Automated Mining of Structured Knowledge from Text in the Era of Large Language Models

Yunyi Zhang
University of Illinois Urbana-Champaign
Urbana, IL, USA
yzhan238@illinois.edu

Ming Zhong
University of Illinois Urbana-Champaign
Urbana, IL, USA
mingz5@illinois.edu

Siru Ouyang
University of Illinois Urbana-Champaign
Urbana, IL, USA
siruo2@illinois.edu

Yizhu Jiao
University of Illinois Urbana-Champaign
Urbana, IL, USA
yizhuj2@illinois.edu

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

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

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

ABSTRACT

Massive amount of unstructured text data are generated daily, ranging from news articles to scientific papers. How to mine structured knowledge from the text data remains a crucial research question. Recently, large language models (LLMs) have shed light on the text mining field with their superior text understanding and instruction-following ability. There are typically two ways of utilizing LLMs: fine-tune the LLMs with human-annotated training data, which is labor intensive and hard to scale; prompt the LLMs in a zero-shot or few-shot way, which cannot take advantage of the useful information in the massive text data. Therefore, it remains a challenge on automated mining of structured knowledge from massive text data in the era of large language models.

In this tutorial, we cover the recent advancements in mining structured knowledge using language models with very weak supervision. We will introduce the following topics in this tutorial: (1) introduction to large language models, which serves as the foundation for recent text mining tasks, (2) ontology construction, which automatically enriches an ontology from a massive corpus, (3) weakly-supervised text classification in flat and hierarchical label space, (4) weakly-supervised information extraction, which extracts entity and relation structures.

CCS CONCEPTS

• **Information systems → Data mining**; • **Computing methodologies → Natural language processing**; *Information extraction*; *Classification and regression trees.*

KEYWORDS

Text Mining, Weak Supervision, Large Language Models

ACM Reference Format:
Yunyi Zhang, Ming Zhong, Siru Ouyang, Yizhu Jiao, Sizhe Zhou, Linyi Ding, and Jiawei Han. 2024. Automated Mining of Structured Knowledge from Text in the Era of Large Language Models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), August 25–29, 2024, Barcelona, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3637528.3671469

KDD '24, August 25–29, 2024, Barcelona, Spain

ACM ISBN 979-8-4007-0490-1/24/08
https://doi.org/10.1145/3637528.3671469

1 INTRODUCTION

The mission of KDD research is to develop new principles and methodologies for effective mining of knowledge from massive amounts of data. Since the majority of such data are in the form of unstructured text, mining structured knowledge from text becomes a core research problem in KDD. With the enormous volume and complex, context-sensitive semantics of text data, it is very costly to rely on human annotation for such mining. However, for a long time, developing automated (i.e., unsupervised or weakly supervised) approaches to accomplish this task remains a major challenge.

With the recent development of representation learning and large language models (LLMs), new and powerful methods have been or are being developed for effective and automated mining of structured knowledge from text. This tutorial provides a comprehensive overview of the recent advancements in this research frontier. We will summarize recent research on exploring the power of representation learning and large language models for automated mining of structured knowledge from massive corpora.

We will first introduce the key concepts, primitives, and recent developments of representation learning and large language models, which serves as the foundation for understanding recent research on transforming unstructured text into structured knowledge. Then we will introduce three frontiers along this line of research: (1) automated ontology construction and enrichment, (2) weakly-supervised text classification in flat and hierarchical label space, and (3) weakly-supervised information extraction, which extracts entity and relation structures and construct knowledge graphs automatically. Finally, we will discuss how such mined, structured knowledge may impact the applications of large language models in question answering, knowledge discovery, and trustworthiness analysis.

2 PRELIMINARY FOR LARGE LANGUAGE MODELS

This section delves into the foundational aspects of Large Language Models (LLMs), covering their architectural designs, training methodologies, and prompting strategies during inference.

2.1 Architectures

The architectures of LLMs can be broadly categorized into three types: encoder-only, encoder-decoder, and decoder-only models. Each type serves distinct purposes and is optimized for different kinds of tasks within the fields of text mining and NLP.

Encoder-only LMs. Encoder-only language models, such as BERT [15] and its variants (e.g., RoBERTa [56], DeBERTa [26]), are designed primarily for understanding tasks. They consist solely of an encoder network, processing input text to predict a class label or generate embeddings. These models have been instrumental in advancing natural language understanding [107, 108], serving as a backbone for tasks like text classification, sentiment analysis, and information extraction.

Enocder-Decoder LMs. Encoder-decoder models, including BART [48], and T5 [84], incorporate both encoder and decoder networks, enabling them to excel at generation tasks. These models first encode the input text into an intermediate representation, and the decoder then uses to generate output text. This architecture is especially effective for tasks requiring a deep understanding of context and the ability to produce coherent, extended text, such as translation and text summarization.

Decoder-only LMs. Decoder-only models, such as GPT [5, 81] and its successors (e.g., PaLM [18], Llama [19]), are optimized for generative tasks. These models predict the next token in a sequence based on the previous tokens, making them well-suited for tasks like text completion and open-ended dialogue. Their architecture allows for flexible application across a wide range of generative tasks, from simple text extension to complex content creation.

2.2 Training Paradigm

The training of LLMs unfolds in stages: initial pre-training to grasp language basics, followed by supervised fine-tuning for specific applications or desired output styles, and finally, aligning with human preferences and ethical standards.

Pre-training Tasks. The initial phase of pre-training is crucial for LLMs to grasp the basics of language [77, 81]. During this stage, models are exposed to vast amounts of text data, learning to predict accurately in various contexts without direct supervision. Key tasks include *Masked Language Modeling (MLM)*, notably employed by BERT [15], involves obscuring parts of the text for the model to predict, fostering a deep grasp of context. *Autoregressive Language Modeling*, used by GPT [81], trains the model to predict the next token in a sequence, honing its generative capabilities. *Denoising*, as seen in T5 [84] and BART [48], challenges models to correct intentionally introduced textual errors, enhancing their comprehension and correction skills. *Mixture of Experts* (MoE), such as those used in Switch Transformer [20], distribute tasks across specialized model components to efficiently scale model capacity.

Supervised Fine-tuning. Following pre-training, models undergo *supervised fine-tuning* to specialize in particular applications. This step adjusts the pre-trained parameters to optimize performance for specific text mining and NLP tasks, ranging from classification to generation. Fine-tuning can also involve *multi-task fine-tuning*, which improves the model's ability to generalize across different types of data. *Parameter-efficient fine-tuning* techniques such as Low-Rank Adaptation (LoRA) [28] are utilized, minimizing computational costs while maintaining the model performance.

Human Alignment. To align LLM outputs with human values, the refinement stage begins with *instruction tuning* [73, 90], a specialized form of supervised fine-tuning. Unlike general fine-tuning aimed at improving performance on specific tasks, instruction tuning enables LLMs to engage in dialogues with humans, complete tasks based on human prompts, and produce outputs in the desired style. This step is pivotal in transitioning models from text completion engines to interactive systems that can understand and execute complex instructions. Following instruction tuning, *Reinforcement Learning from Human Feedback (RLHF)* [73] tailors the models' behavior. Through RLHF, employing algorithms like Proximal Policy Optimization (PPO) [91] and Direct Preference Optimization (DPO) [83], models are trained based on human preferences, steering them toward generating responses that are ethical, relevant, and contextually appropriate, thereby ensuring their outputs are closely aligned with human expectations.

2.3 Prompting Strategy

A myriad of prompting techniques have been explored to tailor these models for specific tasks or to navigate complex scenarios effectively. These strategies exploit the intrinsic abilities and knowledge accumulated by LLMs, pushing the boundaries of what can be achieved through few-shot or zero-shot approaches.

In-Context Learning (ICL) [5] allows LLMs to adapt to new tasks by presenting them with a minimal set of examples within the prompt. This method utilizes the model's extensive pre-training on diverse text data, enabling it to generate responses tailored to the specifics of the task at hand without the need for additional training. The selection of in-context demonstrations plays a critical role in leveraging this capability, with strategies that focus on retrieving semantically similar examples to a given query showing notable improvements in model performance [70].

Multi-step Reasoning. Beyond ICL, the development of reasoning strategies has further expanded the capabilities of LLMs, enabling them to engage in more complex problem-solving processes. Strategies such as *Chain-of-Thought* [120], *Self-Consistency* [113] prompting guide the model through multi-step reasoning, allowing it to explore various pathways to arrive at an answer. These strategies not only improve the output accuracy but also enhance transparency and insight into the reasoning process. The progression from simple task execution to complex reasoning demonstrates the growing proficiency of LLMs in handling tasks that require a deeper level of understanding and cognitive engagement.

These advancements in prompting techniques represent the evolving interaction between human users and LLMs, showcasing how adaptability and improved task execution can be achieved without further training.

