

Automated Construction of Theme-specific Knowledge Graphs

Linyi Ding

University of Illinois Urbana-Champaign
IL, USA
linyid2@illinois.edu

Jinfeng Xiao

University of Illinois Urbana-Champaign
IL, USA
jxiao13@illinois.edu

Sizhe Zhou

University of Illinois Urbana-Champaign
IL, USA
sizhez@illinois.edu

Jiawei Han

University of Illinois Urbana-Champaign
IL, USA
hanj@illinois.edu

ABSTRACT

Despite widespread applications of knowledge graphs (KGs) in various tasks such as question answering and intelligent conversational systems, existing KGs face two major challenges: **limited information granularity** and **deficiency in timeliness**. These hinder considerably the retrieval and analysis of in-context, fine-grained, and up-to-date knowledge from KGs, particularly in highly specialized themes (e.g., specialized scientific research) and rapidly evolving contexts (e.g., breaking news or disaster tracking).

To tackle such challenges, we propose **theme-specific knowledge graph** (i.e., **ThemeKG**), a KG constructed from a theme-specific corpus, and design an unsupervised framework for ThemeKG construction (named **TKGCon**). The framework takes raw theme-specific corpus and generates a high-quality KG that includes salient entities and relations under the theme. Specifically, we start with an **entity ontology** of the theme from Wikipedia, based on which we then generate candidate relations by Large Language Models (LLMs) to construct a **relation ontology**. To parse the documents from the theme corpus, we first map the extracted entity pairs to the ontology and retrieve the candidate relations. Finally, we incorporate the context and ontology to consolidate the relations for entity pairs.

We observe that directly prompting GPT-4 for theme-specific KG leads to inaccurate entities (such as "two main types" as one entity in the query result) and unclear (such as "is", "has") or wrong relations (such as "have due to", "to start"). In contrast, by constructing the theme-specific KG step by step, our model outperforms GPT-4 and could consistently identify accurate entities and relations. Experimental results also show that our framework excels in evaluations compared with various KG construction baselines.

CCS CONCEPTS

• Computing methodologies → Information extraction; • Information systems → Entity relationship models.

KEYWORDS

Knowledge graph, theme-specific, information extraction

1 INTRODUCTION

Knowledge graphs (KGs) [26] integrate collections of real-world entities connected by semantically meaningful relations. They usually store structured factual knowledge that allows easy access and information retrieval. KGs have been used in a variety of applications including recommendation systems [23], question answering

Theme : Electrical Vehicle Batteries

Lead-acid batteries suffer from relatively short cycle lifespan. There are two main types of lead-acid batteries: automobile engine starter batteries, and deep cycle batteries. Automobile engine starter batteries are designed to provide high charge rates. Lead-acid batteries are used in electric vehicles, with the notable exception of some early battery electric vehicles, such as the Detroit Electric which used a nickel-iron battery.

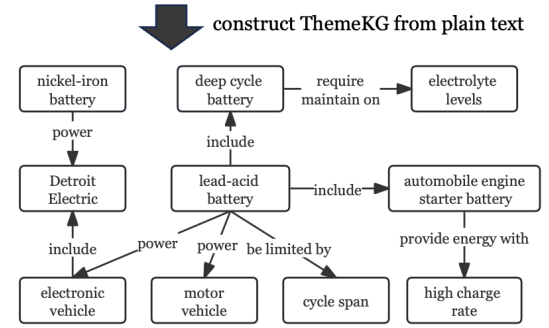


Figure 1: TKGCon: Given a set of theme-specific documents, automatic construction of a theme-specific knowledge graph.

[45], intelligent conversational systems [33], and medical concept modeling [17]. Existing knowledge graphs can be categorized into *generic open-world KGs* including Wikidata¹ and *domain-specific KGs* including UMLS [9].

Despite the broad applications of knowledge graphs, there are two major issues attached to the existing KGs, even in the current era of large language models (LLMs). The first issue is the **limited information granularity of existing KGs**. Existing KGs, including the domain-specific ones, often integrate numerous sources of texts and cover comprehensive information on a topic. They are designed for general public and do not address detailed, fine-grained information for theme-specific researchers.

Example 1 (EV battery). Taking *electric vehicle battery* as the theme, Wikidata contains limited and coarse-grained knowledge triples such as (*lead-acid battery*, *subclass of*, *rechargeable battery*). More finer-grained entities (e.g., *automobile engine starter batteries*) are not included (while it is one type of *lead-acid battery*). To inquire details of such entities (e.g., usage and features), one has to rely on

¹See https://www.wikidata.org/wiki/Wikidata:Main_Page

manual retrieval of documents. Also, LLMs could easily hallucinate on this kind of “long-tail” knowledge of a specific theme [29].

The second issue is the **lack of timeliness in existing KGs**. It is hard for a KG to keep pace with the dynamics of the real world, especially for rapid changing events, since such updates often require huge efforts of human/expert annotation and guidance.

Example 2 (Hamas attack on Israel). When a piece of news broke out, such as *Hamas attack on Israel* (Oct. 7, 2023), readers may like to know and digest international responses, but the existing KGs may not be updated timely. The knowledge of LLMs (e.g., GPT-4 [1]) is also largely limited within the training corpus, unable to conduct smart reasoning based on the up-to-date events and information.

To handle the above issues, this paper proposes a novel concept, **theme-specific knowledge graph (ThemeKG)**, that is, a knowledge graph based on a theme-specific corpus and supporting fine-grained, theme-based analysis. This idea, though attractive, poses a major challenge: Manual construction of such graphs is too costly—we need quality, automated ThemeKG construction.

With the emergence of LLMs, one may wonder whether we can simply use LLM to do so. Unfortunately, directly prompting GPT-4 can easily lead to factually incorrect triples or improper triples. In our experiment, GPT4 generates the triple (*lead-acid battery, have, two main types*) containing inaccurate entity and unclear relation, and the triple (*WhiteHouse, lit up in colors of, the Israel flag*) containing overly specific and implicit relation. To construct KGs with salient, clear and informative entities and relations, we design an unsupervised framework for Theme-specific Knowledge Graph construction (named **TKGCon**), which will take raw documents of the given theme as input and produce a quality KG on the theme with no annotations or pre-defined relations required.

Considering the limited scope of a theme-specific corpus, we leverage the commonsense knowledge from Wikipedia and LLMs to generate a theme ontology (including the entity ontology and relation ontology) to guide the construction. Specifically, we collect the hierarchical categories related to the theme in Wiki to construct high-level “entity ontology”. Then we utilize LLMs to generate potential relation candidates for pairs of entity categories to form the “relation ontology”. Guided by the theme ontology, our framework types the recognized entity mentions to the ontology and retrieves the candidate relations from LLMs. Finally, we incorporate the candidate relations with contextual information to consolidate the relations further. Note also that there are numerous works on KG constructions but few focus on the complete framework starting from the raw texts (which is the most typical scenario) and then yielding the final KG. An example is shown in Figure 1.

