A Survey on Open Information Extraction from Rule-based Model to Large Language Model

Anonymous ACL submission

Abstract

Open Information Extraction (OpenIE) represents a crucial NLP task aimed at deriving structured information from unstructured text, unrestricted by relation type or domain. This survey paper provides an overview of OpenIE technologies spanning from 2007 to 2024, emphasizing a chronological perspective absent in prior surveys. It examines the evolution of task settings in OpenIE to align with the advances in recent technologies. The paper categorizes OpenIE approaches into rule-based, neural, and pre-trained large language models, discussing each within a chronological framework. Additionally, it highlights prevalent datasets and evaluation metrics currently in use. Building on this extensive review, the paper considers how traditional OpenIE research can inspire future IE research in the LLM era, aiming to provide insights into the past, present, and future of OpenIE methodologies and applications.

1 Introduction

Open Information Extraction (OpenIE) aims to extract structured information from unstructured text sources (Niklaus et al., 2018), typically outputting relationships as triplets (arg_1, rel, arg_2). As illustrated in Figure 1, unlike standard IE, which relies on predefined categories to identify relationships, OpenIE operates without such constraints, enabling the extraction of diverse and unforeseen relations. This flexibility makes OpenIE especially valuable for rapidly evolving Natural Language Processing (NLP) tasks such as question answering, search engines, and knowledge graph completion (Han et al., 2020), as well as for handling large-scale and dynamic data sources like web data.

Since its inception in 2007, the field of OpenIE has witnessed continual advancements. Initially utilizing basic linguistic tools, OpenIE models have progressively integrated more complex syntactic and semantic features, while preserving

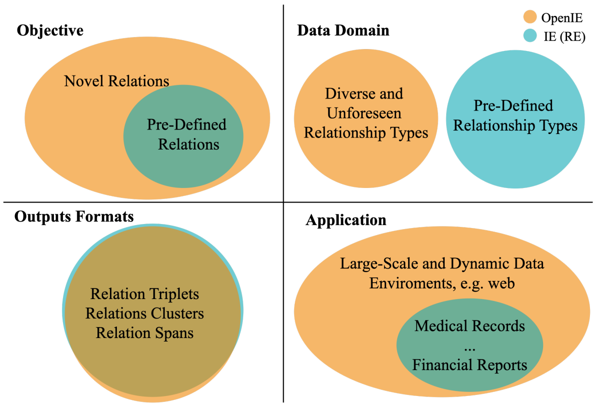

Figure 1: Comparison of OpenIE and standard relation extraction.

the intuitive task of directly extracting relational triplets from text. The advent of neural models in 2019 marks a paradigm shift for OpenIE research, where systems employing Transformer-based architectures like BERT (Devlin et al., 2019) significantly enhance feature extraction capabilities. To accommodate the technological shift, a variety of methods and task settings have evolved within diversified OpenIE approaches.

The emergence of Large Language Models (LLMs) in 2023 has marked another revolutionary phase, steering OpenIE toward a generative method of information extraction. The robust generalization abilities of these models not only advance the technical prowess of OpenIE systems but also facilitate a convergence of methodologies and task settings – revisiting the original, straightforward ***text* → *relational triplet*** format. This transition also fosters potential integration with standard IE tasks, pointing toward a promising future where extraction tasks are tackled through a unified, multi-task approach.

As a result, there has been a decline in OpenIE research in the LLM era. *Is OpenIE research going to its end? How can traditional OpenIE research inspire IE research in the LLM era?* Previous surveys largely focus on pre-LLM era models or limit their scope to methodological insights (Gamallo, 2014; Vo and Bagheri, 2018; Zouaq et al., 2017; Glauber and Claro, 2018; Niklaus et al., 2018; Zhou et al., 2022). While recent studies (Xu et al., 2023b) delve into information extraction in the LLM era, they largely bypass OpenIE, concentrating instead on standard IE tasks. We aim to bridge this gap by providing a holistic survey of the OpenIE field from a chronological view, addressing the two research questions above.

From a chronological perspective, we summarize all task settings (Section 2), data (Section 3), evaluation metrics (Section 4), and mainstream methods (Section 5) before and after LLM era. We use a single table to summarize mainstream methods and results from different periods. We emphasize the co-evolution between models and task settings, and the various sources of information used to address Open challenges. Based on this, we compare the ideas and relative strengths and weaknesses of large models and traditional models (Section 6.1), review the impact of large language models on open information extraction and traditional methods (Section 6.2), and explore future directions (Section 6.3).

2 Task Settings

We categorize OpenIE task settings into three groups: Open Relation Triplet Extraction (***ORTE***), Open Relation Span Extraction (***ORSE***) and Open relation clustering (***ORC***). ***ORTE*** is the classic task setting, while ***ORSE*** and ***ORC*** settings are variations developed to cater to diverse models with the advancement of NLP techniques. For all three task settings, openness is shown in the absence of restraints on relation types. Figure 2 depicts the workflow for each task setting.

ORTE Task: Text → Relational Triplet
Banko et al. (2007) initially defines open information extraction as an unsupervised task that automatically extracts $(entity_1, relation, entity_2)$ triplets from a vast corpus of unstructured web text, where $entity_1$, $entity_2$ and $relation$ consist of selected words from input sentences. Although the term *triplet* is more commonly used, the actual extraction tasks are not always limited to triplets and can involve more diverse n-ary relations, such as condition, temporal information, etc. This task setting, irrespective of the learning method or the forms of input and output, represents the most idealized configuration.

ORSE Task: Entities + Text → Relation Span
Different from the first setting, open relation span extraction finds relational spans according to previously extracted predicates and entities, aiming to partition complex tasks into easier ones to improve model performance. However, it should be clear that errors in entity extraction steps can accumulate in two-stage pipelines. See Open Relation Extraction (***ORSE***) in Fig.2 for an example.

ORC Task: Entities + Text→ Clustering without Explicit Relation Span or Label
Open relation clustering (***ORC***), also widely known as open relation extraction, clusters relation instances (h, t, s), where h and t denote head entity and tail entity respectively, and s denotes the sentence corresponding to two entities. Different from the ***ORTE***, ***ORC*** does not extract relation from text but uses text between two entities to represent the relation. Clustering similar relations is a step forward in labeling specific relations to each relation instance. These task settings outlined above are distinctly characterized by era-specific traits and methodologies, further discussed in Section 5.

3 Datasets

Table 1 lists some popular and promising OpenIE datasets grouped by their creating methods.

Question Answering (QA) derived datasets are converted from other crowd-sourced QA datasets. OIE2016 (Stanovsky and Dagan, 2016) is one of the most popular OpenIE benchmarks, which leverages QA-SRL (He et al., 2015) annotations. Additional datasets extend from OIE2016, such as AW-OIE (Stanovsky et al., 2018), Re-OIE2016 (Zhan and Zhao, 2020) and CaRB (Bhardwaj et al., 2019). LSOIE (Solawetz and Larson, 2021), is created by converting the QA-SRL 2.0 dataset (FitzGerald et al., 2018) to a large-scale OpenIE dataset, which claims to be 20 times larger than the next largest human-annotated OpenIE dataset.

Crowdsourced datasets are created from direct human annotation, including WiRe57 (Léchelle et al., 2019), SAOKE dataset (Sun et al., 2018), and BenchIE dataset (Gashteovski et al., 2021).

Knowledge Base (KB) derived datasets are established by aligning triplets in KBs with text in the corpus. Several works (Mintz et al., 2009; Yao et al., 2011) have aligned the New York Times corpus (Sandhaus, 2008) with Freebase (Bollacker

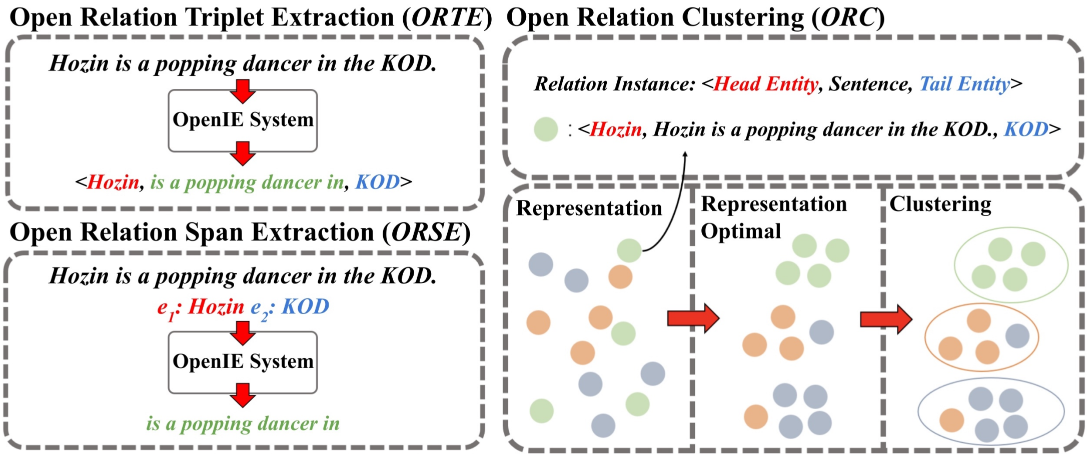

Figure 2: An overview of workflow processes in OpenIE task settings.

et al., 2008) triplets, resulting in several variations of the same dataset, NYT-FB. Others are created by aligning relations of given entity pairs(ElSahar et al., 2018), such as TACRED(Zhang et al., 2017), FewRel (Han et al., 2018), T-REx (ElSahar et al., 2018), T-REx SPO and T-REx DS (Hu et al., 2020). COER (Jia et al., 2018), a large-scale Chinese KB dataset, is automatically created by an unsupervised open extractor.

Instruction-based datasets transform IE tasks into tasks requiring instruction-following, thus harnessing the capabilities of LLMs. Strategies include integrating existing IE datasets into a unified-format (Wang et al., 2023a; Lu et al., 2022), and deriving others from Wikidata and Wikipedia such as INSTRUCTOPENWIKI (Lu et al., 2023), INSTRUCTIE (Gui et al., 2023), and Wikidata-OIE (Wang et al., 2022b).

Overall, KB derived datasets are mostly used in ***ORC*** task settings, whereas QA derived, crowdsourced, and instruction-based datasets are usually used in ***ORTE*** and ***ORSE*** task settings. We provide more detailed descriptions in Appendix C.

Dataset	#Tuple	Domain
	QA Derived	
OIE2016 (2016)	10,359	Wiki, Newswire
Re-OIE2016 (2020)	NR	Wiki, Newswire
CaRB (2019)	NR	Wiki, Newswire
AW-OIE (2018)	17,165	Wiki, Wikinews
LSOIE-wiki (2021)	56,662	Wiki, Wikinews
LSOIE-sci (2021)	97,550	Science
	Crowdsourced	
WiRe57 (2019)	343	Wiki, Newswire
SAOKE^{zh} (2018)	NR	Baidu Baike
BenchIE^{en} (2021)	136,357	Wiki, Newswire
BenchIE^{de} (2021)	82,260	Wiki, Newswire
BenchIE^{zh} (2021)	5,318	Wiki, Newswire
	KB Derived	
NYT-FB (2008; 2008; 2009; 2011)	39,000	NYT, Freebase
TACRED (2017)	119,474	TAC KBP
FewRel (2018)	70,000	Wiki, Wikidata
T-REx (2018)	11M	Wiki, Wikidata
COER^{zh} (2018)	1M	Baidu Baike, Chinese news
	Instruction-Based	
INSTRUCTOPENWIKI (2023)	19M	Wiki, Wikidata
Wikidata-OIE (2022b)	27M	Wiki, Wikidata

Table 1: Statistics of popular OpenIE datasets. "NR" stands for "Not Reported". Non-English datasets are indicated with superscripts.

