A Comprehensive Survey on Automatic Knowledge Graph Construction

LINGFENG ZHONG and JIA WU, School of Computing, Macquarie University, Australia
QIAN LI, School of Computer Science and Engineering, Beihang University, China
HAO PENG, School of Cyber Science and Technology, Beihang University, China
XINDONG WU, Research Center for Knowledge Engineering, Zhejiang Lab, China and Key Laboratory of Knowledge Engineering with Big Data (the Ministry of Education of China), Hefei University of Technology, China

Automatic knowledge graph construction aims at manufacturing structured human knowledge. To this end, much effort has historically been spent extracting informative fact patterns from different data sources. However, more recently, research interest has shifted to acquiring conceptualized structured knowledge beyond informative data. In addition, researchers have also been exploring new ways of handling sophisticated construction tasks in diversified scenarios. Thus, there is a demand for a systematic review of paradigms to organize knowledge structures beyond data-level mentions. To meet this demand, we comprehensively survey more than 300 methods to summarize the latest developments in knowledge graph construction. A knowledge graph is built in three steps: knowledge acquisition, knowledge refinement, and knowledge evolution. The processes of knowledge acquisition are reviewed in detail, including obtaining entities with fine-grained types and their conceptual linkages to knowledge graphs; resolving coreferences; and extracting entity relationships in complex scenarios. The survey covers models for knowledge refinement, including knowledge graph completion, and knowledge fusion. Methods to handle knowledge evolution are also systematically presented, including condition knowledge acquisition, condition knowledge graph completion, and knowledge dynamic. We present the paradigms to compare the distinction among these methods along the axis of the data environment, motivation, and architecture. Additionally, we also provide briefs on accessible resources that can help readers to develop practical knowledge graph systems. The survey concludes with discussions on the challenges and possible directions for future exploration.

CCS Concepts: • **Information systems → Graph-based database models**; **Information retrieval**; • **Computing methodologies → Knowledge representation and reasoning;**

This article is supported by the Australian Research Council (ARC) Projects Nos. DE200100964, LP210301259, and DP230100899, and the National Key R&D Program of China through grant 2021YFB1714800, and the National Natural Science Foundation of China (NSFC) under grant 62120106008.

Authors' addresses: L. Zhong and J Wu (Corresopnding author), 4 Research Park Drive, School of Computing, Faculty of Science and Engineering, Macquarie University, Sydney, NSW 2109, Australia; e-mails: lingfeng.zhong@hdr.mq.edu.au, jia.wu@mq.edu.au; Q. Li, School of Computer Science and Engineering, Beihang University, 37 Xueyuan Road, Haidian, Beijing, China; e-mail: liqian@act.buaa.edu.cn; H. Peng, School of Cyber Science and Technology, Beihang University, 37 Xueyuan Road, Haidian, Beijing, China; e-mail: penghao@buaa.edu.cn; X. Wu (Corresopnding author), Research Center for Knowledge Engineering, Zhejiang Lab, Kechuang Avenue, Zhongtai Street, Yuhang District, Hangzhou, Zhejiang, China and Key Laboratory of Knowledge Engineering with Big Data (the Ministry of Education of China), Hefei University of Technology, 193 Tunxi Road, Baohe District, Hefei, Anhui, China; e-mail: xwu@hfut.edu.cn.

0360-0300/2023/11-ART94 $15.00
https://doi.org/10.1145/3618295

Additional Key Words and Phrases: Knowledge graph, deep learning, information extraction, knowledge graph completion, knowledge fusion, logic reasoning

ACM Reference format:
Lingfeng Zhong, Jia Wu, Qian Li, Hao Peng, and Xindong Wu. 2023. A Comprehensive Survey on Automatic Knowledge Graph Construction. *ACM Comput. Surv.* 56, 4, Article 94 (November 2023), 62 pages.
https://doi.org/10.1145/3618295

1 INTRODUCTION

Techniques for constructing **knowledge graphs** (**KGs**) have been a long-lasting research trend. In recent years, academic communities have seen substantial literature publications presenting KG construction techniques [1][2][3][4][5], along with emerging applications of KG construction covering tasks of recommendation systems [6], dialogue systems [7], and fake news detection [8] et al. KG construction techniques have been recognized as the paving stone for AI [4].

Trustworthy knowledge graphs provide well-organized human knowledge as auxiliary information to enable knowledge awareness. However, many high-quality KG systems have relied on crowd-sourcing strategies for construction, like Freebase [9] and Wikidata [10]. Hence, a systematic solution that can automatically build a knowledge graph from unstructured or semi-structured data offers a massive boost to what is a very arduous manual process for practical purposes.

A knowledge graph is defined as a semantic graph consisting of edges and nodes that depicts knowledge of real-world objects. Within these structures, a knowledge tuple is the minimum knowledge-carrying group. The tuples comprise two nodes representing concepts connected by an edge representing a relationship. Thus, constructing a knowledge graph is the task of discovering the elements that constitute a knowledge graph in a domain-specific or open-domain area. Early in the study of this discipline, researchers were mostly focusing on scratching out factual tuples from semi-structured or unstructured textual data as patterned knowledge mentions. Information extraction systems like TextRunner [11] and Knowitall [12] are the milestones for early knowledge graph construction, driven by designated rules or clustering. Unfortunately, these designs are not sufficiently equipped with background knowledge, and thus suffer from two major defects: (1) insubstantial, traditional information extraction systems do not create or distinguish entities from different expressions, which prevents knowledge aggregation; (2) uninformative, traditional information extraction systems only extract information from syntactic structures without capturing the semantic denotations in the given expressions. Furthermore, conventional rule-based information extraction systems also require heavy feature engineering and extra expert knowledge. Wu et al. [13] point out that if a KG system does not organize nodes and edges with background knowledge about concepts, it is merely a data graph.

Regarding this issue, researchers then recourse to well-partitioned acquisition sub-tasks for arranging semantic knowledge structures. The most classic paradigm is the pipeline that first discovers and links conceptual entities, resolves coreference mentions, then extracts relationships among entities. The general procedure of knowledge graph construction is displayed in Figure 1.

More recently, deep learning methods have given rise to tremendous breakthroughs in **natural language processing** (**NLP**), and these breakthroughs have fed applications for knowledge graph construction in a range of respects. Numerous deep learning models have delivered good performances with tasks like named entity recognition [14][15], entity typing [16][17], entity linking [18][19], coreference resolution [20], and relation extraction [21][22]. Additionally, deep knowledge representation models have also been developed that can refine knowledge graphs. The refinements include completing corrupt tuples, discovering new tuples in a built knowledge graph via its inner graph structure, and merging graphs from different sources to construct new

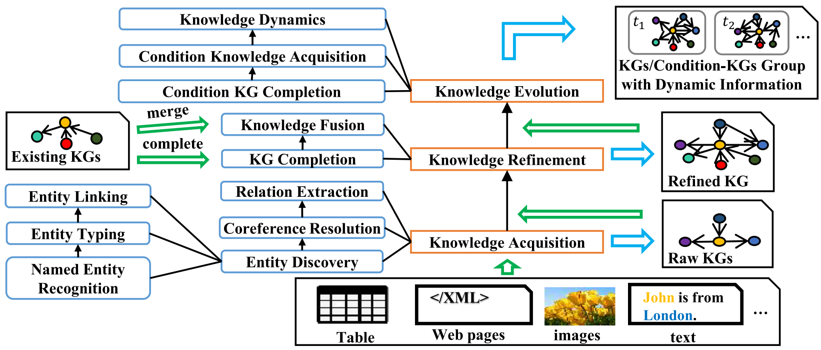

Fig. 1. The general process of constructing a knowledge graph. In this diagram, semi-structured or unstructured input data is manufactured into a raw knowledge graph by acquiring knowledge. Then the knowledge will be refined to complete the knowledge graph or enrich it with other existing knowledge graphs. If the input is only an existing knowledge graph, it will be directly handled by the knowledge refinement process. Last, the knowledge evolution process will try to obtain a group of knowledge graphs/conditional knowledge graphs that contains dynamic information about the graph's evolution.

knowledge graphs. At present, many knowledge bases,[1] such as TransOMCS [23], ASER [24] and huapu [25] have put automatic KG construction methods into practice.

Further, with advances in the pre-training of deep learning models, such as the pre-trained BERT model [26] and some of the massive-scale **graph convolution network** (**GCN**) models, KG construction tasks are being applied to more complicated scenarios in the big data environment. Beyond systems that deal with heterogeneous data, like web pages and table forms, more attention is being paid to effective methods of tackling complex data for example, jointly unifying multiple acquisition sub-tasks or solutions that harvest knowledge graphs from long-term contexts [27], noisy data [28], or low-resource data [29].

In terms of knowledge graph refinement tasks, interpretable reasoning has become a prevalent trend. Researchers are seeking solutions that merge cross-lingual knowledge and derive new relationships between nodes through logic and reasoning. Researchers are also focusing on knowledge graphs for conditional knowledge, such as temporal knowledge graphs [30] and generic condition knowledge graphs [31]. Active learning [32], asking human users about unknown data for collection, is another significant direction for handling knowledge from autonomous communities.

Further, Wu et al. [33] summarize the essence of big data environments with the HACE theorem. The term, "HACE", stands for heterogeneous, autonomous, complex, and evolving characteristics of big data sources. We present the specific challenges facing knowledge discovery in such environments for KG construction in Figure 2.

1.1 Major Differences and Contributions

Many surveys have provided an overview of knowledge graphs and their applications. For example, Hogan et al. [5] provide an encyclopedic survey for the knowledge graph, while Paulheim [1] looks into methods that refine and fill knowledge graphs. Other surveys summarize methods for acquiring knowledge from unstructured or semi-structured data. Wu et al. [3] review competitive tools and models for KG construction sub-task over texts including relation extraction, named entity recognition, and coreference resolution, while Yan et. al [2] browse methods for different data types like web pages, table forms, and so on. Deep learning approaches for jointly extracting

[1]Knowledge base (KB) and knowledge graph (KG) are identical terms in this survey.

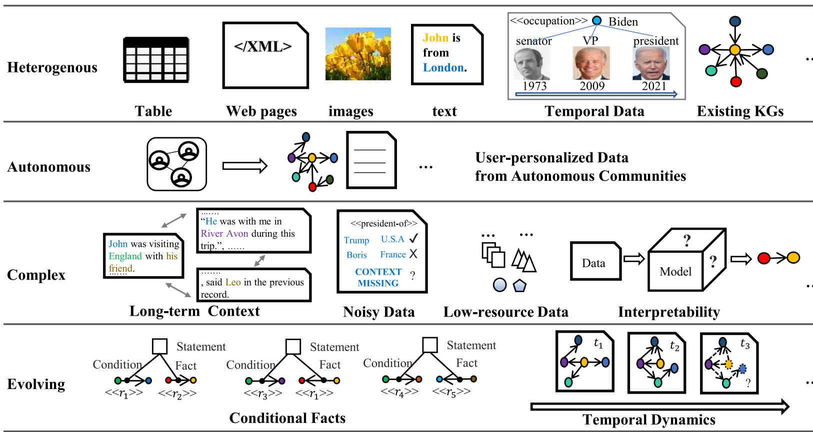

Fig. 2. An illustration of the challenges framed by the HACE environments. In terms of heterogeneous data, knowledge graph construction with semi-structured and unstructured data is outlined in Section 3. Methods of refining existing knowledge graphs are detailed in Section 4. Methods for obtaining temporal data are described in Section 5.1. Section 6.5 presents a discussion on multi-modal knowledge graphs. In terms of complex data, long-term contexts and their involvement with multiple knowledge graph construction tasks are discussed in Sections 3.1.2, 3.2, and 3.3.6. Methods for tackling noisy data are mainly presented in Section 3.3.3. Model interpretability is covered in Section 4.1.3. In terms of evolving data, recent work in knowledge evolution is presented in Section 5, and research on autonomous data is discussed in Section 6.2.

Table 1. A Comparison between Existing Survey on Knowledge Graph Construction

Survey	Year	Topic	KA				KGR			Target Data				Resource	
			Ent	Rel	CO	Cond	KGC	TKGC	KF	Web	Tab	Sent	Doc	Tool	Dataset
Our Survey	2022	Overall Process of KG Construction in HACE environment	● blue	● blue	● blue	● light blue	● blue	● blue	● blue	● blue	● blue	● blue	● blue	● blue	● blue
Paulheim [1]	2017	KG Refinement	● grey	● light blue	–	–	● blue	–	–	● grey	● grey	● light blue	–	–	● light blue
Yan et al. [2]	2018	KG Application, Construction	● blue	● blue	● grey	–	● blue	–	–	● light blue	● grey	● blue	● grey	● blue	● blue
Wu et al. [3]	2019	Raw KG Construction	● light blue	● light blue	● light blue	–	–	–	–	● grey	–	● blue	–	● light blue	● grey
Ji et al. [4]	2020	KG Application, Representation, Acquisition	● blue	● blue	–	–	● blue	● blue	● light blue	–	–	● blue	–	● light blue	● blue
Arora [36]	2020	KG Completion	–	–	–	–	● blue	–	–	–	–	–	–	–	● grey
Nayak et al. [34]	2021	Relation Triples Extraction	● blue	● blue	–	–	–	–	–	● grey	–	● blue	● grey	● grey	● light blue
Pawar et al. [35]	2021	Joint Extraction of Entities and Relations	● blue	● blue	–	–	–	–	–	–	–	● blue	–	–	● grey
Hogan et al. [5]	2021	Overview of KG	● blue	● blue	–	–	● blue	● grey	–	● light blue	● blue	● blue	–	● light blue	● blue
Cai et al. [37]	2022	Temporal KG	–	–	–	–	● blue	● blue	–	–	–	–	–	–	● grey

*-:Not covered, ● grey:1-5 references covered, ● light blue:6-14 references covered, ● blue: 15+ references covered.
*Ent: Entity, Rel: Relationship, CO: Coreference, Cond: Condition (timestamp or prerequisite).
*Web: Web page, Tab: Table forms, Sent: Sentence, Doc: Textual document.
*KA: Knowledge Acquisition, KGR: Knowledge Graph Refinement, KF: Knowledge Fusion.
*KGC: Knowledge Graph Completion, TKGC: Temporal Knowledge Graph Completion.

entities with their relationships are reviewed in [34], [35]. Some surveys also focus on acquiring knowledge from existing knowledge graphs. Prior work such as [4] and [36] also covers the methods for knowledge representation learning and knowledge graph completion, while Cai et al. [37] dive into temporal knowledge graphs. Table 1 compares previous work with this survey.

Unlike other surveys, we go deeper into the paradigms of the recent models for knowledge graph construction, arranging our work according to different stages and aspects of the HACE environments. We also present practical resources and discuss future challenges and directions with data, models, and architectures. Hence, our contributions are summarized as follows:

— We introduce the process of knowledge graph construction and various knowledge graphs by giving formal definitions and classifications. We also summarize necessary information on KG-related resources, including practical knowledge graph datasets and construction tools, covering published years, citations, and access links for readers to compare.

— We comprehensively analyze models for knowledge graph construction in different scenarios from knowledge acquisition to knowledge graph refinement according to their task backgrounds and challenges. We summarize the motivations and designs of classical and novel models, then primarily delineate the pragmatics in terms of their architectures and improvements.
— We discuss knowledge graph construction in HACE big data environments, including noisy, document-level data, and low-resource data. Then we review achievements for acquiring model interpretability and evolutionary condition knowledge. Finally, we summarize the major challenges and directions that impact KG construction tasks.

