year-conference graph shows a coefficient of 0.32. These values suggest a positive linear correlation between year and number of publications, with a particularly strong trend in Fig. 2. The correlation coefficient approaching 0.7 further reinforces this positive relationship. We can conclude that the correlation is positive, indicating an increasing interest in RE over time.

In terms of application areas, our analysis reveals that relation extraction has found application in diverse domains, including media, academia, economics, and even unconventional realms such as heritage conservation and mathematics. This broad applicability underscores the technique's usefulness and emphasizes the need for new systems capable of autonomously identifying novel relations in an ever-changing landscape.

In the upcoming subsections, we lay the groundwork for addressing current challenges while providing an exhaustive review of language models used for RE. Our approach is to align our findings with our RQ in a concerted effort to contribute meaningfully to ongoing research in the field of RE.

7.1 RQ1: What are the challenges of RE that are being solved by systems that leverage language models?

We have identified areas for improvement and further exploration in the following challenges:

- **Document-level RE**: While progress has been made in sentence-level RE, document-level RE remains challenging. Extending the capabilities of language models to capture relationships and context across entire documents deserves further attention. With their ability to capture broader contexts, LLMs present a promising avenue for exploration in this area (Xue et al. 2024).
- **Multimodal RE**: Integrating information from diverse modalities, such as text and images, poses a distinct challenge. Enhancing language models to extract relations from multimodal data effectively requires fusing models for text and pictures. Despite some progress, there is still room for improvement, particularly in benchmark datasets. During the survey period, only one dataset in the multimodal domain was developed (Zheng et al. 2021).

Table 5 State-of-the-art results for sentence-level RE using the TACRED dataset

References	Year	Model	F1	LM/LLM
Wang et al. (2022a)	2022	DeepStruct	76.8	GLM
Hu et al. (2022)	2022	UNIST	75.5	RoBERTa
Baek and Choi (2022)	2022	RE-MC	75.4	RoBERTa
Han et al. (2022)	2022	GEN-PT	75.3	T5
Lyu and Chen (2021)	2021	BERT	75.2	BERT

Table 6 State-of-the-art results for document-level RE in the DocRED dataset

References	Year	Model	F1	LM/LLM
Ma et al. (2023)	2023	DREEAM	67.53	RoBERTa
Tan et al. (2022)	2022	KD-Rb-l	67.28	RoBERTa
Xu et al. (2021)	2021	SSAN-RoBERTa-large	65.92	RoBERTa
Xiao et al. (2022)	2022	SAIS-RoBERTa-large	65.11	RoBERTa
Xie et al. (2022)	2022	Eider-RoBERTa-large	64.79	RoBERTa

Table 7 State-of-the-art results for Zero-Shot RE models in the FewRel dataset

References	Year	Model	F1	LM/LLM
Tran et al. (2023)	2023	ESC-ZSRE	81.68	BERT
Chia et al. (2022)	2022	RelationPrompt	79.96	GPT2+BART
Ouchi et al. (2022)	2022	IDL	62.61	BERT
Najafi and Fyshe (2023)	2023	OffMML-G(+negs)	61.3	T5
Chen and Li (2021)	2021	ZS-BERT	57.25	BERT

- **Multilingual RE**: Handling relations in languages not encountered during training or extracting relations with no training examples remains challenging. While many language models can generalize to other languages, their performance may still need to be bettered. Developing systems and training strategies, especially in transfer learning (Hennig et al. 2023), is essential to address this challenge effectively.
- **Few-Shot and Low-Resource Relation Extraction**: This category encompasses areas with a scarcity of training examples, such as medical or biological datasets, and known domains but with previously unseen relations. LLMs can offer valuable solutions in these domains (Xu et al. 2023a), leveraging their ability to generate new content for training strategies and learn complex relations and contexts. Significant improvements have been observed in distant supervision RE, where automatically labeled datasets are created from knowledge bases, and LLMs can contribute to further enhancements.
- **Open Relation Extraction**: Developing techniques to identify and characterize relations without predefined categories is an area that requires further investigation. Unsupervised techniques, such as clustering (Zhao et al. 2021, 2023b), hold promise in addressing this challenge.

RE grapples with challenges related to ambiguity and polysemy, where entities and relations frequently exhibit multiple meanings across different contexts. LMs, with their con-

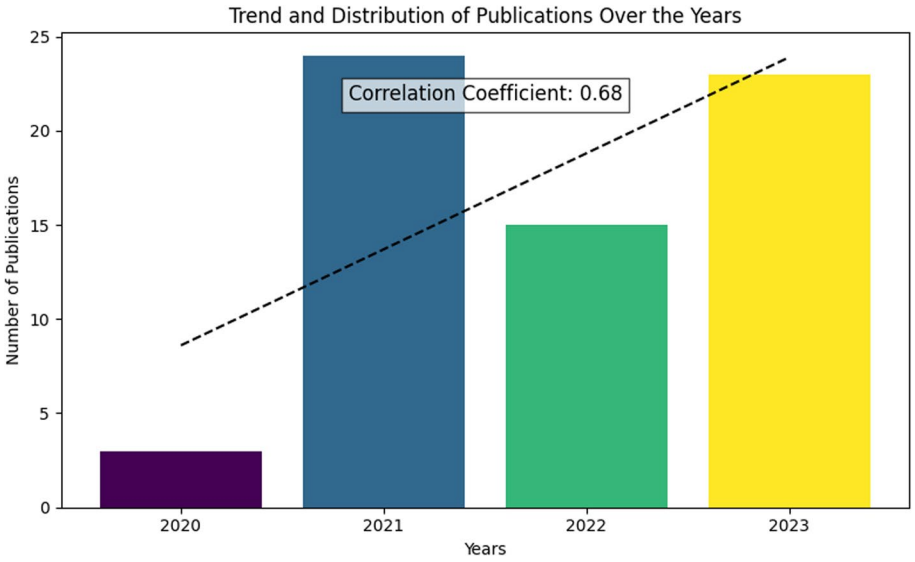

Fig. 2 Trend and distribution of publications over the years

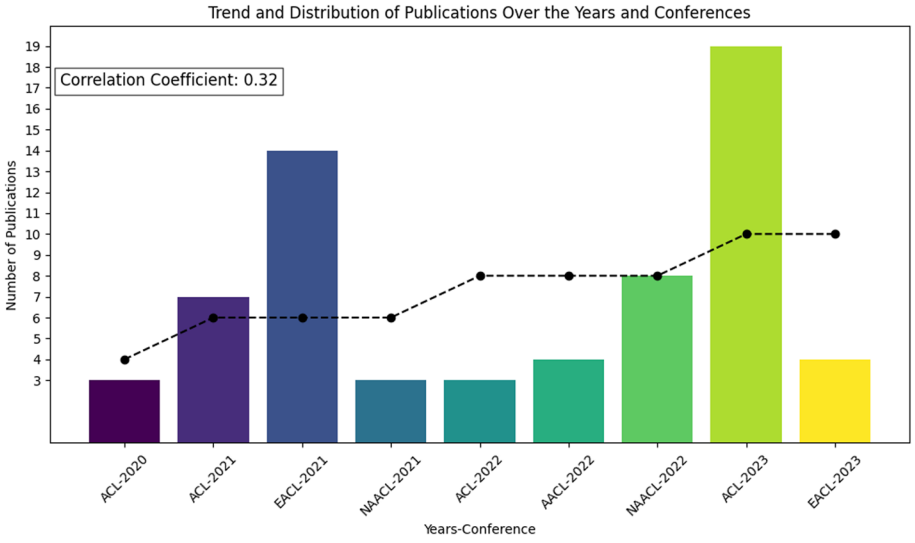

Fig. 3 Trend and distribution of publications over the years and conferences

textual understanding, play a crucial role in capturing the nuanced nature of language and disambiguating entities and relations based on the surrounding context. However, there is room for improvement. A promising avenue for further research is enhancing performance through the use of similarities or refining semantic searches. Finally, RE techniques must contend with negations and expressions of uncertainty in language. Ongoing research focuses on developing methods to handle the negation of relations or the presence of "none

of the above" relations, reflecting the persistent challenges in dealing with negations and uncertain language expressions.

7.2 RQ2: What are the most commonly used language models for the RE problem?

Table 8 illustrates the adoption of various language models and word embeddings for state-of-the-art relation extraction solutions in ACL conferences. It is worth mentioning that we have considered the base model if a paper does not explicitly specify the submodel used. Also, it is necessary to mention that we have condensed all the papers created over the survey period in this table. In this table, we have discussed the papers proposing a cutting-edge technique from Sect. 5 and the papers proposing a dataset from Sect. 4.

It is evident that BERT stands out as the predominant model across the literature, featuring prominently in 45 different papers, constituting almost 55% of the recent papers about extraction. The second most widely used model is RoBERTa achieving state-of-the-art results, specially in areas like document-level relation extraction. When examining the percentage of papers that incorporate or at least mention LLMs such as T5 or GPT in their methodology, we find that only 8.5% of papers leverage these language models.

BERT is one of the oldest models studied in this survey. Its popularity may be influenced by the fact that it was one of the first language models widely adopted. To address this potential bias, we performed an analysis that normalizes the number of times a paper references a model by the number of years since model release. The formula used is tu/y, where tu represents the frequency of use as shown in Table 8, and y denotes the number of years from model release to the end of the survey in 2023. In case of BERT, was released in 2018 so, $y = 2023-2018$, 5. This approach yields a factor of 9, which remains higher than newer models like T5 or GPT. Furthermore, it is essential to consider potential issues on the use and adoption between BERT-based models and other large LLMs like GPT or T5. However, thanks to platforms like Huggingface's Transformers library (Wolf et al. 2019), most models have been made widely available, making them widely accessible in their pre-trained forms. Given that researchers predominantly rely on transfer learning or fine-tuning rather than extensive training from scratch, we can conclude that access issues are minimal or nonexistent.

To provide a more accurate comparison, considering that BERT was released significantly earlier than other language models, we analyzed the data from the most recent year of the survey, 2023. By focusing on 2023, we ensure a fair evaluation of the models. We normalized the percentages and counts based on the release year of each model that had at least one citation in the 2023 survey papers. The results are presented in Table 9.

This underscores the prevalence of BERT-based methods being the most widespread in most literature. The preference for BERT-based models in the realm of RE can be attributed to their inherent characteristics that align seamlessly with the nature of RE systems:

– BERT is pre-trained using two objectives: MLM and NSP. The MLM aspect predicts missing words in a sentence, while NSP helps the model understand relationships between consecutive sentences. This process aligns closely with the essence of relation extraction (Sect. 2.1), wherein the objective is to identify a relation between two entities, a concept closely mirrored in BERT's training strategies.