In summary, this work claims the following contributions:

- (1) To our best knowledge, we are the **first to introduce the theme-specific knowledge graph construction setting** along with **two newly constructed theme-specific KG construction datasets**.
- (2) We propose an **unsupervised framework that integrates the free-form generation power of LLMs and the structured guidance from a general Wiki ontology to construct a fine-grained, corpus-based, theme-coherent, and hallucination-repellent ThemeKG**. More importantly, **our framework can**

be easily adapted to any other themes with the only requirement of a theme-specific raw corpus.

- (3) Quantitative experiments and qualitative analysis show that our proposed method has achieved **advanced entity recognition and relation extraction performance** while being **coherent to themes**.

2 RELATED WORK

2.1 Information Extraction

2.1.1 Relation Extraction. Relation extraction (RE) is the task of recognizing semantic relationships between entities in the texts [16, 19, 25, 36, 38, 53, 57]. Textual pattern-based approaches leverage relation-specific patterns and syntactic parsing to extract relations [2, 37]. Prompt-tuning-based approaches construct well-curated prompts to distill the hidden relational knowledge from large language models (LLMs) [7, 8, 20]. Recent advances mainly lie within the scope of applying LLMs to assist the few-shot extraction of relations [51, 52] or synthesis of relational triples [46]. One line of research distinguishes itself by focusing on solving RE with weak or zero supervision [43, 56, 61]. To maximize models’ relation understanding under scarce training signals, several works utilize entity typing information which allows the model to better induce the relations [8, 22, 61].

2.1.2 Open Information Extraction. Among existing RE efforts, open information extraction (OIE) [10, 37, 54] is a line of work that connects closely to our task setting, as OIE extracts relation triples from text without requiring pre-defined relation sets. Traditional OIE models are based on linguistic features and statistic methods [5, 18, 37]. Recently, neural OIE models [11, 24, 31, 40, 59] learn the extraction patterns and show promising results with supervised learning. Unsupervised OIE methods [13] mainly utilize the power of PLM/LLMs. Our work further extends unsupervised OIE to a new setting that requires the extracted entities and relations to be relevant to a specific theme, which is especially challenging when no labeled data is available.

2.2 KG Construction

2.2.1 LLM-Enhanced KG Construction. Large language models [4, 39] have been shown as versatile tools for various natural language processing applications. Recent studies [6, 12, 35, 48, 55, 62] have explored leveraging LLMs for KG construction. Tahir [47] utilizes prompting methods to convert plain text to a KG. Zhu et al. [62] proposes AutoKG which adopts a multi-agent-based approach employing LLMs for KG construction and reasoning while some other studies directly query LLM to reason the relations between certain entities [6]. Ye et al. [60] introduces CooperKGC which establishes a collaborative processing network based on LLM to integrate entities, relations and extracted events for KG construction. Jiang et al. [27] generates query prompts to retrieve external information and probe knowledge from PLM for KG completion.

2.2.2 Domain-Specific KG Construction. In recent years, constructing domain-specific KGs has gained significant attention [49]. Traditional methods involve information extraction from diverse sources [58], while recent advancements leverage LLMs for entity and relation recognition. Specialized KGs tailored to specific domains,

such as geospatial or finance, have been constructed using domain-specific ontologies and data sources, with emphasis on noise filtering in extraction, ontology alignment and schema integration. [15] introduces a KG construction method tailored for online marketing which replaces the relation generation of LLM with a relation filtering process empowered by domain-specific knowledge. They also leverage progressive prompting augmentation for entity expansion. [30] focuses on the biomedical domain and constructs a domain ontology for validating gene-disease relations. They further fine-tune the biomedical KG using LLMs based on more recent articles and new knowledge bases.

Most works for domain-specific KG construction are costly in human annotations and ontology construction but are difficult to generalize to other domains. Our work focuses on automatically constructing more fine-grained, theme-specific KGs, and the framework could be easily applicable to other themes.

3 METHOD

3.1 Task Formulation

3.1.1 Theme-specific Knowledge Graph (ThemeKG). Compared with existing knowledge graphs including general KGs and domain-specific KGs (focusing on a broader domain, such as biomedical sciences, sports, and finance), the theme-specific knowledge graphs (i) focus on narrower themes (such as EV-battery and Hamas-attack-on-Israel) that may need in-depth exploration for special purposes; (ii) contain more detailed knowledge including finer-grained entities and triples. For example, (*deep cycle batteries, be power source of, forklifts*) can be a fine-grained triple in ThemeKG, while a relevant triple in general KGs could be (*EV batteries, be used in, electric vehicles*). These finer-grained triples might be seen as long-tail knowledge and discarded in previous KGs.

3.1.2 Theme-specific Knowledge Graph Construction. Given a specific theme and a set of documents \mathcal{D} with each document $d \in \mathcal{D}$ describing relevant content about the theme, our task aims to extract the theme-related knowledge triples from \mathcal{D} in the format of (e_i, r_k, e_j) , where $e_i, e_j \in \mathcal{E}$ are salient theme-specific entities recognized from \mathcal{D} , and $r_k \in \mathcal{R}$ are open-vocabulary phrases that describes the relation of two entities. Here, \mathcal{E} denotes the entity set, and \mathcal{R} denotes the relation set under the theme.

Example 3 (“EV battery” ThemeKG construction). Given the theme “EV-battery” and the following text: “*Deep cycle batteries are used to provide continuous electricity to run electric vehicles like forklifts*”, the output of theme-specific knowledge graph construction may include the following possible knowledge triples: (*deep cycle batteries, provide, continuous electricity*), (*deep cycle batteries, be power source of, electric vehicles*), (*deep cycle batteries, be power source of, forklifts*), (*electric vehicles, include, forklifts*). Figure 1 also shows a set of triples (presented as a knowledge graph) extracted from theme corpus by our TKGCon approach.

3.2 Overall Framework

To handle the task of theme-specific knowledge graph construction, our proposed TKGCon includes two main parts: *Theme Ontology Construction* and *Theme KG Construction*. The overall framework is shown in Figure 2.

- **Theme ontology construction:** Given a specific theme, we construct an ontology including both entities and relations. Specifically, we first collect a hierarchy of entity categories relevant to the theme from Wikipedia² as the *entity ontology* (Sec. 3.3.1). Then, for each pair of entity categories in the entity ontology, we generate potential candidate relations by querying LLMs to construct the relation ontology (Sec. 3.3.2).
- **Theme KG construction:** Given the theme-specific documents, we first mine the possible entity mentions and map them to the closest category on the entity ontology (Sec. 3.4.1). Then, for each pair of entity mentions, we retrieve the candidate relations (from the relation ontology) based on the entity categories that they map to. Finally, we incorporate the contextual information to decide the relation between these two entity mentions (Sec. 3.4.2).

3.3 Theme Ontology Construction

Directly constructing ThemeKG from plain text can be challenging since the theme-specific corpus may lack essential commonsense knowledge for open-vocabulary information extraction. Without any guidance, the extracted triples can contain ambiguous entities (such as “two main types”) or unclear relations (such as “are”) or be inconsistent with commonsense knowledge. To circumvent these issues, this paper leverages the advantages of large-scale general Wikipedia and LLM’s reasoning ability to construct a *theme ontology* before extracting the KG triples. The theme ontology includes an *entity ontology* constructed from Wikipedia and a *relation ontology* with high-quality relation candidates generated by LLMs.