Task Setting	Evaluation Metrics
ORTE	Recall, AUC, F1
ORSE	F1
ORC	ARI, B^3, V-measure

Table 2: Core evaluation metrics of each task setting.

4 Evaluation

Evaluation metrics for OpenIE models vary by task setting. In the ***ORTE*** and ***ORSE*** settings, models are assessed using precision, recall, F1 score, and AUC, potentially employing various scoring functions. In the ***ORC*** setting, performance is evaluated using B^3 (Bagga and Baldwin, 1998), V-measure (Rosenberg and Hirschberg, 2007), and ARI (Hubert and Arabie, 1985).

To compare the extracted and golden triplets, various datasets employ different matching strategies, typically categorized into **token-level** and **fact-level** scorers. Token-level scorers focus on individual tokens to ensure precision and semantic accuracy, accommodating linguistic variability (Stanovsky and Dagan, 2016), enhancing conciseness (Léchelle et al., 2019), and adapting to complex model outputs like those from LLMs (Han et al., 2023). Fact-level scorers assess the informational faithfulness of extractions to ensure reliable knowledge extraction, validating semantic and information integrity (Sun et al., 2018; Gashteovski et al., 2021; Li et al., 2023a) to enhance OpenIE evaluations comprehensively. Further details are discussed in Appendix D.

From the perspective of task formulation, token-level scorers are better suited for open relation span extraction (***ORSE***), where outputs are succinct, and labeling models in open relation triplet extraction (***ORTE***), whose outputs are precise tokens derived from the inputs. Conversely, fact-level scorers are more appropriate for generative models in ***ORTE***, particularly LLMs, whose outputs exhibit diversity and necessitate semantic evaluation.

5 A Chronological Review of Mainstream Methods

The research approaches for Open IE have undergone three significant changes along with technological advancements. We categorize these periods into three eras: the pre-neural era, dominated by rule-based and statistic-based methods; the neural model era, primarily based on neural networks; and the LLMs era, characterized by the use of LLMs. Chronologically, we will discuss the key models and methods from each period and explore their connections. More details about model implementation is provided in Appendix.A

5.1 Pre-neural Model Era

In the beginning, OpenIE systems were developed to create a universal model capable of extracting relation triplets through shallow features, such as Part-of-Speech (POS) that do not have lexical information, for instance, characterizing a verb based on its context. Traditional machine learning models, such as Naive Bayes (Rish et al., 2001) and Conditional Random Field (Sutton et al., 2012), are used to train on shallow features (Yates et al., 2007; Wu and Weld, 2010; Zhu et al., 2009). Using only lexical features will lead to problems of incoherent and uninformative relations. Therefore, lexical features and syntactic features are used to mitigate such problems (Schmitz et al., 2012; Qiu and Zhang, 2014; Mausam, 2016). Later, rule-based models take advantage of hand-written patterns and rules to match relations (Fader et al., 2011; Akbik and Löser, 2012). To extract relations in a fine-grained way, clause-based models determine the set of clauses and identify clause types before extracting relations (Del Corro and Gemulla, 2013; Schmidek and Barbosa, 2014; Angeli et al., 2015).

5.2 Neural Model Era

Sequence Labeling. RnnOIE (Stanovsky et al., 2018) is the first neural method, which formulates ORTE task as a sequence labeling problem where inputs a sequence of tokens $\{x_1, x_2, ..., x_n\}$ and outputs a sequence of labels $\{l_1, l_2, ..., l_n\}$ with the same length n as input. RnnOIE uses a Bi-LSTM to process input features, including word embeddings, POS tags, and indicated predicates. A Softmax classifier tags a BIO label for the last layer hidden state of each token, after which relation triplets are constructed. Since one sentence usually contains more than one relation triplet, many approaches propose to avoid encoding and labeling the same input several times (Kolluru et al., 2020a; Bowen et al., 2021; Vasilkovsky et al., 2022). SMiLe-OIE (Dong et al., 2022) improves the model in an information-source view, using GCNs and multi-view learning to incorporate constituency and dependency information and aggregating semantic features and syntactic features by concatenating BERT and graph embeddings.

The sequence labeling paradigm is characterized by its computational efficiency, especially for large-scale text processing. It yields readily interpretable output, as each token associates itself with a specific role, such as subject, relation, object, spatial information, etc. It is limited by treating tokens in isolation, potentially failing to capture global context and complex relationships that extend beyond single tokens or cross sentences. Additionally, its output format may not adequately represent the nuanced variability of natural language.

Sequence to Sequence Generation. Cui et al. (2018) casts OpenIE as a sequence-to-sequence (S2S) generation problem and proposes NeuralOIE, an encoder-decoder model generating a sequence of relation triplets conditioned by the input sentence. Facing unknown token openness problem, NeuralOIE uses the attention-based coping mechanism to enlarge the vocabulary. IMoJIE(Kolluru et al., 2020b) is an iterative generative OpenIE model that uses a BERT encoder to keep encoding previous generated relation triplets and generates the next triplet with an LSTM decoder until an "EndOfExtractions" token is reached.

The S2S paradigm excels in capturing complex relationships, as it considers the broader contextual information. It is adaptable to various languages

	Representative Approach	OIE16 F1	OIE16 AUC	Re-OIE16 F1	Re-OIE16 AUC	CaRB F1	CaRB AUC	FewRel ARI	FewRel B^3	FewRel V	TACRED ARI	TACRED B^3	TACRED V
Pre-Neural (*ORTE*) 2007 - 2018	OLLIE (Schmitz et al., 2012)	38.6	20.2	49.5	31.3	41.1	22.4	-	-	-	-	-	-
	ClausIE (Del Corro and Gemulla, 2013)	58.0	36.4	64.2	46.4	44.9	22.4	-	-	-	-	-	-
	OPENIE4 (Mausam, 2016)	58.8	40.8	68.3	50.9	51.6	29.5	-	-	-	-	-	-
	PropS (Stanovsky and Dagan, 2016)	54.4	32.0	64.2	43.3	31.9	12.6	-	-	-	-	-	-
Neural Era (*ORTE*) 2018 - 2022	RnnOIE (Stanovsky et al., 2018)	62.0	48.0	-	-	49.0	26.1	-	-	-	-	-	-
	OpenIE6 (Kolluru et al., 2020a)	-	-	-	-	52.7	33.7	-	-	-	-	-	-
	SpanOIE (Zhan and Zhao, 2020)	69.4	49.1	77.0	65.8	48.5	-	-	-	-	-	-	-
	IMoJIE (Kolluru et al., 2020b)	-	-	-	-	53.5	33.3	-	-	-	-	-	-
	MacroIE (Bowen et al., 2021)	-	-	-	-	54.8	36.3	-	-	-	-	-	-
	DetIE$_{LSOIE}$ (Vasilkovsky et al., 2022)	-	-	-	-	43.0	27.2	-	-	-	-	-	-
	DetIE$_{IMoJIE}$ (Vasilkovsky et al., 2022)	-	-	-	-	52.1	36.7	-	-	-	-	-	-
	SMiLe-OIE (Dong et al., 2022)	-	-	-	-	53.8	34.9	-	-	-	-	-	-
Neural Era (*ORSE*) 2018 - 2022	Multi^{2}OIE (Ro et al., 2020)	-	-	83.9	74.6	52.3	32.6	-	-	-	-	-	-
	GEN2OIE (Kolluru et al., 2022)	-	-	-	-	54.4	32.3	-	-	-	-	-	-
	GEN2OIE (label-rescore)	-	-	-	-	54.5	38.9	-	-	-	-	-	-
	OIE@OIA (Wang et al., 2022d)	71.6	54.3	85.3	76.9	51.1	33.9	-	-	-	-	-	-
	DragonIE (Yu et al., 2022)	-	-	-	-	55.1	36.4	-	-	-	-	-	-
	ChunkOIE(SaC-OIA-SP) (Dong et al., 2023)	-	-	-	-	53.6	35.5	-	-	-	-	-	-
	ChunkOIE(SaC-CoNLL)	-	-	-	-	53.2	34.7	-	-	-	-	-	-
Neural Era (*ORC*) 2018 - 2022	RSN (Wu et al., 2019)	-	-	-	-	-	-	45.3	58.9	70.8	45.9	63.1	64.3
	RSN-CV (Wu et al., 2019)	-	-	-	-	-	-	54.2	63.8	72.4	-	-	-
	SelfORE (Hu et al., 2020)	-	-	-	-	-	-	64.7	67.8	78.3	44.7	54.1	61.9
	RSN-BERT (Zhao et al., 2021)	-	-	-	-	-	-	53.2	70.9	78.1	75.6	83.4	85.9
	RoCORE (Zhao et al., 2021)	-	-	-	-	-	-	70.9	79.6	86	81.2	86	88.8
	OHRE (Zhang et al., 2021a)	-	-	-	-	-	-	64.2	70.5	76.7	-	-	-
	MatchPrompt (Wang et al., 2022c)	-	-	-	-	-	-	66.5	72.3	82.2	75.3	83.0	84.5
	PromptORE (Genest et al., 2022)	-	-	-	-	-	-	43.4	48.8	71.8	-	-	-
	CaPL (Duan et al., 2022)	-	-	-	-	-	-	79.4	81.9	88.9	82.9	87.3	89.8
	ASCORE (Zhao et al., 2023)	-	-	-	-	-	-	67.6	73.5	83.5	78.1	78	83.1
LLM Era (*ORTE*) 2022 -	IELM GPT-2$_{XL}$ (Wang et al., 2022b)	-	-	35.0	-	22.7	-	-	-	-	-	-	-
	GPT-3.5-TURBO ICL (Ling et al., 2023)	65.1	-	67.9	-	52.1	-	-	-	-	-	-	-
	ChatGPT n-shot (Qi et al., 2023a)	-	-	-	-	55.3	-	-	-	-	-	-	-

Table 3: Performance of OpenIE models. For B^3 and V measures, F1 scores are reported. Rows filled with colors represent models of different task settings: = ***ORSE***, = ***ORSE***, = ***ORC***.

and domains but more demand in datasets and training times. The flexible output, while better-addressing openness challenges, also poses challenges for downstream applications requiring standardized output structures.

Two-Stage Open Relation Extraction. Taking advantage of the remarkable representation capability of PLMs such as BERT, many researchers refine the model architecture into two stages to achieve more effective extractions. Multi^{2}OIE (Ro et al., 2020) is a two-stage labeling method. Its first stage is to label all predicates upon BERT-embedded hidden states instead of locating predicates with syntactic features. The second stage is to extract the arguments associated with each identified predicate by using a multi-head attention mechanism. The intermediate representation can be other formats such as chunk sequence (Kolluru et al., 2022) and directed acyclic (Yang et al., 2022).