Table 8 Most widespread models

Model	Sub-model	Times used	Papers that used the model
BERT	Base	45	Zhao et al. (2023a, b, c, d), Hennig et al. (2023), Zheng et al. (2023a), Gururaja et al. (2023), Xu et al. (2023a), Zhang et al. (2023), Zhou et al. (2023), Sun et al. (2023), Yuan et al. (2023), Hu et al. (2023), Ye et al. (2022), Qin and Joty (2022), Jie et al. (2022), Bassignana and Plank (2022), Wang et al. (2021), Xie et al. (2021), Tian et al. (2021b), Ma et al. (2021), Yang et al. (2021b), Huang et al. (2021a), Nan et al. (2020), Yu et al. (2020), Pouran Ben Veyseh et al. (2020), Brody et al. (2021), Eberts and Ulges (2021), Zhao et al. (2021), Eyal et al. (2021), Han et al. (2021), Zhang et al. (2021b), Yan et al. (2021, 2022) Hao et al. (2021), Hu et al. (2021a, b), Sui et al. (2021), Zhou and Chen (2022), Teru (2023), Zhong and Chen (2021), Xiao et al. (2022), Xu et al. (2022), Liang et al. (2022), Popovic et al. (2022), Liu et al. (2022b)
RoBERTa	Large	10	Zhang et al. (2023), Liu et al. (2022a), Jie et al. (2022), Huang et al. (2021a), Brody et al. (2021), Yang and Song (2022), Zhou and Chen (2022), Ma et al. (2023), Guo et al. (2023), Xiao et al. (2022)
SciBERT	scvocab	7	Ye et al. (2022), Bassignana and Plank (2022), Yan et al. (2021, 2022), Zhong and Chen (2021), Xiao et al. (2022), Tiktinsky et al. (2022)
GloVe		4	Huang et al. (2021b), Kruiper et al. (2020), Nan et al. (2020), Christopoulou et al. (2021)
ALBERT	xxlarge	4	Ye et al. (2022), Wang et al. (2021), Yan et al. (2021), Zhong and Chen (2021)
RoBERTa	Base	4	Zhang et al. (2023), Brody et al. (2021), Eyal et al. (2021), Wang et al. (2022b)
CNN-based-encoder		3	Dai et al. (2021b), Brody et al. (2021), Yuan and Eldardiry (2021)
T5		3	Wadhwa et al. (2023), Arcan et al. (2022), Najafi and Fyshe (2023)
BERT	Large	3	Tian et al. (2021b), Zhou and Chen (2022), Liang et al. (2022)
Span-BERT		3	Brody et al. (2021), Liang et al. (2022), Xu and Choi (2022)
GPT	3.5	2	Xu et al. (2023a), Wadhwa et al. (2023)
GPT	2	2	Xu et al. (2023a), Eyal et al. (2021)
Word2Vec		2	Pouran Ben Veyseh et al. (2020), Zhang et al. (2021b)
PubMedBERT		2	Zhou et al. (2023), Tiktinsky et al. (2022)
Word2Vec	Skip-Gram	1	Shahbazi et al. (2020)
BioLinkBERT	Large	1	Xu et al. (2023b)
AnchiBERT		1	Yang et al. (2023)
KLUE		1	Yang et al. (2023)
mBART	50	1	Huguet Cabot et al. (2023)
ViT	Base	1	Wu et al. (2023)
BERT	Chinese	1	Jie et al. (2022)
RoBERTa	Chinese	1	Jie et al. (2022)
ELMO		1	Kruiper et al. (2020)
LUKE	Base	1	Brody et al. (2021)

Table 8 (continued)

Model	Sub-model	Times used	Papers that used the model
M-BERT		1	Seagnti et al. (2021)
Sentence-BERT		1	Chen and Li (2021)
BlueBERT		1	Tiktinsky et al. (2022)
BioBERT		1	Tiktinsky et al. (2022)

Table 9 2023 Papers by model variant with normalization by year

Model	Variant	Release year	Count	%	Norm. count	Norm. %	References
BERT	Base	2018	9	45	**1.80**	**11.54**	Zhao et al. (2023a, b, d), Hennig et al. (2023), Zheng et al. (2023a), Gururaja et al. (2023), Xu et al. (2023a), Zhao et al. (2023c), Zhang et al. (2023)
RoBERTa	Large	2019	2	10	**0.50**	**3.21**	Zhang et al. (2023), Ma et al. (2023)
T5	–	2020	2	10	**0.50**	**3.21**	Wadhwa et al. (2023), Najafi and Fyshe (2023)
GPT	3, 5	2022	2	10	**1.00**	**6.41**	Xu et al. (2023a), Wadhwa et al. (2023)
PubMedBERT	–	2020	1	5	**0.25**	**1.60**	Zhou et al. (2023)
BioLinkBERT	Large	2022	1	5	**0.50**	**3.21**	Xu et al. (2023b)
AnchiBERT	–	2023	1	5	**1.00**	**6.41**	Yang et al. (2023)
mBART	50	2020	1	5	**0.25**	**1.60**	Huguet Cabot et al. (2023)
ViT	Base	2021	1	5	**0.33**	**2.12**	Wu et al. (2023)

- During MLM training, BERT randomly masks specific tokens in each sequence, enhancing the model's generalizability. This characteristic proves to be particularly beneficial for few-shot problems in RE.
- BERT uses unique tokens ([CLS] and [SEP]) to indicate the beginning and separation of sentences. Those tokens ease the delineation of head and tail entities in relation extraction. Many state-of-the-art results in RE are built upon adaptations of these tokens in the pre-training strategy.

Most papers on language adoption focus on English. However, there is also significant development in Asian languages. Notably, there are specific BERT models tailored for Chinese, such as ChineseBERT, and models like KLUE-RE and AnchiBERT, designed for Korean and Hanja, respectively, are emerging.

Finally, it is interesting to note the ongoing presence of traditional word embedding techniques, such as GloVe and Word2Vec. However, these methods, along with CNN-based approaches, are primarily associated with the earlier years of the survey. Specifically, GloVe appears in four papers: two from 2020 and two from 2021, while Word2Vec is represented by only one paper from 2021. As for CNN-based approaches, as reflected in Table 4, there is a noticeable trend in their use during the early period of the survey. However, their usage has

declined in recent years, indicating a reduced interest compared to BERT-based techniques in contemporary research.

7.3 RQ3: Which datasets are used as benchmarks for RE using language models?

Table 10 presents the utilization of datasets in ACL conferences. Addressing RQ3 directly, the most prevalent benchmark datasets for RE have been TACRED, DocRed, and FewRel. Examining datasets with over ten recent cutting-edge RE research mentions reveals an intriguing pattern.

Firstly, in traditional sentence-level RE, TACRED serves as the most robust benchmark. Researchers are actively exploring new techniques, including comparisons between joint models and pipeline approaches, methods for sentence encoding, and advancements in accurately marking the tails and heads of relations for each model using this dataset.

Secondly, the domain of document-level relation extraction is gaining significant attention. DocRed, as a dataset, is instrumental in testing novel approaches and assessing recent advances. Researchers leverage this dataset to compare against baselines, improving the ongoing evolution of methodologies in document-level RE.

Lastly, the FewRel dataset has emerged as a benchmark for few-shot relation extraction. This dataset plays a crucial role in evaluating the capabilities of both standard and LLMs. The discussions surrounding the effectiveness of LMS and the foray of LLMs in capturing previously unseen relations underscore the dataset's importance in shaping the discourse on few-shot relation extraction.

Exploring the long tail of datasets with only a single mention in cutting-edge relation extraction reveals a diverse landscape. These datasets, spanning diverse domains such as mathematics, dialogues, natural disasters, drugs, and heritage, signal the broad interest in employing relation extraction techniques across various academic and societal domains. The presence of datasets in these unique and specialized areas is compelling evidence of the technique's applicability and relevance to a wide array of fields within academia and society.

We conducted a thorough analysis, focusing on datasets with a minimum of five references in the literature, to provide an in-depth exploration of the landscape of RE (Table 11). Our systematic examination covers various dimensions, including the number and types of relations and relevant statistical metrics. We meticulously consider factors such as the source of information, domain specificity, nature of relations, and other pertinent attributes for each dataset. By delving into the intricacies of these datasets, our goal is to offer a comprehensive overview that illuminates their diverse characteristics, providing valuable insights and contributing to a nuanced understanding of the field of RE.

7.4 RQ4: Are new LLMs like GPT or T5 useful in RE versus widespread models such as BERT?

To address RQ4, we analyzed the performance of various models on three benchmark datasets in Sect. 6 and reviewed the most widespread models used in Table 8. The results shed light on the effectiveness of new LLMs like GPT and T5 compared to well-established models like BERT.

Table 10 Most widespread benchmarks datasets

Dataset	Times used	Papers that used the dataset
TACRED	18	Zhao et al. (2021, 2023b, c, d), Xu et al. (2023a), Qin and Joty (2022), Liu et al. (2022a), Ma et al. (2021), Brody et al. (2021), Eyal et al. (2021), Hu et al. (2021a, b), Yang and Song (2022), Zhou and Chen (2022), Teru (2023), Wang et al. (2022b), Liang et al. (2022), Liu et al. (2022b)
DocRED	15	Chen et al. (2023), Zhang et al. (2023), Wadhwa et al. (2023), Sun et al. (2023), Huang et al. (2021b), Huang et al. (2021a), Nan et al. (2020), Eberts and Ulges (2021), Eyal et al. (2021), Ma et al. (2023), Guo et al. (2023), Xiao et al. (2022), Xu et al. (2022), Xu and Choi (2022), Popovic et al. (2022)
FewRel	12	Zhao et al. (2023a, b, c, d), Qin and Joty (2022), Liu et al. (2022a), Yang et al. (2021b), Brody et al. (2021), Zhao et al. (2021), Han et al. (2021), Najafi and Fyshe (2023), Chen and Li (2021)
SemEval 2010—Task 8	10	Xu et al. (2023a), Zheng et al. (2023b), Tian et al. (2021b), Yuan and Eldardiry (2021), Hu et al. (2021a, b), Yang and Song (2022), Teru (2023), Wang et al. (2022b), Liang et al. (2022)
ACE05	9	Zheng et al. (2023b), Ye et al. (2022), Wang et al. (2021), Tian et al. (2021b), Pouran Ben Veyseh et al. (2020), Fu and Grishman (2021), Yan et al. (2021, 2022), Zhong and Chen (2021)
SciERC	9	Zheng et al. (2023b), Ye et al. (2022), Bassignana and Plank (2022), Wang et al. (2021), Pouran Ben Veyseh et al. (2020), Yan et al. (2021), Yan et al. (2022), Zhong and Chen (2021), Popovic et al. (2022)
NYT	8	Wadhwa et al. (2023), Xie et al. (2021), Ma et al. (2021), Dai et al. (2021b), Yan et al. (2021), Hao et al. (2021), Sui et al. (2021), Christopoulou et al. (2021)
Tacred-Revisited	5	Xu et al. (2023a), Yang and Song (2022), Zhou and Chen (2022), Wang et al. (2022b), Liang et al. (2022)
Re-DocRED	4	Zhang et al. (2023), Zhou et al. (2023), Sun et al. (2023), Guo et al. (2023)
ACE04	4	Ye et al. (2022), Wang et al. (2021), Yan et al. (2021), Zhong and Chen (2021)
Re-TACRED	4	Xu et al. (2023a), Yang and Song (2022), Zhou and Chen (2022), Wang et al. (2022b)
MNRE	3	Zheng et al. (2023a), Wu et al. (2023), Hu et al. (2023)
GDA	3	Huang et al. (2021b), Nan et al. (2020), Xiao et al. (2022)
CDR	3	Huang et al. (2021b), Nan et al. (2020), Xiao et al. (2022)
Wiki-ZSL	3	Zhao et al. (2023a), Najafi and Fyshe (2023), Chen and Li (2021)
FB-NYT	3	Shahbazi et al. (2020), Yuan and Eldardiry (2021), Liu et al. (2022b)
ChemProt	2	Xu et al. (2023a, b)
ConLL	2	Zheng et al. (2023b), Wadhwa et al. (2023)
ADE	2	Wadhwa et al. (2023), Yan et al. (2021)
DialogRE	2	Yu et al. (2020), Xu et al. (2022)
DDI	1	Xu et al. (2023b)
GAD	1	Xu et al. (2023b)
HistRED	1	Yang et al. (2023)
MultiTacred	1	Hennig et al. (2023)
SRED-FM	1	Huguet Cabot et al. (2023)
RED-FM	1	Huguet Cabot et al. (2023)
Twitter-2015	1	Zheng et al. (2023a)
Twitter-2017	1	Zheng et al. (2023a)
RISeC	1	Gururaja et al. (2023)