1.2 Organization of the Survey

We organize our survey as follows. Section 2 gives the background for knowledge graph construction, including definitions and resources for KG constructions (KG project datasets in Section 2.2 and tools in Section 2.3). Section 3 introduces the methods for handling tasks that obtain entities and relationships from various data types and environments. Section 4 reviews the methods for refining knowledge graphs with external structured data. We portray recent achievements and trends in evolutionary knowledge graphs, including conditional knowledge graphs and temporal knowledge graphs in Section 5. Finally, we envisage future directions and development in Section 6 while concluding the article in Section 7.

2 BACKGROUND

2.1 Definitions

Many contributions have been made to formally define knowledge graphs. Wang et al. [38] model a knowledge graph as a multi-relation graph, where the nodes are entities and the edges represent different types of relationships. However, the previous definition does not consider the semantic structures in a knowledge graph. Ehrlinger and Woß [39] further emphasize that a knowledge graph arranges information into an ontology and then enlightens novel knowledge discovery with "a reasoner". To specifically protrude the essence component supporting knowledge-level information, Wu et al. [3] define a knowledge graph as a semantic graph where the nodes represent concepts (entities/attributes/facts), and the edges represent relationships that connect the nodes while drawing on background knowledge about the concepts and relations.

Definition 1 (Knowledge Graph). A knowledge graph $\mathcal{G}$ is defined as $\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{F}_k\}$, where $\mathcal{E}$ and $\mathcal{R}$ represent sets of concepts and relations, respectively. In this survey, concepts can be regarded as entities/attributes. $\mathcal{T}$ is the set of factual triples, where a standard binary fact is a triple $(h, t, r) \in \mathcal{T}, h, t \in \mathcal{E}, r \in \mathcal{R}$. An n-ary relation triple will be formed as $(e_1, \dots, e_n, r)$, where $e_1, \dots, e_n \in \mathcal{E}$. $\mathcal{F}_k$ is a set function representing the background knowledge that constrains potential facts to be knowledge-level informative, and we have $\mathcal{T} \subset \mathcal{F}_k(\{\mathcal{E}, \mathcal{R}\})$. In practice, background knowledge can be seen as a rule set, schema, or set of implicit math principles.

Definition 2 (Knowledge Graph Construction). Knowledge graph construction f is a procedure that maps a data source into a knowledge graph: $f : D \times f_k(D) \rightarrow \mathcal{G}$, where D is the set of data sources, and $f_k(D)$ is background knowledge of the data target, which can be domain knowledge. Notably, knowledge graph construction is usually unable to continue without background knowledge that is provided by pre-designed rules or a language model of representations.

2.2 Practical Knowledge Graph Datasets

In this section, we review the representative practical knowledge graph projects (datasets), including encyclopedia knowledge graphs, linguistic knowledge graphs, commonsense knowledge graphs, enterprise support knowledge graphs, domain-specific knowledge graphs, and federated knowledge graphs. Complete information about the knowledge graph datasets is provided in Appendix B.1.

Encyclopedia knowledge graphs systematically cover factual or event knowledge from different domains. Many researchers have developed knowledge graph structures from manually-built online encyclopedias. For example, DBpedia [40] (developed from Wikipedia) is a fundamental encyclopedia knowledge graph, while Freebase [9] incorporates automatic extraction tools to obtain more content. Probase [41] (supported by Microsoft Concept Graph), as a concept encyclopedia knowledge graph, creatively depicts knowledge of uncertainties containing conflicting information in the form of probabilistic models. XLore [42] (a sub-project of THUKC), as a multi-lingual encyclopedia knowledge graph, establishes entity links across multi-lingual content via deep learning approaches.

Linguistic knowledge graphs deliver knowledge of the human language to provide basic semantics as ontologies or external features. WordNet [43] is a classical widely-used knowledge graph dictionary for linguistic study, providing synonymy or hyponymy relationships among words. With these tools, developers create high-performance word embeddings based on well-built linguistic knowledge graphs for downstream applications.

Commonsense knowledge graphs depict widely-accepted knowledge of common understandings. OpenCyc [44] is one of the earliest of these attempts, which encodes knowledge concepts, rules, and common sense ideas in the form of CycL.

Enterprise knowledge graphs effectively support the business of enterprise activities. **Google Knowledge Graph** (**GKG**) attempt at a knowledge graph application. It has delivered knowledge support in responding to user queries [45] is the earliest practical with semantically-related content since 2012.

Domain-specific knowledge graphs serve multiple professional research fields. Drugbank [46] provides an insight into pharmacology with protein and drug information. Creatively, Huapu [25] has built a high-quality semantic network of Chinese family trees by digitizing thousands of ancestral genealogy books and fusing multiple web sources.

Federated knowledge graphs prevent data leaks via federated learning strategies. Privacy protection plays a critical role in knowledge bases, especially where massive user data have been sourced from different providers. Knowledge communities develop federated knowledge graphs to utilize the abundant multi-source data features. GEDmatch[2] is a representative genealogy knowledge graph that collects user-provided information based on DNA tests supported by a distributed federated knowledge model that tracks the data with privacy desensitization.

2.3 Knowledge Graph Construction Tools

In this section, we review some of the tools commonly used for KG construction. These tools mainly support construction sub-tasks, such as pre-processing, knowledge acquisition, and knowledge refinement. Complete information about tools is detailed in Appendix B.2.

Data pre-processing tasks remove noise like advertisements. For instance, WebCollector [47] is a representative data pre-processing toolkit that automatically filters non-content noise, such as advertisements and layout information, while retaining the page content via integrated algorithms.

Knowledge acquisition tools support sub-tasks that obtain entities and relations to form a raw knowledge graph (consisting of fact triples). Early toolkits directly extracted fact triples through rules, with one notable early achievement being KnowItAll [12]. For industrial applications,

[2] https://www.gedmatch.com/

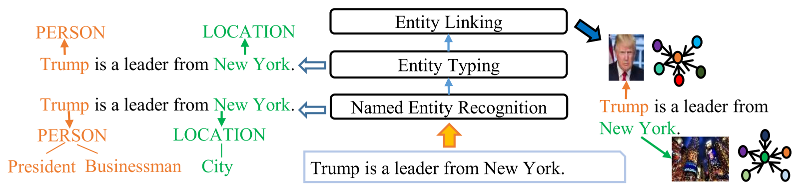

Fig. 3. The entity discovery process.

OpenCalais[3] provides customized knowledge acquisition solutions to users. Deep learning toolkits provide high-performance techniques for users. Further, gBuilder [48] delivers an end-to-end KG construction solution with customizable neural architectures, an active learning interface, and schemas to integrate existing knowledge databases. The tool is significant, for it lets users flexibly practice multiple deep-learning construction strategies without coding.

Knowledge refinement tools refine an existing raw knowledge graph by completing it or merging it with other knowledge graphs. Tools based on deep learning are popular solutions. For example, OpenKE [49] provides multiple knowledge representation models for knowledge graph completion, while OpenEA [50] leverages similar methods for merging structures in knowledge graphs.

3 KNOWLEDGE ACQUISITION

Knowledge acquisition is the general process of collecting elements from multi-structured data to build a knowledge graph. It includes entity discovery, coreference resolution, and relation extraction. Entity discovery tasks recognize entity mentions within data. Coreference resolution tasks then locate referred mention pairs, followed by relation extraction tasks, which link entities with their semantic relationships.

3.1 Entity Discovery

Entity discovery acquires a subset of concepts from semi-structured or structured data that can constitute the nodes of a knowledge graph. The general procedure of entity discovery includes named entity recognition, entity typing, and entity linking tasks. Named entity recognition tasks discover strings that refer to semantic entities and then classify them into general types (e.g., person, location, country, company). Entity typing tasks categorize the found entities into specific types (e.g., actor, artist, brand). Entity linking associates a discovered entity with a possible node in the knowledge graph. If there are no available nodes for linking, a corresponding entity node will be created to represent the newly found entity. Figure 3 depicts an overview of the general process.

3.1.1 Named Entity Recognition from Semi-structured Data vs. Unstructured Data. **Named entity recognition** (**NER**) tasks tag named entities in semi-structured or unstructured data with their positions and classifications. Semi-structured data are enveloped by semantic hints related to property-attribute structures, while unstructured data only contains texts.

Statistic-based approaches treat NER tasks as a sequential classification tagging task that tags entities according to the BIES scheme (beginning, intermedia, ending, single) and their types. For unstructured NER tasks, the key hypothesis is that a tag for each word only depends on the previous words. Hence, applications built on hidden Markov models [51] and **conditional random fields** (**CRF**) models [52], which capture neighborhood-dependencies, are popular NER designs.

[3]https://www.refinitiv.com/en/products/intelligent-tagging-text-analytics

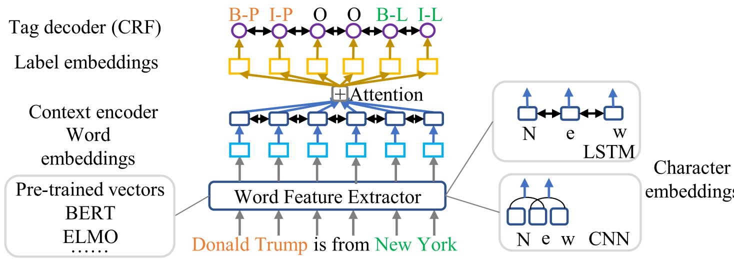

Fig. 4. Illustration of the standard architecture for deep-learning-based named entity recognition. When a sentence is inputted, such NER models will output tagged entity words with positional information and rough entity classifications.

With semi-structured table data, researchers often use CRF variants to tackle the two-dimensional features of the attributes that relate to the entities, such as 2D-CRF [53]. Thus, they extract multiple attributes of each entity in a 2D structure. **Dynamic conditional random field** (**DCRF**) [54] infers potential attribute-entity interaction via the dynamic Bayesian network, and hierarchical CRF [55] models semi-structured data into a hierarchical tree for joint extraction. Further, Finn and Kushmerick [56] develop a SVM-based model to locate the boundaries of entities in texts. Previous models are exemplars of quick NER solutions, however heavily rely on handcraft features.

Deep learning thus becomes another popular trend in named entity recognition, especially for text NER tasks. Deep learning approaches typically treat NER configurations as a seq2seq model (word sequences to label sequences). These models aggregate contextual and lexical feature embeddings according to the input, and context encoders then output word type tags through tag decoders such as a CRF structure or a softmax structure [58]. The following deep learning-based NER models stick to the standard architecture pictured in Figure 4.

CNN and RNN structures are the classic NER designs. CNN structures mainly focus on local features for capturing entities. Colloabert et al. [59] are the first to employ a CNN with a CRF output layer as a unified solution for entity detection, utilizing lexical features via embeddings. Pure CNNs cannot effectively comprehend long-term dependency. Thus, researchers turn to RNN structures to digest global contextual features in long sentences, such as the uni-directional RNN for biomedical entity recognition presented in [60]. However, RNNs may suffer from context bias with later upcoming words [14]. Hence, many models consider bi-directional RNNs, such as the Bi-LSTM-CRF-based in [14] and the GRU-based NER model in [61]. Meanwhile, training procedures of recursive models cannot be completely parallelized. To this end, IDCNN [62] improves upon the CNN with dilated convolutions that enlarge the perception field by omitting some of the input to enhance the generalization ability to RNN-based designs. Combinations of character and word encoders incorporate the advantages of both classical neural structures, such as a design comprising a CNN for character embedding and an LSTM for word embedding [15], extricating neural models from the reliance on handcraft lexical features. From the perspective of more recent practices of NERs, conventional black-box neural models remain limited, for they do not explicitly emphasize task-specific features while disregarding prior background knowledge in the big data environment.

To this end, more efforts design attention mechanisms that project salient interactions for global contexts. Luo et al. [63] introduce the word-level soft attention to enhance named entity recognition. Gregoric et al. [64] employ the word-word self-attention for named entity recognition.

Further, researchers propose graph convolution networks utilizing prior linguistic knowledge of graph structures to handle contexts for NER tasks. For example, Cetoli et al. [65] propose a

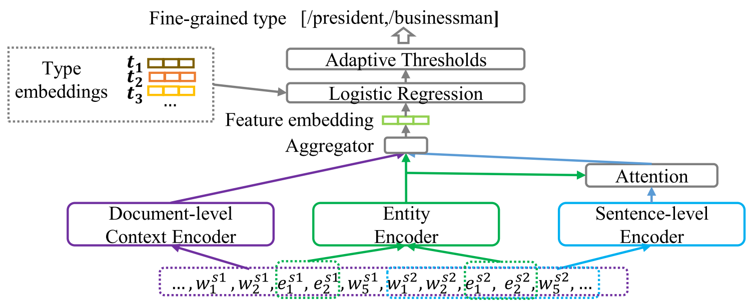

Fig. 5. Illustration of the ET model [57] based on multi-scale feature extraction.

GCN framework that encodes the LSTM-proceed features via GCN structures with a syntactic dependency tree. Pre-trained language models are also developed from big datasets to extract representations of implicit background knowledge for breakthrough-level entity recognition. Such models include Elmo [66], Ltp [67], and LUKE [68].

3.1.2 Entity Typing. **Entity typing (ET)** tasks provide fine-grained and ultra-grained type information for entities such as scientists, clubs, and hotels. Information loss occurs if ET tasks are not performed, e.g., Donald Trump is a politician and a businessman. Semi-structured tables provide hints for fine-grained types in the captions. For example, the caption "soccer players in England" is a fine-grained entity type that suggests players who play soccer in a specific area. However, tagging proper fine-grained entity types in different contexts for unstructured data can be intricate to achieve.

ET tasks require comprehending fine-grain semantics for typing. A deep learning model serves this by overcoming two-fold challenges: (1) interpreting infrequent fine-grain types, and (2) preventing overly-specific typing. Some specific types can be imbalanced or infrequent. For this reason, Shimaoka et al. [69] propose an LSTM-based attentive neural network for infrequent entity typing that relies on hierarchical label encoding integrated with mention and context representations to exploit fine-grained contextual features. Overly-specific type annotations derive correct types but do not fit the current data context. Xu et al. [16] apply an out-of-context loss function to the entities with multiple labels for filtering overly-specific data noise which assumes that the type label which scores the highest probability during training is correctly tagged. To further explore context scenarios, Zhang et al. [57] introduce document-level representations to provide a global context for discovering entities. Sentence-level contextual representations are also utilized to align identical entity representations appearing in different sentences. An adaptive probability threshold then generates labels of the entity types in different contexts. We specify this model in Figure 5.

3.1.3 Entity Linking from Semi-structured Data vs Unstructured Data. EL tasks, also called entity disambiguation, link entity mentions to their corresponding objects in a knowledge graph. A textual mention can have different references, e.g., the text "Tesla" may refer to the car, the corporation, or the scientist. Entity linking connects mentions in different data backgrounds with the contextual information of their respective nodes. With semi-structured data, entity linking identifies entities using semantic hints from column heads, type labels, cell texts of tables, and hyperlinks. With unstructured text, EL models focus on the contextual representations of entity mentions.