Various intermediate representations are used to enhance the pipeline's performance. OIE@OIA (Wang et al., 2022d) is an adaptable OpenIE system that employs the method of Open Information expression (OIX) by parsing sentences into Open Information Annotation (OIA) Graphs. It consists of two components: an OIA generator that converts sentences into OIA graphs and a set of adaptors that trained to for versatile extraction formats. By using different intermediate representations, Chunk-OIE (Dong et al., 2023) introduces the Chunk sequence (SaC) as an intermediate representation layer while Yu et al. (2022) introduces directed acyclic graph (DAG) as a minimalist intermediate expression.

Open Relation Clustering. The clustering-based approaches are divided into relation representation and clustering. Some studies label clusters: Wang et al. (2022c) and (Genest et al., 2022) introduce an unsupervised prompt-based algorithm, MatchPrompt, which clusters sentences by leveraging representations from masked relation tokens within a prompt template. Its superb performance against traditional unsupervised methods indicates that leveraging the semantic expressive power of pre-trained models is very important.

SelfORE (Hu et al., 2020) propose a self-supervised learning method for learning better feature representations for clustering. SelfORE is composed of three sections: (1) encode relation instances by leveraging BERT (Devlin et al., 2019) to obtain relation representations; (2) apply adaptive clustering based on updated relation representations from (1) to assign each instance to a cluster with high confidence. In this way, pseudo labels are generated. (3) pseudo labels from (2) are used as supervision signals to train the relation classifier and update the encoder in (1). Repeat (2) until converge. Based on similar self-supervised approaches, many works propose to reduce irrelevant information in relation representation (Zhao et al., 2021), create pseudo labels (Duan et al., 2022), and introduce human intervention during training to address the challenge of poorly clustered samples (Zhao et al., 2023). During relation clustering, using complete input (sentence) representations as relation representations often leads to a significant decline in clustering performance when multiple relations exist within a single input. Semi-supervised learning has shown the best results. However, the effectiveness largely depends on the quality of the annotated data.

Apart from labeled data, **knowledge bases** also benefit OpenIE by generating positive and negative instances. OHRE (Zhang et al., 2021b) proposes a top-down hierarchy expansion algorithm to cluster and label relation instances based on the distance between the KB hierarchical structure. Existing relations are labeled with KB elements, and novel relations are labeled as children relations of existing ones. Using a structured KB can determine the broad category of a cluster's relations, partially addressing cluster labeling issues. The KB structure can also define relation boundaries during clustering. However, errors in the KB can affect clustering accuracy, and building a high-quality KB still requires significant human effort.

5.3 Large Language Models Era

The recent evolution and emergence of Large Language Models (LLMs), such as GPT-4 (OpenAI, 2024), ChatGPT (OpenAI, 2023), and Llama 2 (Touvron et al., 2023), have significantly advanced the field of NLP. Their remarkable capabilities in text understanding, generation, and generalization have led to a surge of interest in generative IE methods (Qi et al., 2023b; Xu et al., 2023b). Recent studies have employed LLMs for OpenIE tasks by transforming input text through specific instructions or schemas. This approach facilitates tasks such as triplet extraction and relation classification under the structured language generation framework. It allows for a versatile task configuration where diverse forms of input text can be processed to generate structured relational triplets uniformly.

Zero-Shot. Wang et al. (2022b) propose IELM, a benchmark for assessing the zero-shot performance of GPT-2 (Radford et al., 2019) by encoding entity pairs in the input and extracting relations associated with each entity pair. On large-scale evaluation on various OpenIE benchmark tasks, research has shown that the zero-shot performance of leading LLMs, such as ChatGPT, still falls short of the state-of-the-art supervised methods (Han et al., 2023; Qi et al., 2023b), specifically on more challenging tasks (Li et al., 2023a). This shortfall is partly because LLMs struggle to distinguish irrelevant context from long-tail target types and relevant relations (Ling et al., 2023; Han et al., 2023).

Fine-Tuning and Few-Shot. Consequently, efforts have been made to fine-tune pre-trained LLMs or employ in-context learning prompting strategies to utilize and enhance the instruction-following ability of LLMs. For example, Lu et al. (2023) addresses open-world information extraction, including unrestricted entity and relation detection, as an instruction-following generative task, and develops PIVOINE, a fine-tuned information extraction LLM that generates comprehensive entity profiles in JSON format. To minimize the need for extensive fine-tuning of LLMs, Ling et al. (2023) proposes various in-context learning strategies for performing relation triplet generation to improve the instruction-following ability of LLMs, and introduces an uncertainty quantification module to increase the confidence in the generated answers. Qi et al. (2023a) proposes to construct a consistent reasoning environment by mitigating the distributional discrepancy between test samples and LLMs. This strategy aims to improve the few-shot reasoning capability of LLMs on specific OpenIE tasks.

6 Discussion

This section reviews the diverse sources of information used by OpenIE models and discusses current limitations and future prospects, offering a comprehensive overview of the field's evolving trajectory.

6.1 Co-Evolution of OpenIE: Task Settings and Model Capabilities

In this section, we unveil the connection between task settings and model capabilities in handling various features and information, demonstrating the intertwined development of both aspects.

Input-based information refers to features explicitly or implicitly present in the input unstructured text. Early OpenIE models extensively utilized explicit information such as ***shallow syntactic information***, including part of speech (POS) tags and noun-phrase (NP) chunks (Banko et al., 2007; Wu and Weld, 2010; Fader et al., 2011). This approach is reliable, yet it does not capture all relation types (Stanovsky et al., 2018), leading to the increasing use of ***deep dependency information***, which reveals word dependencies within sentences (Vo and Bagheri, 2018; Elsahar et al., 2017). Subsequent OpenIE models have emphasized the use of ***semantic information*** to grasp literal meanings and linguistic structures, thereby enhancing the expression of relations despite the risk of over-specificity (Vashishth et al., 2018; Wu et al., 2018). Recent models, including pre-trained language models, combine syntactic and semantic information to improve accuracy (Hwang and Lee, 2020; Ni et al., 2021). Further details in Appendix E.1.

External information supplements OpenIE systems to enhance model performance. Early systems employ ***expert rules***, including heuristic rules that integrate domain knowledge and assist in error tracing and resolution, based on syntactic analyses like POS-tagging (Chiticariu et al., 2013; Fader et al., 2011). Following this, the integration of ***hierarchical information*** from knowledge bases (KBs) advances knowledge representation learning. This integration provides structured hierarchies and detailed factual knowledge, supporting more organized relation extraction and data augmentation (Xie et al., 2016; Zhang et al., 2021b; Fangchao et al., 2021). With the developments of LLMs recently, the ***pre-trained knowledge*** within these models is utilized, encapsulating extensive relational data (Jiang et al., 2020; Petroni et al., 2020) and enabling efficient retrieval with well-designed instructions. The strong generalization capabilities of LLM-based approaches allow them to embrace ***open-world knowledge***, making them more robust and adaptable to various tasks and real-world applications. Further details in Appendix E.2.

6.2 Transforming OpenIE: The Impact of LLMs

When comparing the **performance of LLMs with pre-LLM approaches**, we see that LLMs have significantly advanced the task of OpenIE, often outperforming traditional methods. Zero-shot LLMs have achieved impressive and state-of-the-art (SOTA) results in various scenarios when evaluated on classic metrics such as token-level scorers (Li et al., 2023a; Wang et al., 2022b). However, these models struggle with long-tail and more challenging tasks (Gao et al., 2023). A major challenge for LLMs, compared to pre-LLM approaches like sequence tagging, is the issue of hallucination, which frequently occurs in various natural language generation tasks (Ji et al., 2023), making faithfulness and reliability significant concerns. Traditional generative-based approaches from the pre-LLM era also suffer from errors such as redundant and incorrect extractions (Schneider et al., 2017; Zhou et al., 2022), known as intrinsic hallucination. In contrast, LLM-based methods face the risk of both intrinsic hallucination and generating information unsupported by the original context or additional references, known as extrinsic hallucination (Zhu et al., 2023; Ren et al., 2023; Li et al., 2023a). Despite these challenges, few-shot learning and fine-tuning can help mitigate issues related to long-tail challenges and hallucination to some extent. Additionally, until fundamental improvements in LLMs fully address these shortcomings, incorporating traditional approaches as supplementary supervisors when using LLM-based methods could potentially enhance reliability.

We also observe trends in developing **universal paradigms for tackling various IE tasks.** Recent advancements and the robust generalization capabilities of LLMs have led to the exploration of universal frameworks designed to address all IE tasks (UIE). These frameworks aim to leverage the shared capabilities inherent in IE, while also uncovering and learning from the dependencies between various tasks (Xu et al., 2023b). This approach marks a significant shift from focusing on isolated subtasks, such as OpenIE, to a more integrated methodology that seeks a comprehensive understanding of the domain. The prevailing trajectory in developing universal IE frameworks is to establish unified, structured schemas, either natural language-based (Wang et al., 2022a; Lu et al., 2022; Lou et al., 2023) or code-based (Li et al., 2023d; Guo et al., 2023b; Sainz et al., 2023), to transform various IE tasks into a uniform task of structural information extraction while preserving the flexibility to adapt to the unique aspects of different tasks. More details on these approaches are provided in Appendix B.

Is OpenIE research going to its ends? **LLMs bridge the gap between standard IE and OpenIE.** LLMs are naturally suited for OpenIE, even under zero-shot scenarios, as they address both standard IE and OpenIE within the same task setting. In this setting, schemas and templates are designed to extract desired structural information.

The primary difference is that standard IE schemas include more restrictions to limit the set of relations and entities. The flexibility and strong performance of LLMs in tackling various IE tasks through zero-shot and few-shot prompting, without requiring model updates, is attributed to their robust generalization ability acquired through pre-training. With this generalization capability, addressing both standard IE and OpenIE may not require fundamentally different methods; the main distinction lies in schema design. This significantly blurs the boundaries between standard IE and OpenIE. In the future, OpenIE might be viewed as a more complex and challenging scenario within IE tasks, rather than being distinctly separate from standard IE. Though we refrain from making a definitive conclusion, we can foresee OpenIE potentially merging into the broader scope of standard IE.

6.3 Future Directions

Although we see the momentum of blurred gaps between OpenIE and standard IE with the impact of LLM, the fundamental task itself remains. Then *how can traditional OpenIE research inspire IE research in the LLM era?* Following we discuss future directions draw from reflections on a chronological perspective.

OpenIE datasets are growing but remain small and narrow in scope. Insights from traditional OpenIE research suggest that future expansions are needed to include more languages, domains, and broader sources. LLMs offer the opportunity to improve this through their capabilities in synthesizing and augmenting data. While synthesized datasets have been extensively explored within the domain of standard IE (Zhang et al., 2023a; Xu et al., 2023a), with researchers claiming that the proposed methods can be adapted for OpenIE (Josifoski et al., 2023), there is a notable gap regarding comprehensive studies on synthesized datasets for OpenIE. Addressing this gap could facilitate the creation of cross-domain datasets and the integration of existing datasets and tasks.

As discussed in Section 6.2, LLMs enable the exploration of various IE tasks with universal frameworks (UIE). Despite advances, most LLM-based UIE systems focus on standard IE tasks and often overlook OpenIE, a complex challenge within the IE spectrum. LLMs are inherently suited for OpenIE due to their extensive pre-trained knowledge. Therefore, the **primary challenge of LLMs** lies not in extracting relational information but in accurately interpreting and following task-specific instructions, as well as mitigating hallucination. Integrating traditional approaches into LLM-based frameworks might address these current shortcomings of LLMs. Additionally, these approaches can provide insights for developing more robust, faithful, and reliable fundamental LLMs.