Table 10 (continued)

Dataset	Times used	Papers that used the dataset
EFGC	1	Gururaja et al. (2023)
MSCorpus	1	Gururaja et al. (2023)
Wiki80	1	Xu et al. (2023a)
BioRED	1	Zhou et al. (2023)
MATRES	1	Yuan et al. (2023)
EventStoryLine	1	Yuan et al. (2023)
Causal-TimeBank	1	Yuan et al. (2023)
MAWPS	1	Jie et al. (2022)
Math23k	1	Jie et al. (2022)
MathQA	1	Jie et al. (2022)
SVAMP	1	Jie et al. (2022)
SemEval 2018 Task 7	1	Bassignana and Plank (2022)
SKE	1	Xie et al. (2021)
FOBIE	1	Kruiper et al. (2020)
SF	1	Yu et al. (2020)
SPOUSE	1	Pouran Ben Veyseh et al. (2020)
ERE	1	Fu and Grishman (2021)
FUNDS	1	Zhang et al. (2021b)
GIDS	1	Hao et al. (2021)
SMiLER	1	Seagnti et al. (2021)
RE-QA	1	Najafi and Fyshe (2023)
KBP37	1	Teru (2023)
WikiDistant	1	Christopoulou et al. (2021)
DrugCombination	1	Tiktinsky et al. (2022)
DWIE	1	Xu and Choi (2022)

Despite their promising capabilities, LLMs, including GPT and T5, are used less frequently than models such as RoBERTa and BERT. This phenomenon could be attributed to the inherent suitability of these language models for downstream tasks like chatbots, where their remarkable ability to capture extensive contextual information is more prominently leveraged. The current landscape suggests that LLMs do not play a central role in advancing state-of-the-art performance in extraction tasks. However, it is essential to note that the field is dynamic, and the influence of these models may evolve with ongoing research and advancements.

An intriguing observation arises when comparing the insights from Sect. 6 with the utilization trends outlined in Table 8. Notably, state-of-the-art results are predominantly achieved by models based on RoBERTa. Surprisingly, RoBERTa, despite its demonstrated strength in terms of results, is underused in contrast to the widespread adoption of BERT. This discrepancy in adoption could be due to BERT's adaptability, which encourages a more profound exploration of diverse approaches for encoding relations and comparison against other baseline models. In contrast, RoBERTa's superior performance might result in fewer variations in its application, as it already delivers robust results without requiring extensive modifications.

The landscape of relation extraction is currently marked by a nuanced interplay between language models' performance and practical adoption. While LLMs show promise, their

Table 11 Statistics and a summary of the most widely used datasets for Relation Extraction

Dataset	RE task	Dataset based on	Relations	Example relations	Relation mentions	Corpus size	Train	Dev	Test
TACRED	Multipurpose—sentence level RE	Newswire and web text manually annotated using Amazon Mechanical Turk crowdsourcing. Covers common relations between people, organizations, and locations	43	Per:schools_attended and org:members	21,784	106,264	68,124	22,631	15,509
DocRED	Document-Level RE	Wikipedia data, both manually and distant supervised annotated	96	Educated_at, spouse, creator, publication_date..	155,535	Manually labelled: 5053 Distant supervised: 101,873	Manually labeled: 3053 Distant supervised: 101,873	1000	1000
FewRel	Few Shot RE	70,000 sentences on relations derived from Wikipedia and annotated by crowdworkers	100	Member_of, capital_of, birth_name	58,267	70,000	44,800	6400	14,000
SemEval 2010—Task 8	Multipurpose—sentence level RE	Semantic relations between pairs of nominals of general purpose	9	Entity-Origin, entity-destination, cause-effect.	6674	10,717	8000	–	2717
ACE05	Multilingual—sentence level RE	1800 files encompassing diverse English, Arabic, and Chinese genres. These texts have been meticulously annotated for entities, relations, and events. This extensive collection is the entirety of the training data utilized during the 2005 Automatic Content Extraction (ACE) technology evaluation for these languages	18	Person–social, organization–affiliation, agent–artifact	7105	10,573	7273	1765	1535
SciERC	Multipurpose—low resource	Collection of 500 scientific abstracts annotated in terms of entities and relations	7	usef_for, hyponym_of, feature_of, conjunction	4716	500	–	–	–
NYT-FB	Multipurpose—distant supervision RE	New York Times news annotated by distant supervision	52	Nationality, place_of_birth	142,823	717,219	455,771	–	172,448

utilization in cutting-edge relation extraction remains modest. The interplay of factors such as adaptability, performance, and the specific demands of downstream tasks shapes the landscape, suggesting a need for ongoing exploration and fine-tuning of these models about extraction.

To ensure this analysis is as up-to-date as possible, we used the advanced search feature of Web of Science to perform four different queries. Each query included "relation extraction" in the title and searched for specific models and their variants in the abstract or topic. For example, the query for T5 was: *TI=("relation extraction") AND (TS=(T5 OR "Text-To-Text Transfer Transformer") OR AB=(T5 OR "Text-To-Text Transfer Transformer"))*. For BERT, we found 31 papers published in the last 2 years, 2023 and 2024. In contrast, we found only three papers each for T5 and GPT. These results support more robust conclusions regarding the predominance of BERT in our survey findings.

7.5 Synthesis

We have identified two primary methods of RE: joint and pipeline. Joint models are designed to simultaneously identify entities and relationships, capturing the inherent interdependence between these two tasks. In contrast, the pipeline approach follows a sequential process, starting with NER to identify entities, followed by at least a classification stage for these recognized entities. The pipeline approach can involve up to three stages, with an additional step of relation identification between NER and relation classification, as depicted in Tables 2 and 3. While this two-step or three-step process allows for more specialized handling of each task, it also introduces the risk of error accumulation. In recent years, joint approaches have gained more attention, mainly due to concerns that the pipeline method is prone to error propagation.

Future trends in the realm of new language models, including sub-models and LLMs, point toward increasingly complex and intricate models with more parameters. For instance, LLaMA 3 has 70 billion parameters (Dubey et al. 2024), GPT-4 has 1.8 trillion parameters (OpenAI et al. 2023), and Gemini supports a context window of 128,000 tokens with billions of parameters (Team et al. 2023). These models will seem small compared to future developments in our rapidly evolving space. In the context of RE, these advanced models, with their extended context windows, will be particularly valuable for document-level tasks. They can maintain entire documents within their context, improving the handling of extensive information. Furthermore, in multilingual and multimodal settings, the vast context capabilities of these LLMs will enhance understanding across different languages and modalities. Additionally, these models can offer significant advantages for few-shot or low-resource scenarios by augmenting evidence for underrepresented domains and improving generalization with minimal data.

Before concluding this section, it is essential to address this research's ethical considerations and global impact. All reviewed papers adhere to ethical standards by utilizing anonymized or publicly available data, ensuring their contributions to society are made responsibly. They are committed to minimizing bias and avoiding potential ethical issues, reflecting a conscientious approach to research and its broader implications. Finally, the environmental impact is minimal, considering the CO2 emissions and energy consumption associated with the research discussed. While the reviewed papers do not explicitly provide this information, it is reasonable to infer that only foundational work and the initial train-

ing of LLMs have a significant environmental footprint. Most of the studies leverage pre-trained models, as a result avoiding the resource-intensive process of training from scratch, which is the most demanding stage in terms of energy consumption.

8 Conclusion

This paper reviews cutting-edge relation extraction, explicitly focusing on models leveraging language models. Our analysis encompasses 137 papers, spanning dataset proposals, novel techniques, and excluded studies.

Our research focuses on 65 papers, which we have rigorously categorized into eight sub-tasks reflecting the latest advances in RE. From these papers, we have extracted crucial insights into training strategies, novel methodologies, and the utilization of language models, which have emerged as a cornerstone of contemporary RE research.

Our findings highlight the dominance of BERT-based approaches to extraction, underscoring their efficacy in achieving state-of-the-art results. Notably, BERT has emerged as the most widely utilized model for extraction to date. Additionally, emerging LLMs like T5 exhibit promise, particularly in the context of few-shot relation extraction, showcasing their ability to capture previously unseen relations. However, these models are primarily suited for tasks such as text generation and question answering; their capability to handle large context windows suggests a promising future for developing RE techniques based on these models.

Language models are highly suitable for RE because they effectively capture the semantic nuances and complex relationships within text. By leveraging these models, research has been able to address the inherent challenges of RE, such as ambiguity and context sensitivity, providing robust solutions that enhance the accuracy and efficiency of extracting relations across diverse domains.

Finally, throughout our exploration, we have identified FewRel, DocRed, and TACRED as pivotal benchmark datasets, aligning with the three primary themes driving contemporary research about extraction. FewRel serves as a benchmark for few-shot relation extraction, DocRed facilitates advancements in document-level relation extraction, and TACRED remains a stalwart in traditional sentence-level relation extraction. Collectively, these datasets offer a comprehensive evaluation ground for models addressing the diverse challenges in cutting-edge relation extraction.

Acknowledgements The research reported in this paper was supported by the DesinfoScan project: Grant TED2021-129402B-C21 funded by MCIN /AEI/ 10.13039/ 501100011033 and, by the European Union NextGenerationEU/PRTR, and FederaMed project: Grant PID2021-123960OB-I00 funded by MCIN/ AEI/ 10.13039/501100011033 and by ERDF A way of making Europe. Finally, the research reported in this paper is also funded by the European Union (BAG-INTEL project, grant agreement no. 101121309).

Author contributions JA Diaz Garcia and JA Diaz-Lopez wrote the main manuscript. The original idea between the manuscript was provided by JA Diaz-Lopez. JA Diaz-Lopez reviewed the main manuscript.

Funding Funding for open access publishing: Universidad de Granada/CBUA.

Data availability No datasets were generated or analysed during the current study.

Competing interests The authors declare no competing interests.