Statistical approaches, especially SVM-based models and probabilistic-based models, are the general solutions for semi-structured and unstructured data. SVM models treat entity linking as a

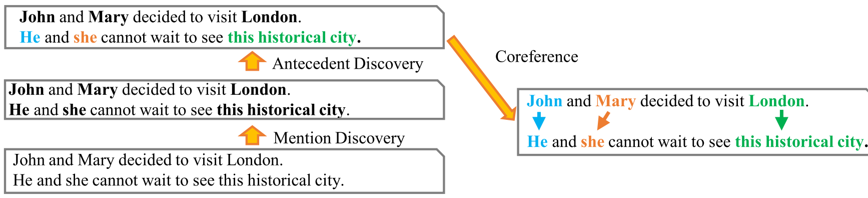

Fig. 6. The coreference resolution process. First, mentions are detected. Then the antecedents of the mentions are selected and matched to co-referred pairs. Noticeably, coreference resolution tasks can be performed on documents with multiple sentences, while handling identical mentions in a compound sentence.

classification task. Here, Mulwad et al. [70] develop a model based on SVMRanker that determines which potential nodes can link to a target entity. More flexible models based on probabilistic graphs construct a probabilistic graph of mentions in tables, then link entities by calculating the semantic factors of nodes. Limaye et al. [71], for instance, construct a factor graph for collective entity linking based on the TF-IDF algorithm that calculates the term frequency of entity labels with cell-text pairs and type labels with column-head pairs. Some models incorporate external knowledge bases to improve EL tasks. For example, TabEL [72] improves its factor graph by leveraging the hyperlinks in Wikipedia to estimate semantic relatedness features before collective classification for disambiguation. Wu et al. [73] propose an approach for enhancing entity linking with the "same-as" edges in multiple knowledge bases. Efthymiou et al. [74] systematically exploit semantic features for entity linking. Their approach integrates vector representations of an entity's context, minimal entity context, and schematic structures shared between KBs and tables.

Deep learning models can achieve better performance than statistic-based models. As an example, GENRE [75] proposes a generative seq2seq solution to handle EL in massive unstructured data. This design elegantly encodes a text via a pre-trained autoencoder and decodes candidate corresponding entities by autoregressive beam search, which leads to breakthroughs in different EL tasks. However, such a design consumes excessive computation resources. We recommend developers consider pre-train models for EL only when complex environments are involved.

3.2 Coreference Resolution

Coreference expressions often appear in unstructured text. As such, **coreference resolution** (**CO**) tasks detect mentions that refer to the same entities (including aliases and pronouns). A mention will be a singleton if no other mentions refer to it. Given some unstructured sentences, such tasks will output co-referred word span pairs. Figure 6 presents this process.

3.2.1 Statistic-based Models. Early attempts to capture co-referred linguistic objects focused on the statistical features of entities, mentions, and antecedents.

Graph-based and tree-based models are popular designs for converting coreference resolution into a statistic-based partition task. These models construct a hypergraph from the given document, in which each edge can link more than two nodes for modeling coreferences among multiple mentions. Cai and Strube [76] learn statistical features to weigh edges and obtain coreference partitions via clustering algorithms. Sapena et al. [77] further employ relaxed labeling to interpret coreferences. Researchers have also simplified graphs for coreference resolution to adapt to tree-based methods. For example, Bean and Riloff [78] introduce a decision tree model to distinguish anaphoric mentions combined with context features. Fernandes et al. [79] leverage a voted perceptron algorithm to detect mention-pair coreference trees. Previous methods provide lightweight solutions based on intriguing feature engineering. Nevertheless, the performance of

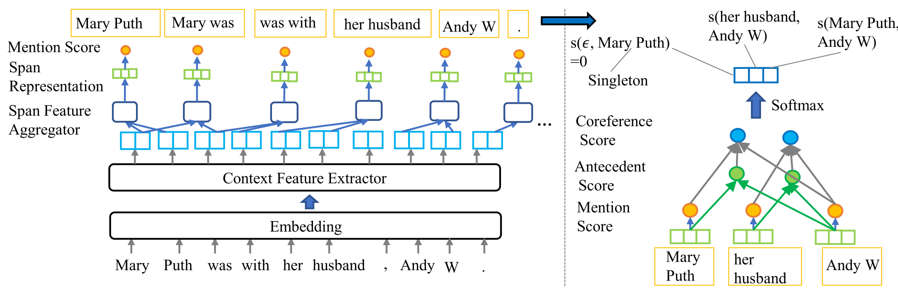

Fig. 7. Architecture of the standard deep end-to-end CO model [20]. A deep learning model performs a two-stage procedure to tackle coreference resolution tasks: (1) Mention detection, which discovers entity mentions as spans from texts; (2) Coreference detection, which scores the antecedents in the span to match coreference mention pairs as outputs. Spans include combinations of all word sequences. This figure displays simplified results.

such designs may drop when handling long-distance anaphorical pairs. It is also hard to design statistical features generalizable for diversified expressions.

3.2.2 Deep Learning-based Models. Deep learning models automatically convert a document input into word representations to collect features for detecting complicated coreference mention pairs in more diversified scenarios. Meanwhile, more tagged data and computation resources are needed for training them.

Many early models are based on CNNs. Xi et al. [80] incorporate distant features with hierarchical mention-pair features and score mention pairs via a softmax layer to resolve coreferences, especially long-distance anaphors. Wu et al. [81] develop a CO model to effectively handle coreference and singleton expressions with abundant multi-scale contexts, incorporating context feature combinations of antecedents, mentions, and mention pairs features via convolution and concatenation.

RNN and its variants better extract global features between word mention pairs than CNN-based methods. Wiseman et al. [82] propose the early feasible RNN-based CO model. Lee et al. [20] further develop an end-to-end LSTM-based model, detecting internal dependencies within mentions spans to comprehend global contexts that surround the spans. Figure 7 depicts this standard CO model. Gu et al. [83] apply a cluster modification algorithm to LSTM to rule out dissimilar pairs.

Attention mechanisms further model semantic interactions for CO tasks by emphasizing generalizable features. A good example of this is the Bi-LSTM structure enhanced with the word-level attention presented in [20]. However, many different coreference resolution-specific attention mechanisms have been developed to exploit coreference features. These include: the biaffine attention model for CO tasks [84] that captures word span interactions for detecting linked expressions; and the mutual attention model [85] that incorporates syntactic features with interactive features between dependency structures and antecedents for word spans. Further, Clark and Manning et al. [86] employ a RL-based strategy to enhance the robustness of their neural CO model in various expressions, which uses a heuristic policy network to filter out wrong coreference matching actions.

3.3 Relation Extraction

Relation extraction (**RE**) tasks extract relational facts from unstructured or semi-structured data to indicate interactions and properties among entities. Relation extraction, as a downstream task, is often called **relation classification** (**RC**). Binary relation extraction extracts relation triples

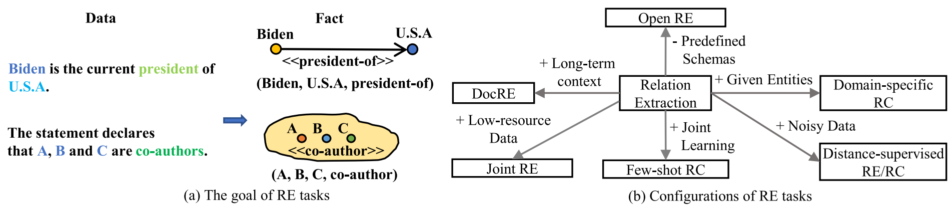

Fig. 8. Relation extraction. The goal of relation extraction tasks is to extract factual triples from data and find edges to link nodes together. For n-ary relations, the link is a super-edge that covers multiple nodes.

between entity pairs, while n-ary relation extraction obtains relation triples over multiple entities, such as co-authors. Relation extraction endows a knowledge graph with semantic links. Figure 8 presents an overview of the relation extraction tasks.

3.3.1 Domain-specific Relation Classification from Unstructured Sentence-level Data. Given unstructured sentences with conceptual (entity) mentions, domain-specific RC tasks label the given mentions with relation tags in a pre-defined relation set given the context of the sentences. Kernel methods and deep learning frameworks typically handle relation classification as a multi-label single-class classification task.

SVM kernel-based methods employ the feature vectors of words to train a classifier for supervised RC tasks on unstructured text. These models map specific semantic objects onto a feature space via a kernel function for classification, such as with a lexical-kernel-based SVM (with POS and entity tags) [96], a dependency-tree-kernel based SVM [97], or a shallow-parse-tree-kernel based SVM [98]. However, a high-performance kernel function can be hard to design.

To this end, deep learning-based frameworks are designed to automatically collect entity-related contextual information for RC tasks. Given a sentence that needs its relations classified $\{x_1, x_2, x_{eh}, \ldots, x_{et}, x_n\}$, let x_{eh} and x_{et} stand for head and tail entities, respectively. Deep learning models will generate a representation for each word: $\{\mathbf{w}_1, \ldots, \mathbf{e}_h, \ldots, \mathbf{e}_t, \mathbf{x}_n\}$, then the feature extractor will derive a vector $\mathbf{r}$ to indicate the probability of each relation type.

Classical convolution-based model, such as a feature-based CNN combined [87] with lexical features and a max-pooling strategy, focus on relations expressed in local contexts within neighborhood words of entities. Nguyen and Grishman [88] use multiscale convolution windows to enhance local feature aggregation. CNN-based designs are portrayed in Figure 9(a). For relational facts requiring a whole long sentence to convey, LSTM-based frameworks capture long-distance reliances between each part of a sentence to gain global context awareness to obtain these facts. Zhou et al. [22] use a BiLSTM that employs inter-word attention to capture the long-distance dependencies of relations, while Miwa and Bansal [91] incorporate tree structures into an LSTM framework. Architectures with LSTMs are presented in Figure 9(b). The recursive parts of the above frameworks slow down the training process. Thus, many designs have also incorporated global context features into CNN structure via attention mechanisms to model salient interactions, such as Attention-CNN [89] selecting entity-relevant contexts with the word-level attention and Multi-level CNN [90] developing an input attention mechanism with attention-based pooling. Figure 9(c) outlines the structure of attention units. Further, to incorporate models with auxiliary information, more recent studies have explored graph-level contexts via GCNs and extracting background knowledge via pre-trained models. Examples of this approach include EPGNN [92], which proposes an entity pair graph for a GCN (with a pre-trained BERT model); AGGCN [93], which integrates a multi-head attention mechanism for graph convolution; and RIFRE [94], which further employs a heterogeneous graph network to merge high-order features. GCN frameworks are presented in Figure 9(d).

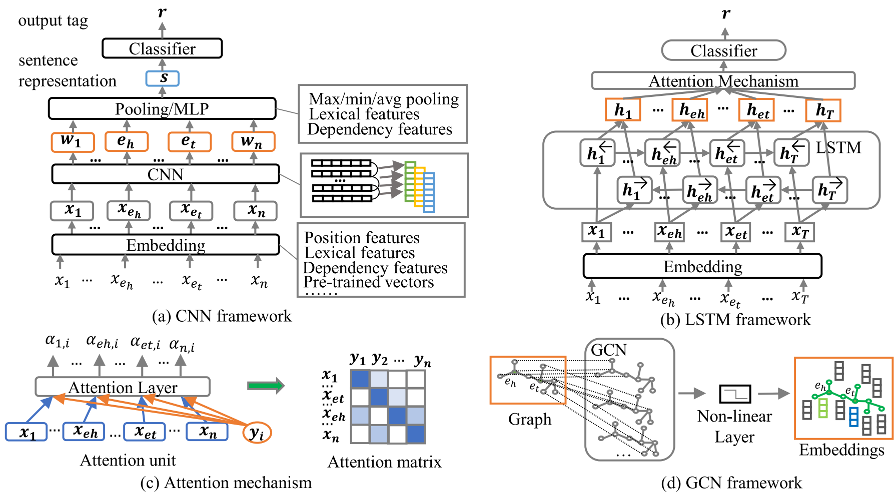

Fig. 9. Some frameworks of classic RC models. In this figure, (a), (b), and (d) display the architectures of CNNs, LSTMs, and GCNs, while (c) displays the structure of the widely-used attention mechanism.

Table 2. Comparison of Designs of Classical and Recent Sentence-level RC Models

Category	Model	Architecture	Background Information
Local context-aware	CNN [87]	CNN + max pooling	WordNet hypernyms, position features
	Multi-kernel CNN [88]	multi-kernel CNN + max pooling	Position embeddings
	Attention-CNN [89]	CNN + word-level attention + MLP	Pre-trained word-vectors, position embeddings, WordNet, POS
	Multi-level Attention [90]	word-level input attention + CNN + attention-based pooling	Pre-trained word-vectors, position embeddings
Global context-aware models	BiLSTM + Att [22]	BiLSTM + word-level attention	Pre-trained word-vectors
	TreeLSTM [91]	BiTreeLSTM + compound label embedding	SPTree, WordNet, position embeddings
Graph context-aware models	EPGNN [92]	BERT + CNN (sentence encoder)/GCN (topological encoder)	Pre-defined entity pair graph, pre-trained model, position embeddings
	AGGCN [93]	GCN + Multi-head Attention + DC + FF	pre-trained word-vectors, position features, dependency graph
	RIFRE [94]	BERT + HGCN	pre-trained model
Task conversion-based	QA [95]	BERT + Span Prediction Model	Converted questions/answers, pre-trained model

*The methods are arranged in terms of publication years in each category.

Cohen et al. [95] convert relation classification into a question-answering task and incorporated BERT embeddings for classification. Table 2 compares key design aspects of the popular models.

Some tasks require a model to handle *n*-ary relationships between multiple entities. To this end, semantic role labeling solutions have been devised to decompose *n*-ary relations into binary ones. Examples include NNF [99] and dependency path embedding [100].

Adding more nodes or extending neural network depth does little help to a conversational neural network model for fitting lengthy texts, noisy data, or other complicated big-data scenarios. It is worth reminding the developers about this "no-silver-bullet" principle. Constantly searching for novel architectures and learning strategies for adapting to data environments is necessary.

3.3.2 Open Relation Extraction from Semi-structured Data vs Unstructured Data. Open RE tasks discover facts from unstructured data without pre-defined relation types. These techniques detect nominal words (as the subject or object) and verbal phrases (as the predicate) from free text to form knowledge triples like (subject, predicate, object).

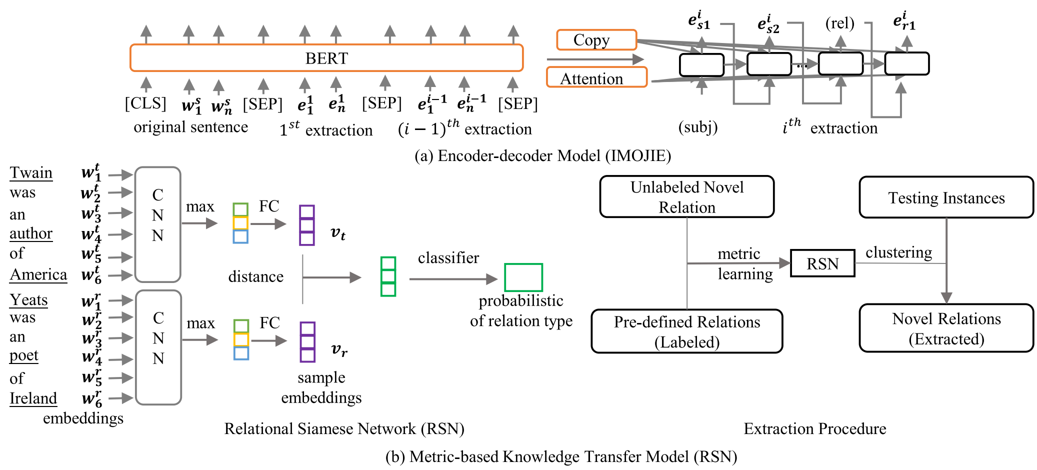

Fig. 10. Two paradigms of deep learning-based open relation extraction. In this figure, (a) shows the IMOJIE model [101], which extracts facts via an encoder-decoder design. (b) portrays a model [102] that uses the RSN to compare relational patterns, then leverages clustering to collect the relations.