More comprehensive automatic metrics are needed to evaluate LLM-based approaches. As discussed in Sections 2 and 4, task settings and corresponding evaluation metrics develop hand-in-hand. Now the changes brought by LLM calls for a more holistic and update-to-date evaluation metrics. The changes brought by LLMs call for more holistic and up-to-date evaluation metrics. Current efforts explore different options, as noted in Section 4 and Appendix D, but aspects such as faithfulness still rely heavily on human evaluation and lack a commonly accepted metric. Developing new, comprehensive automatic evaluation methods that capture nuanced aspects of OpenIE output, such as semantic coherence, factual accuracy, and information completeness, will lead to more robust and reliable LLM-based OpenIE systems. These metrics can address the unique challenges posed by LLMs, including their propensity for generating diverse and open-ended outputs.

Latency, cost, and distillation. Reviewing the development of models for OpenIE, we see the trend that recent development in LLMs introduce a more expensive system with higher latency, especially using close sourced LLMs such as GPT. Although the rapid iteration of models shows cuts on cost and latency, more effective solutions could be possible with knowledge distillation from LLMs onto specialized SLMs, revisiting the prior neural model era we discussed.

7 Conclusion

We systematically survey the development of OpenIE from a chronological perspective, highlighting historical trends in task settings and model development. We draw important connections and derive lessons from the influence of technology on task settings, examining the advantages and disadvantages of both past and present methods. Furthermore, we explore the increasingly blurred distinctions between OpenIE and standard IE. For researchers in LLMs, past work should not be overlooked; instead, it should serve as a valuable resource for future inquiries.

Limitations

Our survey primarily concentrates on the chronological evolution of OpenIE technologies and their alignment with significant milestones in NLP development. Consequently, we have not covered multi-domain and multi-lingual datasets or methodologies extensively. While we do address some non-English datasets, specifically Mandarin, and briefly mention multilingual models in Appendix A and model applications across various domains in Appendix B.3, these discussions are not the focal point of our analysis. This limitation is intentional in order to maintain a clear focus on the historical progression of the field rather than the breadth of dataset diversity or the adaptability of methodologies across languages and domains.

Another potential limitation is our survey's emphasis on the macro aspects of the OpenIE field rather than detailed, micro-level analysis of specific methodologies. As outlined in Section 1, many existing surveys already cover methodologies and models from the pre-LLM era, and we felt that redundant elaboration on these would not add significant value. Post-LLM, despite substantial research leveraging LLMs for standard IE tasks, there is still a scarcity of studies specifically applying LLMs to OpenIE tasks. This scarcity has constrained our ability to conduct an in-depth survey focused exclusively on LLM methodologies within OpenIE. Nonetheless, from the existing work on LLMs in standard IE and UIE, detailed in Appendix B, we observe emerging trends that warrant a macro-level analysis. Our approach of integrating and reviewing the field through a historical lens is essential to provide a comprehensive view, enabling a clearer understanding of the task and aiding in the development of a more defined future roadmap.

References

Zhila A and Gelbukh A. 2013. Comparison of open information extraction for english and spanish.

Alan Akbik and Alexander Löser. 2012. Kraken: N-ary facts in open information extraction. In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction (AKBC-WEKEX)*, pages 52–56.

Arthur Amalvy, Vincent Labatut, and Richard Dufour. 2023. Learning to rank context for named entity recognition using a synthetic dataset. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 10372–10382.

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D Manning. 2015. Leveraging linguistic structure for open domain information extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 344–354.

Amit Bagga and Breck Baldwin. 1998. Entity-based cross-document coreferencing using the vector space model. In *COLING-ACL*.

Michele Banko, Michael J Cafarella, Stephen Soderland, Matthew Broadhead, and Oren Etzioni. 2007. Open information extraction from the web. In *IJCAI*.

Sangnie Bhardwaj, Samarth Aggarwal, and Mausam. 2019. Carb: A crowdsourced benchmark for open ie. In *EMNLP*.

Zhen Bi, Jing Chen, Yinuo Jiang, Feiyu Xiong, Wei Guo, Huajun Chen, and Ningyu Zhang. 2024. Codekgc: Code language model for generative knowledge graph construction. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 23(3).

Kurt D. Bollacker, Colin Evans, Praveen K. Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *SIGMOD Conference*.

Yu Bowen, Yucheng Wang, Tingwen Liu, Hongsong Zhu, Limin Sun, and Bin Wang. 2021. Maximal clique based non-autoregressive open information extraction. In *Conference on Empirical Methods in Natural Language Processing*.

Chenran Cai, Qianlong Wang, Bin Liang, Bing Qin, Min Yang, Kam-Fai Wong, and Ruifeng Xu. 2023. In-context learning for few-shot multimodal named entity recognition. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2969–2979, Singapore. Association for Computational Linguistics.

Feng Chen and Yujian Feng. 2023. Chain-of-thought prompt distillation for multimodal named entity recognition and multimodal relation extraction. *Preprint*, arXiv:2306.14122.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. 2021. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*.

Laura Chiticariu, Yunyao Li, and Frederick R. Reiss. 2013. Rule-based information extraction is dead! long live rule-based information extraction systems! In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 827–832, Seattle, Washington, USA. Association for Computational Linguistics.

Janara Christensen, Stephen Soderland, and Oren Etzioni. 2011. An analysis of open information extraction based on semantic role labeling. In *Proceedings of the sixth international conference on Knowledge capture*, pages 113–120.

Lei Cui, Furu Wei, and M. Zhou. 2018. Neural open information extraction. In *ACL*.

Luciano Del Corro and Rainer Gemulla. 2013. Clausie: clause-based open information extraction. In *Proceedings of the 22nd international conference on World Wide Web*, pages 355–366.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv*, abs/1810.04805.

Kuicai Dong, Aixin Sun, Jung jae Kim, and Xiaoli Li. 2023. Open information extraction via chunks. *ArXiv*, abs/2305.03299.

Kuicai Dong, Aixin Sun, Jung-Jae Kim, and Xiaoli Li. 2022. Syntactic multi-view learning for open information extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 4072–4083. Association for Computational Linguistics.

Bin Duan, Shusen Wang, Xingxian Liu, and Yajing Xu. 2022. Cluster-aware pseudo-labeling for supervised open relation extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1834–1841, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Hady Elsahar, Elena Demidova, Simon Gottschalk, Christophe Gravier, and Frederique Laforest. 2017. Unsupervised open relation extraction. In *European Semantic Web Conference*, pages 12–16. Springer.

Hady ElSahar, Pavlos Vougiouklis, Arslen Remaci, Christophe Gravier, Jonathon S. Hare, Frédérique Laforest, and Elena Paslaru Bontas Simperl. 2018. T-rex: A large scale alignment of natural language with knowledge base triples. In *International Conference on Language Resources and Evaluation*.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the 2011 conference on empirical methods in natural language processing*, pages 1535–1545.

Liu Fangchao, Lingyong Yan, Hongyu Lin, Xianpei Han, and Le Sun. 2021. Element intervention for open relation extraction. *ArXiv*, abs/2106.09558.

Nicholas FitzGerald, Julian Michael, Luheng He, and Luke Zettlemoyer. 2018. Large-scale qa-srl parsing. In *ACL*.

Luis Galárraga, Geremy Heitz, Kevin P. Murphy, and Fabian M. Suchanek. 2014. Canonicalizing open knowledge bases. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*.

Pablo Gamallo. 2014. An Overview of Open Information Extraction (Invited talk). In *3rd Symposium on Languages, Applications and Technologies*, volume 38 of *OpenAccess Series in Informatics (OASIcs)*, pages 13–16, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Jun Gao, Huan Zhao, Changlong Yu, and Ruifeng Xu. 2023. Exploring the feasibility of chatgpt for event extraction. *arXiv preprint arXiv:2303.03836*.

Kiril Gashteovski, Mingying Yu, Bhushan Kotnis, Caroline Lawrence, Goran Glavas, and Mathias Niepert. 2021. Benchie: Open information extraction evaluation based on facts, not tokens. *ArXiv*, abs/2109.06850.

Pierre-Yves Genest, Pierre-Edouard Portier, Elöd Egyed-Zsigmond, and Laurent-Walter Goix. 2022. Promptore-a novel approach towards fully unsupervised relation extraction. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 561–571.

Rafael Glauber and Daniela Barreiro Claro. 2018. A systematic mapping study on open information extraction. *Expert Syst. Appl.*, 112:372–387.

Akshay Goel, Almog Gueta, Omry Gilon, Chang Liu, Sofia Erell, Lan Huong Nguyen, Xiaohong Hao, Bolous Jaber, Shashir Reddy, Rupesh Kartha, Jean Steiner, Itay Laish, and Amir Feder. 2023. Llms accelerate annotation for medical information extraction. In *Proceedings of the 3rd Machine Learning for Health Symposium*, volume 225 of *Proceedings of Machine Learning Research*, pages 82–100. PMLR.

Honghao Gui, Jintian Zhang, Hongbin Ye, and Ningyu Zhang. 2023. Instructie: A chinese instruction-based information extraction dataset. *arXiv preprint arXiv:2305.11527*.

Yucan Guo, Zixuan Li, Xiaolong Jin, Yantao Liu, Yutao Zeng, Wenxuan Liu, Xiang Li, Pan Yang, Long Bai, Jiafeng Guo, and Xueqi Cheng. 2023a. Retrieval-augmented code generation for universal information extraction. *Preprint*, arXiv:2311.02962.

Yucan Guo, Zixuan Li, Xiaolong Jin, Yantao Liu, Yutao Zeng, Wenxuan Liu, Xiang Li, Pan Yang, Long Bai, Jiafeng Guo, et al. 2023b. Retrieval-augmented code generation for universal information extraction. *arXiv preprint arXiv:2311.02962*.

Ridong Han, Tao Peng, Chaohao Yang, Benyou Wang, Lu Liu, and Xiang Wan. 2023. Is information extraction solved by chatgpt? an analysis of performance, evaluation criteria, robustness and errors. *arXiv preprint arXiv:2305.14450*.

Xu Han, Tianyu Gao, Yankai Lin, Hao Peng, Yaoliang Yang, Chaojun Xiao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2020. More data, more relations, more context and more openness: A review and outlook for relation extraction. *arXiv preprint arXiv:2004.03186*.

Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Y. Yao, Zhiyuan Liu, and Maosong Sun. 2018. Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *EMNLP*.

Tom Mesbah Harting, Sepideh Mesbah, and Christoph Lofi. 2020. Lorem: Language-consistent open relation extraction from unstructured text. *Proceedings of The Web Conference 2020*.

Luheng He, Mike Lewis, and Luke Zettlemoyer. 2015. Question-answer driven semantic role labeling: Using natural language to annotate natural language. In *EMNLP*.

Xuming Hu, Lijie Wen, Yusong Xu, Chenwei Zhang, and S Yu Philip. 2020. Selfore: Self-supervised relational feature learning for open relation extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3673–3682.

Zhiting Hu, Po-Yao Huang, Yuntian Deng, Yingkai Gao, and Eric P. Xing. 2015. Entity hierarchy embedding. In *ACL*.

Lawrence J. Hubert and Phipps Arabie. 1985. Comparing partitions. *Journal of Classification*, 2:193–218.

Hyunsun Hwang and Changki Lee. 2020. Bert-based korean open information extraction. *KIISE Transactions on Computing Practices*.