References

Agosti M, Di Nunzio G, Marchesin S, Silvello G et al (2018) A relation extraction approach for clinical decision support. In: Proceedings of the CIKM 2018 workshops co-located with 27th ACM international conference on information and knowledge management (CIKM 2018), Torino, Italy, 22 October 2018, vol 2482

Aguilar J, Beller C, McNamee P, Van Durme B, Strassel S, Song Z, Ellis J (2014) A comparison of the events and relations across ace, ere, tac-kbp, and framenet annotation standards. In: Proceedings of the second workshop on EVENTS: definition, detection, coreference, and representation, pp 45–53

Alt C, Gabryszak A, Hennig L (2020a) 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 TACRED revisited annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pp 1558–1569. https://doi.org/10.18653/v1/2020.acl-main.142, https://aclanthology.org/2020.acl-main.142

Alt C, Gabryszak A, Hennig L (2020b) Tacred revisited: a thorough evaluation of the tacred relation extraction task. arXiv preprint. arXiv:2004.14855

Amini MR, Kananizadeh N, Ganjtabesh M (2019) MathQA: towards interpretable math word problem solving with operation-based formalisms. In: Proceedings of the 57th annual meeting of the Association for Computational Linguistics, pp 6196–6206

Arcan M, O'Halloran R, Robin C, Buitelaar P (2022) Towards bootstrapping a chatbot on industrial heritage through term and relation extraction. In: Hämäläinen M, Alnajjar K, Partanen N, Rueter J (eds.) Proceedings of the 2nd international workshop on natural language processing for digital humanities. Association for Computational Linguistics, Taipei, Taiwan, pp 108–122. https://aclanthology.org/2022.nlp4dh-1.15

Baek HR, Choi YS (2022) Enhancing targeted minority class prediction in sentence-level relation extraction. Sensors 22(13):4911

Bassignana E, Plank B (2022) What do you mean by relation extraction? A survey on datasets and study on scientific relation classification. In: Louvan S, Madotto A, Madureira B (eds) Proceedings of the 60th annual meeting of the Association for Computational Linguistics: student research workshop. Association for Computational Linguistics, Dublin, Ireland, pp 67–83. https://doi.org/10.18653/v1/2022.acl-srw.7, https://aclanthology.org/2022.acl-srw.7

Becker KG, Barnes KC, Bright TJ, Wang SA (2004) The genetic association database. Nat Genet 36(5):431–432

Beltagy I, Lo K, Cohan A (2019) SciBert: a pretrained language model for scientific text. arXiv preprint. arXiv:1903.10676

Bengio Y, Ducharme R, Vincent P (2001) A neural probabilistic language model. Adv Neural Inf Process Syst 3:1137–1155

Bennet DT, Bennet PS, Thiagarajan P et al (2024) Content based classification of short messages using recurrent neural networks in nlp. In: 2024 International conference on artificial intelligence, computer, data sciences and applications (ACDSA). IEEE, pp 1–6

Bravo À, Piñero J, Queralt-Rosinach N, Rautschka M, Furlong LI (2015) Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. BMC Bioinform 16:1–17

Brody S, Wu S, Benton A (2021) Towards realistic few-shot relation extraction. In: Moens MF, Huang X, Specia L, Yih SWt (eds) Proceedings of the 2021 conference on empirical methods in natural language processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pp 5338–5345. https://doi.org/10.18653/v1/2021.emnlp-main.433, https://aclanthology.org/2021.emnlp-main.433

Cai PX, Fan YC, Leu FY (2022) Compare encoder–decoder, encoder-only, and decoder-only architectures for text generation on low-resource datasets. In: Advances on broad-band wireless computing, communication and applications: proceedings of the 16th international conference on broad-band wireless computing, communication and applications (BWCCA-2021). Springer, pp 216–225

Chen CY, Li CT (2021) ZS-BERT: towards zero-shot relation extraction with attribute representation learning. In: Toutanova K, Rumshisky A, Zettlemoyer L, Hakkani-Tur D, Beltagy I, Bethard S, Cotterell R, Chakraborty T, Zhou Y (eds) Proceedings of the 2021 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Online, pp 3470–3479. https://doi.org/10.18653/v1/2021.naacl-main.272, https://aclanthology.org/2021.naacl-main.272

Chen ZY, Chang CH, Chen YP, Nayak J, Ku LW (2019) UHop: An unrestricted-hop relation extraction framework for knowledge-based question answering. In: Burstein J, Doran C, Solorio T (eds) Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: human language technologies, vol 1 (long and short papers). Association for Computational Linguistics, Minneapolis, pp 345–356. https://doi.org/10.18653/v1/N19-1031, https://aclanthology.org/N19-1031/

Chen H, Chen B, Zhou X (2023) Did the models understand documents? Benchmarking models for language understanding in document-level relation extraction. In: Rogers A, Boyd-Graber J, Okazaki N (eds) Proceedings of the 61st annual meeting of the Association for Computational Linguistics (vol 1: long papers). Association for Computational Linguistics, Toronto, pp 6418–6435. https://doi.org/10.18653/v1/2023.acl-long.354, https://aclanthology.org/2023.acl-long.354

Chia YK, Bing L, Poria S, Si L (2022) RelationPrompt: leveraging prompts to generate synthetic data for zero-shot relation triplet extraction. In: Muresan S, Nakov P, Villavicencio A (eds) Findings of the association for computational linguistics: ACL 2022. Association for Computational Linguistics, pp 45–57. https://doi.org/10.18653/v1/2022.findings-acl.5

Cho K, van Merriënboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: encoder–decoder approaches. In: Wu D, Carpuat M, Carreras X, Vecchi EM (eds) Proceedings of SSST-8, eighth workshop on syntax, semantics and structure in statistical translation. Association for Computational Linguistics, Doha, pp 103–111. https://doi.org/10.3115/v1/W14-4012, https://aclanthology.org/W14-4012

Christopoulou F, Miwa M, Ananiadou S (2021) Distantly supervised relation extraction with sentence reconstruction and knowledge base priors. In: Toutanova K, Rumshisky A, Zettlemoyer L, Hakkani-Tur D, Beltagy I, Bethard S, Cotterell R, Chakraborty T, Zhou Y (eds) Proceedings of the 2021 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Online, pp 11–26. https://doi.org/10.18653/v1/2021.naacl-main.2, https://aclanthology.org/2021.naacl-main.2

Conneau A, Khandelwal K, Goyal N, Chaudhary V, Wenzek G, Guzmán F, Grave E, Ott M, Zettlemoyer L, Stoyanov V (2020) Unsupervised cross-lingual representation learning at scale. In: Jurafsky D, Chai J, Schluter N, Tetreault J (eds) Proceedings of the 58th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747, https://aclanthology.org/2020.acl-main.747/

Coppersmith G, Leary R, Crutchley P, Fine A (2018) Natural language processing of social media as screening for suicide risk. Biomed Inform Insights 10:1178222618792860

Cui L, Yang D, Yu J, Hu C, Cheng J, Yi J, Xiao Y (2021) Refining sample embeddings with relation prototypes to enhance continual relation extraction. In: Zong C, Xia F, Li W, Navigli R (eds) Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (vol 1: long papers). Association for Computational Linguistics, Online, pp 232–243. https://doi.org/10.18653/v1/2021.acl-long.20, https://aclanthology.org/2021.acl-long.20

Dai Q, Inoue N, Reisert P, Takahashi R, Inui K (2019) Incorporating chains of reasoning over knowledge graph for distantly supervised biomedical knowledge acquisition. In: Proceedings of the 33rd Pacific Asia conference on language, information and computation. Waseda University, pp 19–28

Dai Q, Inoue N, Takahashi R, Inui K (2021a) Two training strategies for improving relation extraction over universal graph. arXiv preprint. arXiv:2102.06540

Dai Q, Inoue N, Takahashi R, Inui K (2021b) Two training strategies for improving relation extraction over universal graph. In: Merlo P, Tiedemann J, Tsarfaty R (eds.) Proceedings of the 16th conference of the European Chapter of the Association for Computational Linguistics: main volume. Association for Computational Linguistics, Online, pp 3673–3684. https://doi.org/10.18653/v1/2021.eacl-main.321, https://aclanthology.org/2021.eacl-main.321

Dauphin YN, Fan A, Auli M, Grangier D (2017) Language modeling with gated convolutional networks. In: Precup D, Teh YW (eds) Proceedings of the 34th international conference on machine learning research, vol 70. PMLR, International Convention Centre, Sydney, pp 933–941. http://proceedings.mlr.press/v70/dauphin17a.html

de Winter J (2024) Can ChatGPT be used to predict citation counts, readership, and social media interaction? An exploration among 2222 scientific abstracts. Scientometrics. https://doi.org/10.1007/s11192-024-04939-y

Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint. http://arxiv.org/abs/1810.04805

Diaz-Garcia JA, Gutiérrez-Batista K, Fernandez-Basso C, Ruiz MD, Martin-Bautista MJ (2024) A flexible big data system for credibility-based filtering of social media information according to expertise. Int J Comput Intell Syst 17(1):93

Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, Letman A, Mathur A, Schelten A, Yang A, Fan A et al (2024) The llama 3 herd of models. arXiv preprint. arXiv:2407.21783

Eberts M, Ulges A (2021) An end-to-end model for entity-level relation extraction using multi-instance learning. In: Merlo P, Tiedemann J, Tsarfaty R (eds) Proceedings of the 16th conference of the European chapter of the Association for Computational Linguistics: main volume. Association for Computational Linguistics, Online, pp 3650–3660. https://doi.org/10.18653/v1/2021.eacl-main.319, https://aclanthology.org/2021.eacl-main.319

Eyal M, Amrami A, Taub-Tabib H, Goldberg Y (2021) Bootstrapping relation extractors using syntactic search by examples. In: Merlo P, Tiedemann J, Tsarfaty R (eds) Proceedings of the 16th Conference of the European chapter of the Association for Computational Linguistics: main volume. Association for Computational Linguistics, Online, pp 1491–1503. https://doi.org/10.18653/v1/2021.eacl-main.128, https://aclanthology.org/2021.eacl-main.128

Freedman D, Pisani R, Purves R (2007) Statistics, 4th edn. W. W. Norton & Company, New York

Fu L, Grishman R (2021) Learning relatedness between types with prototypes for relation extraction. In: Merlo P, Tiedemann J, Tsarfaty R (eds) Proceedings of the 16th conference of the European chapter of the Association for Computational Linguistics: main volume. Association for Computational Linguistics, Online, pp 2011–2016. https://doi.org/10.18653/v1/2021.eacl-main.172, https://aclanthology.org/2021.eacl-main.172

Gardent C, Shimorina A, Narayan S, Perez-Beltrachini L (2017) Creating training corpora for nlg micro-planning. In: 55th annual meeting of the Association for Computational Linguistics (ACL)

Gillioz A, Casas J, Mugellini E, Abou Khaled O (2020) Overview of the transformer-based models for NLP tasks. In: 2020 15th Conference on computer science and information systems (FedCSIS). IEEE, pp 179–183

Gu Y, Tinn R, Cheng H, Lucas M, Usuyama N, Liu X, Naumann T, Gao J, Poon H (2021) Domain-specific language model pretraining for biomedical natural language processing. ACM Trans Comput Healthcare (HEALTH) 3(1):1–23

Guo J, Kok S, Bing L (2023) Towards integration of discriminability and robustness for document-level relation extraction. In: Vlachos A, Augenstein I (eds) Proceedings of the 17th conference of the European chapter of the Association for Computational Linguistics. Association for Computational Linguistics, Dubrovnik, Croatia, pp 2606–2617. https://doi.org/10.18653/v1/2023.eacl-main.191, https://aclanthology.org/2023.eacl-main.191

Gurulingappa H, Rajput AM, Roberts A, Fluck J, Hofmann-Apitius M, Toldo L (2012) Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. J Biomed Inform 45(5):885–892