3.3.1 Entity Ontology Construction from Wikipedia. Wikipedia is a large online database containing rich commonsense knowledge. Although it lacks fine-grained details for theme-specific tasks, the categorization function in Wikipedia³ can provide a high-quality, human-curated hierarchy of high-level concepts/categories about the theme in a top-down manner. For any entity (such as *flooded lead-acid battery*) that is not included in Wikipedia, we can still find its category path (*battery* → *rechargeable battery* → *lead-acid battery*) from the hierarchy and use it as external knowledge of the theme. This hierarchy is referred as the “entity ontology” where the nodes are high-level concepts/categories from Wiki.

We show a concrete example for entity ontology construction. Given a specific theme such as *electrical vehicle battery*, we choose the relevant concepts in the categorization of Wikipedia as initial root nodes: *Battery (Electricity)* and *Vehicles*. Then we collect their sub-trees on the pages of Wiki Category as the basic entity ontology for the theme, including the nodes like *battery inventors*, *rechargeable batteries*, *battery chargers*. Each child node in the ontology denotes a sub-category. Note that the original ontology from Wiki may contain irrelevant connections, and thus we filter out wrong connections by the similarity of parent and child nodes calculated by a pre-trained transformer, all-MiniLM-L6-v2⁴ [50] (which will be used again later). In the step of ThemeKG construction, the entity ontology can be further expanded as entities from new categories are discovered in the documents.

²<https://www.wikipedia.org/>

³<https://en.wikipedia.org/wiki/Wikipedia:Contents/Categories>

⁴<https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>

Ontology Construction (Theme: Electric Vehicle Battery)

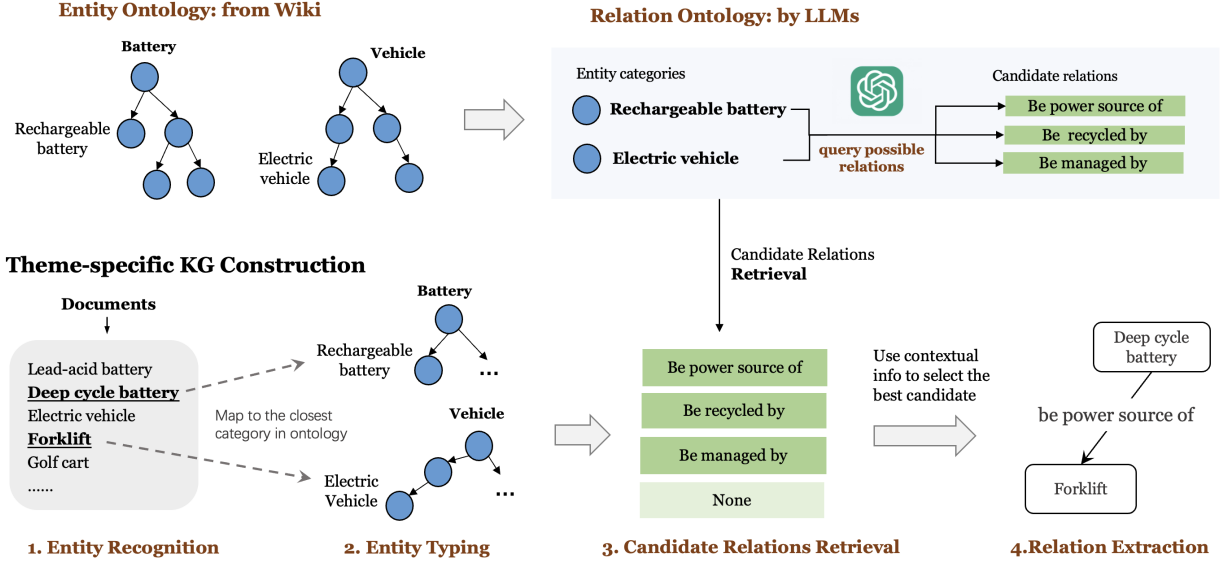


Figure 2: The overall framework of TKGCon consists of (i) ontology construction and (ii) ThemeKG construction. For (i), it leverages the large general Wikipedia and GPT-4’s reasoning ability to obtain high-quality entity ontology and relation ontology for the given theme. For (ii), we first process the theme documents, with phrase mining by SpaCy, entity typing by ZOE to retrieve candidate relations. Finally, the candidate relations generated by LLMs are further filtered with contextual information to consolidate final relations.

3.3.2 Relation Ontology Construction by LLMs. Directly generating relations for an entity pair under a theme-specific context can result in undesired levels of variations or even mistakes. For example, the generated relations can be overly general (e.g., “is”), overly specific (e.g., “be the oldest and cheapest battery of”), overly implicit (e.g., “lit up in colors of”), or wrong (e.g., “start” in the triple (charging rate, start, engine)). We observe that relations generated from entity categories are usually more robust than those generated from theme-specific entities. To construct a KG with salient, clear, and informative relations, we make the following hypothesis:

HYPOTHESIS 1. Give entity categories E_1 and E_2 from Wiki ontology, the set of possible relations from entity $e_1 \in E_1$ to entity $e_2 \in E_2$, is a subset of possible relations from category E_1 to category E_2 .

For example, we know that the possible relations from category *rechargeable battery* to category *electric vehicles* could be {“be power source of”, “be recycled from”, “be managed by”}. Then, we can reason that the relation from $e_1 \in E_1$ (e.g., *deep cycle battery*, which is a type of *rechargeable battery*) to $e_2 \in E_2$ (e.g., *forklift*, which is a type of *electric vehicles*) should be within the set.

Thus, we construct the “relation ontology” in the following approach. For a given pair of entity categories (E_1, E_2) in the entity ontology collected from Wiki, we follow the hypothesis and query LLM to reason the potential candidate relations⁵. The final relation ontology is a map from category pairs to candidate relation sets: $(E_1, E_2) \rightarrow \mathcal{R}$. The prompt we used is as follows:

Given the theme [Theme], what are the possible relations from [Category 1] to [Category 2]? List Answers in the format: ([Category 1], ____, [Category 2])

Here, the commonsense knowledge in LLM can compensate for the deficiency of limited theme-specific data. For an up-to-date specific theme (such as *Hamas-attack-on-Israel*), the approach can reason the potential relations as long as the entity’s affiliated category is known to the pre-trained LLM. Note that, we also limit the depth of Wiki entity ontology to prevent overly fine-grained entity categories, which may cause hallucination during querying LLM.

3.4 Theme KG Construction

We observe that end-to-end methods for KG construction tend to miss triples especially for document-level extraction. Our framework can effectively avoid the issue by following a pipeline that first recognizes the entities and then extracts their relations.