Statistical approaches are trending solutions for open relation extraction. In terms of relation detection, models based on probabilistic-graph are popular designs for allowing contextual information to flow through semi-structured structures or unstructured free text. Mulwad et al. [103] put forward a probabilistic graph on semi-structured tables and semantic message passing for tagging relationships. Chen and Cafarella [104] leverage a module based on the CRF structure with a frame finder to tag cells with their location labels (such as left, middle, and right). From this, a hierarchical tree is built where relation triples can be recovered through parent-child structures. Researchers have also applied heuristic probabilistic models to relation classification with text. As an example, StatSnowball [105] employs Markov logic networks to identify relationships. Previous methods rely on empirical features to match relational patterns, with which, uninformative triples will still be generated when coping with intricate utterances or rare expressions.

Deep learning models have thus been developed to handle open RE tasks. A common framework is an encoder-decoder model designed to acquire factual patterns. CopyAttention [118] includes a mechanism to copy words from input to output sequences via a neural bootstrapping strategy. However, CopyAttention cannot adjust the count of extracted triples fitting the contexts of sentences with varied lengths, where redundant triples emerge [101]. Hence, IMOJIE [101] improves CopyAttention with BERT-LSTM structures while incorporating an unsupervised aggregation scheme to perform iterative extraction adapting to multi-scale contexts. See also Figure 10(a). Previous solutions have taken critical steps to retrieve relational triples in open domains, while rare relations in complex expressions remain disregarded. Hence, more efforts attempt to transfer supervised knowledge to a model so as to adapt known relations to obtain unsupervised low-resource relations. In this vein, Wu et al. [102] develop a metric learning-based solution that combines **Relation Siamese Net** (**RSN**) with the clustering strategy to discover new facts. See also Figure 10(b).

3.3.3 Distant Supervised Relation Extraction/Classification. Laborious label tagging on large datasets for fully-supervised configurations is formidable. To cope with this problem, Mintz et al. [119] develop the distant supervision strategy to annotate relation labels with an external knowledge base. This measure assumes that entity pairs appearing in different sentences reflect the same

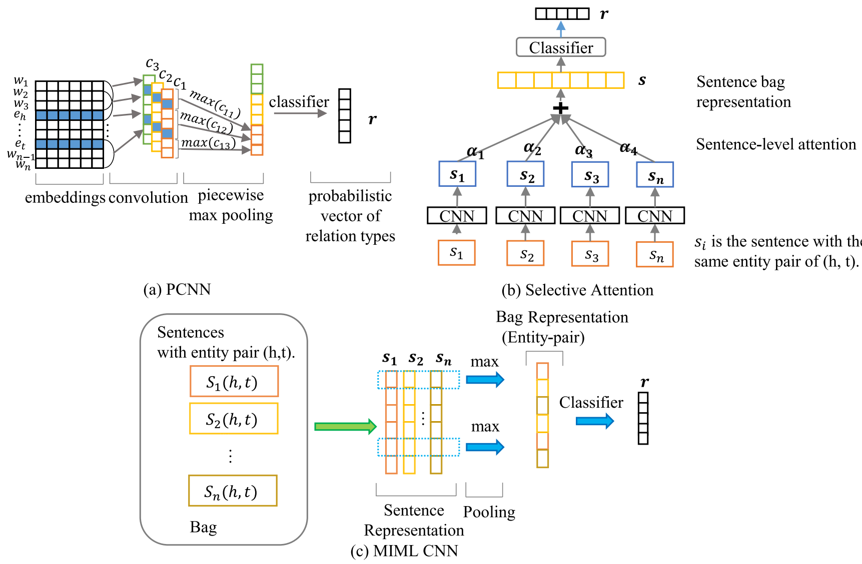

Fig. 11. Milestone models for distant supervision. (a) depicts [21]. (b) depicts [106]. (c) depicts [107].

relationships that link them in the knowledge base. However, distant supervision does not fully consider data contexts, hence, inevitably suffers from noise.

Early methods try to overcome data noises by enhancing the conventional feature extractors. **Piecewise convolution** (**PCNN**) [21] divides a sentence into three separate pieces for convolution according to entities to preserve critical contextual features. See also Figure 11(a). Transition matrix structure [108] learns incorrect patterns to divorce noise features. More models primarily improve learning strategies to promote robustness to noises. Qin et al. [109] leverage the reinforcement learning strategy to remove wrongly-labeled samples, while DSGAN [110] picks reliable samples for training via adversarial learning. However, early models with enhanced extractors or improved learning strategies remain limited, for they only consider surficial features to treat inescapable noises without a wide systematic view of distant-supervised instances.

More recently, efforts have thus focused on designing instance selectors to compare reliable features of instances in sample bags. For instance, Riedel et al. [28] develop the "at-least-one" hypothesis for **multi-instance multi-label Learning** (**MIML**). The hypothesis holds that at least one of the samples containing the same entity pairs will express the given distantly-supervised relation (i.e., the sample is correct). Based on that, selective attention [106] presents a classic design that groups sentences labeled with the same relation tags. See also Figure 11(b). MIML CNN [107] uses a CNN to proceed with each sentence bag, then leverages a cross-sentence pooling operation to derive an entity-pair representation for multi-label relation modeling. See also Figure 11(c). Ji et al. [115] combine entity descriptions to enhance the MIML CNN. Another direction for implementing instance-level feature extraction is instance-level attention mechanisms. The contribution of each sentence representation is then scored across different groups with the same relation tags. Last, an attention-weighted contextual representation is generated for each relation type. Many models extend this idea with MIML designs. For noisy long-tail labels, HAtt [112] proposes a selective hierarchical attention mechanism to intensify decoder features in each instance bag. Another example is Intra/Inter-Bag Attention [113]. This method captures the sentence features of inner relations and outer bag-relation interactions via compound attention mechanisms and cross-relation attention [114], where Baye's rule is used to acquire the global similarities of bags of

Table 3. Comparison of Popular Models for Distant Supervision RE/RC Tasks

Category	Model	Architecture	Background Information
Enhanced feature extractor-based	PCNN [21]	PCNN + pooling	position embeddings
	TM [108]	PCNN + Transition matrix + Bag embedding	position embeddings
Enhanced learning strategy-based	RL-based [109]	RL-based data redistributor + CNN/PCNN + Result-driven reward	position embeddings
	DSGAN [110]	GAN + CNN/PCNN + Attention	position embeddings
	CCL-CT [111]	CNN/PCNN + Self Att + [NetAtt + SelfAtt] + CCL-CT	position embeddings
Instance feature-based	Lin et al. [106]	CNN/PCNN + selective attention + max-pooling	position embeddings
	MIML CNN [107]	CNN (sentence) + Cross-sentence max-pooling	position embeddings
	HAtt [112]	CNN/PCNN + selective hierarchical attention	position embeddings, relation hierarchy
	Ye et al. [113]	CNN/PCNN + Intra/inner bag attention	pre-trained model
	Yuan et al. [114]	PCNN (Sentence) + Cross-relation Cross-bag Selective Attention	position embeddings
Background information-enhanced	MIML CNN + ED [115]	description embeddings + MIML CNN	entity description, position embeddings
	RESIDE [116]	Bi-GRU (sentence) + Syntactic GCNN	Dependency graph, external KB
	Zhang et al. [117]	CNN/PCNN (sentence) + GCN + Knowledge-aware attention	external KG, position embeddings

different relation types. Further, researchers find previous models cannot integrate multiple MIML networks to deliberate correct features. Thus, Huang et al. [111] adapt collaborative learning to incorporate the proposed combined framework of NETAtt/NetMax to handle the interaction contexts.

Deep learning approaches also consider external knowledge that overrides abnormal features to improve distance supervised relation extraction, such as incorporating the KG embeddings of entities into models [117]. RESIDE [116] further uses a syntactic graph with side information for GCN-based representations. We compare these popular achievements in Table 3.

3.3.4 Few-shot Relation Classification. Low-resource scenarios, specifically, few-shot and zero-shot RC tasks, require a deep learning model to learn from a few examples, where the majority of knowledge types in a knowledge base could be expressed due to the long-tail phenomenon [123]. Generally, researchers have tried to amplify the usable characteristics of these low-resource configurations through three methods: metric learning, meta-learning, and domain adaptation.

Meta-learning enhances optimizers by reserving conveyable meta-information from limited supervision. **Model-agnostic machine learning** (**MAML**) [124] improves batch learning through a two-stage multiple gradient descent. Here, the model is trained to estimate the gradients of each relation type before all the sample types undergo general optimization with the estimations. Task-sensitive meta-information of respective relation types is hoarded in partial gradient values. Gradient estimation through separate backpropagation is also applied in more flexible manners. To intimate the paces for how the brain naturally comprehends concepts, MetaNet [125] utilizes the fast-slow mechanism to obtain high-order implicit relational meta-features of samples and the specific task. Both the meta learner and the base learner contain a group of slow weights and fast weights for optimization. Another critical problem elicited is catastrophic forgetting. To

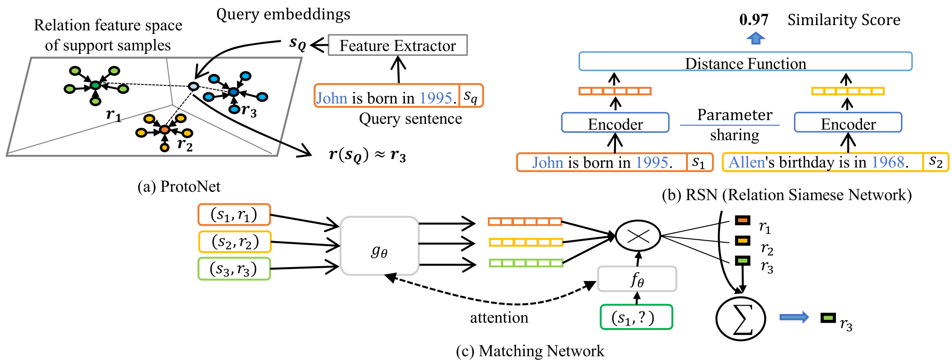

Fig. 12. The metric-based few-shot relation classification models. In this figure, (a) shows ProtoNet [120], which compares the distances of a query sample among support vectors of different relation types. (b) shows RSN [121], which calculates the similarity of sample pairs. (c) shows Matching Network [122], which uses an attention mechanism to tag a query by matching it with different-tagged support samples.

deal with this issue, Wu et al. [126] develop a curriculum meta-learning strategy that reviews samples in order and preserves the learned features in a memory mechanism. Pure meta-learning strategies enhance the capability for absorbing features in a few samples. Meanwhile, such ideas cease to conceive the essence of low-resource relation types, achieving insufficient advances.

Metric learning thus aims at finding metric spaces with which to profile different relation types for contrast. To determine the relation types in a query sample, ProtoNet [120], averages the embeddings of each relation type in a support set as a prototypical support vector. See also Figure 12(a). However, the previous method may not distinguish inter-class essentials from entangled features. To this end, LM-ProtoNet [127] concatenates phrase embeddings with sentence embeddings induced by a CNN to build fine-grained support vectors, which utilizes the marginal triple loss to decouple inter-class non-meta-features while improving inner-class cohesion. Noises in low-resource samples attract impertinent meta-features, reducing model robustness. To overcome this, Gao et al. [128] combine feature-level attention with instance-level attention to emphasize reliable prototypical features. Furthermore, RSN [121] compares the similarity of the sample embeddings. See also Figure 12(b). Matching Network [122] presents an attention-based embedding strategy for classification, calculating the attention scores of the query sample for each support sample through vector multiplication. See also Figure 12(c). As for long-tail rare relations, **multi-Level matching and aggregation network** (**MLMAN**) [129], aggregates the local and instance features by combining the support vectors with the query vector to match the correct long-tail class label for the query sample.

In terms of unseen types, few-shot domain adaptation maps unseen labels for classification. BERT-PAIR pairs with domain adaptation strategies for unseen "none-of-the-above" types. Gao et al. [130] discuss domain adaptation for few-shot relation classification as a game process for searching domain-invariant features. They implement domain adaptation via adversarial training. More domain adaptation strategies for few-shot relation classification can be found in [29].

3.3.5 Joint Relation Extraction Models. Conventional pipeline-based models suffer from error propagation in each stage, which also undermine inter-task interactions. Joint RE models are thus proposed.

Parameter-sharing strategies merge neural architectures for various types of tasks. They share weights and use different output layers to fetch entities that have relationships. Zheng et al. [131]

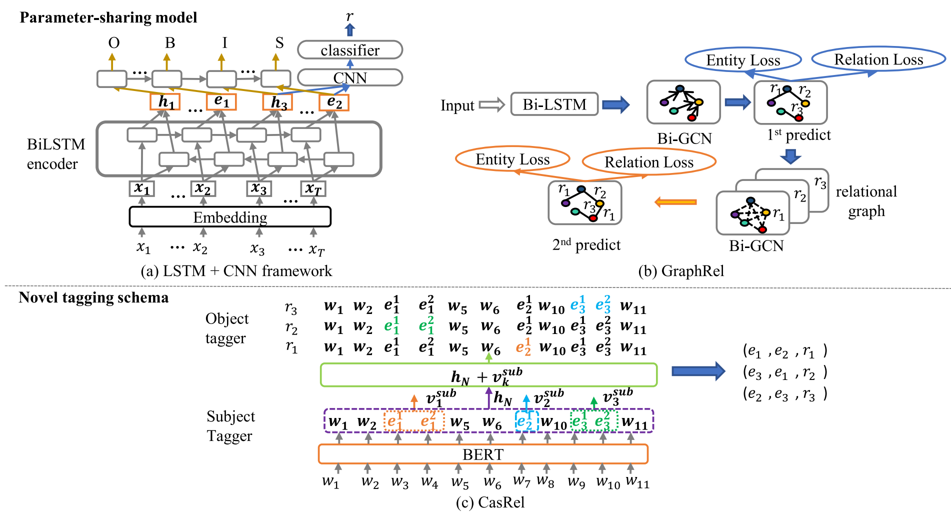

Fig. 13. The joint extraction model paradigms. (a) depicts [131], (b) depicts [132] and (c) depicts [133].

merge double BiLSTM layers for NER and RC tasks to share parameters, then use a CNN and an LSTM Network to label relationships and entities respectively. See also Figure 13(a). Miwa and Bansal [91] also integrate the dependency features associated with NER and RC with a combination of a Bi-LSTM and a Bi-TreeLSTM layer. Some models focus on delicate strategies for distributing cross-task characteristics. The GraphRel model [132], for example, intuitively leverages two-phase supervision to dedicate cross-task interaction through two respective BiGCN layers. This framework incorporates a dependency graph with a relation-entity graph to exploit deep features. See also Figure 13(b). Sharing parameters mix cross-task features to alleviate error cascading in training procedures. However, such ideas cannot extend to complex supervision signals with overlapped labels.

To handle overlapping labels, novel tagging schemes set joint decoding targets for the output layers with compound labeling. Zheng et al. [134] extend BIES labels with the relationship types and roles of a word (e.g., the subject or object of a sentence) to develop a sequence tagging task comprising named entity recognition and relation classification. Wei et al. [133] intuitively label all object candidates of a subject entity via a cascade map function for each relation type to contain overlapping mentions. See also Figure 13(c). Further, Wang et al. [135] develop a hand-shaking scheme to alleviate exposure bias within overlapping entities. Bekoulis et al. [136] devise a multi-head selection mechanism to explore all entity/relation combinations. Li et al. [137] turn entity-relation tagging into a multi-turn question-answering problem, leveraging the **machine reading comprehensive** (**MRC**) model for long-range semantics between entities. Unlike previous schemes, KGGen [138] directly generates triples via an encoder-decoder/generator structure based on a pre-trained model combined with adversarial learning, which overcomes feature reliance on entity co-occurrence information.