Gururaja S, Dutt R, Liao T, Rosé C (2023) Linguistic representations for fewer-shot relation extraction across domains. 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). Association for Computational Linguistics, Toronto, Canada, pp 7502–7514. https://doi.org/10.18653/v1/2023.acl-long.414, https://aclanthology.org/2023.acl-long.414

Han X, Sun L (2016) Global distant supervision for relation extraction. In: Proceedings of the AAAI conference on artificial intelligence, vol 30

Han X, Zhu H, Yu P, Wang Z, Yao Y, Liu Z, Sun M (2018) FewRel: a large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In: Riloff E, Chiang D, Hockenmaier J, Tsujii J (eds) Proceedings of the 2018 conference on empirical methods in natural language processing. Association for Computational Linguistics, Brussels, Belgium, pp 4803–4809. https://doi.org/10.18653/v1/D18-1514, https://aclanthology.org/D18-1514

Han X, Gao T, Yao Y, Ye D, Liu Z, Sun M (2019) Opennre: an open and extensible toolkit for neural relation extraction. arXiv preprint. arXiv:1909.13078

Han X, Gao T, Lin Y, Peng H, Yang Y, Xiao C, Liu Z, Li P, Zhou J, Sun M (2020) More data, more relations, more context and more openness: a review and outlook for relation extraction. In: Wong KF, Knight K, Wu H (eds) Proceedings of the 1st Conference of the Asia-Pacific chapter of the Association for Computational Linguistics and the 10th international joint conference on natural language processing. Association for Computational Linguistics, Suzhou, China, pp 745–758. https://aclanthology.org/2020.aacl-main.75

Han J, Cheng B, Lu W (2021) Exploring task difficulty for few-shot relation extraction. In: Moens MF, Huang X, Specia L, Yih SWt (eds) Proceedings of the 2021 conference on empirical methods in natural language processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pp 2605–2616. https://doi.org/10.18653/v1/2021.emnlp-main.204, https://aclanthology.org/2021.emnlp-main.204

Han J, Zhao S, Cheng B, Ma S, Lu W (2022) Generative prompt tuning for relation classification. In: Goldberg Y, Kozareva Z, Zhang Y (eds) Findings of the Association for Computational Linguistics: EMNLP 2022. Association for Computational Linguistics, Abu Dhabi, pp 3170–3185. https://doi.org/10.18653/v1/2022.findings-emnlp.231, https://aclanthology.org/2022.findings-emnlp.231

Hao K, Yu B, Hu W (2021) Knowing false negatives: an adversarial training method for distantly supervised relation extraction. In: Moens MF, Huang X, Specia L, Yih SWt (eds) Proceedings of the 2021 conference on empirical methods in natural language processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pp 9661–9672. https://doi.org/10.18653/v1/2021.emnlp-main.761, https://aclanthology.org/2021.emnlp-main.761

Hendrickx I, Kim SN, Kozareva Z, Nakov P, Séaghdha DO, Padó S, Pennacchiotti M, Romano L, Szpakowicz S (2019) Semeval-2010 task 8: multi-way classification of semantic relations between pairs of nominals. arXiv preprint. arXiv:1911.10422

Hennig L, Thomas P, Möller S (2023) MultiTACRED: A multilingual version of the TAC 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). Association for Computational Linguistics, Toronto, Canada, pp 3785–3801. https://doi.org/10.18653/v1/2023.acl-long.210, https://aclanthology.org/2023.acl-long.210

Hossain E, Rana R, Higgins N, Soar J, Barua PD, Pisani AR, Turner K (2023) Natural language processing in electronic health records in relation to healthcare decision-making: a systematic review. Comput Biol Med 155:106649

Hu X, Zhang C, Ma F, Liu C, Wen L, Yu PS (2021a) Semi-supervised relation extraction via incremental meta self-training. In: Moens MF, Huang X, Specia L, Yih SW (eds) Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, Punta Cana, Dominican Republic, pp 487–496. https://doi.org/10.18653/v1/2021.findings-emnlp.44, https://aclanthology.org/2021.findings-emnlp.44

Hu X, Zhang C, Yang Y, Li X, Lin L, Wen L, Yu PS (2021b) Gradient imitation reinforcement learning for low resource relation extraction. In: Moens MF, Huang X, Specia L, Yih SW (eds) Proceedings of the 2021 conference on empirical methods in natural language processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pp 2737–2746. https://doi.org/10.18653/v1/2021.emnlp-main.216, https://aclanthology.org/2021.emnlp-main.216

Hu X, Guo Z, Teng Z, King I, Yu PS (2023) Multimodal relation extraction with cross-modal retrieval and synthesis. In: Rogers A, Boyd-Graber J, Okazaki N (eds) Proceedings of the 61st annual meeting of the Association for Computational Linguistics (volume 2: short papers). Association for Computational Linguistics, Toronto, Canada, pp 303–311. https://doi.org/10.18653/v1/2023.acl-short.27, https://aclanthology.org/2023.acl-short.27

Huang K, Qi P, Wang G, Ma T, Huang J (2021a) Entity and evidence guided document-level relation extraction. In: Proceedings of the 6th workshop on representation learning for NLP (RepL4NLP-2021). Association for Computational Linguistics, Online, pp 307–315. https://doi.org/10.18653/v1/2021.repl4nlp-1.30, https://aclanthology.org/2021.repl4nlp-1.30

Huang Q, Zhu S, Feng Y, Ye Y, Lai Y, Zhao D (2021b) Three sentences are all you need: Local path enhanced document relation extraction. In: Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on natural language processing (volume 2: short papers). Association for Computational Linguistics, Online, pp 998–1004. https://doi.org/10.18653/v1/2021.acl-short.126, https://aclanthology.org/2021.acl-short.126

Huang JY, Li B, Xu J, Chen M (2022) Unified semantic typing with meaningful label inference. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Seattle, United States, pp 2642–2654. https://doi.org/10.18653/v1/2022.naacl-main.190, https://aclanthology.org/2022.naacl-main.190

Huguet Cabot L, Tedeschi S, Ngonga Ngomo AC, Navigli R (2023) [CDATA[\text{RED}^{{\rm fm}}]]RED$^{\rm fm}$: a filtered and multilingual 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). Association for Computational Linguistics, Toronto, Canada, pp 4326–4343. https://doi.org/10.18653/v1/2023.acl-long.237, https://aclanthology.org/2023.acl-long.237

Jat S, Khandelwal S, Talukdar P (2018) Improving distantly supervised relation extraction using word and entity based attention. arXiv preprint. arXiv:1804.06987

Jaume G, Ekenel HK, Thiran JP (2019) FUNSD: a dataset for form understanding in noisy scanned documents. In: 2019 international conference on document analysis and recognition workshops (ICDARW), vol 2. IEEE, pp 1–6

Jiang Y, Zaporojets K, Deleu J, Demeester T, Develder C (2020) Recipe instruction semantics corpus (risec): Resolving semantic structure and zero anaphora in recipes. In: Proceedings of the 1st conference of the Asia-Pacific chapter of the Association for Computational Linguistics and the 10th international joint conference on natural language processing, pp 821–826

Jiang AQ, Sablayrolles A, Mensch A, Bamford C, Chaplot DS, Casas Dl, Bressand F, Lengyel G, Lample G, Saulnier L et al (2023) Mistral 7b. arXiv preprint. arXiv:2310.06825

Jie Z, Li J, Lu W (2022) Learning to reason deductively: math word problem solving as complex relation extraction. In: Muresan S, Nakov P, Villavicencio A (eds) Proceedings of the 60th annual meeting of the Association for Computational Linguistics (volume 1: long papers). Association for Computational Linguistics, Dublin, Ireland, pp 5944–5955. https://doi.org/10.18653/v1/2022.acl-long.410, https://aclanthology.org/2022.acl-long.410

Joshi M, Chen D, Liu Y, Weld DS, Zettlemoyer L, Levy O (2020) Spanbert: Improving pre-training by representing and predicting spans. Trans Assoc Comput Linguistics 8:64–77

Kadlec R, Schmid M, Bajgar O, Kleindienst J (2016) Text understanding with the attention sum reader network. In: Erk K, Smith NA (eds) Proceedings of the 54th annual meeting of the Association for Computational linguistics (volume 1: long papers). Association for Computational Linguistics, Berlin, Germany, pp 908–918. https://doi.org/10.18653/v1/P16-1086, https://aclanthology.org/P16-1086

Khachatrian H, Nersisyan L, Hambardzumyan K, Galstyan T, Hakobyan A, Arakelyan A, Rzhetsky A, Galstyan A (2019) BioRelEx 1.0: biological relation extraction benchmark. In: Proceedings of the 18th BioNLp workshop and shared task. Association for Computational Linguistics, Florence, Italy, pp 176–190. https://doi.org/10.18653/v1/W19-5019, https://aclanthology.org/W19-5019

Kim Kjærulff S, Wich L, Kringelum J, Jacobsen UP, Kouskoumvekaki I, Audouze K, Lund O, Brunak S, Oprea TI, Taboureau O (2012) Chemprot-2.0: visual navigation in a disease chemical biology database. Nucleic Acids Res 41(D1):D464–D469

Koncel-Kedziorski R, Roy S, Wang Y (2016) Parsing algebraic word problems into equations. In: Proceedings of the 54th annual meeting of the Association for Computational Linguistics (volume 1: long papers), pp 2237–2247

Kruiper R, Vincent J, Chen-Burger J, Desmulliez M, Konstas I (2020) In Layman's terms: semi-open relation extraction from scientific texts. In: Proceedings of the 58th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp 1489–1500. https://doi.org/10.18653/v1/2020.acl-main.137, https://aclanthology.org/2020.acl-main.137

Kumar P, Manikandan S, Kishore R (2024) AI-driven text generation: a novel gpt-based approach for automated content creation. In: 2024 2nd international conference on networking and communications (ICNWC). IEEE, pp 1–6

Kumar P (2024) Large language models (LLMS): survey, technical frameworks, and future challenges. Artif Intell Rev 57(10):260

Lan Y, He S, Liu K, Zeng X, Liu S, Zhao J (2021) Path-based knowledge reasoning with textual semantic information for medical knowledge graph completion. BMC Med Inform Decis Mak 21:1–12

Lebanoff L, Song K, Liu F (2018) Adapting the neural encoder–decoder framework from single to multi-document summarization. In: Riloff E, Chiang D, Hockenmaier J, Tsujii J (eds) Proceedings of the 2018 conference on empirical methods in natural language processing. Association for Computational Linguistics, Brussels, Belgium, pp 4131–4141. https://doi.org/10.18653/v1/D18-1446, https://aclanthology.org/D18-1446

Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J (2020) Biobert: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 36(4):1234–1240

Levy O, Seo M, Choi E, Zettlemoyer L (2017) Zero-shot relation extraction via reading comprehension. In: Levy R, Specia L (eds) Proceedings of the 21st conference on computational natural language learning (CoNLL 2017). Association for Computational Linguistics, Vancouver, Canada, pp 333–342. https://doi.org/10.18653/v1/K17-1034, https://aclanthology.org/K17-1034

Li J, Sun Y, Johnson RJ, Sciaky D, Wei CH, Leaman R, Davis AP, Mattingly CJ, Wiegers TC, Lu Z (2016) Biocreative v CDR task corpus: a resource for chemical disease relation extraction. Database 2016