3.4.1 Entity Recognition and Typing. Given a set of documents about the given theme, our TKGCon first extracts the possible entity mentions, which will be mapped to the categories on the entity ontology later. Due to the limited data and unseen fine-grained entities, we find that most existing phrase/entity extraction methods such as AutoPhrase [44] have low recall rates. To ensure a high recall rate, our approach first gets noun chunks by SpaCy [21] as candidate entity mentions. We then filter out the noisy mentions (e.g. “itself”, “the features”) or noisy modifiers (e.g. “most common” in the phrase “most common vehicle batteries”) by a set of rules.

- (1) The valid mentions should contain at least one noun and no pronouns/stopwords according to the POS.

⁵We query in both directions: (entity 1, relation, entity 2) and (entity 2, relation, entity 1).

- (2) The phrases that have a high frequency in large corpus⁶ and low theme coherence (similar calculation could refer to Eqn. (2)) are treated as noisy mentions (e.g. "cases").
- (3) The non-nouns that have high frequency and low co-occurrence with other words should be the noisy modifiers (e.g. "actual").

With the entity mentions from theme documents, we then map them to the closest Wiki category. We consider two cases.

- **Case 1:** For a mention that directly matches a Wikipedia page title, we consider the page title as an entity and the categories at the bottom of the page as the associated categories. We then calculate two scores for determining which category from the set is the best match: (i) *self coherence* between entity e and category E ; (ii) *theme coherence* between theme t and category E .

$$C_{self} = \cos(\text{Rep}(E), \text{Rep}(e)), \quad (1)$$

$$C_{theme} = \cos(\text{Rep}(E), \text{Rep}(t)). \quad (2)$$

The representations $\text{Rep}(\cdot)$ are provided by the pre-trained transformer [50]. After obtaining C_{theme} , we first use a threshold to filter out categories that are irrelevant to the theme. If there is no category left after the filtering, we consider the entity mention is also irrelevant to the theme; Otherwise, we sort the multiplication value $C_{self} \times C_{theme}$ for each remaining category in the set and find the best match that corresponding to the max value.

- **Case 2:** If the mention does not have a direct match in Wikipedia, we do the following. For the entity mention e and context c , we adapt ideas from ZOE [41] to type the entity based on the entity ontology. Specifically, for every category E in the ontology, we use the pre-trained ELMo [42] to calculate the context consistency score $P(E | e, c)$ as ZOE does. If the score reaches a lower threshold, meaning that we could not find a proper match in the existing entity ontology. In this case, we then apply explicit semantic analysis [14] to fast retrieve relevant Wikipedia categories as the candidate set. Similarly, we adopt the self coherence and theme coherence scores to determine the best-matched category or consider the entity irrelevant to the theme if none of the categories in the candidate set has high C_{theme} .

3.4.2 Relation Retrieval and Extraction. Given two entity mentions, we introduce how to extract their relations by incorporating the contextual information. For a pair of entity mentions (e_1, e_2) , assuming their typed (i.e., best matched) categories are (E_1, E_2) , we first retrieve candidate relations from theme relation ontology. To improve the accuracy, diversity, and coverage, we also retrieve the relations based on the parent nodes of E_1 and E_2 in the entity ontology and merge these candidate relations together.

Then, we select the best suitable relation based on the context (neighboring sentences that include the entities) by LLMs since relation extraction requires semantic understanding and reasoning abilities. Before doing so, we add the *none* relation into the candidate relations in case the pair of entity mentions cannot find suitable relation in candidate set according to the context. The prompt we use is as follows:

Please choose the most proper relation in the candidate set for [entity1] to [entity2] according to the context. If all the relations in the candidate set are not suitable, please choose none. The

Table 1: Statistics of Datasets

Datasets	Documents	Entities	Relations	Triples
EVB	20	130	64	330
HAI	20	142	77	425

output format should be (entity1, relation, entity2). Context: [context]. Relation candidates: [relation1, relation2, ..., none]

If the output relation is *none*, then either (i) two entities do not have factual relations from the document; or (ii) during relation ontology construction, the LLMs have missed the relation between these two entities (the cases are rare). For both cases, without further guidance, we query LLMs again to directly extract the entity relation based on the context or output a *none*. This prompt is shown in Sec 4.5. If LLM gives *none* as the answer, we consider it as case (i); Otherwise, we generate a triple (entity1, relation, entity2) and add this new relation to enrich the theme relation ontology.

Finally, we organize these valid triples to get a theme-specific knowledge graph (ThemeKG).

4 EXPERIMENT

Our framework aims to construct theme-specific KGs from scratch with limited documents related to the given theme. The task is closely related to document-level open information extraction. This section evaluates the performance of our TKGCon against existing baseline models on constructing two theme-specific KGs.

4.1 Dataset Collection

To verify the validity of our framework, we choose two themes and collect corpus for theme-specific KG construction: *EV battery (EVB)* and *Hamas-attack-on-Israel (2023) (HAI)*. Specifically, for each theme, we collected the documents from online sources with four criteria: (i) the relevance to the given theme, (ii) the reliability of the source contents, (iii) the granularity of description details, and (iv) the diversity of information about the theme.

Electrical vehicle batteries (EVB) is a traditional theme in the industry while the specific terms or other details of this theme are hardly included in existing KGs. *Hamas-attack-on-Israel (2023)* is an up-to-date theme in the press with a number of new entities (e.g. "Nova music festival") not included in existing knowledge bases. For each theme, we mainly collect 20 high-quality and representative documents related to the theme from various online databases (e.g. New York Times, CRS Reports⁷). Then, we annotate the entities and triples at the document level from two corpora to generate the gold theme-specific knowledge graphs for evaluation. The statistics of two datasets EVB and HAI can be found in Table 1.

The datasets can be used as a benchmark for theme-specific knowledge graph construction or sub-tasks like entity recognition and open relation extraction.

4.2 Evaluation Metrics

Given the target theme-specific KG, $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$ and the generated KG, $\hat{\mathcal{G}} = (\hat{\mathcal{E}}, \hat{\mathcal{R}}, \hat{\mathcal{T}})$ with $\mathcal{E}, \hat{\mathcal{E}}$ denoting the set of gold and recognized entities, $\mathcal{R}, \hat{\mathcal{R}}$ denoting the set of gold and generated

⁶<https://github.com/IlyaSemenov/wikipedia-word-frequency>

⁷<https://crsreports.congress.gov/>

Table 2: Comparison with baselines on KG construction.