Ganesh Jawahar, Benoît Sagot, and Djamé Seddah. 2019. What does bert learn about the structure of language? In *ACL 2019-57th Annual Meeting of the Association for Computational Linguistics*.

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38.

Shengbin Jia, E Shijia, Ling Ding, Xiaojun Chen, and Yang Xiang. 2022. Hybrid neural tagging model for open relation extraction. *Expert Systems with Applications*, 200:116951.

Shengbin Jia, E. Shijia, Maozhen Li, and Yang Xiang. 2018. Chinese open relation extraction and knowledge base establishment. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 17:1 – 22.

Zhengbao Jiang, Frank F Xu, Jun Araki, and Graham Neubig. 2020. How can we know what language models know? *Transactions of the Association for Computational Linguistics*, 8:423–438.

Martin Josifoski, Marija Sakota, Maxime Peyrard, and Robert West. 2023. Exploiting asymmetry for synthetic training data generation: SynthIE and the case of information extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1555–1574, Singapore. Association for Computational Linguistics.

Keshav Kolluru, Vaibhav Adlakha, Samarth Aggarwal, Mausam, and Soumen Chakrabarti. 2020a. Openie6: Iterative grid labeling and coordination analysis for open information extraction. *ArXiv*, abs/2010.03147.

Keshav Kolluru, Samarth Aggarwal, Vipul Rathore, Mausam, and Soumen Chakrabarti. 2020b. Imojie: Iterative memory-based joint open information extraction. *ArXiv*, abs/2005.08178.

Keshav Kolluru, Mohammed Muqeeth, Shubham Mittal, Soumen Chakrabarti, and Mausam. 2022. Alignment-augmented consistent translation for multilingual open information extraction. In *Annual Meeting of the Association for Computational Linguistics*.

Bhushan Kotnis, Kiril Gashteovski, Daniel Onoro Rubio, Vanesa Rodríguez-Tembrás, Ammar Shaker, Makoto Takamoto, Mathias Niepert, and Carolin (Haas) Lawrence. 2022. Milie: Modular & iterative multilingual open information extraction. In *ACL*.

William Léchelle, Fabrizio Gotti, and Philippe Langlais. 2019. Wire57 : A fine-grained benchmark for open information extraction. *ArXiv*, abs/1809.08962.

Nadège Lechevrel, Kata Gábor, Isabelle Tellier, Thierry Charnois, Haïfa Zargayouna, and Davide Buscaldi. 2017. Combining syntactic and sequential patterns for unsupervised semantic relation extraction. In *DMNLP Workshop@ ECML-PKDD*, pages 81–84.

Bo Li, Gexiang Fang, Yang Yang, Quansen Wang, Wei Ye, Wen Zhao, and Shikun Zhang. 2023a. Evaluating chatgpt's information extraction capabilities: An assessment of performance, explainability, calibration, and faithfulness. *arXiv preprint arXiv:2304.11633*.

Jinyuan Li, Han Li, Zhuo Pan, Di Sun, Jiahao Wang, Wenkun Zhang, and Gang Pan. 2023b. Prompting chatgpt in mner: enhanced multimodal named entity recognition with auxiliary refined knowledge. In *The 2023 Conference on Empirical Methods in Natural Language Processing*.

Jinyuan Li, Han Li, Zhuo Pan, Di Sun, Jiahao Wang, Wenkun Zhang, and Gang Pan. 2023c. Prompting ChatGPT in MNER: Enhanced multimodal named entity recognition with auxiliary refined knowledge. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2787–2802, Singapore. Association for Computational Linguistics.

Jinyuan Li, Han Li, Di Sun, Jiahao Wang, Wenkun Zhang, Zan Wang, and Gang Pan. 2024. Llms as bridges: Reformulating grounded multimodal named entity recognition. *arXiv preprint arXiv:2402.09989*.

Peng Li, Tianxiang Sun, Qiong Tang, Hang Yan, Yuanbin Wu, Xuanjing Huang, and Xipeng Qiu. 2023d. CodeIE: Large code generation models are better few-shot information extractors. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15339–15353, Toronto, Canada. Association for Computational Linguistics.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*.

Chen Ling, Xujiang Zhao, Xuchao Zhang, Yanchi Liu, Wei Cheng, Haoyu Wang, Zhengzhang Chen, Takao Osaki, Katsushi Matsuda, Haifeng Chen, et al. 2023. Improving open information extraction with large language models: A study on demonstration uncertainty. *arXiv preprint arXiv:2309.03433*.

Yongbin Liu and Bingru Yang. 2012. Joint inference: A statistical approach for open information extraction. *Applied Mathematics & Information Sciences*, 6(5):627–633.

Jie Lou, Yaojie Lu, Dai Dai, Wei Jia, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. 2023. Universal information extraction as unified semantic matching. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press.

Keming Lu, Xiaoman Pan, Kaiqiang Song, Hongming Zhang, Dong Yu, and Jianshu Chen. 2023. Pivoine: Instruction tuning for open-world entity profiling. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 15108–15127.

Yaojie Lu, Qing Liu, Dai Dai, Xinyan Xiao, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. 2022. Unified structure generation for universal information extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5755–5772, Dublin, Ireland. Association for Computational Linguistics.

Mingyu Derek Ma, Xiaoxuan Wang, Po-Nien Kung, P Jeffrey Brantingham, Nanyun Peng, and Wei Wang. 2023. Star: Boosting low-resource event extraction by structure-to-text data generation with large language models. *arXiv preprint arXiv:2305.15090*.

Mausam Mausam. 2016. Open information extraction systems and downstream applications. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence*, pages 4074–4077.

Simon Meoni, Eric De la Clergerie, and Theo Ryffel. 2023. Large language models as instructors: A study on multilingual clinical entity extraction. In *The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, pages 178–190, Toronto, Canada. Association for Computational Linguistics.

Filipe Mesquita, Jordan Schmidek, and Denilson Barbosa. 2013. Effectiveness and efficiency of open relation extraction. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 447–457.

Julian Michael, Gabriel Stanovsky, Luheng He, Ido Dagan, and Luke Zettlemoyer. 2017. Crowdsourcing question-answer meaning representations. *ArXiv*, abs/1711.05885:560–568.

Mike D. Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *ACL*.

Tao Ni, Qing Wang, and Gabriela Ferraro. 2021. Explore bilstm-crf-based models for open relation extraction. *arXiv preprint arXiv:2104.12333*.

Christina Niklaus, Matthias Cetto, André Freitas, and Siegfried Handschuh. 2018. A survey on open information extraction. In *COLING*.

OpenAI. 2023. Introducing chatgpt. *OpenAI blog*.

OpenAI. 2024. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Harinder Pal et al. 2016. Demonyms and compound relational nouns in nominal open ie. In *Proceedings of the 5th Workshop on Automated Knowledge Base Construction*, pages 35–39.

Judea Pearl. 2000. Causality: Models, reasoning and inference. In *cambridge university press*.

Fabio Petroni, Patrick Lewis, Aleksandra Piktus, Tim Rocktäschel, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2020. How context affects language models' factual predictions. *arXiv preprint arXiv:2005.04611*.

Ji Qi, Kaixuan Ji, Xiaozhi Wang, Jifan Yu, Kaisheng Zeng, Lei Hou, Juanzi Li, and Bin Xu. 2023a. Mastering the task of open information extraction with large language models and consistent reasoning environment. *arXiv preprint arXiv:2310.10590*.

Ji Qi, Chuchun Zhang, Xiaozhi Wang, Kaisheng Zeng, Jifan Yu, Jinxin Liu, Lei Hou, Juanzi Li, and Xu Bin. 2023b. Preserving knowledge invariance: Rethinking robustness evaluation of open information extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5876–5890, Singapore. Association for Computational Linguistics.

Likun Qiu and Yue Zhang. 2014. Zore: A syntax-based system for chinese open relation extraction. In *EMNLP*.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*.

Colin Raffel, Noam M. Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *ArXiv*, abs/1910.10683.

Ruiyang Ren, Yuhao Wang, Yingqi Qu, Wayne Xin Zhao, Jing Liu, Hao Tian, Hua Wu, Ji-Rong Wen, and Haifeng Wang. 2023. Investigating the factual knowledge boundary of large language models with retrieval augmentation. *CoRR*, abs/2307.11019.

Irina Rish et al. 2001. An empirical study of the naive bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46. Citeseer.

Youngbin Ro, Yukyung Lee, and Pilsung Kang. 2020. Multi^2oie: Multilingual open information extraction based on multi-head attention with bert. *ArXiv*, abs/2009.08128.

Andrew Rosenberg and Julia Hirschberg. 2007. V-measure: A conditional entropy-based external cluster evaluation measure. In *EMNLP*.

Oscar Sainz, Iker García-Ferrero, Rodrigo Agerri, Oier Lopez de Lacalle, German Rigau, and Eneko Agirre. 2023. Gollie: Annotation guidelines improve zero-shot information-extraction. *arXiv preprint arXiv:2310.03668*.

Evan Sandhaus. 2008. The new york times annotated corpus. *Linguistic Data Consortium, Philadelphia*, 6(12):e26752.

Jordan Schmidek and Denilson Barbosa. 2014. Improving open relation extraction via sentence re-structuring. In *LREC*, pages 3720–3723.

Michael Schmitz, Stephen Soderland, Robert Bart, Oren Etzioni, et al. 2012. Open language learning for information extraction. In *EMNLP-CoNLL*.

Rudolf Schneider, Tom Oberhauser, Tobias Klatt, Felix A. Gers, and Alexander Löser. 2017. Analysing errors of open information extraction systems. In *WS:2017:54*, pages 11–18, Copenhagen, Denmark. Association for Computational Linguistics.

Heng Tao Shen. 2009. Principal component analysis. In *Encyclopedia of Database Systems*.

Yikang Shen, Shawn Tan, Alessandro Sordoni, and Aaron C. Courville. 2019. Ordered neurons: Integrating tree structures into recurrent neural networks. *ArXiv*, abs/1810.09536.

Jacob Solawetz and Stefan Larson. 2021. Lsoie: A large-scale dataset for supervised open information extraction. In *EACL*.

Gabriel Stanovsky and Ido Dagan. 2016. Creating a large benchmark for open information extraction. In *EMNLP*.

Gabriel Stanovsky, Julian Michael, Luke Zettlemoyer, and Ido Dagan. 2018. Supervised open information extraction. In *NAACL*.

Mingming Sun, Xu Li, Xin Wang, Miao Fan, Yue Feng, and Ping Li. 2018. Logician: A unified end-to-end neural approach for open-domain information extraction. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*.

Charles Sutton, Andrew McCallum, et al. 2012. An introduction to conditional random fields. *Foundations and Trends® in Machine Learning*, 4(4):267–373.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Shikhar Vashishth, Prince Jain, and Partha Pratim Talukdar. 2018. Cesi: Canonicalizing open knowledge bases using embeddings and side information. *Proceedings of the 2018 World Wide Web Conference*.

Michael Vasilkovsky, Anton Alekseev, Valentin Malykh, Ilya Shenbin, Elena Tutubalina, Dmitriy Salikhov, Mikhail Stepnov, Andrei Chertok, and Sergey Nikolenko. 2022. Detie: Multilingual open information extraction inspired by object detection. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*.

Duc-Thuan Vo and Ebrahim Bagheri. 2018. Open information extraction. In *Semantic Computing*, pages 3–8. World Scientific.