Liang Z, Zhang J, Shao J, Zhang X (2021) MWP-BERT: a strong baseline for math word problems. arXiv Preprint. https://doi.org/10.48550/arXiv.2107.13435

Liang X, Wu S, Li M, Li Z (2022) Modeling multi-granularity hierarchical features for relation extraction. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Seattle, United States, pp 5088–5098. https://doi.org/10.18653/v1/2022.naacl-main.375, https://aclanthology.org/2022.naacl-main.375

Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V (2019) ROBERTA: a robustly optimized BERT pretraining approach. arXiv preprint. arXiv:1907.11692

Liu Y, Gu J, Goyal N, Li X, Edunov S, Ghazvininejad M, Lewis M, Zettlemoyer L (2020) Multilingual denoising pre-training for neural machine translation. Trans Assoc Comput Linguist 8:726–742. https://aclanthology.org/2020.tacl-1.47/

Liu F, Lin H, Han X, Cao B, Sun L (2022a) Pre-training to match for unified low-shot relation extraction. In: Muresan S, Nakov P, Villavicencio A (eds) Proceedings of the 60th annual meeting of the Association for Computational Linguistics (volume 1: long papers). Association for Computational Linguistics, Dublin, Ireland, pp 5785–5795. https://doi.org/10.18653/v1/2022.acl-long.397, https://aclanthology.org/2022.acl-long.397

Liu S, Hu X, Zhang C, Li S, Wen L, Yu P (2022b) HiURE: Hierarchical exemplar contrastive learning for unsupervised relation extraction. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Seattle, United States, pp 5970–5980. https://doi.org/10.18653/v1/2022.naacl-main.437, https://aclanthology.org/2022.naacl-main.437

Lu D, Neves L, Carvalho V, Zhang N, Ji H (2018) Visual attention model for name tagging in multimodal social media. In: Proceedings of the 56th annual meeting of the Association for Computational Linguistics (volume 1: long papers), pp 1990–1999

Luan Y, He L, Ostendorf M, Hajishirzi H (2018) Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In: Riloff E, Chiang D, Hockenmaier J, Tsujii J (eds) Proceedings of the 2018 conference on empirical methods in natural language processing. Association for Computational Linguistics, Brussels, Belgium, pp 3219–3232. https://doi.org/10.18653/v1/D18-1360, https://aclanthology.org/D18-1360

Luo L, Lai PT, Wei CH, Arighi CN, Lu Z (2022a) Biored: a rich biomedical relation extraction dataset. Brief Bioinform 23(5):bbac282

Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H, Liu TY (2022b) BioGPT: generative pre-trained transformer for biomedical text generation and mining. Brief Bioinform 23(6):bbac409

Lyu S, Chen H (2021) Relation classification with entity type restriction. In: Zong C, Xia F, Li W, Navigli R (eds) Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, Online, pp 390–395. https://doi.org/10.18653/v1/2021.findings-acl.34, https://aclanthology.org/2021.findings-acl.34

Ma R, Gui T, Li L, Zhang Q, Huang X, Zhou Y (2021) SENT: Sentence-level distant relation extraction via negative training. In: Zong C, Xia F, Li W, Navigli R (eds) Proceedings of the 59th annual meeting of the association for Computational Linguistics and the 11th international joint conference on natural language processing (volume 1: long papers). Association for Computational Linguistics, Online, pp 6201–6213. https://doi.org/10.18653/v1/2021.acl-long.484, https://aclanthology.org/2021.acl-long.484

Ma Y, Wang A, Okazaki N (2023) DREEAM: Guiding attention with evidence for improving document-level relation extraction. In: Vlachos A, Augenstein I (eds) Proceedings of the 17th conference of the European chapter of the Association for Computational Linguistics. Association for Computational Linguistics, Dubrovnik, Croatia, pp 1971–1983. https://doi.org/10.18653/v1/2023.eacl-main.145, https://aclanthology.org/2023.eacl-main.145

Mastropaolo A, Scalabrino S, Cooper N, Palacio DN, Poshyvanyk D, Oliveto R, Bavota G (2021) Studying the usage of text-to-text transfer transformer to support code-related tasks. In: 2021 IEEE/ACM 43rd international conference on software engineering (ICSE). IEEE, pp 336–347

Mathur P, Jain R, Dernoncourt F, Morariu V, Tran QH, Manocha D (2021) Timers: document-level temporal relation extraction. In: Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on natural language processing (volume 2: short papers), pp 524–533

Mikolov T, Zweig G (2012) Context dependent recurrent neural network language model. In: 2012 IEEE workshop on spoken language technology, SLT 2012—proceedings. IEEExplore, pp 234–239. https://doi.org/10.1109/SLT.2012.6424228

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems, vol 26

Mirza P, Tonelli S (2014) An analysis of causality between events and its relation to temporal information. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. pp. 2097–2106

Mostafazadeh N, Grealish A, Chambers N, Allen J, Vanderwende L (2016) Caters: causal and temporal relation scheme for semantic annotation of event structures. In: Proceedings of the fourth workshop on events, pp 51–61

Mysore S, Jensen Z, Kim E, Huang K, Chang HS, Strubell E, Flanigan J, McCallum A, Olivetti E (2019) The materials science procedural text corpus: annotating materials synthesis procedures with shallow semantic structures. arXiv preprint. arXiv:1905.06939

Nagy K, Kapusta J (2021) Improving fake news classification using dependency grammar. PLoS ONE 16(9):e0256940

Najafi S, Fyshe A (2023) Weakly-supervised questions for zero-shot relation extraction. In: Vlachos A, Augenstein I (eds) Proceedings of the 17th conference of the European chapter of the Association for Computational Linguistics. Association for Computational Linguistics, Dubrovnik, Croatia, pp 3075–3087. https://doi.org/10.18653/v1/2023.eacl-main.224, https://aclanthology.org/2023.eacl-main.224

Nan G, Guo Z, Sekulic I, Lu W (2020) Reasoning with latent structure refinement for document-level relation extraction. In: Proceedings of the 58th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp 1546–1557. https://doi.org/10.18653/v1/2020.acl-main.141, https://aclanthology.org/2020.acl-main.141

Nassiri K, Akhloufi M (2023) Transformer models used for text-based question answering systems. Appl Intell 53(9):10602–10635

Ning Q, Wu H, Roth D (2018) A multi-axis annotation scheme for event temporal relations. arXiv preprint arXiv:1804.07828

OpenAI AJ, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman FL, Almeida D, Altenschmidt J, Altman, S, Anadkat S et al (2023) Gpt-4 technical report. arXiv preprint. arXiv:2303.08774

Ouchi H, Watanabe T, Matsumoto Y (2022) Improving discriminative learning for zero-shot relation extraction. In: Proceedings of the 1st workshop on semiparametric methods in NLP: decoupling logic from knowledge. pp 1

Parikh A, Täckström O, Das D, Uszkoreit J (2016) A decomposable attention model for natural language inference. In: Proceedings of the 2016 conference on empirical methods in natural language processing. Association for Computational Linguistics, Austin, pp 2249–2255. https://doi.org/10.18653/v1/D16-1244, https://www.aclweb.org/anthology/D16-1244

Patel S, Khashabi D, Roth D (2021) Svamp: structural variations for assessing mathematical problem-solving. In: Proceedings of the 2021 conference of the North American chapter of the Association for Computational Linguistics: human language technologies, pp 305–316

Peng Y, Yan S, Lu Z (2019) Transfer learning in biomedical natural language processing: an evaluation of BERT and ELMo on ten benchmarking datasets. In: Demner-Fushman D, Cohen KB, Ananiadou S, Tsujii J (eds) Proceedings of the 18th BioNLP workshop and shared task. Association for Computational Linguistics, Florence, Italy, pp 58–65. https://doi.org/10.18653/v1/W19-5006, https://aclanthology.org/W19-5006

Pennington J, Socher R, Manning CD (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 1532–1543

Popovic N, Färber M (2022) Few-shot document-level relation extraction. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Seattle, United States, pp 5733–5746. https://doi.org/10.18653/v1/2022.naacl-main.421, https://aclanthology.org/2022.naacl-main.421

Pouran Ben Veyseh A, Dernoncourt F, Dou D, Nguyen TH (2020) Exploiting the syntax-model consistency for neural relation extraction. In: Jurafsky D, Chai J, Schluter N, Tetreault J (eds) Proceedings of the 58th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp 8021–8032. https://doi.org/10.18653/v1/2020.acl-main.715, https://aclanthology.org/2020.acl-main.715

Qin C, Joty S (2022) Continual few-shot relation learning via embedding space regularization and data augmentation. In: Muresan S, Nakov P, Villavicencio A (eds) Proceedings of the 60th annual meeting of the Association for Computational Linguistics (volume 1: long papers). Association for Computational Linguistics, Dublin, Ireland, pp 2776–2789. https://doi.org/10.18653/v1/2022.acl-long.198, https://aclanthology.org/2022.acl-long.198

Radford A, Narasimhan K, Salimans T, Sutskever I et al (2018) Improving language understanding by generative pre-training. arXiv Preprint

Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ (2020) Exploring the limits of transfer learning with a unified text-to-text transformer. J Mach Learn Res 21(1):5485–5551

Ranjan N, Mundada K, Phaltane K, Ahmad S (2016) A survey on techniques in NLP. Int J Comput Appl 134(8):6–9

Reimers N, Gurevych I (2019) Sentence-Bert: sentence embeddings using siamese bert-networks. arXiv preprint. arXiv:1908.10084

Riedel S, Yao L, McCallum A (2010) Modeling relations and their mentions without labeled text. In: Machine learning and knowledge discovery in databases: European conference, ECML PKDD 2010, Barcelona, Spain, 20–24 September 2010, proceedings, Part III 21. Springer, pp 148–163

Roy A, Pan S (2021) Incorporating medical knowledge in bert for clinical relation extraction. In: Proceedings of the 2021 conference on empirical methods in natural language processing, pp 5357–5366

Seagnti A, Firląg K, Skowronska H, Satława M, Andruszkiewicz P (2021) Multilingual entity and relation extraction dataset and model. In: Merlo P, Tiedemann J, Tsarfaty R (eds) Proceedings of the 16th conference of the European chapter of the Association for Computational Linguistics: main volume. Association for Computational Linguistics, Online, pp 1946–1955. https://doi.org/10.18653/v1/2021.eacl-main.166, https://aclanthology.org/2021.eacl-main.166

Shahbazi H, Fern X, Ghaeini R, Tadepalli P (2020) Relation extraction with explanation. In: Proceedings of the 58th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp 6488–6494. https://doi.org/10.18653/v1/2020.acl-main.579, https://aclanthology.org/2020.acl-main.579

Shen Y, Jin C (2020) Solving math word problems with multi-encoders and multi-decoders. In: Proceedings of the 28th International Conference on Computational Linguistics, pp 2924–2934

Shen Y, Tan S, Sordoni A, Courville A (2018) Ordered neurons: integrating tree structures into recurrent neural networks. arXiv preprint. arXiv:1810.09536

Sui D, Chen Y, Liu K, Zhao J (2021) Distantly supervised relation extraction in federated settings. In: Moens MF, Huang X, Specia L, Yih SW (eds) Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, Punta Cana, Dominican Republic, pp 569–583. https://doi.org/10.18653/v1/2021.findings-emnlp.52, https://aclanthology.org/2021.findings-emnlp.52