Dataset	Method	Entity Metric			Triple Metric			Theme Metric
		Recall	Precision	F1-score	Recall	Precision	F1-score	Coherence
EVB	OpenIE [37]	0.62	0.36	0.46	0.13	0.24	0.17	0.46
	REBEL [24]	0.22	0.80	0.35	0.11	0.80	0.19	0.80
	IMoJIE [31]	0.44	0.49	0.46	0.26	0.45	0.33	0.78
	KG-GPT [47]	0.72	0.69	0.70	0.67	0.64	0.65	0.95
	GPT-4 [1]	0.68	0.71	0.69	0.64	0.65	0.64	0.97
	TKGCon (w/o ontology)	/	/	/	0.67	0.57	0.62	0.92
	TKGCon	0.92	0.80	0.86	0.78	0.73	0.75	0.97
HAI	OpenIE [37]	0.52	0.28	0.36	0.17	0.22	0.19	0.35
	REBEL [24]	0.16	0.87	0.27	0.15	0.75	0.25	0.75
	IMoJIE [31]	0.33	0.39	0.36	0.25	0.31	0.28	0.83
	KG-GPT [47]	0.84	0.79	0.81	0.72	0.69	0.70	0.91
	GPT-4 [1]	0.82	0.80	0.83	0.70	0.72	0.71	0.93
	TKGCon (w/o ontology)	/	/	/	0.75	0.62	0.68	0.88
	TKGCon	0.90	0.88	0.89	0.81	0.75	0.78	0.92

relations, and $\mathcal{T}, \hat{\mathcal{T}}$ denoting the set of gold and extracted triples (e_i, r_k, e_j) . We define the following evaluation metrics from different perspectives to evaluate the quality of generated KG.

- **Entity Metrics.** We use precision, recall, and F1-score to evaluate the extracted entities. Since KG construction is an open task, there is no penalty on precision if the recognized entities are reasonable but not included in the gold set \mathcal{E} . Also, we regard two entities as the same if they are semantically-matched coreference of entities such as *EV battery* and *electrical vehicle batteries*.
- **Triple Metrics.** We use the precision, recall, and F1-scores for evaluating the generated triples. Since the relations can be paraphrased into various expressions such as *subclass of* and *type of*, we apply soft matching [28] by obtaining embeddings of the pre-trained transformer and calculating cosine similarities between the extracted triples and gold triples.
- **Theme Coherence.** The raw documents usually include noisy information. We expect the KG to focus on theme-related entities and relations. Therefore, we propose a metric to evaluate the theme coherence of triples. We calculate the cosine similarity between triples and the theme descriptions by the pre-trained transformer. The triples with high similarities are seen as theme-coherent triples, thus the metric of theme coherence is defined by the rate of theme-coherent triples in all generated triples.

4.3 Baselines

We consider existing open knowledge graph construction methods including two GPT-4 based baselines.

- (1) **OpenIE** [37] is an open information extraction system which extracts relation triples from plain text. This baseline focuses mainly on the syntactic patterns of the text to compensate for the lack of supervision signals.
- (2) **REBEL** [24] is a generative framework that formulates KG construction as a sequence-to-sequence task that extracts all the triples within the input texts based on BART [32]. The model is pre-trained on the general corpus and evaluated on our corpus.
- (3) **IMoJIE** [31] is a BERT-based OpenIE system that generates information extraction using an iterative Seq2Seq model. We use the model that is trained on the CaRB [3] data.

- (4) **KnowledgeGraph GPT (KG-GPT)**⁸ [47] is a tool that converts plain text to knowledge graph with GPT-4. The approach designs prompts to generate the results in the style of RDF Triples⁹.
- (5) We also directly take **GPT-4**¹⁰ [1] as a baseline by prompting it in an end-to-end manner. We use in-context learning to query the theme document one at a time and then combine the generated triples together. The prompt we used is:

Given the theme [Theme], extract all knowledge triples in the format of (entity1, relation, entity2) from the following text: [document].

4.4 Main Results Analysis

We compare our proposed framework TKGCon against baselines on both datasets EVB and HAI. The evaluation results of entities, triples, and theme are present in Table 2.

We first evaluate and compare the extracted **entities** in the output KGs of all models. The results in Table 2 show that our TKGCon can outperform other baseline methods considerably on all metrics. Linguistic-based OpenIE has relatively high recall among non-GPT-based baselines. However, the model lacks an understanding of semantics, and thus the entities extracted in OpenIE are usually incomplete, erroneous and noisy, especially in processing long and complex sentences, which leads to low precision and F1-score. End-to-end KGC methods such as KG-GPT and GPT-4 include overly specific entities (such as "two main types") and are inflexible to optimize. Compared to end-to-end methods, our TKGCon can greatly improve the recall rate of entities by explicit phrase mining and boost the precision with entity filtering in intermediate steps.

Our TKGCon also shows decent improvements over the baselines on **triple** extraction and **theme** coherence. Specifically, OpenIE generates relatively poor quality outputs due to its lack of semantic understanding. While supervised sequence generative methods like REBEL and IMoJIE have slightly better precision and theme coherence, they are found weak in recall rate due to the difficulty of generalizing to unseen relations beyond the training dataset, while

⁸<https://github.com/iAmmarTahir/KnowledgeGraphGPT>

⁹<https://www.w3.org/TR/WD-rdf-syntax-971002/>

¹⁰<https://openai.com/gpt-4>

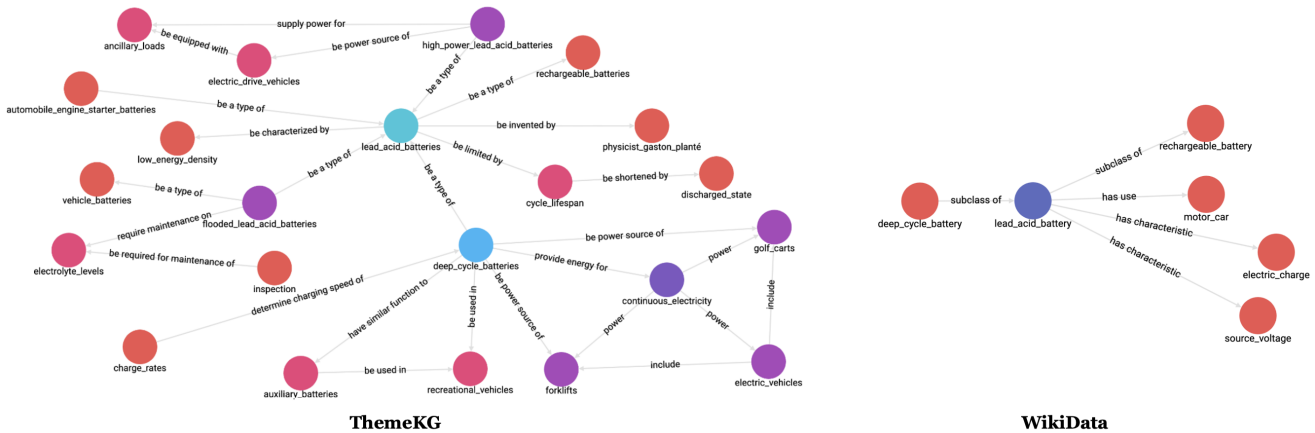


Figure 3: Comparison of ThemeKG and WikiData on EV battery. For the theme *EV battery*, the left side is part of our extracted ThemeKG extracted from theme-specific document corpus. The triples on the right side are retrieved from WikiData of the same topics. Our ThemeKG contains more specific entities and relations of the theme compared to the WikiData KG.

our datasets include a number of fine-grained and theme-specific relations. GPT-4 is capable of comprehension and summarizing, and thus its induced models (GPT-4 and KG-GPT) have significant improvements compared to others. However, in the setting of ThemeKG, these two GPT-4-based models may still generate ambiguous and illogical relations or entities (details and examples can be found in Sec. 4.6). Breaking the ThemeKG construction into multiple stages, our framework TKGCon can effectively improve the recall rate of document-level triples, compared to end-to-end models. With the guidance of theme ontology, TKGCon is able to extract higher quality relations and outperform GPT-based methods in triple F1-score while getting competitive theme coherence on par with GPT-4.