Denny Vrandečić. 2012. Wikidata: A new platform for collaborative data collection. In *Proceedings of the 21st international conference on world wide web*, pages 1063–1064.

Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, and Dawn Song. 2021. Zero-shot information extraction as a unified text-to-triple translation. *arXiv preprint arXiv:2109.11171*.

Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, and Dawn Song. 2022a. DeepStruct: Pretraining of language models for structure prediction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 803–823, Dublin, Ireland. Association for Computational Linguistics.

Chenguang Wang, Xiao Liu, and Dawn Song. 2022b. IELM: An open information extraction benchmark for pre-trained language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8417–8437, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Jiaxin Wang, Lingling Zhang, Jun Liu, Xi Liang, Yujie Zhong, and Yaqiang Wu. 2022c. MatchPrompt:

Prompt-based open relation extraction with semantic consistency guided clustering. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7875–7888, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Xiao Wang, Weikang Zhou, Can Zu, Han Xia, Tianze Chen, Yuansen Zhang, Rui Zheng, Junjie Ye, Qi Zhang, Tao Gui, et al. 2023a. Instructuie: Multi-task instruction tuning for unified information extraction. *arXiv preprint arXiv:2304.08085*.

Xin Wang, Minlong Peng, Mingming Sun, and Ping Li. 2022d. Oie@oia: an adaptable and efficient open information extraction framework. In *Annual Meeting of the Association for Computational Linguistics*.

Xingyao Wang, Sha Li, and Heng Ji. 2022e. Code4struct: Code generation for few-shot structured prediction from natural language. *arXiv preprint arXiv:2210.12810*, 3.

Xingyao Wang, Sha Li, and Heng Ji. 2023b. Code4Struct: Code generation for few-shot event structure prediction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3640–3663, Toronto, Canada. Association for Computational Linguistics.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph and text jointly embedding. In *EMNLP*.

Fei Wu and Daniel S Weld. 2010. Open information extraction using wikipedia. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 118–127.

Ruidong Wu, Yuan Yao, Xu Han, Ruobing Xie, Zhiyuan Liu, Fen Lin, Leyu Lin, and Maosong Sun. 2019. Open relation extraction: Relational knowledge transfer from supervised data to unsupervised data. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 219–228.

Tien-Hsuan Wu, Zhiyong Wu, Ben Kao, and Pengcheng Yin. 2018. Towards practical open knowledge base canonicalization. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*.

Clarissa Castellã Xavier, Vera Lúcia Strube de Lima, and Marlo Souza. 2013. Open information extraction based on lexical-syntactic patterns. *2013 Brazilian Conference on Intelligent Systems*, pages 189–194.

Ruobing Xie, Zhiyuan Liu, and Maosong Sun. 2016. Representation learning of knowledge graphs with hierarchical types. In *IJCAI*.

Wang Xinwei and Zhou Hui. 2020. Open information extraction for waste incineration nimby based on bert network in china. *Journal of Physics: Conference Series*.

Benfeng Xu, Quan Wang, Yajuan Lyu, Dai Dai, Yongdong Zhang, and Zhendong Mao. 2023a. S2ynRE: Two-stage self-training with synthetic data for low-resource relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8186–8207, Toronto, Canada. Association for Computational Linguistics.

Derong Xu, Wei Chen, Wenjun Peng, Chao Zhang, Tong Xu, Xiangyu Zhao, Xian Wu, Yefeng Zheng, and Enhong Chen. 2023b. Large language models for generative information extraction: A survey. *arXiv preprint arXiv:2312.17617*.

Huifan Yang, Da-Wei Li, Zekun Li, Donglin Yang, and Bin Wu. 2022. Open relation extraction with non-existent and multi-span relationships. Technical report, EasyChair.

Limin Yao, Aria Haghighi, Sebastian Riedel, and Andrew McCallum. 2011. Structured relation discovery using generative models. In *EMNLP*.

Alexander Yates, Michele Banko, Matthew Broadhead, Michael J Cafarella, Oren Etzioni, and Stephen Soderland. 2007. Textrunner: open information extraction on the web. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, pages 25–26.

Bowen Yu, Zhenyu Zhang, Jingyang Li, Haiyang Yu, Tingwen Liu, Jian Sun, Yongbin Li, and Bin Wang. 2022. Towards generalized open information extraction. *Preprint*, arXiv:2211.15987.

Junlang Zhan and Hai Zhao. 2020. Span model for open information extraction on accurate corpus. In *AAAI*.

Kai Zhang, Yuan Yao, Ruobing Xie, Xu Han, Zhiyuan Liu, Fen Lin, Leyu Lin, and Maosong Sun. 2021a. Open hierarchical relation extraction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5682–5693.

Kai Zhang, Yuan Yao, Ruobing Xie, Xu Han, Zhiyuan Liu, Fen Lin, Leyu Lin, and Maosong Sun. 2021b. Open hierarchical relation extraction. In *NAACL*.

Ruoyu Zhang, Yanzeng Li, Yongliang Ma, Ming Zhou, and Lei Zou. 2023a. LLMaAA: Making large language models as active annotators. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13088–13103, Singapore. Association for Computational Linguistics.

Ruoyu Zhang, Yanzeng Li, Yongliang Ma, Ming Zhou, and Lei Zou. 2023b. Llmaaa: Making large language models as active annotators. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13088–13103.

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. 2017. Position-aware attention and supervised data improve slot filling. In *Conference on Empirical Methods in Natural Language Processing*.

Jun Zhao, Tao Gui, Qi Zhang, and Yaqian Zhou. 2021. A relation-oriented clustering method for open relation extraction. In *EMNLP*.

Jun Zhao, Yongxin Zhang, Qi Zhang, Tao Gui, Zhongyu Wei, Minlong Peng, and Mingming Sun. 2023. Actively supervised clustering for open relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4985–4997, Toronto, Canada. Association for Computational Linguistics.

Shaowen Zhou, Bowen Yu, Aixin Sun, Cheng Long, Jingyang Li, Haiyang Yu, Jian Sun, and Yongbin Li. 2022. A survey on neural open information extraction: Current status and future directions. *arXiv preprint arXiv:2205.11725*.

Jun Zhu, Zaiqing Nie, Xiaojiang Liu, Bo Zhang, and Ji-Rong Wen. 2009. Statsnowball: a statistical approach to extracting entity relationships. In *Proceedings of the 18th international conference on World wide web*, pages 101–110.

Yutao Zhu, Huaying Yuan, Shuting Wang, Jiongnan Liu, Wenhan Liu, Chenlong Deng, Zhicheng Dou, and Ji-Rong Wen. 2023. Large language models for information retrieval: A survey. *arXiv preprint arXiv:2308.07107*.

Amal Zouaq, Michel Gagnon, and Ludovic Jean-Louis. 2017. An assessment of open relation extraction systems for the semantic web. *Information Systems*, 71:228–239.

A Open IE Methodologies in Details

A Chronological Overview of Open IE methods are summarized in Figure 3.

A.1 Open Relation Triplet Extraction

A.1.1 Labeling

OpenIE6 (Kolluru et al., 2020a) adopts a novel Iterative Grid Labeling (IGL) architecture, with which OpenIE is modeled as a 2-D grid labeling problem. Each extraction corresponds to one row in the grid. Iterative assignments of labels assist the model in capturing dependencies among extractions without re-encoding.

Owing to the outstanding performance of PLMs, many researchers extend the sequence labeling task to other problems. MacroIE(Bowen et al., 2021) reformulates the OpenIE as a non-parametric process of finding maximal cliques from the graph. It uses a non-autoregressive framework to mitigate the issue of enforced order and error accumulation during extraction. DetIE (Vasilkovsky et al., 2022) casts the task to a direct set prediction problem. This encoder-only model extracts a predefined number of possible triplets (proposals) by generating multiple labeled sequences in parallel, and its order-agnostic loss based on bipartite matching ensures the predictions are unique.

A.2 Open Relation Span Extraction

GEN2OIE (Kolluru et al., 2022) extends to a generative paradigm operating in two stages. It first generates all possible relations from input sentences. Then, it produces extractions for each generated relation. This generative approach allows for overlapping relations and multiple extractions with the same relation.

Jia et al. (2022) propose a hybrid neural network model (HNN4ORT) for open relation tagging. The model employs the Ordered Neurons LSTM (Shen et al., 2019) to encode potential syntactic information for capturing associations among arguments and relations. It also adopts a novel Dual Aware Mechanism, integrating Local-aware Attention and Global-aware Convolution. QuORE (Yang et al., 2022) is a framework to extract single/multi-span relations and detect non-existent relationships, given an argument tuple and its context. The model uses a manually defined template to map the argument tuple into a query. It concatenates and encodes the query together with the context to generate sequence embedding, with which this framework dynamically determines a sub-module (Single-span Extraction or Query-based Sequence Labeling) to label the potential relation(s) in the context.

Inspired by OIA, Chunk-OIE (Dong et al., 2023) introduces the concept of Sentence as Chunk sequence (SaC) as an intermediate representation layer, utilizing chunking to divide sentences into related non-overlapping phrases. Yu et al. (2022) introduce directed acyclic graph (DAG) as a minimalist expression of open fact in order to reduce the extraction complexity and improves the generalization behavior. They propose DragonIE which leverages the sequential priors to reduce the complexity of function space (edge number and type) in the previous graph-based model from quadratic to linear, while avoiding auto-regressive extraction in sequence-based models.

A.3 Open Relation Clustering

Lechevrel et al. (2017) select core dependency phrases to capture the semantics of the relations between entities. The design rules are based on the length of the dependency phrase in the dependency path, which sometimes contains more than one dependency phrase that uses all terms and brings in irrelevant information. Each relation instance is clustered on the basis of the semantics of core dependency phrases. Finally, clusters are named by the core dependency phrase most similar to the center vector of the cluster.

Instead of directly cutting less irrelevant information, Elsahar et al. (2017) propose a more resilient approach based on the shortest dependency path. The model generates representations of relation instances by assigning a higher weight to word embedding of terms in the dependency path and then reduces feature dimensions by PCA (Shen, 2009). Although the model ignores noisy terms in the dependency path, re-weighting is a forward-looking idea resembling the subsequent attention mechanism.

The key idea of Fangchao et al. (2021) is based on blocking backdoor paths from a causal view (Pearl, 2000). The intervened context is generated by a generative PLM, while entities are intervened by placing them with three-level hierarchical entities in KB. Model parameters are optimized by those intervened instances via contrastive learning. The learned model encodes each instance into its representations, before using clustering algorithms.

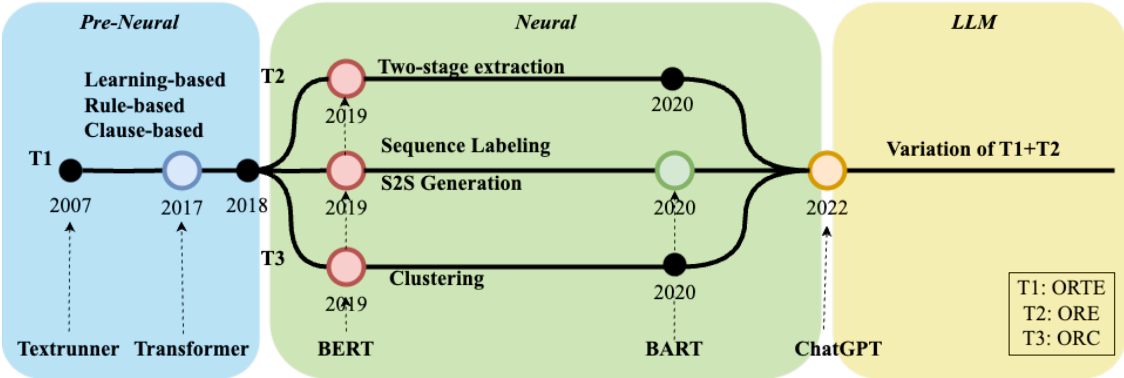

Figure 3: Chronological overview of Open IE methods.