Suissa O, Elmalech A, Zhitomirsky-Geffet M (2022) Text analysis using deep neural networks in digital humanities and information science. J Am Soc Inf Sci 73(2):268–287

Sun P, Yang X, Zhao X, Wang Z (2018) An overview of named entity recognition. In: 2018 International conference on Asian language processing (IALP). IEEE, pp 273–278

Sun Q, Huang K, Yang X, Hong P, Zhang K, Poria S (2023) Uncertainty guided label denoising for document-level distant relation extraction. 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). Association for Computational Linguistics, Toronto, Canada, pp 15960–15973. https://doi.org/10.18653/v1/2023.acl-long.889, https://aclanthology.org/2023.acl-long.889

Tan Q, Xu L, Bing L, Ng HT, Aljunied SM (2022) Revisiting docred-addressing the false negative problem in relation extraction. In: Proceedings of the 2022 conference on empirical methods in natural language processing, pp 8472–8487

Tan X, Pergola G, He Y (2023) Event temporal relation extraction with Bayesian translational model. In: Vlachos A, Augenstein I (eds) Proceedings of the 17th conference of the European chapter of the Association for Computational Linguistics Association for Computational Linguistics, Dubrovnik, Croatia, pp 1125–1138.. https://doi.org/10.18653/v1/2023.eacl-main.80, https://aclanthology.org/2023.eacl-main.80

Team G, Anil R, Borgeaud S, Wu Y, Alayrac JB, Yu J, Soricut R, Schalkwyk J, Dai AM, Hauth A et al (2023) Gemini: a family of highly capable multimodal models. arXiv preprint. arXiv:2312.11805

Teru K (2023) Semi-supervised relation extraction via data augmentation and consistency-training. In: Vlachos A, Augenstein I (eds) Proceedings of the 17th conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, Dubrovnik, Croatia, pp 1112–1124. https://doi.org/10.18653/v1/2023.eacl-main.79, https://aclanthology.org/2023.eacl-main.79

Tezgider M, Yildiz B, Aydin G (2022) Text classification using improved bidirectional transformer. Concurr Comput Pract Exp 34(9):e6486

Thomas A, Sivanesan S (2022) An adaptable, high-performance relation extraction system for complex sentences. Knowl-Based Syst 251:108956

Tian H, Yang K, Liu D, Lv J (2021a) Anchibert: a pre-trained model for ancient chinese language understanding and generation. In: 2021 International joint conference on neural networks (IJCNN). IEEE, pp 1–8

Tian Y, Chen G, Song Y, Wan X (2021b) Dependency-driven relation extraction with attentive graph convolutional networks. In: Zong C, Xia F, Li W, Navigli R (eds) Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on natural language processing (volume 1: long papers). Association for Computational Linguistics, Online, pp 4458–4471. https://doi.org/10.18653/v1/2021.acl-long.344, https://aclanthology.org/2021.acl-long.344

Tiktinsky A, Viswanathan V, Niezni D, Meron Azagury D, Shamay Y, Taub-Tabib H, Hope T, Goldberg Y (2022) A dataset for n-ary relation extraction of drug combinations. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Seattle, United States, pp 3190–3203. https://doi.org/10.18653/v1/2022.naacl-main.233, https://aclanthology.org/2022.naacl-main.233

Tjong Kim Sang EF, De Meulder F (2003) Introduction to the CoNLL-2003 shared task: language-independent named entity recognition. In: Proceedings of the seventh conference on natural language learning at HLT-NAACL 2003, pp 142–147, https://aclanthology.org/W03-0419

Touvron H, Lavril T, Izacard G, Martinet X, Lachaux MA, Lacroix T, Rozière B, Goyal N, Hambro E, Azhar F et al (2023) Llama: open and efficient foundation language models. arXiv preprint. arXiv:2302.13971

Tran TT, Le P, Ananiadou S (2020) Revisiting unsupervised relation extraction. In: Jurafsky D, Chai J, Schluter N, Tetreault J (eds.) Proceedings of the 58th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp 7498–7505. https://doi.org/10.18653/v1/2020.acl-main.669, https://aclanthology.org/2020.acl-main.669/

Tran VH, Ouchi H, Shindo H, Matsumoto Y, Watanabe T (2023) Enhancing semantic correlation between instances and relations for zero-shot relation extraction. J Nat Lang Process 30(2):304–329

Vakil N, Amiri H (2022) Generic and trend-aware curriculum learning for relation extraction. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Seattle, United States, pp 2202–2213. https://doi.org/10.18653/v1/2022.naacl-main.160, https://aclanthology.org/2022.naacl-main.160

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in neural information processing systems, vol 30. Curran Associates, Inc., pp 5998–6008. http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf

Veyseh APB, Dernoncourt F, Dou D, Nguyen TH (2020) Exploiting the syntax-model consistency for neural relation extraction. In: Proceedings of the 58th annual meeting of the Association for Computational Linguistics, pp 8021–8032

Wadhwa S, Amir S, Wallace B (2023) Revisiting relation extraction in the era of large language models. 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). Association for Computational Linguistics, Toronto, Canada, pp 15566–15589. https://doi.org/10.18653/v1/2023.acl-long.868, https://aclanthology.org/2023.acl-long.868

Walker C, Strassel S, Medero J, Maeda K (2006) Ace 2005 multilingual training corpus. Linguistic Data Consortium, Philadelphia, vol 57, p 45

Wang S, Liu C, Zhu KQ (2017) Deep neural solver for math word problems. In: Proceedings of the thirty-first AAAI conference on artificial intelligence, pp 1170–1176

Wang H, Xiong W, Yu M, Guo X, Chang S, Wang WY (2019) Sentence embedding alignment for lifelong relation extraction. In: Burstein J, Doran C, Solorio T (eds) Proceedings of the 2019 conference of the North American chapter of the Association for Computational linguistics: human language technologies, volume 1 (long and short papers). Association for Computational Linguistics, Minneapolis, pp 796–806. https://doi.org/10.18653/v1/N19-1086, https://aclanthology.org/N19-1086/

Wang Y, Sun C, Wu Y, Zhou H, Li L, Yan J (2021) UniRE: a unified label space for entity relation extraction. In: Zong C, Xia F, Li W, Navigli R (eds) Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on natural language processing (volume 1: long papers). Association for Computational Linguistics, Online, pp 220–231. https://doi.org/10.18653/v1/2021.acl-long.19, https://aclanthology.org/2021.acl-long.19

Wang C, Liu X, Chen Z, Hong H, Tang J, Song D (2022a) DeepStruct: Pretraining of language models for structure prediction. In: Muresan S, Nakov P, Villavicencio A (eds) Findings of the Association for Computational Linguistics: ACL 2022. Association for Computational Linguistics, Dublin, Ireland, pp 803–823. https://doi.org/10.18653/v1/2022.findings-acl.67, https://aclanthology.org/2022.findings-acl.67

Wang Y, Chen M, Zhou W, Cai Y, Liang Y, Liu D, Yang B, Liu J, Hooi B (2022b) Should we rely on entity mentions for relation extraction? Debiasing relation extraction with counterfactual analysis. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American chapter of the Association for computational linguistics: human language technologies. Association for Computational Linguistics, Seattle, United States, pp 3071–3081. https://doi.org/10.18653/v1/2022.naacl-main.224, https://aclanthology.org/2022.naacl-main.224

Weng Y, Chen X, Chen L, Liu W (2020) GAIN: graph attention & interaction network for inductive semi-supervised learning over large-scale graphs. IEEE Trans Knowl Data Eng 34(9):4257–4269

Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M et al (2019) Huggingface's transformers: state-of-the-art natural language processing. arXiv preprint. arXiv:1910.03771

Wu S, Fei H, Cao Y, Bing L, Chua TS (2023) Information screening whilst exploiting! multimodal relation extraction with feature denoising and multimodal topic modeling. 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). Association for Computational Linguistics, Toronto, Canada, pp 14734–14751. https://doi.org/10.18653/v1/2023.acl-long.823, https://aclanthology.org/2023.acl-long.823

Xiao Y, Zhang Z, Mao Y, Yang C, Han J (2022) SAIS: supervising and augmenting intermediate steps for document-level relation extraction. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Seattle, United States, pp 2395–2409. https://doi.org/10.18653/v1/2022.naacl-main.171, https://aclanthology.org/2022.naacl-main.171

Xie C, Liang J, Liu J, Huang C, Huang W, Xiao Y (2021) Revisiting the negative data of distantly supervised relation extraction. In: Zong C, Xia F, Li W, Navigli R (eds) Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on natural language processing (volume 1: long papers). Association for Computational Linguistics, Online, pp 3572–3581. https://doi.org/10.18653/v1/2021.acl-long.277, https://aclanthology.org/2021.acl-long.277

Xie Y, Shen J, Li S, Mao Y, Han J (2022) Eider: empowering document-level relation extraction with efficient evidence extraction and inference-stage fusion. In: Muresan S, Nakov P, Villavicencio A (eds) Findings of the association for computational linguistics: ACL 2022. Association for Computational Linguistics, pp 257–268. https://doi.org/10.18653/v1/2022.findings-acl.23

Xu X, Cai H (2021) Ontology and rule-based natural language processing approach for interpreting textual regulations on underground utility infrastructure. Adv Eng Inform 48:101288

Xu L, Choi J (2022) Modeling task interactions in document-level joint entity and relation extraction. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Seattle, United States, pp 5409–5416. https://doi.org/10.18653/v1/2022.naacl-main.395, https://aclanthology.org/2022.naacl-main.395

Xu Y, Li M, Cui L, Huang S, Wei F, Zhou M (2020) Layoutlm: pre-training of text and layout for document image understanding. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp 1192–1200

Xu W, Chen K, Mou L, Zhao T (2022) Document-level relation extraction with sentences importance estimation and focusing. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Seattle, United States, pp 2920–2929. https://doi.org/10.18653/v1/2022.naacl-main.212, https://aclanthology.org/2022.naacl-main.212

Xu B, Wang Q, Lyu Y, Dai D, Zhang Y, Mao Z (2023a) S2ynRE: two-stage self-training with synthetic data for low-resource relation extraction. 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). Association for Computational Linguistics, Toronto, Canada, pp 8186–8207. https://doi.org/10.18653/v1/2023.acl-long.455, https://aclanthology.org/2023.acl-long.455

Xu J, Ma MD, Chen M (2023b) Can NLI provide proper indirect supervision for low-resource biomedical relation extraction? 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). Association for Computational Linguistics, Toronto, Canada, pp 2450–2467. https://doi.org/10.18653/v1/2023.acl-long.138, https://aclanthology.org/2023.acl-long.138

Xue L, Zhang D, Dong Y, Tang J (2024) AutoRE: Document-level relation extraction with large language models. In: Cao Y, Feng Y, Xiong D (eds) Proceedings of the 62nd annual meeting of the Association for Computational Linguistics (volume 3: system demonstrations). Association for Computational Linguistics, Bangkok, Thailand, pp 211–220. https://doi.org/10.18653/v1/2024.acl-demos.20, https://aclanthology.org/2024.acl-demos.20/

Yamakata Y, Mori S, Carroll JA (2020) English recipe flow graph corpus. In: Proceedings of the twelfth language resources and evaluation conference, pp 5187–5194