3.3.6 Document-level Relation Extraction. Entities in a document can express relationships via complex cross-sentence contexts, which defeats most of the traditional sentence-level context encoders. Novel architectures have therefore been conceived to capture document-level contexts.

Intra-sentence semantic passages are critical to document-level RE tasks. As such, researchers initially developed variants of an LSTM fitting graph structure to handle long-term dependencies, such as Graph LSTM [139] and Graph-state LSTM [140]. Previous immature methods only superficially profile obscure graph-level contexts with the absence of novel cross-sentence linguistic features, impeding further extraction.

To this end, more approaches develop new static document-level linguistic graphs to handle inter-sentence contexts via GCN-based models. For instance, Sahu et al. [141] introduce coreference edges and adjacent word edges to form a homogeneous document graph. As for *n*-ary relation extraction, AGGCN [93] links the roots of dependency trees of adjacent sentences via attention-guided GCN layers, which also overcomes the reliance on semantic role labeling. Christopoulou et al. [143] employ mention/sentence/entity (M, S, E) nodes to create a heterogeneous semantic graph distinguishing various linguistic roles, while reasoning via the EoG interference layer using the above intermedia node structures. However, static graph-based designs can be blinded when complicated inference is required.

Researchers then devise dynamic document graph models for high-order reasoning. Many models leverage dynamic edges. GP-GNNs [142] deduces hidden semantic logic with dynamic edge weights in a fully-connected graph for reasoning. LSR [145] regards graph structures as a latent variable to iteratively refine links and weights for constituting logical features from contexts. Xu et al. [147] consider reconstructing dependency paths to reweight relational entity pairs. GraphRel [132] jointly extracts entities and relationships via a two-stage procedure that incorporates static dependencies with dynamic relation-weighted graphs to enhance multi-hop reasoning. Some models also consider feature extraction with multiple graphs. For example, Zeng et al. [146] design a heterogeneous mention-level graph with an entity-level graph for multi-hop inference. Previous advances have not been explicitly guided by detailed features that pave the path to correct decisions, which will be hampered by intricate expressions.

To this end, more models are devised to reason with evidence. Zhang et al. [149] develop a rationale graph with external tagged co-occurrence evidence features for capturing long-term relational dependencies. Dynamic graphs with alterable nodes have also been considered within the realms of complex reasoning. DyGIE [144] prunes mute low-confident entity spans nodes through gate mechanisms for document-level feature exploitation. Other approaches seek to understand the common sense rules implicit in the different contexts. Discriminative reasoning network (DRN) [148] recognizes common-sense relationships while performing intra-sentence reasoning through heterogeneous graph representation features. The method is based on the assumption that multi-scale contexts with syntactic structures contain distinguishable common-sense features. Background common-sense features can also be acquired from pre-trained models like COMET [153]. However, understanding how common sense is expressed over contexts and how it unfolds in human logic remains challenging.

Empirical graph-level features still yield frustrating performances on diverse documents. Non-graph-based ideas are thus proposed for looking beyond. Previous GCN models unify the representations of each entity occurring in various parts, which invites dummy context throughout a document. ATLOP [151] thus introduces localized context pooling to distill the entity-relevant features contextualized in different parts with BERTs while using an **adaptive threshold loss** (**ATL**) for decoding with relation label interactions. Incorporating external knowledge from other trained models also enriches available features. As such, Tan et al. [152] incorporate a distant-supervised extraction model via the label-based **knowledge distillation** (**KD**) strategy, while improving upon ATL with an adaptive focus loss (**AFL**). However, how this model prevents noise transmission is not clear. Further, U-Net treats a document as a visual semantic matrix [150], employing a U-shaped segmentation for document-level reasoning via a multilayer convolution. We compare all typical milestone models in Table 4. Currently, little is clarified about how document-level contexts should be organized. More future efforts should be devoted to this challenge.

Table 4. Comparison of Model Designs for Document-level Relation Extraction

Category	Model	Word Encoder	Long-context Encoder	Inference	Cross-sentence Feature
Statistic Graph-based	Graph LSTM[139]	Embedding	Graph-LSTM	Softmax	Root-linked cross-sentence dependency tree
	Graph-state LSTM [140]	FFN	Graph-state LSTM	Softmax	Root-linked cross-sentence dependency tree
	AGGCN [93]	LSTM	GCN + Multi-head Attention	DC + softmax	Root-linked cross-sentence dependency tree
	Sahu et al. [141]	embeddings	GCN	MIL-based	Coreference/Adjacent sentence edges
Dynamic Graph-based Model	GP-GNNs [142]	Bi-LSTM	GCNs	MLP + softmax	Inter-node graph with generated parameters
	EoG [143]	BiLSTM	GCNN	Node-feature aggregation + softmax	Sentence-mention-entity pair graph
	GraphRel [132]	BiLSTM	BiGCN	Threshold-based	Relation-weighted graph
	DyGIE [144]	ELmo + BiLSTM	GCN + Span enumeration	FFN	Dynamic span graph
	LSR [145]	BiLSTM/BERT	GCN	FFN + GCN + DC	Weighted dependency graphs
	GAIN [146]	LSTM	GCN	FFN + Attention	hMG + EG
	Xu et al. [147]	BiLSTM	AGGCN [93]	LSTM + softmax	Reconstructed hetergenous S-M-E graph [143]
	DRN [148]	BiLSTM	GAIN [146]	Aggregation + MLP	Hetergenous document-level meta-paths
	RARE [149]	BERT	R-GCN	MLP + softmax	Rationale graph, pre-trained model
Others	U-Net [150]	BERT	2D-Conv	Matrix-based	Feature visualization
	ATLOP [151]	BERT	Localized Context Pooling	Group bilinear + ATL	Pre-train model
	Tan et al. [152]	BERT	Localized Context Pooling	Group bilinear + AFL + KD	Pre-train/Distance-supervised model

*The methods are arranged in terms of publication years in each category.
*Some models may consider other word encoders. Here, we present the basic settings.

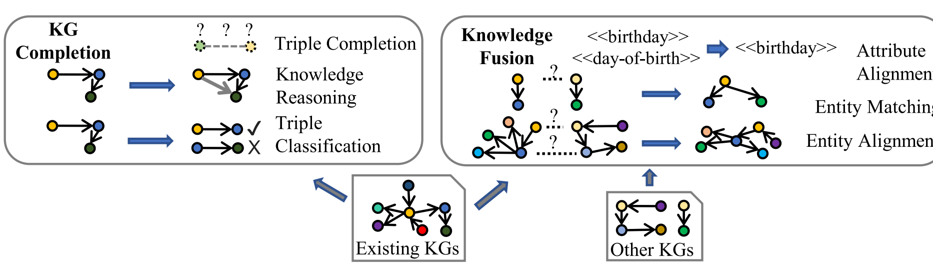

Fig. 14. Illustration of knowledge refinement.

4 KNOWLEDGE GRAPH REFINEMENT FROM STRUCTURED DATA

Raw knowledge graphs constructed from unstructured or semi-structured data can be sparse, and the knowledge triples can be incomplete or corrupted. Knowledge graph refinement repairs these problems through background semantics or by populating knowledge triples with additional knowledge graphs (structured data). The sub-tasks of knowledge graph refinement include knowledge graph completion and knowledge fusion. The general procedure is shown in Figure 14.

4.1 Knowledge Graph Completion

Knowledge graph completion fills in incomplete triples while deriving new triples from completed ones. In terms of completed triples, knowledge graph completion evaluates the accountability of each triple through triple classification. By accountability, we mean the correctness of the triples.

4.1.1 Embedding-based Triple Completion. A triple completion model searches for elements that can fill missing parts formulated as (h, ?, t) or (?, r, t) (entity prediction), and (h, ?, t) (relation prediction). For practical applications, empirical embedding-based models can achieve feasible performance with fewer parameters for tasks in small/medium KGs, while feature-enhanced embedding models incorporate external advanced semantics for high-performance completion in larger complex KG structures.

An empirical embedding-based model leveraged well-designated distribution feature spaces induced from shallow knowledge structures for completion. For example, the TransE-based model [154] searches the head entity h, the tail entity t, and the relation r, whose representations approach $\mathbf{h} + \mathbf{r} = \mathbf{t}$ to complete a triple. Later, researchers discovered that the previous symmetrical TransE model does not consider one-to-many relationships. Focuses then turned to import hyperspace structures (matrices) with distance-based translation models for link prediction, such as TranSparse [155]. Some models, such as TuckerER [156] and NTN [157], consider matching entity pair representations to a latent relational tensor semantic space for predictions with larger graphs. However, diversified knowledge structures can still be underestimated by them.

More recently, researchers have concentrated on feature-enhanced embedding models incorporated with complicated semantic knowledge structures. HAKE [158], for instance, uses the polar coordinate system to model semantic hierarchies in knowledge graphs, like hypernyms, hyponyms, and the apposition of an entity's ontological associations, which differentiate various-layered entity vectors by mold and angle constraints. CAKE [159] boosts negative sampling with common sense rules. Many models, such as SimKGC [160] and HaLE [161], optimize negative sampling for low-dimension embeddings via contrastive learning. CAFE [162] introduces a neighborhood sub-graph feature set to enhance relevant link information. Further, there has been more interest in decomposing the semantic constituents of knowledge representations with the sub-structures of knowledge graphs via semantic diffusion mechanisms of GCNs. DisenKGAT [163] discerns the high-order neighbor node features of a knowledge graph by disentangling the representation components into distinct semantics implied in the sub-graph structures. The hypothesis behind these models is that a large knowledge graph should contain sufficient subsets that can be reduced into k components to reason about linked entity nodes. Further, COMET [153] innovatively converts triples to word tokens for generative prediction tasks. Here, incomplete triples are filled by the transformer-based frameworks pre-trained with commonsense knowledge. Compared to the previous techniques, the generative model flexibility achieves breakthroughs in complex scenarios. However, extra computation costs remain the bottlenecks for these models.

4.1.2 Relation Path Reasoning. Relation path reasoning deduces new facts through completed triple sequences as support evidence, such as "(B, lives-in, Seattle) ⟵ (A, works-in, Microsoft), (Microsoft, located-in, Seattle)".

Many designs leverage neural networks that capture global features to find reasonable paths. Path-RNN [164] recursively aggregates relation path features for multi-hop reasoning. The chains-of-reason model [165] enhances a path-RNN with attention mechanisms to emphasize multiple-path dependencies with type information in the entities. The above pure neural architectures gather long-term dependencies in triples via black-box recursive feature extractors, while eluding to model the logic within sequences in an explicit manner.

Researchers then model the relation path reasoning tasks as a Markov decision process to traverse the KG environment for explicitly recognizing logical constraints. Deep **reinforcement learning** (**RL**) achieves this idea by learning a policy agent that assesses each selection step and expands the reasoning path. DeepPath [166] models the state space as (pre-trained) translation-based representations of entities and their induced relations. The taken actions then find the best matching relation labels via the feature space of entity pairs. Rewards for actions are calculated by a binary function. However, low-quality evaluations by the binary reward function will mean a RL-based model that is not well generalized to handle incomplete knowledge structures [167]. To this end, Lin et al. [167] devise a soft reward shaping function based on the vector spaces of relations and entities, while Li et al. [168] employ multiple agents to select entity pairs and relations. M-Walk [169] leverages an RNN to capture chronological state dependencies among pathing decisions. Chen et al. [170] further unify path-reasoning and path-finding tasks via variational encoding. RL-based models have surpassed non-RL RNN-based designs in tasks reasoning over length paths, however, vanilla policy networks cannot flexibly monitor a model to track global semantics in intricate relational paths.

Thus, more methods use attention mechanisms to augment features for RL models. ADRL [171] leverages a self-attention mechanism to emphasize neighborhood entity-relation interaction features. Similarly, Wang et al. [172] introduce a graph attention mechanism to enhance knowledge features. Recent research interest has been drawn into incorporating neural structures that handle intricate semantic features, such as the Hierarchical policy network of Zheng et al. [173], and DAPath [174], which incorporates a distant-aware mechanism to issue rewards via path length features. MemoryPath [175] is an attention-based memory component that preserves knowledge features for reinforcement learning and alleviates the model's reliance on pre-trained embeddings.

4.1.3 Interpretable Relation Reasoning. Interpretability serves to make a machine-learning model understandable to human users [176], and this plays a critical role in assessing a model's reliability. Interpretation models include pre-hoc models that contain self-explained procedures to reason over results and post-hoc models that provide indirect inspection mechanisms to analyze how features contribute to the outputs.

Pre-hoc reasoning models that comprise transparent decision processes can be self-interpreted through their inner structures by introspection. Logic rule-mining approaches such as AMIE [177] and RLvLR [178] can feed back the logic rules to explain linkage decisions to users. Some models only contain some components that are interpretable to humans (e.g., the learned rules). Users can observe these learned rules as side information when reasoning with rule-finding approaches based on neural models such as NeuralLP [179], pLogicNet [180], and ExpressGNN [181]. However, these neural networks are still black-boxed. Mainstream partial pre-hoc models also include models based on random-walk (probabilistic values for potential paths), reinforcement learning (reward values for each action), and attention (attention score for salient correlation).

Post-hoc interpretation methods develop proxies to probe into implicit features in black-box models like matrices and neural network frameworks. Some proxies extract rules or learn a probabilistic distribution to reproduce a model. Carmona et al. [182] train Bayesian networks with first-order logic to extract rules from embedding models. OXKBC [183] generates plausible explanation paths through the similarities between relationships and entities. Model simplification cannot decompose the features of non-linear neural models that are entwined. One solution is to conduct a sensitivity analysis to exploit the deep features. The analysis would involve imposing small perturbations on the models so as to observe how the output changes. These changes reveal influential features. GNNExplainer [184] explores sub-graph structures that affect single-instance and multi-instance predictions. CRIAGE [185] generates false facts to evaluate model performance and to locate obtrusive fact triples for each relation.

4.1.4 Triple Classification. Triple classification aims at distinguishing triples with surety from abnormal (untrue) triples in a knowledge graph. Many semantic models are designed for this task of judging suspicious triples in a knowledge graph that is constantly updated with novel relation types and facts.

Negative triple samples give knowledge representation models expressiveness to judge disordered triples. CKRL [186], for example, proposes an index system to determine reliable triples, including local triple confidence, which compares the distance between a triple and a negative sample; global path confidence, which tests the global resources of the reasoning paths that form a triple; and adaptive path confidence, which scores a local reasoning path deriving a triple. This solution inspects triples systematically to ensure accuracy but is intricate for training.

Embedding-based models are also considered for simplified triple classification. However, many potentially reasonable triples are not covered due to insufficient negative sampling specifically, one-to-many relations [4]. Hence, researchers have leveraged more sophisticated semantic structures to alleviate this issue. In this vein, Dong et al. [187] expand entity embeddings into n-ball structures that are leveraged to incorporate fine-grained type chains as a way to classify triples. Amador-Domínguez et al. [188] add ontological information to enhance model-agnostic expressiveness. Some models focus on advanced neural network embeddings to detect credible triples. For example, R-MeN [189] captures latent dependencies among triples by employing a multi-head attention mechanism that generates memory-based embeddings.