A.4 Neural Model Era: Other Settings

Translation. Wang et al. (2021) cast information extraction tasks into a text-to-triplet translation problem. They introduce DEEPEX, a framework that translates NP-chunked sentences to relational triplets in a zero-shot setting. This translation process consists of two steps: generating a set of candidate triplets and ranking them.

Multilingual. MILIE (Kotnis et al., 2022) is an integrated model of a rule-based system and a neural system, which extracts triplet slots iteratively from simple to complex, conditioning on preceding extractions. The iterative nature guarantees the model to perform well in a multilingual setting. Multi^{2}OIE (Ro et al., 2020) also has a multilingual version based on multilingual-BERT, which makes it able to deal with various languages. Differently, LOREM (Harting et al., 2020) trains two types of models, language-individual models, and language-consistent models and incorporates multilingual, aligned word embeddings to enhance model performance.

B LLMs for IE in general

In Section 5.3, we begin by reviewing the work that utilizes LLMs to address OpenIE. Here, we 1). broaden our scope to introduce some emerging trends and paradigms in universal information extraction. For an in-depth exploration of how LLMs are applied to closed relation extraction and other IE tasks, we refer readers to the survey by Xu et al. (2023b) for comprehensive details. Moreover, we 2). further expand our discussion to explore research that integrates LLMs into IE system pipelines, beyond merely using them for direct IE task solution. We 3). also includes an discussion of current trends in IE dataset using LLMs that shed light on the future of datasets on openIE.

We believe this broader perspective provides readers with a comprehensive understanding of current trends and future directions in OpenIE and generic IE in the LLM era, enhancing their grasp of the field's evolving dynamics.

B.1 Universal Information Extraction

Recent advancements and the robust generalization capabilities of LLMs have led to the exploration of universal frameworks designed to tackle all IE tasks (UIE). These frameworks aim to harness the shared capabilities inherent in IE, while also uncovering and learning from the dependencies that exist between various tasks (Xu et al., 2023b). This approach marks a significant shift from focusing on isolated subtasks such as OpenIE to a more integrated methodology that seeks to understand a more integrated and comprehensive understanding of the domain.

Natural Language-Based Schema. A prevailing trend in developing universal IE frameworks is to establish a unified, structured natural language schema for diverse subtasks, designed for schema-prompting LLMs. For instance, Wang et al. (2022a) introduce DeepStruct, which reformulates various IE tasks as triplet generation tasks, using generalized task-specific prefixes in prompts and pretraining LLMs to comprehend text structures. Lu et al. (2022) propose UIE, encoding different extraction structures uniformly through a structured extraction language and adaptively generating specific extractions with a schema-based prompt strategy. Similarly, Lou et al. (2023) present USM, encoding different schemas and input texts together to enable structuring and conceptualizing, aiming for a single model that addresses all tasks. Building on UIE and USM, Wang et al. (2023a) introduce

InstructUIE, which models various IE tasks uniformly with descriptive natural language instructions for instruction tuning, exploiting inter-task dependencies.

Code-Based Schema. Despite their empirical success, natural language-based approaches face challenges in generating outputs for IE tasks due to the distinct syntax and structure that differ from the training data of LLMs (Bi et al., 2024). In response to these limitations and leveraging recent advancements in Code-LLMs (Chen et al., 2021), researchers have begun to utilize Code-LLMs for structure generation tasks (Wang et al., 2022e), as code, a formalized language, adeptly describes structural knowledge across various schemas universally (Guo et al., 2023b). For instance, Li et al. (2023d) present CodeIE, which translates structured prediction tasks such as NER and RE into code generation, employing Python functions to create task-specific schemas and using few-shot learning to instruct Code-LLMs. Guo et al. (2023b) introduce Code4UIE, utilizing Python classes to define task-specific schemas for diverse structural knowledge universally. Similarly, Sainz et al. (2023) propose GoLLIE, which employs Python classes to encode IE tasks and, in addition, integrates task-specific guidelines as docstrings, enhancing the robustness of fine-tuned Code-LLMs to schemas not encountered during training.

B.2 Role of LLMs in IE System

In addition to directly addressing IE tasks, LLMs have shown utility as specific components within IE system pipelines, including data synthesis for IE model training and knowledge retrieval for downstream IE tasks.

Data Synthesis. A prominent application of LLMs in IE systems is the synthesis of high-quality training data, as data curation through human annotation is time-consuming and labor-intensive. One approach employs LLMs as annotators within a learning loop (Zhang et al., 2023b), while another strategy involves using LLMs to inversely generate natural language text from structured data inputs (Josifoski et al., 2023; Ma et al., 2023), thereby producing large-scale, high-quality training data for IE tasks.

Knowledge Retrieval. Another research direction exploits the capability of LLMs, developed through pre-training, as implicit knowledge bases to generate or retrieve relevant context for downstream IE tasks. For instance, Li et al. (2023b, 2024) employ LLMs to generate auxiliary knowledge improving multimodal IE tasks. Amalvy et al. (2023) demonstrate that pre-trained LLMs possess inherent knowledge of the datasets they work on, and use these models to generate a context retrieval dataset, enhancing NER performance on long documents.

B.3 IE in Different Domains

The development of Information Extraction (IE) has seen significant advancements across various domains, including Multimodal IE, Medical Information Extraction, and the application of Code Models for IE tasks. These developments have been particularly enhanced by the integration of Large Language Models (LLMs), which have improved downstream task performance through their use in model architecture and as tools for annotation and training guidance.

Medical Information Extraction has greatly benefited from the use of LLMs as efficient tools for annotation, as highlighted in research by Goel et al. (2023); Meoni et al. (2023). These applications enhance data quality and contribute to the overall improvement of model performance.

Multimodal IE tasks, such as Multimodal Named Entity Recognition (MNER) and Multimodal Relation Extraction (MRE), have advanced through frameworks that capitalize on the capabilities of LLMs in IE. Cai et al. (2023) propsed to use in-context learning (ICL) ability in ChatGPT to help Few-Shot MNER by employing in-context learning to convert visual data into text and select relevant examples for effective entity recognition. Li et al. (2023c) tackles MNER on social media by efficient usage of generated knowledge and improved generalization, which utilizes ChatGPT as an implicit knowledge base for generating auxiliary knowledge to aid entity prediction. Chen and Feng (2023) distill the reasoning ability of LLMs by using "chain of thought" (CoT) to elicit reasoning capability from LLMs across multiple dimensions to improve MNER and MRE.

Code generative LLMs have found application in performing IE tasks such as Universal Information Extraction (UIE) (Li et al., 2023d; Guo et al., 2023a), Event Structure Prediction (Wang et al., 2023b), and Generative Knowledge Graph (Bi et al., 2024), where researchers convert the structured output in the form of code instead of natural language, and utilize generative LLMs of code (Code-LLMs) by designing code-style prompts and formulating these IE tasks as code generation tasks.

Leveraging LLMs across different domains has not only broadened the scope of IE applications but also significantly improved the effectiveness and efficiency of extraction tasks.

C Datasets

Question Answering (QA) derived datasets are converted from other crowdsourced QA datasets. OIE2016 (Stanovsky and Dagan, 2016) is one of the most popular OpenIE benchmarks, which leverages QA-SRL (He et al., 2015) annotations. AW-OIE (Stanovsky et al., 2018) extends the OIE2016 training set with extractions from QAMR dataset (Michael et al., 2017). The OIE2016 and AW-OIE datasets are the first datasets used for supervised OpenIE. However, because of its coarse-grained generation method, OIE2016 has some problematic annotations and extractions. On the basis of OIE2016, Re-OIE2016 (Zhan and Zhao, 2020) and CaRB (Bhardwaj et al., 2019) re-annotate part of the dataset. LSOIE (Solawetz and Larson, 2021) is created by converting QA-SRL 2.0 dataset (FitzGerald et al., 2018) to a large-scale OpenIE dataset, which claims 20 times larger than the next largest human-annotated OpenIE dataset.

Crowdsourced datasets are created from direct human annotation, including WiRe57 (Léchelle et al., 2019), SAOKE dataset (Sun et al., 2018), and BenchIE dataset (Gashteovski et al., 2021). WiRe57 is created based on a small corpus containing 57 sentences from 5 documents by two annotators following a pipeline. SAOKE dataset is generated from Baidu Baike, a free online Chinese encyclopedia, like Wikipedia, containing a single/multi-span relation and binary/polyadic arguments in a tuple. It is built in a predefined format, which assures its completeness, accurateness, atomicity, and compactness.

Knowledge Base (KB) derived datasets are established by aligning triplets in KBs with text in the corpus. Several works (Mintz et al., 2009; Yao et al., 2011) have aligned the New York Times corpus (Sandhaus, 2008) with Freebase (Bollacker et al., 2008) triplets, resulting in several variations of the same dataset, NYT-FB. FewRel (Han et al., 2018) is created by aligning relations of given entity pairs in Wikipedia sentences with distant supervision, and then filtered by human annotators. ElSahar et al. (2018) propose a pipeline to align Wikipedia corpus with Wikidata (Vrandečić, 2012) and generate T-REx. By filtering triplets and selecting sentences, Hu et al. (2020) create T-REx SPO and T-REx DS. In addition, COER (Jia et al., 2018), a large-scale Chinese knowledge base dataset, is automatically created by an unsupervised open extractor from diverse and heterogeneous web text, including encyclopedia and news. Overall, KB derived datasets are mostly used in open relation clustering task setting, illustrated in Section 5.2, whereas QA derived and crowdsourced datasets are usually used in open relational triplet extraction (Section 5.2) and open relation span extraction task settings (Section 5.2).

Instruction-based datasets transform IE tasks into tasks requiring instruction-following, thus harnessing the capabilities of LLMs. One strategy involves integrating various existing IE datasets into a unified-format benchmark dataset with specifically designed instructions (Wang et al., 2023a; Lu et al., 2022). Alternatively, instruction-based IE datasets such as INSTRUCTOPENWIKI (Lu et al., 2023) and INSTRUCTIE (Gui et al., 2023), or structured IE datasets like Wikidata-OIE (Wang et al., 2022b)—derived from Wikidata and Wikipedia—are created. The first method primarily focuses on ClosedIE tasks, while the second offers more flexibility in generating OpenIE datasets (Lu et al., 2023; Wang et al., 2022b).

Synthesized datasets using LLMs on IE expands significantly compared to previous ones in both the size of the datasets and data qualities. While the methodologies for synthesizing these datasets have been extensively explored within the domain of closed Information Extraction (ClosedIE) (Zhang et al., 2023a; Xu et al., 2023a), where researchers claims the proposed methods can be adapted for OpenIE setting (Josifoski et al., 2023), there remains a notable gap in the literature regarding comprehensive studies on synthesized datasets for OpenIE.