3 ONTOLOGY CONSTRUCTION AND ENRICHMENT

In the weakly-supervised text mining tasks, the form of weak supervision often appears as a few labeled samples, external knowledge, or just the label space with the textual label names or keywords. While the first two still provides some sample-label pairs, the last one does not provide any sample-label correlation. Therefore, the label-name-only weakly-supervised setting, or extremely weakly-supervised setting, requires the least amount of human supervision, while also being challenging as the model needs to fully understand the weak supervision signals before any training.

In this section, we will discuss techniques about the label-name-only weak supervision, from flat to hierarchical label space, including how to enrich the ontology with more class discriminative features (Sect. 3.1 Taxonomy Enrichment), how to construct such ontological structure from a text corpus (Sect. 3.2), and how to update/expand an existing structure (Sect. 3.3).

3.1 Ontology Enrichment

Ontology enrichment aims to enrich each node in the ontology with more discriminative textual features (e.g., keywords).

Enriching a flat structure. One fundamental task of ontology enrichment is to enrich a set of flat classes, which can be done with discriminative topic discovery methods. Unlike recent unsupervised topic discovery methods that directly cluster PLM embeddings [68, 151], discriminative topic discovery aims to find topic-specific key terms for a provided set of topics. For example, "athelets" and "football" are considered discriminative for the Sports topic when compared with Politics and Technology topics. CatE [63] trains a joint word embedding space which not only captures semantic similarity but also enforces topical words of different seed topics to separate in the embedding space. Therefore, the trained embedding space ensures discriminativeness between topics. It additionally assumes a topic-document-word generative process which learns both local and global context. KeyETM [25] extends embedded topic modeling [16] by incorporating user knowledge (topical keywords) as topic-level prior over the vocabulary. SeeTopic [146] specifically tackles the problem of potentially unseen topic names in the text corpus and proposes to utilize the general knowledge of PLMs to encode the out-of-vocabulary seeds. SeedTopicMine [149] studies multiple types of context information: seed-guided text embeddings, PLM-based contextualized embeddings, and topic-indicative sentences. The candidate terms are retrieved and ranked by each type of features and an ensemble ranking mechanism is used to identify the most confident ones according to all features. This process is repeated iteratively by adding enrichment words into each topic and refining the context features.

Enriching a hierarchical structure. Enriching a hierarchy additionally requires modeling structural information. TaxoGen [137] is an unsupervised approach which recursively clusters word embeddings and constructs local corpora for low-level nodes to refine word embeddings. NetTaxo [93] extends it by modeling network structure information associated with the text data. JoSH [69] takes a taxonomy as guidance and trains a joint embedding space that captures word semantic meaning with local context and embeds the structure by preserving relative tree distances between nodes.

3.2 Seed-Guided Taxonomy Construction

To further reduce human effort on curating a taxonomy, seed-guided taxonomy construction is studied to automatically construct a hierarchical structure from a text corpus by taking only a small set of seeds.

Set Expansion. Set expansion is a subtask of taxonomy construction by considering only a flat structure. This task aims to expand a set of seed entities (e.g., United States, China, and Spain) with more entities belonging to the same semantic class (e.g., more countries like Canada and United Kingdom). SetExpan [96] uses the skip-grams and word embedding features from a text corpus to evaluate the similarity of candidate entities with the seed set. It iteratively bootstraps the entity set by find new entities and features in each iteration, and a rand ensemble mechanism is applied to reduce the effect of noisy features. Set-CoExpan [32] additionally generates auxiliary sets by embedding learning and clustering, which are semantically similar to the target entity set. Multiple sets are expanded simultaneously to ensure the expansion quality of the target entity set. CGExpan [147] introduces the text representation power of PLMs by automatically constructing knowledge probing queries. The class-probing queries are used to generate a textual class name for the target set, which is then used to construct entity-probing queries to iteratively expand the seed set. ProbExpan [54] proposes to first refine entity representations with contrastive learning and heuristic-based hard negative selection. Then, the refined representations are used in a probabilistic expansion process with window search and entity re-ranking. FGExpan [122] studies the fine-grained entity set expansion task which aims to expand the seed set according to the finest possible common type. Three scores are combined to infer the finest type, including entity generation score by MLM, type generation score guided by a type taxonomy, and textual entailment score.

Seed-Guided Taxonomy Construction. Given a small seed taxonomy, this task aims to expand it to a more complete taxonomy structure by mining from a text corpus. HiExpan [97] expands an entity taxonomy by decomposing the process into width expansion and depth expansion, followed by a global structure adjustment step. The width expansion is done with an entity set expansion method, and the depth expansion is achieved with an embedding-based method which captures relation using word analogy [80]. CoRel [33] trains a relation transferring module using PLMs to learn the seed parent-child relations, which is applied along multiple paths to expand the seed taxonomy in width and depth. TaxoCom [47] completes a partial topical taxonomy by first learning local discriminative word embeddings and then applying novelty adaptive clustering of embeddings to find novel subtopics.

3.3 Taxonomy Expansion

The taxonomy expansion task assumes an existing taxonomy structure is provided and aims to expand it by inserting new nodes into the taxonomy. TaxoExpan [95] proposes to encode the structural information with local egonets around anchor nodes and a position enhanced graph neural network. The query-anchor matching model is then trained using self-supervision automatically derived from the provided taxonomy and a contrastive loss. STEAM [134] instead samples mini-paths from the taxonomy as anchors and learns to

insert a query node into the paths. Three types of features are considered to capture anchor-query relation, including distributional similarity, contextual information, and syntactic patterns to train a model with multi-view co-training. TEMP [57] fine-tunes a PLM to compare positive and negative paths using a margin ranking loss with tree-distance based dynamic margins. To deal with long-tail entities that cannot be easily extracted from a text corpus, GenTaxo [135] identifies positions in the existing taxonomy that miss an entity and then use a generative model to directly generate new concept. It pretrains a concept name generator with graph-based and relation-based contextual embeddings. TMN [138] proposes to not only find hypernym but also hyponym for a query entity. A triplet matching network is trained to make holistic predictions on (hypernym, query, hyponym) triplets by considering multiple fine-grained signals. QEN [110] proposes a Quadruple Evaluation Network, which utilizes term descriptions as input and considers not only parent-child relations but also sibling relations. TaxoEnrich [35] learns taxonomy-contextualized embedding incorporating both semantic meanings and taxonomic relations and trains two encoders to capture structural information in both vertical and horizontal views. TaxoPrompt [127] utilizes prompt tuning of PLMs to learn taxonomic relations and utilizes random walk algorithm to generate self-supervision data that better capture the global structural information. TaxoComplete [2] trains semantic matching network with self-supervision data consisting both close neighbors and distant neighbors. To better learn the hypernymy relations, it also injects the edge directions into node representations using a direction-aware population module. BoxTaxo [38] proposes to learn box embeddings [104] which have "contained", "intersection", and "disjoint" relations. The box embeddings are learned in a joint view of geometry and probability and are used to decide if a query entity is contained in an anchor node during inference. TaxoInstruct [98] proposes a unified instruction tuning method for the entity set expansion and taxonomy expansion tasks and reformulate the seed-guided taxonomy construction task as a combination of them. It fine-tunes an LLM using self-supervision from an external large taxonomy by constructing task-specific instructions.

4 WEAKLY-SUPERVISED TEXT CLASSIFICATION

One important task for mining structured knowledge from massive unstructured text is to classify text into different categories. To reduce the cost of human annotation and requirement of domain expertise, the weakly-supervised text classification setting is proposed which uses the label name or a small number of examples of each target class as the only supervision signal to train the text classifier. In the section, we will introduce recent studies on weakly-supervised text classification for both flat label space and hierarchical label space.

4.1 Weakly-Supervised Flat Text Classification

Earlier studies train text classifiers in a fully supervised way with substantial amount of training data [130, 142], which is expensive to obtain and hard to scale. Later, the semi-supervised setting is studied to train classifiers with a smaller amount of training samples per class and an unlabeled corpus [10, 124]. However, they still need at least dozens of labels for each target class, which requires domain knowledge and can still be costly if the class distribution is highly imbalanced. To further reduce the requirements of human efforts, the weakly-supervised text classification setting is proposed. Such supervision signals include distant supervision from knowledge bases [99], human-curated rules [3, 6, 85], or a list of keywords [62, 64, 86]. Among these settings, the extremely weakly supervised text classification requires the least amount of supervision signal, which can train the text classifier using the sole class surface name of each class as the only supervision. This line of studies can be classified into keyword-based methods and prompt-based methods.