4.5 Ablation Study

Essentially, our framework could be separated into two sequential stages: *entity recognition* (step 1 in Figure 2) and *ontology-guided relation extraction* (step 2 & 3 & 4 in Figure 2). To validate the effectiveness of the relation ontology, we build a new model variant, named *TKGCon (w/o ontology)*, which leverages entity recognition (step 1) from our framework and then directly queries GPT-4 to generate the relations between entities based on the context. The prompt for querying the relation is as follows:

Extract the relation from [entity1] to [entity2] in the following passage: [context]. Please output in the format of ([entity1], [relation], [entity2]). If no relation from [entity1] to [entity2] is identified based on the context, then output none.

The performance of TKGCon (w/o ontology) has been added to Table 2. Note that, our model and this variant use the same entity recognition step, so their entity metrics are all the same. For all triple metrics and theme coherence, this variant model consistently gives lower performance. Without the guidance and restrictions of theme ontology, the variant model generates incorrect relations when directly querying GPT-4. For example, its generated triples (*auxiliary batteries, be used as, recreational vehicles*) and (*inspection,*

requires inspection of, electrolyte levels) are obviously problematic because the batteries cannot "*be used as*" vehicles and inspection should not have the relation "*requires inspection of*".

Also, TKGCon (w/o ontology) tends to give pointless phrases as relations, such as (*charge rates, provide high, golf carts*) and (*electric vehicles, notable exception, Detroit Electric*). On the contrary, the ontology-based TKGCon can avoid such scenarios with the help of entity ontology and relation ontology, e.g. (*recreational vehicles, be equipped with, auxiliary batteries*), (*inspection, be required for maintenance of, electrolyte levels*).

4.6 Case Study

This section uses the theme *EV battery* as an example and intuitively shows the differences between ThemeKG and Wikidata. Also, we compare the extracted triples of TKGCon to the results of baselines.

4.6.1 Compared to Wikidata. Figure 3 presents the difference of ThemeKG and Wikidata for *EV battery*. Due to space limitations, we only show the parts of the triples related to *lead acid batteries* in our constructed ThemeKG (left side on Figure 3). We also retrieve the triples of relevant topics on Wikidata¹¹ (right side on Figure 3).

We observe that our constructed ThemeKG (from raw *EV battery* documents) includes more **finer-grained entities** such as *flooded lead acid batteries*, *automobile engine starter batteries* which are not presented on the right side. The ThemeKG also contains broader and more **fine-grained knowledge triples** (e.g. *lead acid batteries*, *be limited by*, *cycle lifespan*) about the theme while Wikidata only includes a few coarse-grained triples. Therefore, we claim that ThemeKG could provide broader and more in-depth details on specific themes compared to existing KGs.

4.6.2 *Compared to baselines.* We also sample some extracted triples from TKGCon and compare them with the outputs from OpenIE, GPT-4 in Table 3. We could observe that OpenIE misses many entities and triples relevant to the theme since the method is based

¹¹See https://www.wikidata.org/wiki/Wikidata:Main_Page

Table 3: Examples of triples. The bold refers to correct triples.

Method	Extracted Triples (on same plain text)
OpenIE	(lead acid batteries, have compared, rechargeable batteries), (It, is, first type), (calendar, impede, lifecycle), (lead acid batteries, have, energy density), (features, make along, their cost), (overall lifespan, is in, discharged state), (lead acid batteries, suffer from, cycle lifespan), (deep cycle batteries, are, used), (percentage, provide, charge rates)
GPT-4	(lead-acid battery, is a type of, rechargeable batteries), (lead-acid batteries, have, low energy density), (lead-acid batteries, have, two main types), (lead-acid batteries, suffer from, relatively short cycle lifespan), (lead-acid batteries, have due to, "double sulfation" in the discharged state), (lead-acid batteries, types, deep cycle batteries), (high charge rates, to start, engine), (flooded lead-acid batteries, are, vehicle batteries), (deep cycle batteries, require, multi-stage charging), (deep cycle batteries, used in, electric vehicles like forklifts or golf carts)
TKGCon	(lead acid batteries, be a type of, rechargeable batteries), (lead acid batteries, be characterized by, low energy density), (lead acid batteries, be limited by, cycle lifespan), (cycle lifespan, be shortened by, discharged state), (flooded lead acid batteries, be a type of, vehicle batteries), (automobile engine starter batteries, be different from, deep cycle batteries), (deep cycle batteries, be power source of, forklifts), (deep cycle batteries, be power source of, golf carts), (deep cycle batteries, be used to store, continuous electricity)

on linguistic analysis without understanding of the theme. It also leads to pointless triples when the sentences are long and complex. GPT-4 works relatively better due to the strong ability of text understanding. However, directly prompting GPT-4 makes the entities excessively specific (e.g., "electric vehicles like forklifts or golf carts") or leads to ambiguous (e.g., "is", "have", "were") and wrong relations (e.g., "have due to", "to start").

Our TKGCon breaks the ThemeKG construction task into several steps, which effectively optimizes each step and improves the quality and recall rate of entities. In the example, TKGCon extracts more accurate, e.g., entities "forklifts" and "golf carts" (compared to the excessively specific entity, "electric vehicles like forklifts or golf carts", generated by GPT-4). Additionally, TKGCon retrieves candidate relations from relation ontology, which prevents relation ambiguity and improves the accuracy in relation extraction. For example, our method generates "be power source of" instead of "used in" (generated by GPT-4). Furthermore, benefiting from the ontology, TKGCon improves the consistency of relations while relations generated from GPT-4 are oftentimes unstable (e.g. "is a type of", "types", "are" for the same kind of relation).

4.7 Applications of ThemeKG

The ThemeKG are compressed and structured collections of fine-grained knowledge for specific themes. Besides intuitive access to knowledge, ThemeKG can be used as an external tool to facilitate the downstream tasks. For example, the LLMs may cause hallucinations on the fine-grained details and do not possess up-to-date information for a specific theme. In this case, our ThemeKG can be retrieved and loaded as tools to augment the LLMs.

This section tests the model’s question-answering (QA) ability over a specific theme. We consider the following models.

- **ThemeKG-augmented GPT-4 (TKG+GPT4)** uses ThemeKG (formatted as a list of triples) in GPT-4 for in-context QA since the ThemeKG is more compressed than the entire theme corpus.
- **Vanilla GPT-4** that directly uses GPT-4 for QA.
- **Retrieval-augmented generation with GPT-4 (RAG+GPT4)** [34] is a representative method that retrieving facts from an external knowledge base to augment GPT-4 on the accurate,

Table 4: An example of ThemeKG-augmented GPT-4 on QA task. The bold phrases are the correct answers.