4.2 Knowledge Fusion

Real-world knowledge is usually open for updates. In most scenarios, users should be able to add external knowledge to enrich existing external knowledge graphs. In this way, knowledge fusion is designed to merge semantically-equivalent elements such as "Trump" and "Donald Trump" so as to integrate new knowledge within novel concepts or facts. The sub-tasks of knowledge fusion include attribute alignment, entity matching with small-scale incoming triples, and entity alignment with a complete knowledge graph.

4.2.1 Attribute Alignment. An attribute triple indicates a property of a concept with a description value like a color, date, number, or character string. Users may use different terms to refer to the same attribute, such as "birthday" and "date of birth", where synonyms may lead to semantic sparse. Attribute alignment is thus purposed to unify attribute notations.

Many methods focus on aligning the semantic embeddings of attributes, with the linguistic premise being that two attribute names should be identical if their embeddings are close to each other. Some models leverage the similarity between attribute name strings to generate distributional embeddings, such as in [190] and [191]. Yang et al. [192] leverage a bag-of-words model to learn the contextual embeddings of attributes. Similarly, JAPE [193] leverages the Skip-gram model for attribute embedding to model co-occurring attributes that are frequently used together to describe an entity, such as "latitude" and "longitude" for a position. Previous methods provide fast alignment solutions via shallow embedding learning. However, an attribute may not carry particularly informative data, like a telephone number, which can be challenging when attempting to generate knowledge-level representations.

Some models then consider using neural networks to generate deep embeddings based on contextual values, which also provide side information for entity alignment tasks, such as embeddings of definitions and descriptions. For example, AttrE [194] embeds each character of an attribute value with an LSTM framework so as to compose an attribute embedding for predicting potential phases in monolingual expressions. The approach incorporates an attribute-name predicate alignment strategy to handle unseen attributes.

4.2.2 Entity Matching with Small-scale Knowledge Graph. In its preliminary stages, a knowledge base only contains a few triple mentions with insufficient information for rigorous concepts. Therefore, entity-matching models integrate multi-source knowledge with the available linguistic information in small-scale data.

The more recent models treat entity matching as a classification task. For example, Magellan [195] integrates multiple similarity functions with random forest, such that the approach also considers numerical attributes. MSejrKu [196] explores the feasibility of leveraging the classifier layer including the logic regression and MLP classifier to judge identical entity pairs. DeepMatcher

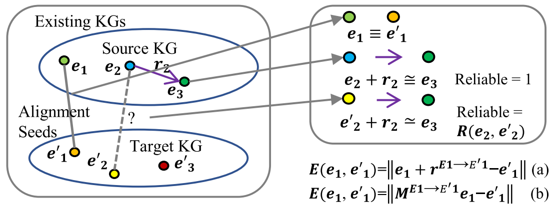

Fig. 15. IPtranE [200]. IPtranE scores entity pairs via (a) translation models and (b) linear transformation models, and merges identical pairs via hard or soft alignment.

[197] is a deep learning system that incorporates an RNN structure with attention mechanisms to represent attribute words for entity matching. Compared to conventional models, models based on deep learning are better at handling noise in text, especially the concept-enrichment tasks [198] with WordNet. These methods exploit useful information in small-scale KG structures. However, they cease to cope with unique information inputted from autonomous communities.

Unique personalized properties such as nicknames and telephone numbers, can be critical for entity matching but only known by users. Strategies to detect unique missing parts and ask users to fill them in are necessary. Active learning methods [199] that judge information or solve conflicts by querying users are the most reliable solutions and are indispensable in these scenarios.

4.2.3 Entity Alignment with Large-scale Knowledge Graph. Large-scale knowledge graphs usually comprise sufficient property information and graph structures that can form knowledge-aware structures with conceptual entities and relational links. Entity alignment tasks aim at integrating structured data with well-built large-scale knowledge graphs containing semantic structures at the knowledge level.

KG embedding-based models learn inter-graph entity mappings for EA tasks via seed entities that have the knowledge embeddings of triples. Sun et al. [201] point out that vanilla negative samples for link prediction can impair the ability to distinguish different entities of the same type. Hence, they use near entities in the feature space of a corresponding target entity to generate negative samples. IPTransE [200] is an iterative joint embedding strategy for knowledge representation and learning entity mappings. It leverages a path translation embedding approach to embed different relation paths linking the same entity pair. These are regarded as links with identical effects. A soft alignment strategy is then used to alleviate matching errors. See Figure 15. MultiEA [202] considers the multi-view features of entity graph attributes, links, and neighbor nodes. BootEA [201] includes a bootstrapped "likely alignment" labeling algorithm that iteratively adds reliable seeds for aligning. In cross-lingual scenarios, MtransE [203] generates axis calibration and translation vectors to model feature space invariance in different languages. Additionally, some models consider self-supervision strategies to exploit seed information, such as SS-AGA [204] and SelfKG [205]. Previous KG embedding methods have effectively considered distributional features to align entities as fast solutions. However, these practical ideas underestimate entity attributes, while high-order knowledge semantics are not well-considered.

More entity alignment models thus use attribute representation for feature augment, especially for the challenge of aligning entities that do not possess surface or structural distribution features. KDCoE [206], for example, leverages a co-training strategy with description attributes. JarKA [207] models interactions among attributes in a sparse multi-lingual knowledge graph to infer equivalent entities. Some models leverage deep learning-based neural networks for attribute context embeddings. For example, AttrE [194] leverages an LSTM to derive the dependency features of attribute

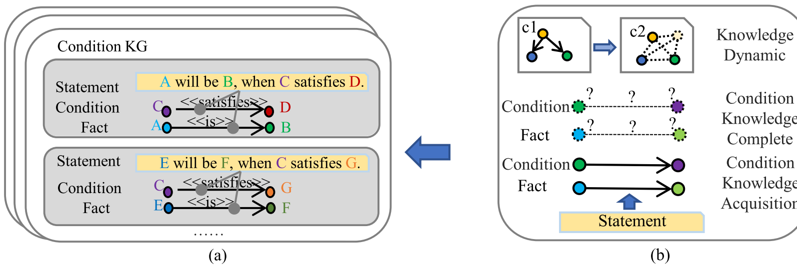

Fig. 16. Knowledge evolution. Evolution analysis tasks presented in (b) manufacture data into groups of knowledge graphs (either conditional or fact knowledge graphs) displayed in (a) to portray knowledge under various dynamic conditions.

values. Unlike previous methods, JAPE [193] consolidates attribute embeddings with overlay relationship graph structures to capture cross-lingual disparities.

Researchers consider graph convolution to explore advanced semantic graph structures for alignment. GCN-Align [208] is the first to propose a GCN-based framework for entity alignment tasks. Since then, recent research has focused on complicated graph semantics using GCN-based models. For instance, RNM [209] matches neighborhood node features to compare entity pairs. RDGCN [210] leverages a dual relation graph to solve contradictory representations in triangular entity graph structures. Further, distinctive semantic sub-graphs for alignment in large-scale KG structures also contain critical features for EA. Here, **graph matching neural network (GMNN)** [211] builds a topic entity graph that links neighboring nodes to merge identical entities. AttrGNN [212] partitions a knowledge graph according to attribute triple types to understand heterogeneous entity information.

Recent research direction also aims at modeling high-order global cross-graph interaction. MuGNN [213], for example, proposes a cross-knowledge graph attention mechanism with a multi-channel GNN encoder that can model inter-graph structural features consistently. Similarly, GTEA [214] involves a joint graph attention mechanism to fuse cross-graph relational information.

5 KNOWLEDGE EVOLUTION

Recently, researchers have focused on how knowledge evolves given environmental conditions. Conditional knowledge graphs serve this goal by reflecting facts established under certain conditions. A conditional tuple is formulated as (h, r, t, γ), where γ can be a prerequisite triple of a fact. Many researchers have studied this in its simplified case, as a temporal knowledge graph, where γ is some kind of temporal information (like a timestamp) - for example (Biden, job, vice president, 2009–2017), (Biden, job, president, 2020-). Figure 16 shows a schematic of knowledge evolution.

5.1 Condition Knowledge Acquisition

Many scientific facts are established upon certain conditions, especially in the biomedical field. Early efforts have not comprehensively considered this scenario in a systematic view. Specifically, Jiang et al. [31] note that the traditional extraction systems merge conditional information into entities to form factual triples, which will compromise entity linking. Further, the same tokens can be both subjects and objects of different tuples in an unstructured statement. Hence, Jiang et al. [215] develop a new tagging schema to describe conditional tuples formatted as "B/I-XYZ", where "BI" stands for positional information (begin/intermediate), "X" is the logic role (fact/condition), "Y" marks the tuple role (subject/object), and "Z" denotes the constituent type (concept/attribute/predicate). Based on that, conditional knowledge extraction achieves three goals: it

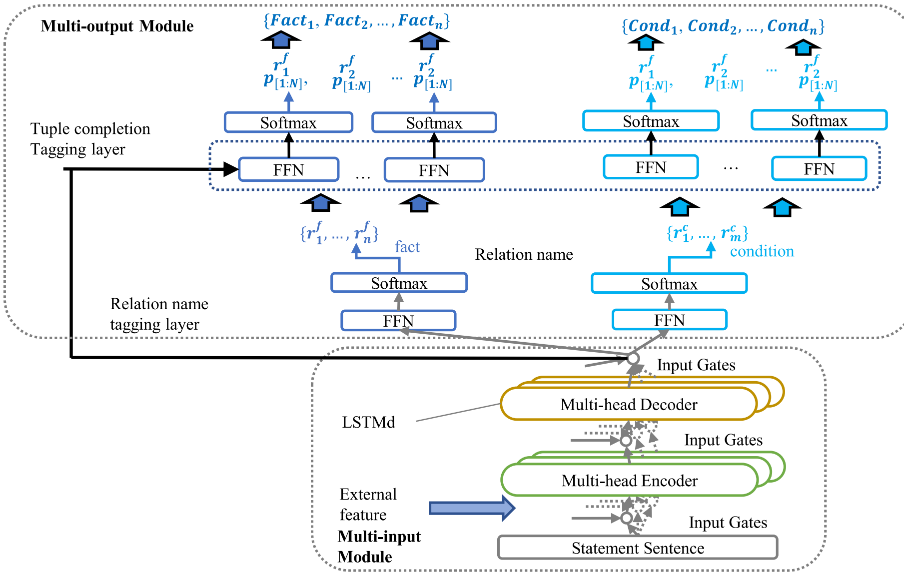

Fig. 17. The architecture of MIMO model [31] for extracting facts with conditions over texts.

extracts fact tuples, it collects conditional tuples, and it connects fact conditions. They then devise a joint extraction method based on the **multi-input multi-output** (**MIMO**) sequence labeling to tackle this problem. Their MIMO model leverages a relation name tagging layer that denotes the relationship tags for each token via factual and conditional tagging sub-layers, respectively. A tuple completion tagging layer is then used to distinguish the logic roles of each token with different relationship names. However, Zheng et al. [216] point out that the MIMO tagging schema cannot effectively handle overlapping triples. Therefore, their devised model leverages hierarchical parsing to simplify the multi-output schema in MIMO models into a one-output schema. Figure 17 illustrates the MIMO model.

Modeling generalized conditions remains challenging, thus current popular trend in conditional knowledge extraction is temporal knowledge extraction, where a conditional triple is simplified into time. Many previous models leverage RNN structures to capture temporal dependencies and therefore identify the temporal relationships within sentences, such as [217] and [218]. In terms of extracting fine-grained temporal knowledge, Vashishtha et al. [219] model events, states, and durations to match their timeline via multiple stacked attention layers. Recent research has improved solutions to handle document-level temporal knowledge extraction. For instance, TIMERS [220] is a rhetoric-aware graph for GCN models to interpret an intricate contiguous "elementary discourse unit" through the document's expressions. Here, an elementary discourse unit is the minimal semantic unit involved in temporal activities.

5.2 Condition Knowledge Graph Completion

Condition Knowledge Graph Completion tasks fill incomplete triples in a Condition KG, such as (h, ?, t, γ), (h, r, ?, γ), and (h, r, t, ?). Note that, in this section, our main focus is on methods for completing temporal knowledge graphs. Matrix embedding-based methods fit for temporal completion in small-scale KGs, while tensor-based methods can perform better for large-scale KGs. GCN-based designs obtain complex chronological features to reason over temporal relational paths.

Researchers can predict incomplete temporal tuples by temporal information matrix embedding models. TTransE [221] extends TransE with temporal embedding vectors. HyTE [222] treats the timestamp as a hyperplane for matching entity and relation embeddings. Another direction is temporal-aware embeddings. In this stream, the LSTM-based model [223] interprets time-encoding sequences, while the CNN-based model [224] captures the temporal consistency of contexts.

Temporal knowledge graph representations can be regarded as tensor structures along the temporal dimension, which means tensor decomposition can be used to complete temporal knowledge graphs. The main solutions for tensor decomposition include canonical polyadic decomposition and Tucker decomposition. Canonical polyadic decomposition uses the sum of several one-rank tensors to approach a target tensor. Many temporal knowledge graph completion models use canonical polyadic decomposition, e.g., [225] and [226]. Tucker decomposition factorizes a target tensor into the multiplication between a kernel tensor and multiple tensors along each dimension of the target tensor. Shao et al. [227] develop a model based on Tucker decomposition to interpret temporal semantic associations that increase the flexibility of representations that include timestamps. SpliMe [228] obtains time-viewed entity embeddings via a static model.

Another critical topic for temporal knowledge graph completion is temporal knowledge reasoning. Recent research interest has focused on GCN-based methods. Here, Han et al. [30] exploit historical contexts by expanding a query-dependent interference subgraph based on edge attention scores. Jung et al. [229] achieve multi-hop temporal reasoning via edge-based attention propagation, while Liu et al. [230] enhance temporal knowledge graph reasoning via a model based on reinforcement learning. Moreover, facts in a timeline cannot ignore temporal dependencies, such as "born-in" before "works-at". Jiang et al. [231] defines a scoring function that contains an asymmetric matrix to preserve temporal ordering constraints for reasoning.

Filling in incomplete general conditional tuples is open for further exploration. Tuples may contain more than one condition, such as chemical reactions that only occur within a certain temperature range. A systematic solution should be put into these complex scenarios. We suggest that readers also consider causality discovery methods [232].

5.3 Knowledge Dynamic

Many researchers have contributed to the literature on knowledge dynamics. A good proportion uses RNN structures to understand diachronic dependencies so as to predict state changes. For example, Know-evolve [233] involves a multivariate temporal point process with an enhanced RNN structure that learns a temporal evolutionary representation function. RE-NET [234] incorporates a neighborhood aggregator to seize concurrent interactions between entity nodes. Models have also been designed that contain evolutionary representations, such as MGraph [235] and DyERNIE [236]. Gracious et al. [237] systematically construct a neural latent space model that combines the evolutionary information of a heterogeneous knowledge graph. Yan et al. [238] improve a GCN model's ability to capture topology-invariant features. The idea is to align nodes in different temporal knowledge graph snapshots and build a dynamic profile of concepts.

How knowledge evolves when different kinds of conditions change remains challenging take the conditions needed to end the COVID-19 outbreak as an example. We recommend that readers refer to causality feature selection methods [232] along with experts and multi-source evidence.

6 DISCUSSIONS ON KNOWLEDGE GRAPH CONSTRUCTION

Researchers have contributed various solutions to different aspects of knowledge graph construction. However, some challenging issues and research directions are still open for further discussion.