Keyword-based methods. WeSTClass [64] first models each class as a distribution in an embedding space, and then sample words from the class distribution to generate pseudo documents for each class. Then pseudo documents are used as supervision to train a text classifier, followed by self-training with soft labeling [125] to iteratively enhance the classifier performance with its own predictions. LOTClass [67] uses masked language modeling (BERT) to find replacement tokens for each occurrence of seed word, which are aggregated into a class vocabulary. Then MLM is used to find replacements for each token in to corpus to find those "class-indicative" tokens that match with any of the class vocabulary. Finally, a PLM-based classifier is first fine-tuned with masked topic prediction objective, followed by self-training similar to WeSTClass. X-Class [116] proposes to use contextualized embeddings to first expand label names with more keywords, which are used to estimate class representations and class-oriented document representations. Pseudo labels are assigned based on representation similarity, which are then used to fine-tune a classifier. ClassKG [140] builds keyword graph to learn subgraph annotators for pseudo labeling. It is an iterative framework by using the classifier predictions to extract more keywords to enrich the weak supervision, and repeats the process until converge. Dong et al. [17] claims the performance of keyword-based method is limited with the bias introduced in the matching process and proposes random deletion in the training process to debias pseudo data. MEGClass [43] further studies the contribution of different text granularities by learning contextualized sentence representations, which are then used to calculate document representations with an attention mechanism. A feedback method is also introduced to refine the representations iteratively.

Prompt-based methods. Because keyword-based features can only generate pseudo labels with limited quality given that their meaning are highly dependent on their contexts, the prompt-based methods are also proposed to acquire pseudo labeled documents by exploiting the contextualized power of PLMs. NPPrompt [152] is a zero-shot method that first construct a set of verbalizers for each class using PLM embeddings, and the embedding similarity is used as weight of the retrieved verbalizer for MLM-based prompting. PIEClass [144] further studies prompting methods of discriminative PLMs for pseudo labeling. It also proposes an iterative ensemble training method that combines two different PLM fine-tuning methods, namely head-token fine-tuning and prompt-based fine-tuning, that complement each other to iteratively expand the pseudo labels while ensure the quality. PIEClass is the first weakly-supervised approach to achieve comparable performance to a fully-supervised baseline on the sentiment classification task. CARP [102] studies

zero-shot and few-shot text classification with LLMs' reasoning ability. It follows the chain-of-thought prompting method by first asking the LLM to identify indicative clues within the input text before predicting its label.

Besides, LOPS [61] shows that selecting pseudo labels in the correct order can improve the performance and proposes to use learning-based confidence scores to decide the order. FuTex [145] studies the weakly-supervised classification of scientific papers by incorporating in-paper structure (i.e., sections, paragraphs) and cross-paper structure (i.e., paper citation network). Wang et al. [118] introduces the first benchmark of the weakly-supervised text classification task, which consists of 11 datasets from 4 different domains with standardized train-test splits.

4.2 Weakly-Supervised Hierarchical Text Classification

Given a label space structured as a taxonomy, the hierarchical text classification task aims to classify input text into a path or multiple nodes on the label taxonomy. Compared to flat text classification where the label space is typically small (e.g., with less than 20 classes), the hierarchical text classification task is more challenging because of its large and structured label space.

Most studies tackle the hierarchical text classification task in the fully-supervised [39, 115] or semi-supervised settings [23, 121]. Previous studies can be classified into local approaches and global approaches. The local approaches train multiple text classifiers for each node or level of the label taxonomy and make the final predictions recursively [4, 119]. The global approaches propose to learn the global structure with a single text classifier [8, 39, 75, 115, 117]. These methods normally requires a substantial amount of annotated training data and domain expertise

The weakly-supervised hierarchical text classification task is also studied to save annotation efforts. WeSHClass [65] extends WeSTClass by first modeling the label hierarchy with a mixture of distribution in an embedding space and sampling vectors from the distributions as input to a pre-trained LSTM model to generate pseudo documents. Then, it utilizes the pseudo data to train local text classifier for each internal node and then trains a global classifier by ensembling local classifiers with soft-labeling self-training. HiMeCat [143] additionally studies the metadata acompanied with the text data by learning a joint representation space for label hierarchy, metadata, and text data with a hierarchical generative model. The learned distributions are then used to generate augmented documents to enrich the weak supervision signals, and a text classifier is trained recursively for each internal node. TaxoClass [94] uses a pre-trained textual entailment model to estimate document-class similarity, based on which a taxonomy-based top-down search method is proposed to obtain the confident core classes for each document. Here, a document's core classes are defined as the set of classes that most accurately describe the document. They are then used to construct pseudo training samples to train a multi-label text matching network. TELEClass [148] proposes to enrich the raw label taxonomy with class-indicative features to help better class understanding. Additionally, it tailors LLMs for the hierarchical label space. An LLM is used to select pseudo labels for each document from a set of candidate classes retrieved from the label hierarchy. To deal with long-tail and fine-grained classes in the taxonomy, the LLM is also prompted to generate pseudo documents conditioned on paths sampled from the taxonomy.

5 WEAKLY-SUPERVISED INFORMATION EXTRACTION

Mining structure for entities is another important task for text mining. In this section, we will introduce recent weakly-supervised methods on entity recognition and typing, relation extraction, and comprehensive knoelwdge structuring.

5.1 Entity Mining

Mining of structured knowledge at an entity level aims to extract and identify the types of entities within their respective contexts. Being an elementary building block for texts, it could be integrated into more advanced structures such as knowledge graphs (KG). The task of entity-level text mining could be categorized as named entity recognition (NER) and fine-grained entity typing (FET).

5.1.1 Named Entity Recognition. NER is a typical sequence labeling task that assigns an entity label to each token in the sequence. There are usually less than 10 labels for NER datasets containing coarse-grained tags such as "location" and "person". In this section, we focus on the existing NER frameworks with weak supervision [30], which could be divided into the following three categories based on how they deal with noisy distant supervision.

Incorporating Distant Label Uncertainty. One solution is to introduce uncertainty expressions to distantly supervised labels, e.g., dictionary-matching results. AutoNER [92] explores learning NER model using only dictionaries. It transcends the conventional sequence labeling framework by introducing the "Tie or Break" tagging strategy. This innovative method enhances the model's ability to utilize noisy distance supervision effectively by determining whether adjacent tokens belong to the same entity or should be separated. Although AutoNER achieves performance gains, it still utilizes limited information from incomplete dictionaries. PaTNER [111] was further proposed to automatically mine the entity naming principles to automatically expand the input dictionaries. PaTNER was particularly useful when it comes to domain-specific NER tasks such as biomedical or technical domains. ETAL [7] further designs a method with pseudo-labeling to search for highly confident entities that maximize the probability of BIO sequences.

Noise-Tolerant Tuning. The semi-supervised scheme [125] has proved to be an effective method for effectively leveraging unlabeled data with limited labeled data as the distant supervision. BOND [55] leverages the power of PLMs, specifically RoBERTa, to improve the performance of NER. It is a two-stage algorithm where the model is firstly trained on distantly-labeled data with early stopping, and then a teacher-student framework is employed to iteratively self-train the model. RoSTER [66] aims to exclude the influence of incomplete and noisy labels. It first proposes a noise-robust learning scheme with a new loss function and a noisy label removal step for training NER in a distantly supervised manner. Then a self-training method was created by leveraging PLMs to further enhance the contextualized generalization ability. Instead of focusing on the representations of entities, X-NER [76] further investigates

the contextualized information of entities being replaced in the sentence. The top-ranked entity spans are then treated as pseudo-labels to train a NER tagger.

Leverating External Knowledge. ChemNER [112] provides the first fine-grained chemistry NER dataset with 65 types. Building upon the dataset, ChemNER designs a flexible KB-Matching for domain-specific entities and then uses the ontology as the guidance for multi-type disambiguation to train a sequence labeling model. SpanNER [21] separates span detection and type prediction, using external class descriptions to construct class representations for matching detected spans, though its model designs differ from the backbone pre-trained model BERT. Similarly, SDNET [9] pre-trains a T5 model on silver entities from Wikipedia, to help universally describe mentions using concepts and map novel entity types to concepts. Then the model is fine-tuned on few-shot examples to adaptively recognize entities on-demand. SEE-Few [131] expands seeded entities using external tools and applies an entailment framework to efficiently learn from a few examples. More recently, LLMs have been investigated in the field of NER with its immense parametric knowledge store. However, it is shown that LLMs perform poorly on this sequence labeling task in zero-shot settings [79]. To this end, GPT-NER [109] further explores how to leverage LLMs for better NER systems, by bridging the gap between sequence labeling and text generation. By instructing LLMs using in-context learning techniques to generate special tokens for entity recognition, GPT-NER largely boosts LLMs' ability in NER tasks.

5.1.2 Fine-grained Entity Typing. Compared with NER, FET focuses on a much more fine-grained classification of a given entity. The label space of FET usually entails dozens and even hundreds of entity types, which is organized as a hierarchical ontology structure (Sect. 3). Due to the large label space and accurate annotation requirement, weakly-supervised frameworks are invented to deal with the data scarcity issue, which could be briefly divided into the following categories.