Question	Which countries support Hamas or condemn Israel in the Hamas attack on Israel in Oct 2023?
Vanilla GPT4	I’m sorry, but as of my knowledge cutoff date in march 2023, i do not have information on specific events that occurred in october 2023.
RAG+GPT4	In the Middle East and North Africa , most countries either condemned Israel or offered full-throated support to Hamas. North Korea is also mentioned as condemning Israel.
TKG+GPT4	1. Iran , 2. Persian Gulf countries , 3. North Korea , 4. most Middle East countries , 5. most North Africa countries

up-to-date information. We take the raw theme corpus as the external database of RAG for QA.

Table 4 lists an example of the theme *Hamas-Attack-on-Israel (2023)* (redundant sentences of QA output are all removed). The question is "Which countries support Hamas or condemn Israel in the Hamas attack on Israel in Oct 2023?". To answer the question requires information integration from multiple documents. Apparently, GPT-4 alone cannot deal with the up-to-date information. The RAG+GPT4 retrieves parts of relevant facts while misses information such as text mentioning *Iran* and *Persian Gulf Countries*. Our TKG+GPT4 can directly integrate information from triples, e.g. (*Iran, provide assistance to, Hamas*) and (*North Korea, condemn, Israel*), without summarizing and reasoning over the documents. Essentially, the ThemeKG could empower LLMs with timely, fine-grained information and also improve their abilities of reasoning and summarizing over the theme.

In future work, we will further explore other utilities of ThemeKGs in more scenarios, such as personalized recommendation and tool learning with different ThemeKGs as plugins.

5 CONCLUSION

This paper proposes an automated KG construction method (named TKGCon) for constructing a theme-specific knowledge graph (ThemeKG) from a raw theme corpus. Existing general or domain-specific KGs could be limited in information granularity and timeliness, while our focused ThemeKG contains fine-granular, timely, theme-specific entities and relations, providing more theme-relevant details (such as specialized terms or key events in recent news) for downstream applications. Our approach leverages well-curated theme ontology information in Wikipedia as well as the strong reasoning ability of LLM for relation candidate generation. Experiments on *EV battery* and *Hamas-attack-on-Israel* show that TKGCon can automatically generate fine-grained, high-quality ThemeKG in comparison with various existing KG construction approaches.

REFERENCES

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altmerschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- [2] David S. Batista, Bruno Martins, and Mário J. Silva. 2015. Semi-supervised bootstrapping of relationship extractors with distributional semantics. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 499–504, Lisbon, Portugal. Association for Computational Linguistics.
- [3] Sangnie Bhardwaj, Samarth Aggarwal, et al. 2019. Carb: A crowdsourced benchmark for open ie. 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 6262–6267.
- [4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- [5] Matthias Cetto, Christina Niklaus, André Freitas, and Siegfried Handschuh. 2018. Graphene: Semantically-linked propositions in open information extraction. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2300–2311.
- [6] Jiao Chen, Luyi Ma, Xiaohan Li, Nikhil Thakurdesai, Jianpeng Xu, Jason HD Cho, Kaushiki Nag, Evren Korpoglu, Sushant Kumar, and Kannan Achan. 2023. Knowledge graph completion models are few-shot learners: An empirical study of relation labeling in e-commerce with lms. *arXiv preprint arXiv:2305.09858*.
- [7] Xiang Chen, Lei Li, Ningyu Zhang, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2022. Relation extraction as open-book examination: Retrieval-enhanced prompt tuning. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2443–2448.
- [8] Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2022. Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction. In *Proceedings of the ACM Web conference 2022*, pages 2778–2788.
- [9] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.
- [10] Kuicai Dong, Zhao Yilin, Aixin Sun, Jung-jae Kim, and Xiaoli Li. 2021. Docoie: A document-level context-aware dataset for openie. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2377–2389.
- [11] Zhiyuan Fan and Shizhu He. 2023. Efficient data learning for open information extraction with pre-trained language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13056–13063.
- [12] Junfeng Fang, Shuai Zhang, Chang Wu, Zhiyuan Liu, Sihang Li, Kun Wang, Wenjie Du, Xiang Wang, and Xiangnan He. 2024. Molte: Towards molecular relational modeling in language models.
- [13] Yufeng Fu, Xiangge Li, Ce Shang, Hong Luo, and Yan Sun. 2023. Zoie: A zero-shot open information extraction model based on language model. In *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 784–789. IEEE.
- [14] Evgeniy Gabrilovich, Shaul Markovitch, et al. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *IJCAI*, volume 7, pages 1606–1611.
- [15] Chunjing Gan, Dan Yang, Binbin Hu, Ziqi Liu, Yue Shen, Zhiqiang Zhang, Jinjie Gu, Jun Zhou, and Guannan Zhang. 2023. Making large language models better knowledge miners for online marketing with progressive prompting augmentation. *arXiv preprint arXiv:2312.05276*.
- [16] Yanglei Gan, Rui Hou, Run Lin, Qiao Liu, Yuxiang Cai, Wannian Gao, et al. 2023. Synergistic anchored contrastive pre-training for few-shot relation extraction. *arXiv preprint arXiv:2312.12021*.
- [17] Junyi Gao, Chaoqi Yang, Joerg Heintz, Scott Barrows, Elise Albers, Mary Stapel, Sara Warfield, Adam Cross, and Jimeng Sun. 2022. Medml: fusing medical knowledge and machine learning models for early pediatric covid-19 hospitalization and severity prediction. *Science*, 25(9).
- [18] Kiril Gashteovski, Rainer Gemulla, and Luciano Del Corro. 2017. Minie: Minimizing facts in open information extraction. In *Conference on Empirical Methods in Natural Language Processing*, pages 2620–2630. ACL.
- [19] Shashank Gupta, Xuguang Ai, and Ramakanth Kavuluru. 2023. Comparison of pipeline, sequence-to-sequence, and gpt models for end-to-end relation extraction: experiments with the rare disease use-case. *arXiv preprint arXiv:2311.13729*.
- [20] Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. 2021. Ptr: Prompt tuning with rules for text classification. *arXiv preprint arXiv:2105.11259*.
- [21] Matthew Honnibal and Ines Montani. 2017. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 7(1):411–420.
- [22] Wei Hu, Bowen Jin, Minhao Jiang, Sizhe Zhou, Zhaonan Wang, Jiawei Han, and Shaowen Wang. 2024. Geospatial topological relation extraction from text with knowledge augmentation. In *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)*.
- [23] Jin Huang, Wayne Xin Zhao, Hongjian Dou, Ji-Rong Wen, and Edward Y Chang. 2018. Improving sequential recommendation with knowledge-enhanced memory networks. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, pages 505–514.
- [24] Pere-Lluís Huguet Cabot and Roberto Navigli. 2021. REBEL: Relation extraction by end-to-end language generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2370–2381, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- [25] Monika Jain, Raghava Mutharaju, Ramakanth Kavuluru, and Kuldeep Singh. 2024. Revisiting document-level relation extraction with context-guided link prediction. *arXiv preprint arXiv:2401.11800*.
- [26] S Ji, S Pan, E Cambria, P Marttinen, and PS Yu. 2022. A survey on knowledge graphs: Representation, acquisition and application. *IEEE transactions on neural networks and learning systems*, 33(2).
- [27] Pengcheng Jiang, Shivam Agarwal, Bowen Jin, Xuan Wang, Jimeng Sun, and Jiawei Han. 2023. Text augmented open knowledge graph completion via pre-trained language models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 11161–11180, Toronto, Canada. Association for Computational Linguistics.
- [28] Yizhu Jiao, Sha Li, Yiqing Xie, Ming Zhong, Heng Ji, and Jiawei Han. 2022. Open-vocabulary argument role prediction for event extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.
- [29] Nikhil Kandpal, Haikang Deng, Adam Roberts, Eric Wallace, and Colin Raffel. 2023. Large language models struggle to learn long-tail knowledge. In *International Conference on Machine Learning*, pages 15696–15707. PMLR.
- [30] Md Rezaul Karim, Lina Molinas Comet, Md Shajalal, Oya Beyan, Dietrich Rebholz-Schuhmann, and Stefan Decker. 2023. From large language models to knowledge graphs for biomarker discovery in cancer. *arXiv preprint arXiv:2310.08365*.
- [31] Keshav Kolluru, Samarth Aggarwal, Vipul Rathore, Soumen Chakrabarti, et al. 2020. Imojie: Iterative memory-based joint open information extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5871–5886.
- [32] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880.
- [33] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Kuttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.
- [34] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Kuttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.
- [35] Dawei Li, Zhen Tan, Tianlong Chen, and Huan Liu. 2024. Contextualization distillation from large language model for knowledge graph completion. *arXiv preprint arXiv:2402.01729*.
- [36] Junpeng Li, Zixia Jia, and Zilong Zheng. 2023. Semi-automatic data enhancement for document-level relation extraction with distant supervision from large language models. *arXiv preprint arXiv:2311.07314*.
- [37] 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.
- [38] Shiao Meng, Xuming Hu, Aiwei Liu, Shu’ang Li, Fukun Ma, Yawen Yang, and Lijie Wen. 2023. Rapl: A relation-aware prototype learning approach for few-shot document-level relation extraction. *arXiv preprint arXiv:2310.15743*.
- [39] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.
- [40] Kevin Song Pei, Ishan Jindal, and Kevin Chang. 2023. Abstractive open information extraction. In *The 2023 Conference on Empirical Methods in Natural Language Processing*.
- [41] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*.
- [42] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.
- [43] Meng Qu, Xiang Ren, Yu Zhang, and Jiawei Han. 2018. Weakly-supervised relation extraction by pattern-enhanced embedding learning. In *Proceedings of the 2018 World Wide Web Conference*, pages 1257–1266.