D Evaluation

Token-level Scorers. To allow some flexibility (e.g., omissions of prepositions or auxiliaries), if automated extraction of the model and the gold triplet agree on the grammatical head of all of their elements (predicate and arguments), OIE2016 (Stanovsky and Dagan, 2016) takes it as matched. Léchelle et al. (2019) penalize the verbosity of automated extractions as well as the omission of parts of a gold triplet by computing precision and recall at token-level in WiRe57. Their precision is the proportion of extracted words that are found in the gold triplet, while recall is the proportion of reference words found in extractions. To improve token-level scorers, CaRB (Bhardwaj et al., 2019) computes precision and recall pairwise by creating an all-pair matching table, with each column as extracted triplet and each row as gold triplet. When assessing LLM extracted spans, Han et al. (2023) report the ratio of invalid responses, which include incorrect formats and content not aligned with task-specific prompts. As generative models, LLMs aim to mimic human-like responses and often generate longer text than the gold standard annotations.

Fact-level Scorers. SAOKE (Sun et al., 2018) measures to what extent gold triplets and extracted triplets imply the same facts and then calculates precision and recall. BenchIE (Gashteovski et al., 2021) introduces *fact synset*: a set of all possible extractions (i.e., different surface forms) for a given fact type (e.g., VP-mediated facts) that are instances of the same fact. It takes the informational equivalence of extractions into account by exactly matching extracted triplets with the gold fact synsets. In assessing outputs from LLMs, Li et al. (2023a) have ChatGPT provide justifications for its predictions and use domain expert annotation to verify their faithfulness relative to the input.

E Source of Information

Section ?? provides a brief overview of the sources of information utilized in OpenIE models. This section offers a detailed discussion of each specific information source.

E.1 Input-based Information

Shallow syntactic information such as part of speech (POS) tags and noun-phrase (NP) chunks abstract input sentences into patterns. It is pervasively used in the early work of OpenIE as an essential model feature (Banko et al., 2007; Wu and Weld, 2010; Fader et al., 2011). In rule-based models, those patterns directly determine whether the input text contains certain relations or not (Xavier et al., 2013; A and A, 2013). Shallow syntactic information is reliable because there is a clear relationship between the relation type and the syntactic information in English (Banko et al., 2007). However, merely using shallow syntactic information can not discover all relation types. Subsequent work uses shallow syntactic information as part of the input and incorporates additional features to enhance the model performance (Stanovsky et al., 2018).

Deep dependency information shows the dependency between words in a sentence, which can be used directly to find relations (Vo and Bagheri, 2018). But because dependency analysis is more complex and time-consuming than shallow syntactic analysis, such information source was not popular in early OpenIE studies. It was the second generation of OpenIE models that brought dependency parsing to great attention. Right now, dependency information is still used as part of the model input, though with less popularity and sometimes not directly. Elsahar et al. (2017) make use of the dependency path to give higher weight to words between two named entities, in which way the model only uses dependency information as a supplement and relies more on the semantic meaning to extract information.

Semantic information captures not only linguistic structures of sentences but literal meanings of phrases, which can express more diverse and fitting relations compared to syntactic patterns. However, semantic information can also be too specific and hence lead to the canonicalizing problem (Galárraga et al., 2014; Vashishth et al., 2018; Wu et al., 2018). The second generation of OpenIE models has tried to use semantic information via semantic role labeling, for example EXAMPLAR (Mesquita et al., 2013), or via dependency parsing, for instance OLLIE (Schmitz et al., 2012). There were also attempts to use WordNet output to comprise semantic information (Liu and Yang, 2012). The third generation of OpenIE models typically use the word and sentence representations obtained from pre-trained language models (Kolluru et al., 2020b; Hwang and Lee, 2020; Xinwei and Hui, 2020). These representations contain both syntactic and semantic information (Jawahar et al., 2019). Meanwhile, some OpenIE models use word embeddings from word embedders such as GloVe, ELMo, and Word2Vec to capture semantic information (Ni et al., 2021).

E.2 External Knowledge

Expert rules are knowledge imported in the form of heuristic rules. It is easy for rule-based OpenIE systems to incorporate domain knowledge as well as to trace and fix errors (Chiticariu et al., 2013). Heuristic rules can be employed to avoid incoherent extractions (Fader et al., 2011). For example, verb words between two entities are likely to be the relation. Thus, to alleviate incoherence, a rule can be defined: *If there are multiple possible matches for a single verb, the shortest possible match is chosen.* Based on patterns generated from POS-tagging, dependency parse, and other syntactic analyses, different rules can be created.

Hierarchical information that implicitly exists in languages, which can be explicitly exhibited by knowledge bases, benefits knowledge representation learning (Wang et al., 2014; Lin et al., 2015; Hu et al., 2015; Xie et al., 2016). In addition, KBs contain fine-grained factual knowledge that provides background information and hierarchical structures needed for relation extraction. Compared to traditional clustering, KB can provide hierarchical information that helps represent and cluster relations in a more organized way (Zhang et al., 2021b) and hierarchical factual knowledge for data augmentation (Fangchao et al., 2021).

Pre-trained knowledge of language models, particularly LLMs, exhibit substantial potential to encapsulate relational knowledge (Jiang et al., 2020; Petroni et al., 2020). Unlike smaller models, which require learning from input and external knowledge in a bottom-up manner, LLMs hold extensive, ready-to-use knowledge from pre-training. Consequently, recent efforts aim to direct LLMs to concentrate solely on pertinent knowledge for specific IE tasks.

F Table of Traditional OpenIE Models

Model	Method	Source of Information	Task Setting	Dataset	Evaluation (Result)
TEXTRUNNER (Banko et al., 2007)	Dependency Parser, NP Chunker, CRF, Naive Bayes Classifier	syntactic, dependency	4.1	400 Web	Average Error Rate (12%)
WOE (Wu and Weld, 2010)	TEXTRUNNER, Self-supervised Learning	syntactic, dependency	4.1	300 news 300 Wikipedia 300 Web	Precision-Recall Curve
REVERB (Fader et al., 2011)	Syntactic Constraints, Lexical Contraints, CRF	syntactic, dependency	4.1	500 Web	Precision-Recall Curve, AUC (1.3*WOE^{parse}, 2*TEXTRUNNER)
OLLIE (Schmitz et al., 2012)	REVERB, Bootstrap, Open Pattern Learning	syntactic, dependency	4.1	300 news (from WOE) 300 Wikipedia (from WOE) 300 biology	Precision-Yield Curve, AUC (1.9*WOE^{parse}, 2.7*REVERB)
OPENIE4 (Mausam, 2016)	SRLIE (Christensen et al., 2011), RELNOUN (Pal et al., 2016)	syntactic, dependency	4.1	Not Reported	Precision-Yield Curve, AUC (1.32*OLLIE, 4*REVERB)
ClausIE (Del Corro and Gemulla, 2013)	Dependency Parser, Clause-based Model	syntactic, dependency	4.1	500 Web (from REVERB) 200 Wikipedia 200 news	Precision-Yield Curve, # of correct extractions / # of extractions
RnnOIE (Stanovsky et al., 2018)	Bi-LSTM, Softmax	word emb, POS emb	4.1	OIE2016 WEB NYT PENN	AUC (48), F1 (62) AUC (47), F1 (67) AUC (25), F1 (35) AUC (26), F1 (44)
NeuralOIE (Cui et al., 2018)	LSTM, Copy Attention	word emb	4.1	OIE2016	AUC (27)
IMoJIE (Kolluru et al., 2020b)	BERT, LSTM, CopyAttention	word emb	4.1	CaRB	AUC (33.3), F1 (53.5)
SpanOIE (Zhan and Zhao, 2020)	Bi-LSTM, Span-consistent Greedy Search	word emb, POS emb, dependency relation emb	4.1	OIE2016 Re-OIE2016	AUC (48.9), F1 (68.65) AUC (65.9), F1 (78.50)
Multi^{2}OIE (Ro et al., 2020)	BERT, Multihead Attention	word emb, position emb, avg vector of predicates	4.1	Re-OIE2016 CaRB	AUC (74.6), F1 (83.9) AUC (32.6), F1 (52.3)
OpenIE6 (Kolluru et al., 2020a)	Iterative Grid Labeling, BERT, Self-attention	word emb, dependency feature	4.1	CaRB	AUC (33.7), F1 (52.7)
HNN4ORT (Jia et al., 2022)	ON-LSTM, CNN, Attention	word emb, POS emb, argument emb, local/global features	4.2	Wikipedia NYT Reverb	F1 (79.8) F1 (74.5) F1 (81.7)
UORE (Elsahar et al., 2017)	Re-weight Word Emb, TF-IDF, PCA, HAC	word emb, dependency	4.3	NYT-FB	F1 (41.6)
RSN (Wu et al., 2019)	Relational Siamese Network, CNN, HAC, Louvain	word emb	4.3	FewRel	B^3: P (48.9) R (77.5) F1 (59.9)
SelfORE (Hu et al., 2020)	Bootstrapping Self-supervision, BERT, K-means, Adaptive Clustering	word emb	4.3	NYT+FB	ARI (40.3), B^3: P (49.1) R (47.3) F1 (51.1), V: F1 (46.6) Hom (45.7) Comp (47.6)
				T-REx SPO	ARI (33.7), B^3: P (41.0) R (39.4) F1 (42.8), V: F1 (41.4) Hom (40.3) Comp (42.5)
				T-REx DS	ARI (20.1), B^3: P (32.9) R (29.7) F1 (36.8), V: F1 (32.4) Hom (30.1) Comp (35.1)
OHRE (Zhang et al., 2021b)	CNN, Virtual Adversarial Training, Reconstruction Loss, Dynamic Hierarchical Triplet Loss, Louvain	word emb, hierarchical information	4.3	FewRel Hierarchy	ARI (64.2), B^3: P (64.5) R (77.7) F1 (70.5), V: F1 (76.7) Hom (73.8) Comp (79.9)
				NYT-FB Hierarchy	ARI (31.9), B^3: P (31.4) R (72.3) F1 (43.8), V: F1 (60.0) Hom (49.9) Comp (75.3)
ElementORE (Fangchao et al., 2021)	BERT, T5 (Raffel et al., 2020), Structure Causal Model, K-means	word emb, hierarchical information	4.3	T-REx SPO	ARI (36.6), B^3: P (46.7) R (43.4) F1 (45.0), V: F1 (45.3) Hom (45.4) Comp (45.2)
				T-REx DS	ARI (25.0), B^3: P (40.2) R (45.9) F1 (42.9), V: F1 (47.3) Hom (46.9) Comp (47.8)
RoCORE (Zhao et al., 2021)	Relation-oriented Representation, BERT, K-means	word emb	4.3	FewRel	ARI (70.9), B^3: P (75.2) R (84.6) F1 (79.6), V: F1 (86.0) Hom (83.8) Comp (88.3)
DEEPEX (Wang et al., 2021)	BERT, Attention, Beam Search, Contrastive Pre-training	NP chunks, word emb, triplet emb	4.4	OIE2016 WEB NYT PENN	AUC (58.6), F1 (72.6) AUC (82.4), F1 (91.2) AUC (72.5), F1 (85.5) AUC (81.5), F1 (88.5)

Table 4: Milestone and representative models of pre-LLM era. ("V" denotes "V-measure", and "emb" stands for "embedding".)