Ontology-Guided Methods. Since the FET usually entails a large and structured ontology as the label space, how to fully *interpret* and *leverage* the structure/hierarchy of the ontology can play a decisive role in FET. AFET [87] proposes a method for embedding both clean and noisy entity references individually. The technique leverages a defined type hierarchy to formulate loss functions and integrates them into a unified optimization problem to calculate the embeddings of references and type paths. OntoType [45] is an ontology-guided framework that leverages the weak supervision of pre-trained language models and headwords, which are further used to match the fine-grained types to type ontology. ALIGNIE [31] is a prompt-based method that consists of two modules. One for entity type interpretation that learns to relate entity types with vocabulary using the ontology. Then a type-based instance generator is designed to enrich the few-shot training samples. OnEFET [74] proposes to enrich the original ontology structure with instances and topics. The instances are used for pseudo-training data generation, while the topics are integrated into the attention mechanism to better discriminate fine-grained entity types. The pseudo-training data are used to train an entailment model, which is used iteratively in a top-down style for inference. SEType [150] further works in the field of technology. Different from OnEFET, SEType first enriches the weak supervision by finding more entities for each seen type from an unlabeled corpus using the contextualized representations of pre-trained language models. It then matches the enriched entities to unlabeled text to get pseudo-labeled samples and trains a textual entailment model that can make inferences for both seen and unseen types.

Knowledge-Based Methods. It is found that lack of world knowledge is one of the major disadvantages of existing methods. Therefore, many works are dedicated to integrating external knowledge to enrich the understanding of entity types. Notably, UFET [13] predicts open types without a pre-defined label structure and is trained using a multi-objective approach that combines supervision from the headwords and prior information from entity linking in Wikipedia. ZOE [153] uses a new type taxonomy defined as Boolean functions of Freebase types and determines the type of a given entity reference by linking it to the type-compatible Wikipedia entries. Recently, there are also works that investigate how LLMs perform on the FET task. However, due to the large label space, it is often difficult for LLMs to strictly follow the instructions and predict entity types in the label space [74]. Also, LLMs have problems interpreting the nuanced contextualized information in the input.

Ultra-Fine-Grained FET. There is another line of FET research that further expands the original label space with tens of thousands of types [13]. It is often infeasible to walk over all the types and give predictions using the entailment method as mentioned before. Targeting this setting, BERT-MLMET [14] invents a model that starts with BERT-base and fine-tunes it using supervision from headwords and entity-type hypernyms extracted from Hearst patterns. The resulting model is used to predict ultra-fine entity types and produce fine-grained entity types by means of a simple type mapping process. LITE [49] borrows indirect supervision from NLI to perform entity typing. It also involves a type-ranking module to help with generalizing prediction with disjoint type sets. Denoise-FET [52] to first cluster the large label space into several centroids with embeddings, after which the clusters are treated as additional domains for typing.

5.2 Relation Extraction

Built on top of classified texts and potentially extracted entity structures, relation extraction (RE) aims to identify and classify semantic relationships between entities. To empower model's understanding on structured relational patterns given scarce expert supervision, the weakly-supervised setting has also been extensively adopted which grasps different dimensions of the nuggets of relation.

5.2.1 Relation Instance Synthesis. With the emergence of powerful generative language models, synthesizing relation instances to alleviate the scarcity of high-quality data becomes a more and more promising direction for relation extraction [42]. LLMs, such as the GPT family and LlaMA family [19], are pre-trained for the domain adaptation ability [82]. They have demonstrated to contain factual relation knowledge [78] and follow-up evaluation studies have shown that LLMs are relatively skilled at constrained content generation, story telling, and rationale generation [100].

RelationPrompt [12] trains two sequence-to-sequence models with one serving as the data generator conditioned on the relation label names while the other serving as the relation triplet(s)

extractor conditioned on the input texts. Apart from vanilla generation methods, recent synthesis-based RE works mainly focus on reducing the hallucinations and noise of generated instances. STAR [59] applies a self-refinement by self-reflection approach to verify the synthesized instance with LLMs. DocGNRE [51] leverages a *generate-then-validate* paradigm where GPT serves to generate candidate relation triplets given context and entity list while an NLI module serves to filter them to augment the original dataset. GenRDK [101] takes the names of similar relation groups to generate relation instances step by step. To mitigate the noisy labels, it trains a pre-denoising relation extractor which produces pseudo labels and applies consistency scores for filtering based on cross-document knowledge graph level statistics.

Instead of synthesizing based on relation names, REPaL [155] shows that relation descriptions or definitions offer a more complete coverage of multifaceted relational semantics which involve entity-entity interactions and entity-related specifications. In terms of synthesis methodology, REPaL emphasizes both correctness of synthesized instances and the internal synergy among synthesized instances. Specifically, REPaL adopts a multi-turn generation approach conditioned on the relation definition while incorporating feedback. The feedback is constructed by (1) sampling inference results on the unlabeled corpus with a small language model trained on the synthesized data and (2) examining the generation history. Therefore, the feedback conveys both the bias of downstream fine-tuned relation extractor and the patterns of generated instance.

5.2.2 Relational Reasoning. Relational reasoning is the cornerstone of the relation extraction task as it aims to enhance models' comprehension and reasoning capabilities to derive target relations.

One line of research resort to boosting the models' relational reasoning ability by formulating the relation extraction task into different task formulations that models are more capable at. Obamuyide and Vlachos [72] and Sainz et al. [89] convert RE into the NLI task. REBEL [34] converts the relation triplet extraction process into Seq2seq generation process leveraging the power of decoder-based LMs. SURE [58] verbalizes relation names with templates and formulates RE as the summarization task. There are also various prompt tuning models for RE [11, 24] that incorporate different relation prompts to distill PLM's relational knowledge for better reasoning performance. SumAsk [50], based on LLMs, adopts a *summarize-and-ask* prompting paradigm to formulate relation inference as summarization and formulate the validation process as question answering.

Another line of research tries to aggregate relation indicative evidence to strengthen the reasoning process. REPEL [80] proposes to co-train a dependency path based pattern module with the distributional module for relation extraction. RClus [154] leverages the patterns of relation-specific entity types and relation indicative words in the dependency path to form relation representations. Eider [126], a document-level RE method, tries to extract sentence-level evidence by learning a classifier. SAIS [123] supervises and augments intermediate steps of extracting relation clues which include: coreference resolution of contextual roles, entity typing, pooled evidence retrieval to distinguish entity-pairs with and without supporting sentences, and fine-grained evidence retrieval for more interpretable and relation-specific evidence sentence.

The fast development of powerful LLMs have further pushed the progress of relation reasoning forward, especially by tuning-free reasoning methodologies like in-context learning. Wadhwa et al. [105] demonstrates the promising potential of few-shot prompting of GPT-3 which matches state-of-the-art fully supervised RE models and it also explores a LLM-based data augmentation technique to inject CoT style explanations for fine-tuning. GPT-RE [106] extends the in-context learning approach by leveraging task-aware sentence-level and entity-level representations to conduct kNN retrieval of demonstration examples. The demonstrations are further enriched with gold label-induced explanations to enhance LLM's relational reasoning power.

In parallel, some RE works have noticed LLM's relative insufficiency in relational reasoning which might be related to the low incidence of RE in instruction tuning datasets [114, 139] and hence resort to circumvent this. Ma et al. [60] demonstrates that LLM is better rerankers for information extraction and further propose the LLM-SLM cooperation framework named as *filter-then-rerank* paradigm. Under the *filter-then-rerank* paradigm, SLM contributes to the initial attempt of extracting relations and LLM will rank low-confidence candidates extracted by SLM to yield the final extraction. QA4RE [139] introduces to formulate RE as question answering which aligns RE with more common instruction-tuning tasks and SumAsk [50] adopts the formulation of summarization and QA.

There are other recent works not covered due to page limitations including but not limited to GeoWISE [29] for geospatial topological RE and GenRES [37] for open RE evaluation.

5.3 Comprehensive Knowledge Structuring

Structured knowledge, especially knowledge graphs and databases, have long been the subject of study in knowledge structuring and grounding.

5.3.1 Knowledge Graph Construction. Knowledge graphs (KGs) are collections of relations between real-world entities. Based on the scope of knowledge, the KGs can be categorized into general KGs like Wikidata[1] and domain-specific KGs like UMLS[2]. Conventionally, construction of KGs consists of entity mining (Sect. 5.1), relation extraction (Sect. 5.2), and comprehensive KG construction.

Specifically, OIE4KGC [71] applies an open information extraction tool to take the raw documents as input and extract the triples within this documents. Then triples are then filtered, linked and merged to generate a knowledge graph. To prevent the error propagation of pipelines, generative KG construction powered by language models are proposed. Zeng et al. [136] constructs KGs in an end-to-end manner based on sequence-to-sequence model where entities and relations can be jointly extracted. The model can solve the overlap problem of triples through the copy mechanism. REBEL [34] and ABSA [129] are also generative frameworks that can formulate the triple extraction task as a sequence-to-sequence task. They translate raw text to structured knowledge schema based on pre-trained language models. TAGREAL [36] exploits the implicit knowledge of PLM for KG completion. TAGREAL automatically generates high-quality query prompts by pattern mining methods and retrieves support information to probe the knowledge in PLM.

[1] https://www.wikidata.org/wiki/Wikidata:Main_Page
[2] https://www.nlm.nih.gov/research/umls/index.html

LLMs also have shown the power in KG construction task. KnowledgeGraph GPT [103] directly utilizes prompting to convert plain texts to KG with the power of GPT-4. AutoKG [156] adopts a multi-agent-based approach employing LLMs' inner knowledge for KG construction and reasoning.