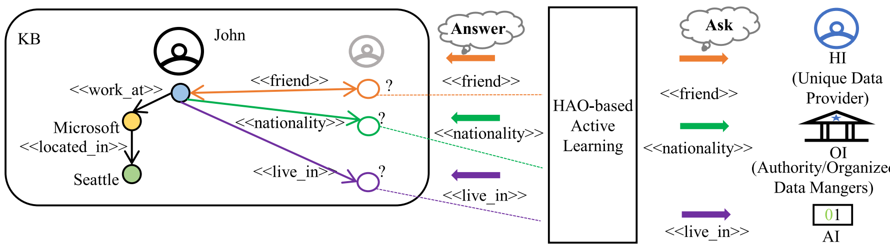

Fig. 18. An HAO-based active learning case for knowledge graph construction. HAO-based active learning models select users with appropriate roles to label uncertain samples. In this case (if the privacy policy allows), John's nationality will be labeled by the authority (OI), while his friends will be found by asking users (HI) in his social network. The AI will then derive his place of living from known facts.

6.1 Cross-lingual Knowledge Graph for Complex or Minority Languages

Building cross-lingual knowledge graphs is a long-term goal referring to integrating imbalanced resources distributed in different languages. Xlore [42] provides an enlightening example of aligning cross-lingual entities via deep learning approaches. However, machine translation remains a formidable bottleneck to many cross-lingual tasks. Firstly, errors and conflicts may emerge while translating complex expressions, compromising efforts for refinement. Secondly, data resources expressed in minority languages may be insufficient for machine learning. As an ultimate example, Icelandic, suffering from the crisis of "digital extinction" [239], may not present some corresponding words for translation. Currently, language experts play an irreplaceable role in these cases. Cross-lingual KG systems may also deliver more implicit knowledge to help these languages overcome the crisis. We believe more ideas to solve such problems by AI will promote cultural diversity.

6.2 Role-aware Human-machine Synergy for Knowledge Graph Construction

Active learning asks users to complete and correct knowledge graphs, providing the ultimate solution for obtaining unknown facts in the open world. Further, an improved strategy that can find appropriate roles to solve related questions will be a beneficial boost. To this end, Wu et al. [13] devise the HAO model to solve different construction problems by having humans and machines collaborate. Here, HAO stands for the collaboration of **human intelligence** (**HI**), **organizational intelligence** (**OI**), and **artificial intelligence** (**AI**). Correspondingly, an HAO-based active learning model will automatically identify these different roles and assigns undetermined data to appropriate users to tag. Such role-aware models will be a promising direction to endow wisdom to knowledge graph construction frameworks. We present an illustration of this significant idea in Figure 18.

6.3 Advanced Semantic and Dynamic in Knowledge Graph Construction Tasks

Recent research has extended to more advanced semantic evaluation tasks, such as detecting equivoque [240] and validating facts with common sense [241] to handle complex lingual phenomena. Interpreting literary expressions, such as similes and metaphors, is a future direction for intelligent KG construction, e.g., "Tom went to heaven in 2008." means " Tom, died-in, 2008". Developing pre-trained models with advanced semantics will be a starting point for high productivity.

Furthermore, many studies have been conducted on the dynamics of temporal knowledge graphs. However, how knowledge semantics evolves with the general associated conditions remains an unexplored field. Diving into heuristic questions such as "How do the professional social networks of medical staff change with the phases of a pandemic?" may help us detect implicit

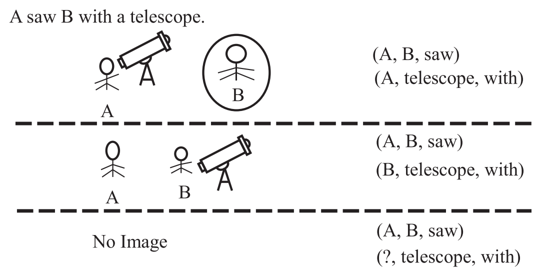

Fig. 19. An example of solving information incompleteness via cross-modal dependency. In this case, the textual expression "A saw B with a telescope." is unclear. An extractor can only determine the relationship between people and the "telescope" by interpreting the side information in the image data.

factors for boosting policy-making in public health. Capturing the dynamics of how associated conditions affect related facts is the ultimate direction for simulating general human knowledge.

6.4 End-to-end Unified Framework for Construction

End-to-end extraction methods, such as GCN-based frameworks, unify the sub-tasks of knowledge acquisition into one unified extraction task, surpassing pipeline designs. However, incorporating knowledge acquisition with knowledge refinement tasks to build an integrated joint model remains a formidable bottleneck. Generative models [242], turning multiple construction tasks into seq2seq generation tasks may also be a potential paradigm to unify more KG-related tasks. However, such designs still require intricate fine-tuning and task-sensitive model structures. Searching for end-to-end frameworks that unify both extraction and the refinement of knowledge graphs could be an enlightening future direction. Providing a high-quality off-the-shelf solution avoids the need for manual adjustments to components and one that considers cross-task semantics would be a worthwhile undertaking. Further, training a framework that unifies the general procedures of knowledge graph construction would be a worth-to-solve challenging multi-task learning problem.

6.5 Multi-modal Knowledge Graph Construction and Completion

Multi-modal knowledge graphs can entirely express and store heterogeneous information for display. They can also be applied to detect fake or low-quality content, such as text with mismatched images (e.g., a document labeled Paris with a picture of London). MMKG [243] is a model for completing a multi-modal knowledge graph by reasoning over image information, while Dost et al. [244] probe into cross-modal entity linking with text and images. How to effectively process more multi-modal information like videos remains a challenge. Another challenging direction is reasoning over multi-modal data. Given an input with a text expressed with ambiguity and a figure illustrating this text, semantic incompleteness in the one-modal text is formidable. Nevertheless, a model combining the figure with the words to comprehend cross-modal dependencies can interpret the correct denotation of the input. We present a specific example in Figure 19.

7 CONCLUSION

With this article, we delivered a comprehensive survey on the topic of **knowledge graph** (**KG**) construction. Specifically, we reviewed the tasks, methods, challenges, and related resources used to construct, refine, and integrate KGs from various data types in different scenarios. To probe into the essential topics for the big data environment, we systematically presented the paragon models for obtaining fine-grained concepts (entity typing), dealing with low-resource knowledge (extraction tasks in few-shot scenarios), understanding large linguistic objects (document-level relation extraction), complex reasoning (logic and interpretable reasoning) and handling conditional structures in knowledge graphs. Moreover, we provided briefs on practical KG toolkits and datasets. In conclusion, knowledge graph construction has become a critical topic for enabling human intelligence in AI applications. In future, the research community will certainly be searching for more paradigms to empower KGs with wisdom in massive heterogeneous, autonomous, complex, and evolving data environments while enhancing collaborations between knowledge communities.

APPENDICES

A APPLICATIONS OF KNOWLEDGE GRAPH CONSTRUCTION

Research communities have been incorporating **knowledge graph** (**KG**) construction techniques into real-world AI applications, besides tasks to build KG systems. Applications of KG construction extend methods to construct KGs into more user-concerned scenarios, integrated with KG side information and task-related approaches. These advances have covered fake news detection, dialogue systems, and more applications. In this section, we review some remarkable achievements.

A.1 Recommendation Systems

A recommendation system predicts how users interact with related objects (such as items and other users). Methods incorporating KB systems for this task have effectively handled the cold-start and data-sparsity problems in big data scenarios.

Current knowledge-enhanced applications also focus on building KG structures in user-item interactive social networks while combining KG side information via embedding-based approaches [245]. DKN [246] performs entity linking on the related news content items to the external **knowledge base** (**KB**) for obtaining a subgraph to learn knowledge embeddings. The model then utilizes an attention-based method to aggregate user embeddings for interpreting user-item interactions for recommendation tasks. To incorporate cross-task features via end-to-end training, KGeRec [247] utilizes feature interaction units to combine the KG construction and the recommendation task to gather deep conceptualized semantics. Meanwhile, previous designs do not fully consider high-order semantics within the graph structures. KRAN [6] proposes a GCN-based framework to obtain neighborhood features that denotes intrinsic user preferences. Here, the presented knowledge-refining attention mechanism captures inter-entity attention coefficients for the recommendation.

A.2 Fake News Detection

Fake news detection judges whether a factual statement reflects the truth. In practice, applications for such tasks apply NER and EL methods to dock on KG systems, then compare knowledge features with contextual features to determine trustworthy contents.

As an early achievement, Pan et al. [248] separately build KGs from real/fake news sources to obtain KG embeddings via the TranE model. The model then performs fact-checking by contrasting the related knowledge embeddings from different content sources. To further incorporate external existing KG information with content features, KAN [249] recognizes entities in contents and aligns them with an external KG, then incorporates feature embeddings via attention mechanisms to detect factual information. However, the previous methods underestimate long-term context characteristics in the content. Furthermore, CompareNet [8] utilizes a heterogeneous document-level graph to capture interactive features among sentences and compare embeddings with high-order contextual features to a KB. More knowledge-aware models also consider multi-modal information for assessment. For example, KMAGCN [250] unifies textual and multi-modal and KG information via an adaptive **graph convolution network** (**GCN**), achieving breakthroughs in different scenarios.

A.3 Dialog Systems

A dialogue system serves the demands for human-machine conversations in natural languages. A KG-based dialogue system traverses a KG system based on user inputs via KG walk approaches and performs task-related knowledge reasoning and semantic augmentation for generating consistent multi-round responses.

Many applications for knowledge-based multi-turn conversation tasks handle this challenge via encoder-decoder architectures. Moon et al. [251] aggregate dialogue-level semantics of inputs via multiple Bi-LSTM encoders and utilize a KG walker (reasoner) based on attention-based pruning for decoding responses. Similarly, KCMC [7] presents a generative seq2seq solution that consolidates hierarchical attention-based dialogue contextual encoder and knowledge-enhanced embedding via dynamic graph attention decoder for knowledge reasoning. KCMC generates fluent user-concerned responses through high-performance copying mechanisms rather than conventional ranking. Further, more efforts also concentrate on diversified users' interests in open-domain conversations. DKRN [252] intriguingly proposes a dynamic knowledge routing strategy of reasoning to retrieve conceptualized information for automatical personalized dialogues. Designing more effective solutions for complex conversations is now a prevailing trend.

B MORE KG-RELATED RESOURCES

B.1 More Practical KG Datasets

In this part, we portray more practical KG projects for readers. We provide information on the KG projects in Table 5.

B.1.1 Encyclopedia KGs. After the early attempt of the DBpedia project [40] (developed from Wikipedia), more KG projects incorporate automatic extraction tools, such as Freebase [9], Wikidata, and CN-DBpedia [253] (developed from Wikipedia, Baidu Baike, Hudong Baike, and automatically-extracted content). Max-Planck-Institution develops YAGO [254] integrates temporal and geographical structures in Wikipedia with WordNet ontology. Minz et al. [119] applied distance supervision Freebase for automatical entity-relationships annotation. KGs of eventualities are also concerned by the research community. CN-Probase [255] extends Probase with concepts in Chinese to comprehend general modes of textual data that involve uncertain occurrences.

B.1.2 Linguistic KGs. Besides WordNet [43], BabelNet [256] extends WordNet with cross-lingual attributes and relations of words from encyclopedias. ConceptNet [257], as a part of the Link Open Data, gathers conceptual knowledge based on crown sourcing, while HowNet [258] manually collects sememe (minimum indivisible semantic units) information of word concepts and attributes. THUOCL [259] records the **document frequency** (**DF**) of the words from the well-filtered web corpus. Developers create high-performance word embeddings based on well-built linguistic KGs for downstream applications.

B.1.3 Commonsense KGs. Besides OpenCyc [44], ASER [24] provides a weighted KG that describes commonsense by modeling entities of actions, states, events, and relationships among these objects, which acquire its nodes via dependency patterns selection and conceptualized by Probase. TransOMCS [23] develops an auto-generated dataset covering 20 commonsense relations obtained from linguistic graphs.

B.1.4 Enterprise Support KGs. Similar to **Google Knowledge Graph** (**GKG**) [45], Facebook Graph Search[4] delivers the powerful semantic search engine of Facebook, providing user-specific answers through the dynamic Facebook social KB.

[4] https://developers.facebook.com/docs/graph-api/

Table 5. The Information of Practical KG Projects

Categorization	Project	KG Inclusion	Year	URL
Encyclopedia KG	YAGO	2B+ facts, 64M+ entities	2007	https://yago-knowledge.org
	Freebase	360M+ fact triples	2007	https://freebase-easy.cs.uni-freiburg.de/dump/
	DBpedia	320 classes with 1, 650 different properties, 247M+ triples	2007	https://github.com/DBpedia/
	CN-DBpedia	9M+ entities, 67M+ triples	2015	http://kw.fudan.edu.cn/cnDBpedia/download/
	Probase	5.4M+ concepts	2010	https://concept.research.microsoft.com/
	Wikidata	96M+items	2012	https://www.wikidata.org/wiki
	CN-Probase	17M+ entities, 33M+ "is-a" relations	2017	http://kw.fudan.edu.cn/apis/cnprobase/
Linguistic KG	WordNet	117 000 synsets	1985	https://wordnet.princeton.edu/
	ConceptNet	34M+ items	1999	https://www.conceptnet.io/
	HowNet	35, 202 concepts, 2, 196 sememes	1999	https://openhownet.thunlp.org/download
	Babelnet	13M nodes	2010	http://babelnet.org/rdf/page/
	THUOCL	157K+ word nodes in 7.3B+ documents	2016	http://thuocl.thunlp.org/
Commonsense KG	OpenCyc	2M+ fact triples	1984	https://sourceforge.net/projects/opencyc/
	ASER	438M+ nodes, 648M+ edges	2020	https://github.com/HKUST-KnowComp/ASER
	TransOMCS	18M+ tuples	2020	https://github.com/HKUST-KnowComp/TransOMCS
Enterprise support KG	Google Knowledge Graph	500B+ facts on 5B+ entities	2012	https://developers.google.com/knowledge-graph
	Facebook Graph Search	dynamic social network of users, User-generated contents	2013	https://developers.facebook.com/docs/graph-api/
Domain-specific KG	Drugbank	14K+ drug entities	2006	https://go.drugbank.com/releases/latest
	AMiner ASN	2M+ paper nodes, 8M+ citation relations	2007	https://www.aminer.cn/aminernetwork
	Huapu	17M+ person nodes	2017	https://www.zhonghuapu.com/
	OAG	369M+ authors, 380M+ papers 92M+ linking relations	2017	https://www.aminer.cn/data/?nav=openData#Open-Academic-Graph
	COVID-19 Concepts	4784 entities, 35172 relation links	2020	http://openkg.cn/dataset/covid-19-concept
	Aminer COVID-19 Open Data	reports, news, research and other achieves of COVID-19	2020	https://www.aminer.cn/data-covid19/
	CPubMed-KG	3.9M+ tuples	2021	https://cpubmed.openi.org.cn/graph/wiki
Federated KG	GEDMatch	1.2M+ DNA profiles	2010	https://www.gedmatch.com/
	OpenKG.cn	200+ datasets from 94 organizations	2015	http://www.openkg.cn

B.1.5 Domain-specific KGs. Besides Drugbank [46], CPubMed-KG[5] innovatively develops a medical KG presented in Chinese. Many KG collection efforts are also contributed to fighting against the COVID-19 pandemic, such as the COVID-19 Concept dataset[6] and Aminer COVID-19 Open Data. As for academic activities, the **Academic Social Network** (**ASN**) of AMiner [260] and **Open Academic Graph** (**OAG**) [261] discloses academic activities including social networks and papers.

B.1.6 Federated KGs. Federation strategies have been applied to more KG systems with sensitive data to build integrated knowledge models while preventing data exchange. Researchers also focus on federated KG platforms. OpenKG.cn [262] as a crowd-sourcing community, provides a knowledge-sharing platform to develop knowledge applications with federated learning while supporting the decentralization of knowledge blockchains.