- [44] Jingbo Shang, Jialu Liu, Meng Jiang, Xiang Ren, Clare R Voss, and Jiawei Han. 2018. Automated phrase mining from massive text corpora. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1825–1837.
- [45] Haitian Sun, Tania Bedrax-Weiss, and William Cohen. 2019. PullNet: Open domain question answering with iterative retrieval on knowledge bases and text. 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 2380–2390, Hong Kong, China. Association for Computational Linguistics.
- [46] Qi Sun, Kun Huang, Xiaocui Yang, Rong Tong, Kun Zhang, and Soujanya Poria. 2024. Consistency guided knowledge retrieval and denoising in llms for zero-shot document-level relation triple extraction. *arXiv preprint arXiv:2401.13598*.
- [47] Ammar Tahir. 2023. Knowledge graph gpt. <https://github.com/iAmmarTahir/KnowledgeGraphGPT>.
- [48] Milena Trajanoska, Riste Stojanov, and Dimitar Trajanov. 2023. Enhancing knowledge graph construction using large language models. *arXiv preprint arXiv:2305.04676*.
- [49] Zaratiana Urchade, Nadi Tomeh, Pierre Holat, and Thierry Charnois. 2024. An autoregressive text-to-graph framework for joint entity and relation extraction. *arXiv preprint arXiv:2401.01326*.
- [50] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- [51] Somin Wadhwa, Silvio Amir, and Byron C Wallace. 2023. Revisiting relation extraction in the era of large language models. *arXiv preprint arXiv:2305.05003*.
- [52] Zhen Wan, Fei Cheng, Zhuoyuan Mao, Qianying Liu, Haiyue Song, Jiwei Li, and Sadao Kurohashi. 2023. GPT-RE: In-context learning for relation extraction using large language models. In *The 2023 Conference on Empirical Methods in Natural Language Processing*.
- [53] Qing Wang, Kang Zhou, Qiao Qiao, Yuepei Li, and Qi Li. 2023. Improving unsupervised relation extraction by augmenting diverse sentence pairs. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12136–12147, Singapore. Association for Computational Linguistics.
- [54] Kaiwen Wei, Yiran Yang, Li Jin, Xian Sun, Zequn Zhang, Jingyuan Zhang, Xiao Li, Linhao Zhang, Jintao Liu, and Guo Zhi. 2023. Guide the many-to-one assignment: Open information extraction via iou-aware optimal transport. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4971–4984.
- [55] Yanbin Wei, Qiushi Huang, James T Kwok, and Yu Zhang. 2024. Kicgpt: Large language model with knowledge in context for knowledge graph completion. *arXiv preprint arXiv:2402.02389*.
- [56] Zeqiu Wu, Xiang Ren, Frank F Xu, Ji Li, and Jiawei Han. 2018. Indirect supervision for relation extraction using question-answer pairs. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 646–654.
- [57] Zhaohui Yan, Songlin Yang, Wei Liu, and Kewei Tu. 2023. Joint entity and relation extraction with span pruning and hypergraph neural networks. *arXiv preprint arXiv:2310.17238*.
- [58] Yunzhi Yao, Shengyu Mao, Xiang Chen, Ningyu Zhang, Shumin Deng, and Huajun Chen. 2022. Schema-aware reference as prompt improves data-efficient relational triple and event extraction. *arXiv preprint arXiv:2210.10709*.
- [59] Hongbin Ye, Honghao Gui, Xin Xu, Huajun Chen, and Ningyu Zhang. 2023. Schema-adaptable knowledge graph construction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*.
- [60] Hongbin Ye, Honghao Gui, Aijia Zhang, Tong Liu, Wei Hua, and Weiqiang Jia. 2023. Beyond isolation: Multi-agent synergy for improving knowledge graph construction. *arXiv preprint arXiv:2312.03022*.
- [61] Sizhe Zhou, Suyu Ge, Jiaming Shen, and Jiawei Han. 2023. Corpus-based relation extraction by identifying and refining relation patterns. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 20–38. Springer.
- [62] Yuqi Zhu, Xiaohan Wang, Jing Chen, Shuofei Qiao, Yixin Ou, Yunzhi Yao, Shumin Deng, Huajun Chen, and Ningyu Zhang. 2023. Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities. *arXiv preprint arXiv:2305.13168*.