Domain-specific KG construction usually requires more human effort on annotations or ontology construction. Rotmensch et al. [88] construct a health KG based on the annotated electronic medical records and the concepts and relations in existing medical KGs. Recent advancements leverage LLMs to facilitate the domain-specific KG construction. PAIR [22] constructs a KG for online marketing, which uses LLM as a relation filter to reduce the search space of pre-defined relation set. PAIR also introduces a progressive prompting augmentation for entity expansion. The prior domain knowledge is injected into LLMs with prompt engineering. In the biomedical domain, Karim et al. [44] first constructs a an ontology for validating gene-disease relations and then proposes BioBERT to create RDF triples. Finally, they applies LLMs to revise inconsistencies or incompleteness in KG. In the e-commerce domain, FolkScope [133] proposes a semi-automatic approach for KG constrcution. FolkScope designs domain-specific prompts to leverages LLMs to probe the intention of user behaviours, which can be aligned to the pre-defined relations.

5.3.2 Database Population. Some recent works focus on the efficient extraction, structuring, and integration of information into tables or databases. These works bridge the gap between unstructured data and structured database requirements. Each of the discussed methods offers a unique perspective on addressing the challenges on schema identification, information extraction, and integration.

Specifically, AVATAR [46] proposes the use of probabilistic database techniques as the formal underpinnings of information extraction systems. RolePred [40] utilizes information redundancy in multiple documents to extract structured tables including the attribute names and the corresponding values of an event type. ODIE [41] adopts instruction tuning for large language models to extract a single table from the texts following user instructions. Specially, the table structures can be either defined by users or inferred by the model automatically. TableLLAMA [141] and TableGPT [53] also fine-tuning large language models with table-related data to improve their ability of processing tables, like row population, column addition, and cell filling. EVAPORATE [1] develops a prototype LLMs-powered system that ingests semi-structured documents and outputs a tabular, structured view of the documents. This system can identify the schema and perform extraction to populate the table. ChatDB [27] builds a symbolic memory framework instantiated as an LLM and a set of SQL databases, where the LLM generates SQL instructions to manipulate the SQL databases, including the read and write operations. DB-GPT [128] utilizes large language models to understand natural language queries and generate complex SQL queries to ingest, structure, and access data with privatization technologies. Also, it uses an adaptive learning mechanism to continuously improve the system based on user feedback. Text2DB [132] emphasizes the integration of information extraction output and the target database (or knowledge base). Given a user instruction, a document set, and a database, it aims to update the database with values from the document set to satisfy the user instruction. To handle this task, it propose an multi LLM-agent framework which calls for existing information extraction tools to populate the database automatically.

6 CONCLUSION

In this tutorial, we have presented a timely overview on the theme of mining structured knowledge from text by exploring various methods developed in representation learning and large language model research. It is convincing that LLMs have substantially enhanced the power and effectiveness of mining massive text data. At the end of the tutorial, we will present not only important research topics on mining structured knowledge from text but also discuss how such mined, structured knowledge may impact the performance of large language models. In particular, we will show how extracted knowledge structures may guide quality question answering, knowledge discovery, and trustworthiness analysis by LLMs. For example, mined knowledge structures from a theme-specific corpus, together with the corpus data, may endow a large language model additional power to conduct in-depth reasoning on specific themes, with less hallucination and better explainability.

ACKNOWLEDGMENTS

Research was supported in part by US DARPA INCAS Program No. HR001121C0165 and KAIROS Program No. FA8750-19-2-1004, National Science Foundation IIS-19-56151, and the Molecule Maker Lab Institute: An AI Research Institutes program supported by NSF under Award No. 2019897, and the Institute for Geospatial Understanding through an Integrative Discovery Environment (I-GUIDE) by NSF under Award No. 2118329. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily represent the views, either expressed or implied, of DARPA or the U.S. Government.

REFERENCES

[1] Simran Arora, Brandon Yang, Sabri Eyuboglu, Avanika Narayan, Andrew Hojel, Immanuel Trummer, and Christopher Ré. 2023. Language Models Enable Simple Systems for Generating Structured Views of Heterogeneous Data Lakes. *Proceedings of the VLDB Endowment* 17, 2 (2023), 92–105.

[2] Ines Arous, Ljiljana Dolamic, and Philippe Cudré-Mauroux. 2023. TaxoComplete: Self-Supervised Taxonomy Completion Leveraging Position-Enhanced Semantic Matching. In *WWW*. 2509–2518.

[3] Sonia Badene, Kate Thompson, Jean-Pierre Lorré, and Nicholas Asher. 2019. Data Programming for Learning Discourse Structure. In *ACL*.

[4] Siddhartha Banerjee, Cem Akkaya, Francisco Perez-Sorrosal, and Kostas Tsioutsiouliklis. 2019. Hierarchical Transfer Learning for Multi-label Text Classification. In *ACL*.

[5] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *NeurIPS*.

[6] Oishik Chatterjee, Ganesh Ramakrishnan, and Sunita Sarawagi. 2018. Data Programming using Continuous and Quality-Guided Labeling Functions. In *AAAI*.

[7] Aditi Chaudhary, Jiateng Xie, Zaid Sheikh, Graham Neubig, and Jaime Carbonell. 2019. A Little Annotation does a Lot of Good: A Study in Bootstrapping Low-resource Named Entity Recognizers. In *EMNLP-IJCNLP*. 5164–5174.

[8] Haibin Chen, Qianli Ma, Zhenxi Lin, and Jiangyue Yan. 2021. Hierarchy-aware Label Semantics Matching Network for Hierarchical Text Classification. In *ACL-IJCNLP*.

[9] Jiawei Chen, Qing Liu, Hongyu Lin, Xianpei Han, and Le Sun. 2022. Few-shot Named Entity Recognition with Self-describing Networks. In *ACL*. 5711–5722.

[10] Jiaao Chen, Zichao Yang, and Diyi Yang. 2020. MixText: Linguistically-Informed Interpolation of Hidden Space for Semi-Supervised Text Classification. In *ACL*.

[11] 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 *WWW*. 2778–2788.

[12] Yew Ken Chia, Lidong Bing, Soujanya Poria, and Luo Si. 2022. RelationPrompt: Leveraging prompts to generate synthetic data for zero-shot relation triplet extraction. *arXiv preprint arXiv:2203.09101* (2022).

[13] Eunsol Choi, Omer Levy, Yejin Choi, and Luke Zettlemoyer. 2018. Ultra-Fine Entity Typing. In *ACL*. 87–96.

[14] Hongliang Dai, Yangqiu Song, and Haixun Wang. 2021. Ultra-Fine Entity Typing with Weak Supervision from a Masked Language Model. In *ACL'21*. 1790–1799.

[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*. 4171–4186.

[16] Adji B. Dieng, Francisco J. R. Ruiz, and David M. Blei. 2020. Topic Modeling in Embedding Spaces. *TACL* (2020).

[17] Chengyu Dong, Zihan Wang, and Jingbo Shang. 2023. Debiasing Made State-of-the-art: Revisiting the Simple Seed-based Weak Supervision for Text Classification. In *EMNLP*. 483–493.

[18] Aakanksha Chowdhery et al. 2023. PaLM: Scaling Language Modeling with Pathways. *J. Mach. Learn. Res.* 24 (2023), 240:1–240:113.

[19] Hugo Touvron et al. 2023. LLaMA: Open and Efficient Foundation Language Models. *arXiv preprint arXiv:2302.13971* (2023).

[20] William Fedus, Barret Zoph, and Noam Shazeer. 2022. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. *J. Mach. Learn. Res.* 23 (2022), 120:1–120:39.

[21] Jinlan Fu, Xuanjing Huang, and Pengfei Liu. 2021. SpanNER: Named Entity Re-/Recognition as Span Prediction. In *ACL-IJCNLP*. 7183–7195.

[22] 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* (2023).

[23] Suchin Gururangan, Tam Dang, Dallas Card, and Noah A. Smith. 2019. Variational Pretraining for Semi-supervised Text Classification. In *ACL*.

[24] Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. 2022. Ptr: Prompt tuning with rules for text classification. *AI Open* 3 (2022), 182–192.

[25] Bahareh Harandizadeh, J. Hunter Priniski, and Fred Morstatter. 2022. Keyword Assisted Embedded Topic Model. In *WSDM'22*.

[26] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. Deberta: decoding-Enhanced Bert with Disentangled Attention. In *ICLR*.

[27] Chenxu Hu, Jie Fu, Chenzhuang Du, Simian Luo, Junbo Zhao, and Hang Zhao. 2023. Chatdb: Augmenting llms with databases as their symbolic memory. *arXiv preprint arXiv:2306.03901* (2023).

[28] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *ICLR*.

[29] 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 *SDM*.