Yan Z, Jia Z, Tu K (2022) An empirical study of pipeline vs. joint approaches to entity and relation extraction. In: He Y, Ji H, Li S, Liu Y, Chang CH (eds) Proceedings of the 2nd conference of the Asia-Pacific chapter of the Association for Computational Linguistics and the 12th international joint conference on natural language processing (volume 2: short papers). Association for Computational Linguistics, Online, pp 437–443. https://aclanthology.org/2022.aacl-short.55

Yan Z, Zhang C, Fu J, Zhang Q, Wei Z (2021) A partition filter network for joint entity and relation extraction. In: Moens MF, Huang X, Specia L, Yih SW (eds) Proceedings of the 2021 conference on empirical methods in natural language processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pp 185–197. https://doi.org/10.18653/v1/2021.emnlp-main.17, https://aclanthology.org/2021.emnlp-main.17

Yang S, Song D (2022) FPC: fine-tuning with prompt curriculum for relation extraction. In: He Y, Ji H, Li S, Liu Y, Chang CH (eds) Proceedings of the 2nd conference of the Asia-Pacific chapter of the Association for Computational Linguistics and the 12th international joint conference on natural language processing (volume 1: long papers). Association for Computational Linguistics, Online, pp 1065–1077. https://aclanthology.org/2022.aacl-main.78

Yang S, Zhang Y, Niu G, Zhao Q, Pu S (2021a) Entity concept-enhanced few-shot relation extraction. arXiv preprint. arXiv:2106.02401

Yang S, Zhang Y, Niu G, Zhao Q, Pu S (2021b) Entity concept-enhanced few-shot relation extraction. In: Zong C, Xia F, Li W, Navigli R (eds) Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on natural language processing (volume 2: short papers). Association for Computational Linguistics, Online, pp 987–991. https://doi.org/10.18653/v1/2021.acl-short.124, https://aclanthology.org/2021.acl-short.124

Yang S, Choi M, Cho Y, Choo J (2023) 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). Association for Computational Linguistics, Toronto, Canada, pp 3207–3224. https://doi.org/10.18653/v1/2023.acl-long.180, https://aclanthology.org/2023.acl-long.180

Yao L, Mao C, Luo Y (2019a) Kg-Bert: Bert for knowledge graph completion. arXiv preprint. arXiv:1909.03193

Yao Y, Ye D, Li P, Han X, Lin Y, Liu Z, Liu Z, Huang L, Zhou J, Sun M (2019b) DocRED: a large-scale document-level relation extraction dataset. In: Proceedings of the 57th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Florence, Italy, pp 764–777. https://doi.org/10.18653/v1/P19-1074, https://aclanthology.org/P19-1074

Ye D, Lin Y, Li P, Sun M (2022) Packed levitated marker for entity and relation extraction. In: Muresan S, Nakov P, Villavicencio A (eds) Proceedings of the 60th annual meeting of the Association for Computational Linguistics (volume 1: long papers). Association for Computational Linguistics, Dublin, Ireland, pp 4904–4917. https://doi.org/10.18653/v1/2022.acl-long.337, https://aclanthology.org/2022.acl-long.337

Yu D, Sun K, Cardie C, Yu D (2020) Dialogue-based relation extraction. In: Proceedings of the 58th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp 4927–4940. https://doi.org/10.18653/v1/2020.acl-main.444, https://aclanthology.org/2020.acl-main.444

Yu J, Yang D, Tian S (2022) Relation-specific attentions over entity mentions for enhanced document-level relation extraction. In: Carpuat M, de Marneffe MC, Meza Ruiz IV (eds) Proceedings of the 2022 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Seattle, United States, pp 1523–1529. https://doi.org/10.18653/v1/2022.naacl-main.109, https://aclanthology.org/2022.naacl-main.109

Yuan C, Eldardiry H (2021) Unsupervised relation extraction: a variational autoencoder approach. In: Moens MF, Huang X, Specia L, Yih SW (eds) Proceedings of the 2021 conference on empirical methods in natural language processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pp 1929–1938. https://doi.org/10.18653/v1/2021.emnlp-main.147, https://aclanthology.org/2021.emnlp-main.147

Yuan C, Huang H, Cao Y, Wen Y (2023) Discriminative reasoning with sparse event representation for document-level event-event relation extraction. 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). Association for Computational Linguistics, Toronto, Canada, pp 16222–16234. https://doi.org/10.18653/v1/2023.acl-long.897, https://aclanthology.org/2023.acl-long.897

Zaporojets K, Deleu J, Develder C, Demeester T (2021) DWIE: an entity-centric dataset for multi-task document-level information extraction. Inform Process Manag 58(4):102563

Zaremba W, Sutskever I, Vinyals O (2014) Recurrent neural network regularization. arXiv preprint. http://arxiv.org/abs/1409.2329

Zhang D, Wang D (2015) Relation classification via recurrent neural network. arXiv preprint. arXiv:1508.01006

Zhang Y, Zhong V, Chen D, Angeli G, Manning CD (2017) 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. Association for Computational Linguistics, Copenhagen, Denmark, pp 35–45. https://doi.org/10.18653/v1/D17-1004, https://aclanthology.org/D17-1004

Zhang Q, Fu J, Liu X, Huang X (2018) Adaptive co-attention network for named entity recognition in tweets. In: Proceedings of the AAAI conference on artificial intelligence, vol 32

Zhang K, Yao Y, Xie R, Han X, Liu Z, Lin F, Lin L, Sun M (2021a) Open hierarchical relation extraction. In: Toutanova K, Rumshisky A, Zettlemoyer L, Hakkani-Tur D, Beltagy I, Bethard S, Cotterell R, Chakraborty T, Zhou Y (eds) Proceedings of the 2021 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Online, pp 5682–5693. https://doi.org/10.18653/v1/2021.naacl-main.452, https://aclanthology.org/2021.naacl-main.452

Zhang Y, Bo Z, Wang R, Cao J, Li C, Bao Z (2021b) Entity relation extraction as dependency parsing in visually rich documents. In: Moens MF, Huang X, Specia L, Yih SWt (eds) Proceedings of the 2021 conference on empirical methods in natural language processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pp 2759–2768. https://doi.org/10.18653/v1/2021.emnlp-main.218, https://aclanthology.org/2021.emnlp-main.218

Zhang R, Li Y, Zou L (2023) A novel table-to-graph generation approach for document-level joint entity and relation extraction. 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). Association for Computational Linguistics, Toronto, Canada, pp 10853–10865. https://doi.org/10.18653/v1/2023.acl-long.607, https://aclanthology.org/2023.acl-long.607

Zhao J, Gui T, Zhang Q, Zhou Y (2021) A relation-oriented clustering method for open relation extraction. In: Moens MF, Huang X, Specia L, Yih SWt (eds) Proceedings of the 2021 conference on empirical methods in natural language processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pp 9707–9718. https://doi.org/10.18653/v1/2021.emnlp-main.765, https://aclanthology.org/2021.emnlp-main.765

Zhao J, Zhan W, Zhao X, Zhang Q, Gui T, Wei Z, Wang J, Peng M, Sun M (2023a) RE-matching: a fine-grained semantic matching method for zero-shot relation extraction. 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). Association for Computational Linguistics, Toronto, Canada, pp 6680–6691. https://doi.org/10.18653/v1/2023.acl-long.369, https://aclanthology.org/2023.acl-long.369

Zhao J, Zhang Y, Zhang Q, Gui T, Wei Z, Peng M, Sun M (2023b) Actively supervised clustering for open relation extraction. 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). Association for Computational Linguistics, Toronto, Canada, pp 4985–4997. https://doi.org/10.18653/v1/2023.acl-long.273, https://aclanthology.org/2023.acl-long.273

Zhao J, Zhao X, Zhan W, Zhang Q, Gui T, Wei Z, Chen YW, Gao X, Huang X (2023c) Open set relation extraction via unknown-aware training. 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). Association for Computational Linguistics, Toronto, Canada, pp 9453–9467. https://doi.org/10.18653/v1/2023.acl-long.525, https://aclanthology.org/2023.acl-long.525

Zhao W, Cui Y, Hu W (2023d) Improving continual relation extraction by distinguishing analogous semantics. 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). Association for Computational Linguistics, Toronto, Canada, pp 1162–1175. https://doi.org/10.18653/v1/2023.acl-long.65, https://aclanthology.org/2023.acl-long.65

Zheng C, Feng J, Fu Z, Cai Y, Li Q, Wang T (2021) Multimodal relation extraction with efficient graph alignment. In: Proceedings of the 29th ACM international conference on multimedia, pp 5298–5306

Zheng C, Feng J, Cai Y, Wei X, Li Q (2023a) Rethinking multimodal entity and relation extraction from a translation point of view. 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). Association for Computational Linguistics, Toronto, Canada, pp 6810–6824. https://doi.org/10.18653/v1/2023.acl-long.376, https://aclanthology.org/2023.acl-long.376

Zheng Y, Hao A, Luu AT (2023b) Jointprop: Joint semi-supervised learning for entity and relation extraction with heterogeneous graph-based propagation. 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). Association for Computational Linguistics, Toronto, Canada, pp 14541–14555. https://doi.org/10.18653/v1/2023.acl-long.813, https://aclanthology.org/2023.acl-long.813

Zhong Z, Chen D (2021) A frustratingly easy approach for entity and relation extraction. In: Toutanova K, Rumshisky A, Zettlemoyer L, Hakkani-Tur D, Beltagy I, Bethard S, Cotterell R, Chakraborty T, Zhou Y (eds) Proceedings of the 2021 conference of the North American chapter of the Association for Computational Linguistics: human language technologies. Association for Computational Linguistics, Online, pp 50–61. https://doi.org/10.18653/v1/2021.naacl-main.5, https://aclanthology.org/2021.naacl-main.5

Zhou W, Chen M (2022) An improved baseline for sentence-level relation extraction. In: He Y, Ji H, Li S, Liu Y, Chang CH (eds) Proceedings of the 2nd conference of the Asia-Pacific chapter of the Association for Computational Linguistics and the 12th international joint conference on natural language processing (volume 2: short papers). Association for Computational Linguistics, Online, pp 161–168. https://aclanthology.org/2022.aacl-short.21

Zhou D, Zhong D, He Y (2014) Biomedical relation extraction: from binary to complex. Comput Math Methods Med 2014(1):298473

Zhou W, Huang K, Ma T, Huang J (2021) Document-level relation extraction with adaptive thresholding and localized context pooling. Proc AAAI Conf Artif Intell 35:14612–14620

Zhou W, Zhang S, Naumann T, Chen M, Poon H (2023) Continual contrastive finetuning improves low-resource relation extraction. 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). Association for Computational Linguistics, Toronto, Canada, pp 13249–13263. https://doi.org/10.18653/v1/2023.acl-long.739, https://aclanthology.org/2023.acl-long.739

Zhu Y, Kiros R, Zemel R, Salakhutdinov R, Urtasun R, Torralba A, Fidler S (2015) Aligning books and movies: towards story-like visual explanations by watching movies and reading books. In: Proceedings of the 2015 IEEe international conference on computer vision (ICCV), ICCV '15. IEEE Computer Society, pp 19–27. https://doi.org/10.1109/ICCV.2015.11

Zhu Y, Yan E, Song IY (2017) A natural language interface to a graph-based bibliographic information retrieval system. Data Knowl Eng 111:73–89