[30] Jiaxin Huang, Chunyuan Li, Krishan Subudhi, Damien Jose, Shobana Balakrishnan, Weizhu Chen, Baolin Peng, Jianfeng Gao, and Jiawei Han. 2021. Few-Shot Named Entity Recognition: An Empirical Baseline Study. In *EMNLP*. 10408–10423.

[31] Jiaxin Huang, Yu Meng, and Jiawei Han. 2022. Few-Shot Fine-Grained Entity Typing with Automatic Label Interpretation and Instance Generation. In *KDD*. 605–614.

[32] Jiaxin Huang, Yiqing Xie, Yu Meng, Jiaming Shen, Yunyi Zhang, and Jiawei Han. 2020. Guiding Corpus-based Set Expansion by Auxiliary Sets Generation and Co-Expansion. In *WWW*. 2188–2198.

[33] Jiaxin Huang, Yiqing Xie, Yu Meng, Yunyi Zhang, and Jiawei Han. 2020. CoRel: Seed-Guided Topical Taxonomy Construction by Concept Learning and Relation Transferring. In *KDD'20*.

[34] Pere-Lluís Huguet Cabot and Roberto Navigli. 2021. REBEL: Relation Extraction By End-to-end Language generation. In *Findings of EMNLP*. 2370–2381.

[35] Minhao Jiang, Xiangchen Song, Jieyu Zhang, and Jiawei Han. 2022. TaxoEnrich: Self-Supervised Taxonomy Completion via Structure-Semantic Representations. In *WWW*. 925–934.

[36] 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 ACL*. 11161–11180.

[37] Pengcheng Jiang, Jiacheng Lin, Zifeng Wang, Jimeng Sun, and Jiawei Han. 2024. GenRES: Rethinking Evaluation for Generative Relation Extraction in the Era of Large Language Models. *arXiv preprint arXiv:2402.10744* (2024).

[38] Song Jiang, Qiyue Yao, Qifan Wang, and Yizhou Sun. 2023. A Single Vector Is Not Enough: Taxonomy Expansion via Box Embeddings. In *WWW*. 2467–2476.

[39] Ting Jiang, Deqing Wang, Leilei Sun, Zhongzhi Chen, Fuzhen Zhuang, and Qinghong Yang. 2022. Exploiting Global and Local Hierarchies for Hierarchical Text Classification. In *EMNLP*. 4030–4039.

[40] Yizhu Jiao, Sha Li, Yiqing Xie, Ming Zhong, Heng Ji, and Jiawei Han. 2022. Open-Vocabulary Argument Role Prediction For Event Extraction. In *Findings of EMNLP*. 5404–5418.

[41] Yizhu Jiao, Ming Zhong, Sha Li, Ruining Zhao, Siru Ouyang, Heng Ji, and Jiawei Han. 2023. Instruct and Extract: Instruction Tuning for On-Demand Information Extraction. In *EMNLP*. 10030–10051.

[42] Martin Josifoski, Marija Sakota, Maxime Peyrard, and Robert West. 2023. Exploiting asymmetry for synthetic training data generation: Synthie and the case of information extraction. *arXiv preprint arXiv:2303.04132* (2023).

[43] Priyanka Kargupta, Tanay Komarlu, Susik Yoon, Xuan Wang, and Jiawei Han. 2023. MEGClass: Extremely Weakly Supervised Text Classification via Mutually-Enhancing Text Granularities. In *Findings of EMNLP*. 10543–10558.

[44] 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* (2023).

[45] Tanay Komarlu, Minhao Jiang, Xuan Wang, and Jiawei Han. 2023. OntoType: Ontology-Guided Zero-Shot Fine-Grained Entity Typing with Weak Supervision from Pre-Trained Language Models. *arXiv preprint arXiv:2305.12307* (2023).

[46] Rajasekar Krishnamurthy, Sriram Raghavan, Shivakumar Vaithyanathan, and Huaiyu Zhu. 2006. Avatar information extraction system. (2006).

[47] Dongha Lee, Jiaming Shen, SeongKu Kang, Susik Yoon, Jiawei Han, and Hwanjo Yu. 2022. TaxoCom: Topic Taxonomy Completion with Hierarchical Discovery of Novel Topic Clusters. In *WWW'22*.

[48] 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 *ACL*. 7871–7880.

[49] Bangzheng Li, Wenpeng Yin, and Muhao Chen. 2022. Ultra-fine entity typing with indirect supervision from natural language inference. *TACL* 10 (2022), 607–622.

[50] Guozheng Li, Peng Wang, and Wenjun Ke. 2023. Revisiting large language models as zero-shot relation extractors. *arXiv preprint arXiv:2310.05028* (2023).

[51] Junpeng Li, Zixia Jia, and Zilong Zheng. 2023. Semi-automatic Data Enhancement for Document-Level Relation Extraction with Distant Supervision from Large Language Models. In *EMNLP*. 5495–5505.

[52] Na Li, Zied Bouraoui, and Steven Schockaert. 2023. Ultra-fine entity typing with prior knowledge about labels: A simple clustering based strategy. *arXiv preprint arXiv:2305.12802* (2023).

[53] Peng Li, Yeye He, Dror Yashar, Weiwei Cui, Song Ge, Haidong Zhang, Danielle Rifinski Fainman, Dongmei Zhang, and Surajit Chaudhuri. 2023. Table-gpt: Table-tuned gpt for diverse table tasks. *arXiv preprint arXiv:2310.09263* (2023).

[54] Yinghui Li, Yangning Li, Yuxin He, Tianyu Yu, Ying Shen, and Hai-Tao Zheng. 2022. Contrastive Learning with Hard Negative Entities for Entity Set Expansion. In *SIGIR*. 1077–1086.

[55] Chen Liang, Yue Yu, Haoming Jiang, Siawpeng Er, Ruijia Wang, Tuo Zhao, and Chao Zhang. 2020. Bond: Bert-assisted open-domain named entity recognition with distant supervision. In *KDD*. 1054–1064.

[56] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692* (2019).

[57] Zichen Liu, Hongyuan Xu, Yanlong Wen, Ning Jiang, HaiYing Wu, and Xiaojie Yuan. 2021. TEMP: Taxonomy Expansion with Dynamic Margin Loss through Taxonomy-Paths. In *EMNLP*. 3854–3863.

[58] Keming Lu, I Hsu, Wenxuan Zhou, Mingyu Derek Ma, Muhao Chen, et al. 2022. Summarization as indirect supervision for relation extraction. *arXiv preprint arXiv:2205.09837* (2022).

[59] 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* (2023).

[60] Yubo Ma, Yixin Cao, YongChing Hong, and Aixin Sun. 2023. Large language model is not a good few-shot information extractor, but a good reranker for hard samples! *arXiv preprint arXiv:2303.08559* (2023).

[61] Dheeraj Mekala, Chengyu Dong, and Jingbo Shang. 2022. LOPS: Learning Order Inspired Pseudo-Label Selection for Weakly Supervised Text Classification. In *EMNLP Findings*.

[62] Dheeraj Mekala and Jingbo Shang. 2020. Contextualized Weak Supervision for Text Classification. In *ACL*. 323–333.

[63] Yu Meng, Jiaxin Huang, Guangyuan Wang, Zihan Wang, Chao Zhang, Yu Zhang, and Jiawei Han. 2020. Discriminative Topic Mining via Category-Name Guided Text Embedding. In *WWW'20*.

[64] Yu Meng, Jiaming Shen, Chao Zhang, and Jiawei Han. 2018. Weakly-Supervised Neural Text Classification. In *CIKM*. ACM, 983–992.

[65] Yu Meng, Jiaming Shen, Chao Zhang, and Jiawei Han. 2019. Weakly-supervised hierarchical text classification. In *AAAI*.

[66] Yu Meng, Yunyi Zhang, Jiaxin Huang, Xuan Wang, Yu Zhang, Heng Ji, and Jiawei Han. 2021. Distantly-Supervised Named Entity Recognition with Noise-Robust Learning and Language Model Augmented Self-Training. In *EMNLP*. 10367–10378.

[67] Yu Meng, Yunyi Zhang, Jiaxin Huang, Chenyan Xiong, Heng Ji, Chao Zhang, and Jiawei Han. 2020. Text Classification Using Label Names Only: A Language Model Self-Training Approach. In *EMNLP*. 9006–9017.

[68] Yu Meng, Yunyi Zhang, Jiaxin Huang, Yu Zhang, and Jiawei Han. 2022. Topic Discovery via Latent Space Clustering of Pretrained Language Model Representations. In *WWW'22*. 3143–3152.

[69] Yu Meng, Yunyi Zhang, Jiaxin Huang, Yu Zhang, Chao Zhang, and Jiawei Han. 2020. Hierarchical Topic Mining via Joint Spherical Tree and Text Embedding. In *KDD'20*.

[70] Sewon Min, Xinxi Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2022. Rethinki Work?. In *EMNLP*. 11048–11064.

[71] Iqra Muhammad, Anna Kearney, Carrol Gamble, Frans Coenen, and Paula Williamson. 2020. Open information extraction for knowledge graph construction. In *Database and Expert Systems Applications: DEXA 2020 International Workshops BIOKDD, IWCFS and MLKgraphs, Bratislava, Slovakia, September 14–17, 2020, Proceedings 31*. 103–113.

[72] Abiola Obamuyide and Andreas Vlachos. 2018. Zero-shot Relation Classification as Textual Entailment. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*. 72–78.

[73] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *NeurIPS*.

[74] Siru Ouyang, Jiaxin Huang, Pranav Pillai, Yunyi Zhang, Yu Zhang, and Jiawei Han. 2023. Ontology Enrichment for Effective Fine-grained Entity Typing. *arXiv preprint arXiv:2310.07795* (2023).

[75] Hao Peng, Jianxin Li, Yu He, Yaopeng Liu, Mengjiao Bao, Yangqiu Song, and Qiang Yang. 2018. Large-Scale Hierarchical Text Classification with Recursively Regularized Deep Graph-CNN. In *WWW*.

[76] Letian Peng, Zihan Wang, and Jingbo Shang. 2023. Less than One-shot: Named Entity Recognition via Extremely Weak Supervision. In *Findings of EMNLP*. 13603–13616.

[77] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. In *NAACL-HLT*. 2227–2237.

[78] Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2019. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066* (2019).

[79] Chengwei Qin, Aston Zhang, Zhuosheng Zhang, Jiaao Chen, Michihiro Yasunaga, and Diyi Yang. 2023. Is ChatGPT a General-Purpose Natural Language Processing Task Solver?. In *EMNLP*. 1339–1384.

[80] Meng Qu, Xiang Ren, Yu Zhang, and Jiawei Han. 2018. Weakly-supervised Relation Extraction by Pattern-enhanced Embedding Learning. In *WWW*. 1257–1266.

[81] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training. (2018).

[82] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[83] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In *NeurIPS*.

[84] Colin Raffel, Noam 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. *J. Mach. Learn. Res.* 21 (2020), 140:1–140:67.

[85] Alexander Ratner, Christopher De Sa, Sen Wu, Daniel Selsam, and Christopher Ré. 2016. Data Programming: Creating Large Training Sets, Quickly. In *NIPS*.

[86] Wendi Ren, Yinghao Li, Hanting Su, David Kartchner, Cassie Mitchell, and Chao Zhang. 2020. Denoising Multi-Source Weak Supervision for Neural Text Classification. In *EMNLP Findings*.

[87] Xiang Ren, Wenqi He, Meng Qu, Lifu Huang, Heng Ji, and Jiawei Han. 2016. Afet: Automatic fine-grained entity typing by hierarchical partial-label embedding. In *EMNLP'16*. 1369–1378.

[88] Maya Rotmensch, Yoni Halpern, Abdulhakim Tlimat, Steven Horng, and David Sontag. 2017. Learning a health knowledge graph from electronic medical records. *Scientific reports* 7, 1 (2017), 5994.

[89] Oscar Sainz, Oier Lopez de Lacalle, Gorka Labaka, Ander Barrena, and Eneko Agirre. 2021. Label Verbalization and Entailment for Effective Zero and Few-Shot Relation Extraction. In *EMNLP*. 1199–1212.

[90] Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal V. Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Févry, Jason Alan Fries, Ryan Teehan, Teven Le Scao, Stella Biderman, Leo Gao, Thomas Wolf, and Alexander M. Rush. 2022. Multitask Prompted Training Enables Zero-Shot Task Generalization. In *ICLR*.

[91] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347* (2017).

[92] Jingbo Shang, Liyuan Liu, Xiang Ren, Xiaotao Gu, Teng Ren, and Jiawei Han. 2018. Learning Named Entity Tagger using Domain-Specific Dictionary. In *EMNLP*.

[93] Jingbo Shang, Xinyang Zhang, Liyuan Liu, Sha Li, and Jiawei Han. 2020. NetTaxo: Automated Topic Taxonomy Construction from Text-Rich Network. In *WWW (WWW '20)*. 1908–1919.

[94] Jiaming Shen, Wenda Qiu, Yu Meng, Jingbo Shang, Xiang Ren, and Jiawei Han. 2021. TaxoClass: Hierarchical Multi-Label Text Classification Using Only Class Names. In *NAACL*.

[95] Jiaming Shen, Zhihong Shen, Chenyan Xiong, Chi Wang, Kuansan Wang, and Jiawei Han. 2020. Taxoexpan: Self-supervised taxonomy expansion with position-enhanced graph neural network. In *WWW'20*. 486–497.

[96] Jiaming Shen, Zeqiu Wu, Dongming Lei, Jingbo Shang, Xiang Ren, and Jiawei Han. 2017. Setexpan: Corpus-based set expansion via context feature selection and rank ensemble. In *ECML-PKDD'17*. 288–304.

[97] Jiaming Shen, Zeqiu Wu, Dongming Lei, Chao Zhang, Xiang Ren, Michelle T Vanni, Brian M Sadler, and Jiawei Han. 2018. Hiexpan: Task-guided taxonomy construction by hierarchical tree expansion. In *KDD'18*. 2180–2189.

[98] Yanzhen Shen, Yu Zhang, Yunyi Zhang, and Jiawei Han. 2024. A Unified Taxonomy-Guided Instruction Tuning Framework for Entity Set Expansion and Taxonomy Expansion. *arXiv preprint arXiv:2402.13405* (2024).

[99] Yangqiu Song and Dan Roth. 2014. On Dataless Hierarchical Text Classification. In *AAAI*.

[100] Jiao Sun, Yufei Tian, Wangchunshu Zhou, Nan Xu, Qian Hu, Rahul Gupta, John Wieting, Nanyun Peng, and Xuezhe Ma. 2023. Evaluating Large Language Models on Controlled Generation Tasks. In *EMNLP*. 3155–3168.

[101] 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 Triplet Extraction. *arXiv preprint arXiv:2401.13598* (2024).

[102] Xiaofei Sun, Xiaoya Li, Jiwei Li, Fei Wu, Shangwei Guo, Tianwei Zhang, and Guoyin Wang. 2023. Text Classification via Large Language Models. In *Findings of EMNLP 2023*. 8990–9005.

[103] Ammar Tahir. 2023. Knowledge Graph GPT. https://github.com/iAmmarTahir/KnowledgeGraphGPT.

[104] Luke Vilnis, Xiang Li, Shikhar Murty, and Andrew McCallum. 2018. Probabilistic Embedding of Knowledge Graphs with Box Lattice Measures. In *ACL*. 263–272.

[105] Somin Wadhwa, Silvio Amir, and Byron Wallace. 2023. Revisiting Relation Extraction in the era of Large Language Models. In *ACL*. 15566–15589.

[106] 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 *EMNLP*. 3534–3547.

[107] Alex Wang, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2019. SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems. In *NeurIPS*. 3261–3275.

[108] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2019. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In *ICLR*.

[109] Shuhe Wang, Xiaofei Sun, Xiaoya Li, Rongbin Ouyang, Fei Wu, Tianwei Zhang, Jiwei Li, and Guoyin Wang. 2023. Gpt-ner: Named entity recognition via large language models. *arXiv preprint arXiv:2304.10428* (2023).

[110] Suyuchen Wang, Ruihui Zhao, Yefeng Zheng, and Bang Liu. 2022. QEN: Applicable Taxonomy Completion via Evaluating Full Taxonomic Relations. In *WWW*. 1008–1017.

[111] Xuan Wang, Yingjun Guan, Yu Zhang, Qi Li, and Jiawei Han. 2020. Pattern-enhanced Named Entity Recognition with Distant Supervision. In *2020 IEEE International Conference on Big Data (Big Data)*. 818–827.

[112] Xuan Wang, Vivian Hu, Xiangchen Song, Shweta Garg, Jinfeng Xiao, and Jiawei Han. 2021. ChemNER: Fine-Grained Chemistry Named Entity Recognition with Ontology-Guided Distant Supervision. In *EMNLP*. 5227–5240.

[113] Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-Consistency Improves Chain of Thought Reasoning in Language Models. In *ICLR*.

[114] Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Atharva Naik, Arjun Ashok, Arut Selvan Dhanasekaran, Anjana Arunkumar, David Stap, et al. 2022. Super-NaturalInstructions: Generalization via Declarative Instructions on 1600+ NLP Tasks. In *EMNLP*. 5085–5109.

[115] Yue Wang, Dan Qiao, Juntao Li, Jinxiong Chang, Qishen Zhang, Zhongyi Liu, Guannan Zhang, and Min Zhang. 2023. Towards Better Hierarchical Text Classification with Data Generation. In *Findings of ACL*. 7722–7739.

[116] Zihan Wang, Dheeraj Mekala, and Jingbo Shang. 2021. X-Class: Text Classification with Extremely Weak Supervision. In *NAACL-HLT*. 3043–3053.

[117] Zihan Wang, Peiyi Wang, Lianzhe Huang, Xin Sun, and Houfeng Wang. 2022. Incorporating Hierarchy into Text Encoder: a Contrastive Learning Approach for Hierarchical Text Classification. In *ACL*. 7109–7119.

[118] Zihan Wang, Tianle Wang, Dheeraj Mekala, and Jingbo Shang. 2023. A Benchmark on Extremely Weakly Supervised Text Classification: Reconcile Seed Matching and Prompting Approaches. In *Findings of ACL*. 3944–3962.

[119] Jonatas Wehrmann, Ricardo Cerri, and Rodrigo C. Barros. 2018. Hierarchical Multi-label Classification Networks. In *ICML*.

[120] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *NeurIPS*.

[121] Huiru Xiao, Xin Liu, and Yangqiu Song. 2019. Efficient Path Prediction for Semi-Supervised and Weakly Supervised Hierarchical Text Classification. In *WWW*. 3370–3376.

[122] Jinfeng Xiao, Mohab Elkaref, Nathan Herr, Geeth De Mel, and Jiawei Han. 2023. Taxonomy-Guided Fine-Grained Entity Set Expansion. In *SDM'23*.

[123] Yuxin Xiao, Zecheng Zhang, Yuning Mao, Carl Yang, and Jiawei Han. 2022. SAIS: Supervising and Augmenting Intermediate Steps for Document-Level Relation Extraction. In *NAACL-HLT*. 2395–2409.

[124] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V. Le. 2020. Unsupervised Data Augmentation for Consistency Training. In *NIPS*.

[125] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. 2020. Self-training with noisy student improves imagenet classification. In *CVPR*. 10687–10698.

[126] Yiqing Xie, Jiaming Shen, Sha Li, Yuning Mao, and Jiawei Han. 2022. Eider: Empowering Document-level Relation Extraction with Efficient Evidence Extraction and Inference-stage Fusion. In *Findings of ACL*. 257–268.

[127] Hongyuan Xu, Yunong Chen, Zichen Liu, Yanlong Wen, and Xiaojie Yuan. 2022. TaxoPrompt: A Prompt-based Generation Method with Taxonomic Context for Self-Supervised Taxonomy Expansion. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*. 4432–4438.

[128] Siqiao Xue, Caigao Jiang, Wenhui Shi, Fangyin Cheng, Keting Chen, Hongjun Yang, Zhiping Zhang, Jianshan He, Hongyang Zhang, Ganglin Wei, et al. 2023. Db-gpt: Empowering database interactions with private large language models. *arXiv preprint arXiv:2312.17449* (2023).

[129] Hang Yan, Junqi Dai, Tuo Ji, Xipeng Qiu, and Zheng Zhang. 2021. A Unified Generative Framework for Aspect-based Sentiment Analysis. In *ACL-IJCNLP*. 2416–2429.

[130] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J. Smola, and Eduard H. Hovy. 2016. Hierarchical Attention Networks for Document Classification. In *NAACL-HLT*.

[131] Zeng Yang, Linhai Zhang, and Deyu Zhou. 2022. SEE-Few: Seed, Expand and Entail for Few-shot Named Entity Recognition. In *COLING*. 2540–2550.

[132] Sizhe Zhou Heng Ji Jiawei Ha Yizhu Jiao, Sha Li. 2024. Text2DB: Integration-Aware Information Extraction with Large Language Model Agents. (2024).

[133] Changlong Yu, Weiqi Wang, Xin Liu, Jiaxin Bai, Yangqiu Song, Zheng Li, Yifan Gao, Tianyu Cao, and Bing Yin. 2023. FolkScope: Intention Knowledge Graph Construction for E-commerce Commonsense Discovery. In *Findings of ACL*. 1173–1191.

[134] Yue Yu, Yinghao Li, Jiaming Shen, Hao Feng, Jimeng Sun, and Chao Zhang. 2020. Steam: Self-supervised taxonomy expansion with mini-paths. In *KDD'20*. 1026–1035.

[135] Qingkai Zeng, Jinfeng Lin, Wenhao Yu, Jane Cleland-Huang, and Meng Jiang. 2021. Enhancing Taxonomy Completion with Concept Generation via Fusing Relational Representations. In *KDD*. 2104–2113.

[136] Xiangrong Zeng, Daojian Zeng, Shizhu He, Kang Liu, and Jun Zhao. 2018. Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism. In *ACL*. 506–514.

[137] Chao Zhang, Fangbo Tao, Xiusi Chen, Jiaming Shen, Meng Jiang, Brian Sadler, Michelle Vanni, and Jiawei Han. 2018. TaxoGen: Unsupervised Topic Taxonomy Construction by Adaptive Term Embedding and Clustering. In *KDD*. 2701–2709.

[138] Jieyu Zhang, Xiangchen Song, Ying Zeng, Jiaze Chen, Jiaming Shen, Yuning Mao, and Lei Li. 2021. Taxonomy Completion via Triplet Matching Network. In *AAAI*.

[139] Kai Zhang, Bernal Jimenez Gutierrez, and Yu Su. 2023. Aligning Instruction Tasks Unlocks Large Language Models as Zero-Shot Relation Extractors. In *Findings of ACL*. 794–812.

[140] Lu Zhang, Jiandong Ding, Yi Xu, Yingyao Liu, and Shuigeng Zhou. 2021. Weakly-supervised Text Classification Based on Keyword Graph. In *EMNLP*. 2803–2813.

[141] Tianshu Zhang, Xiang Yue, Yifei Li, and Huan Sun. 2023. TableLlama: Towards Open Large Generalist Models for Tables. *arXiv preprint arXiv:2311.09206* (2023).

[142] Xiang Zhang, Junbo Jake Zhao, and Yann LeCun. 2015. Character-level Convolutional Networks for Text Classification. In *NIPS*.

[143] Yu Zhang, Xiusi Chen, Yu Meng, and Jiawei Han. 2021. Hierarchical Metadata-Aware Document Categorization under Weak Supervision. In *WSDM'21*. 770–778.

[144] Yunyi Zhang, Minhao Jiang, Yu Meng, Yu Zhang, and Jiawei Han. 2023. PIEClass: Weakly-Supervised Text Classification with Prompting and Noise-Robust Iterative Ensemble Training. In *EMNLP*. 12655–12670.

[145] Yu Zhang, Bowen Jin, Xiusi Chen, Yanzhen Shen, Yunyi Zhang, Yu Meng, and Jiawei Han. 2023. Weakly Supervised Multi-Label Classification of Full-Text Scientific Papers. In *KDD'23*. 3458–3469.

[146] Yu Zhang, Yu Meng, Xuan Wang, Sheng Wang, and Jiawei Han. 2022. Seed-Guided Topic Discovery with Out-of-Vocabulary Seeds. In *NAACL'22*. 279–290.

[147] Yunyi Zhang, Jiaming Shen, Jingbo Shang, and Jiawei Han. 2020. Empower Entity Set Expansion via Language Model Probing. In *ACL'20*. 8151–8160.

[148] Yunyi Zhang, Ruozhen Yang, Xueqiang Xu, Rui Li, Jinfeng Xiao, Jiaming Shen, and Jiawei Han. 2024. TELEClass: Taxonomy Enrichment and LLM-Enhanced Hierarchical Text Classification with Minimal Supervision. *arXiv preprint arXiv:2403.00165* (2024).

[149] Yu Zhang, Yunyi Zhang, Martin Michalski, Yucheng Jiang, Yu Meng, and Jiawei Han. 2023. Effective Seed-Guided Topic Discovery by Integrating Multiple Types of Contexts. In *WSDM'23*.

[150] Yu Zhang, Yunyi Zhang, Yanzhen Shen, Yu Deng, Lucian Popa, Larisa Shwartz, ChengXiang Zhai, and Jiawei Han. 2024. Seed-Guided Fine-Grained Entity Typing in Science and Engineering Domains. *AAAI* (2024).

[151] Zihan Zhang, Meng Fang, Ling Chen, and Mohammad-Reza Namazi-Rad. 2022. Is Neural Topic Modelling Better than Clustering? An Empirical Study on Clustering with Contextual Embeddings for Topics. In *NAACL'22*. 3886–3993.

[152] Xuandong Zhao, Siqi Ouyang, Zhiguo Yu, Ming Wu, and Lei Li. 2023. Pre-trained Language Models Can be Fully Zero-Shot Learners. In *ACL*. 15590–15606.

[153] Ben Zhou, Daniel Khashabi, Chen-Tse Tsai, and Dan Roth. 2018. Zero-Shot Open Entity Typing as Type-Compatible Grounding. In *EMNLP'18*. 2065–2076.

[154] Sizhe Zhou, Suyu Ge, Jiaming Shen, and Jiawei Han. 2023. Corpus-Based Relation Extraction by Identifying and Refining Relation Patterns. In *ECML PKDD*. 20–38.

[155] Sizhe Zhou, Yu Meng, Bowen Jin, and Jiawei Han. 2024. Grasping the Essentials: Tailoring Large Language Models for Zero-Shot Relation Extraction. *arXiv preprint arXiv:2402.11142* (2024).

[156] 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* (2023).