B.2 More Off-the-shelf KG Tools

In this part, we review more off-the-shelf tools for KG construction. The details of these tools are presented in Table 6.

B.2.1 Data Preprocessing. Many web crawlers support data preprocessing tasks that extract informative structures or contents. Besides WebCollector [47], Web Scraper[7] is a user-friendly

[5]https://cpubmed.openi.org.cn/graph/wiki
[6]http://openkg.cn/dataset/covid-19-concept
[7]https://webscraper.io/

Table 6. The Infomation of Off-the-shelf KG Tools

Task	Tool	Year	URL
Data Pre-processing	WebCollector	2016	https://github.com/CrawlScript/WebCollector
	Web Scraper (v0.4.0)	2019	https://webscraper.io/
Knowledge Acquisition	NLTK	2002	https://www.nltk.org/
	StanfordNLP	2002	https://stanfordnlp.github.io/stanfordnlp/
	KnowItAll	2005	https://github.com/knowitall
	TextRunner	2007	https://www.cs.washington.edu/research/textrunner/
	OpenCalais	2008	https://www.refinitiv.com/en/products/intelligent-tagging-tex
	ReVerb	2011	https://github.com/knowitall/reverb
	OLLIE	2012	https://knowitall.github.io/ollie/
	spaCy	2017	https://spacy.io/
	TableMiner+	2017	https://github.com/ziqizhang/sti
	MantisTable	2019	http://mantistable.disco.unimib.it/
	OpenNRE	2019	https://github.com/thunlp/OpenNRE
	gBuilder	2021	http://gbuilder.gstore.cn/
Knowledge Acquisition	Falcon-AO	2008	http://ws.nju.edu.cn/falcon-ao/
	OpenKE	2018	https://github.com/thunlp/OpenKE
	OpenNE	2019	https://github.com/thunlp/OpenNE
	OpenEA	2020	https://github.com/nju-websoft/OpenEA

manual extraction tool for collecting multiple web pages, which provides a user interface to reserve focused web structures and cloud server for massive content extraction.

B.2.2 Knowledge Acquisition. Early toolkits that directly extract fact triples through rules, patterns, and statistic features, which are also known as **Information Extraction** (**IE**) toolkits. Besides KnowItAll [12], more toolkits leverage semi-supervision designs to collect relational information, such as TextRunner [11], ReVerb [263] that produces refined verbal triples via syntactic and lexical information and OLLIE [264] that supports non-verbal triples discovery.

Many NLP applications can direct achieve knowledge acquisition sub-tasks, including NER, RE, and CO tasks, or provide linguistic features for their related applications. NLTK [265] and StanfordNLP [266] are powerful toolkits for knowledge acquisition based on statistic-based algorithms like CRF, MEM, which can also provide background features such as POS tags and NP chunks. TableMiner+ [267] and MantisTable [268] extract knowledge from semi-structured table forms.

Recent developers have been drawn into DL-based toolkits. spaCy [269] is a comprehensive practical NLP toolkit that integrates NeuralCoref for CO tasks, and also provides a trainable **deep-learning** (**DL**) module for specialized **relation extraction** (**RE**) (in spaCy v3). OpenNRE [270] provides various extensible neural network models such as CNN and LSTM to perform supervised RE.

B.2.3 Knowledge Refinement. Besides integrated DL-based toolkits OpenKE [49] and OpenEA [50], OpenNE integrates embedding models such as Node2VEC [271] and LINE [272] to obtain global representations from a complete KG for completions. As for the knowledge fusion (KG merging) task, Falcon-AO [273] utilizes multiple algorithms to measure semantic similarity for aligning concepts in different notations.

C MORE DISCUSSIONS ON KNOWLEDGE GRAPH CONSTRUCTION

In this part, we present more discussions over directions and challenges for KG construction techniques.

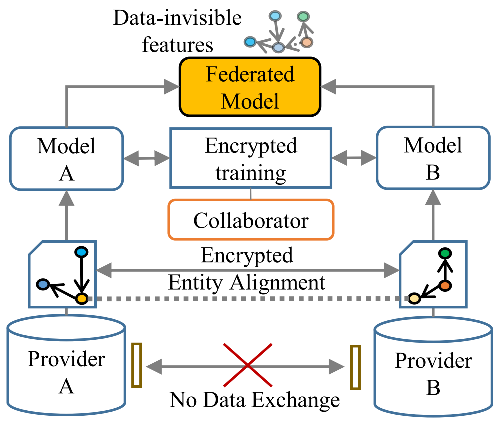

Fig. 20. An illustration of building a federated model from different knowledge providers while protecting privacy. In this procedure, an encrypted entity alignment process is performed before training separate models on multi-source data parts, then a collaborator calculates and aggregates encrypted gradients of each model to prevent leakage. A federated model only reserves data-invisible crowd-sourced knowledge features.

C.1 Strategies for Privacy Protection in Knowledge Graph Completion and Other Applications

Practical KG completion models and other high-performance KG-related applications rely on embedding learning on large datasets. However, these data sources may carry user-provided sensitive data, where privacy violations could happen if machine learning is directly performed. Strategies to prevent privacy violations are thus proposed to serve the mandatory requirement for safe KG applications. Current mainstream research mainly focuses on federated learning and differential privacy to achieve privacy protection.

Federated Learning is an enlightening direction to serve privacy protection while performing machine learning for KG completion and other tasks. A federated setting for KGs that trains model ensembles from multi-sources is one of the popular strategies. Significant advances have been conducted to federated knowledge embeddings to achieve safe KG completion, such as FKGE [274] and FedE [275], which prohibit data exchange while incorporating cross-modal features during training. However, entity alignment for knowledge fusion is a paradoxical bottleneck that impedes federated learning for complex completion tasks and other applications, requiring multi-source KGs to be shared before model learning, which will exchange sensitive information during knowledge fusion. How to create a privacy-reserved super feature space for encrypted entity alignment while federating features is still open for exploration. Designing more privacy-friendly models for constructing KGs is critical for sensitive data scenarios. We illustrate the procedure of developing a federated model in Figure 20.

Even trained on a sanitized database, a KG embedding model serving completion or other purposes may still expose sensitive information about identities when minor variations on these anonymized features are analyzed. Therefore, another challenging direction is proposed to thwart identity detection through such attacks on a sanitized database. **Differential Privacy** (**DP**) [276] is an insightful strategy to anonymize sensitive data so that each near data piece is not distinguishable from the other, thus, privacy attacks by contrasting minor features will be compromised. Currently, efforts have incorporated differential privacy in deep learning models, such as HGM [277] with robust DP strategies. Besides, ensuring the fairness of DP-outputted properties of different individuals or groups is still open for further discussion [278]. The objective of "fairness", requiring anonymized data output can be viewed equally, may have polarized impacts on various scenarios. As for candidate selection tasks, "fairness" can be guaranteed with well-designed DP strategies [279], while researchers find such "fairness" targets may impair privacy-preserving efforts in data of census statistics [280]. More ideas should be considered to handle the tradeoff between privacy and fairness.

C.2 Fabricated Information Detection

The information generated by generative AI or concocted by humans brings a new challenge for KG systems, requiring developers to distinguish unrealistic or unoriginal inputs from reliable data.

Unlike data noises handled by models for triple classification, information faked by humans on purpose is more confusing with hard-to-tell features. Current research works have focused on detecting fake news by comprehending context features, such as CompareNet [8] for text information and KMAGCN [250] for multi-modal information. Explainable models [281] are also considered to serve this goal. Detecting fabricated information from generative AI is more formidable. As for generated images, many models [282] based on deep learning have been devised to tackle face photos manipulated by DeepFake. More attention should also be paid to the generated data in other scenarios like events and sceneries. Furthermore, texts generated by a language model (like ChatGPT [283]) also present unoriginal or low-quality content. DetectGPT [284] proposes a solution by analyzing the probability distribution of a perturbed text to pick model-generated parts. Tools like GPTZero[8] also manage to judge such content by different means. However, these tools still present many inaccuracy judgments. Generative models with adversarial learning mechanisms may compromise the regular efforts for detection to cause this. It is also worth reminding readers that excessive data collection can unintentionally make a human write like an AI. Expert decisions and legislative actions would also be needed to alleviate these problems.

C.3 Incorporating Generative AI and Other Large Pre-trained Models for Knowledge Graph Construction

Large pre-trained models have been leading a significant impact on multiple KG construction tasks and their related applications. KG-BERT [285] incorporates knowledge triples as sequences to train with, achieving breakthroughs in KG completion tasks. DRAGON [286] proposes a self-supervision strategy to consolidate KG with textual information for obtaining deep advanced representations, effectively enhancing model performances on question-answering tasks with complex reasoning. Developing intriguing strategies utilizing large KG structures for pre-training language models to serve KG construction and other applications is promising. Directly extracting intrinsic facts from language models will also be a new direction for interpretable machine learning. Petroni et al. [287] discover that large language models only trained on contextual information like BERT [26] may contain or potentially possess oracle knowledge about entity relations. They propose a LAMA Probe to evaluate explicable relational knowledge within these models by a simplified cloze test. Furthermore, such attempts may also unveil how structured knowledge emerges from context awareness.

Recently, advances in generative AI have also delivered new directions for KG construction. A recent survey [242] has reported generative models that turn multiple KG construction tasks into seq2seq structure prediction tasks, which can better comprehend complex task-related knowledge. These advances serving generative KG construction have provided a promising unified paradigm

[8]https://gptzero.me/

Text	Mary and Henry planned for a trip. They went to England last week. Mary believed the Avon River was the most romantic sight in England, while Henry felt honored to be a compatriot of Shakespeare while walking around their hometown. They were sure that their daughter Lily would like this place as well.	
Inference	Fact	Detail
Coreference Reasoning	(Mary, Entity-Destination, England) (Henry, Entity-Destination, England)	Mary and Henry planned for a trip. They visited England ……
Pattern Recognition	(the Avon River, part-of, England)	$X was …… sight in $Y
Logic Reasoning	(Henry, nationality, England)	(Henry, compatriot-of, Shakespeare) (Shakespeare, nationality, England) ➡ (Henry, nationality, England)
Commonsesnse Reasoning	(Mary, partner, Henry)	(Mary, daughter, Lily) (Henry, daughter, Lily) ➡ (Mary, partner, Henry)

Fig. 21. An example of relation inference over long contexts in a document.

to probe into tasks like entity linking, RE, and KG completion. The survey also points out that training efficiency and the generation quality of models remain open for future improvement.

Another problem is absorbing generated information for KG construction, such as gathering data from commonsense text generation tasks. KG-BART [288] provides a solution based on pre-train models to obtain a text about a concept set from a KB. For example, given {fish, catch, river, net}, the text "fishermen use strong nets to catch plentiful fishes in the river." will be generated by the model. Such accountable data can enrich a KG system with background knowledge in a pre-train model. In terms of AI-generated factoid content, a KG system utilizes it for data augment. Such generated unseen objects like "a WWI battlefield using magic as weapons" are not likely to be reasonable real-world facts, which may not be the appropriate content for display. It is worth pointing out that completing a KG with fabricated data could be misleading or profile users with stereotypes (like generating a man's self-portrait based on his birthplace). Such technical abuses should not be encouraged.

C.4 Long and Intricate Contexts for KG Construction

Intricate cross-sentence or cross-paragraph contexts impede different KG construction sub-tasks for practical use, especially RE tasks. It is worth reminding readers that complex contexts do not merely relate to long-term dependency. Yao et al. [27] point out that four kinds of inferences including pattern recognition, coreference reasoning, logic reasoning, and commonsense reasoning, are also critical to contain high-order contextual semantics. A specific example is presented in Figure 21.

A model that handles complex long contexts should focus on intricate cross-sentence patterns while performing reasoning over multiple linguistic objects. Besides document-level extraction models in Section 3.3.6, Some efforts in Section D.2 also model document-level contexts via heterogeneous models for **entity typing** (**ET**). Noticeably, ambiguous expressions may occur in user-generated texts, which are usually not correctly interpreted by models without external information. Another challenging issue for reasoning is multi-hop reasoning. More linguistic structures should be explored to comprehend tortuous expressions.

Out-of-context expressions requiring background knowledge to handle are bottlenecks for KG construction. The obstacles are mainly two-fold: (1) spontaneous knowledge, and (2) evidence support. Commonsense knowledge spontaneously generated is often utilized to derive new facts, e.g., man and woman who have kids should be couples/partners, despite such convictions sometimes inaccuracy. How to obtain commonsense rules and adapt them to suitable scenarios is an important direction. Meanwhile, many document-level datasets do not contain evidence information for correct logic paths. Efforts like [289] have probed into document-level evidence structures for relation mentions. However, it is not likely to foresee that a model can learn to organize clues correctly to resolve facts in all scenarios (e.g., validating the conclusion in a philosophical book). We believe long-context is not merely an NLP question, and models [290] understanding linguistic expressions will be a critical direction. Furthermore, conditions like temporal and geographical information provided by data sources should also be considered for rigorously comprehending contexts.

D MORE ADVANCES FOR KNOWLEDGE GRAPH CONSTRUCTION

Besides mainstream models for different KG Construction sub-tasks, there are also many other innovative or practical attempts to work on different scenarios. In this part, we present more advances for enlightening readers to design novel construction solutions.

D.1 Rule-based Methods for Knowledge Acquisition

Many early attempts focus on rules that achieve knowledge acquisition or its sub-tasks. Despite inaccuracy in big data environments, rule-based methods are practical solutions to quickly extract massive raw knowledge. These methods also work in scenarios where high-performance computing is not available.

Rule-based approaches [291] are the general solutions for NER. As for semi-structured web data, Wrapper inductions generate rule wrappers to interpret semi-structures such as DOM tree nodes and tags for harvesting entities from pages. Some rule-based solutions are unsupervised, which require no human annotations, such as Omini [292]. As for entities in table forms, many approaches are proposed based on property-attribute layouts of Wikipedia, such as rule-based tools [40][254] for DBpedia, and YAGO. For unstructured data, classic NER systems [293] also rely on manually-constructed rule sets for pattern matching. Semi-supervised approaches are developed to improve rule-based NER by iteratively generating refined new patterns via pattern seeds and scoring, such as Bootstrapping-based NER [294].

Methods focusing rules are the earliest attempts for RE tasks on different data structure kinds, gathering strings that fit in hand-craft templates, e.g., "$PEOPLE is born in $LOCATION." refers to ($PEOPLE, born-in, $LOCATION). However, these unsupervised strategies rely on complex linguist knowledge to label data. Later, researchers concentrate on automatical pattern discovery for triples mining. Semi-supervision design is an enlightening strategy to reduce hand-craft features and data labeling that uncovers more reliable patterns based on a small group of annotated samples, such as DIPRE [295] iteratively extracting patterns with seeds, bootstrapping-based KnowItAll [12] and Snowball [296] equipping DIPRE with confidence evaluation. Some rule-based models consider more lexical objects for mining. OLLIE [264] incorporates lexical structure patterns with relational dependency paths in texts. MetaPAD [297] combines lexical segmentation and synonymous clustering to meta patterns that are sufficiently informative, frequent, and accurate for relational triples. Specifically for semi-structured tables, researchers design table structure-based rules to acquire relationships arranged in rows, columns, and table headers, such as [298]. Furthermore, Some semi-structured extraction systems utilizing distant supervision tolerate potential errors, which directly query external databases like DBpedia and Wikipedia to acquire relationships for the found entities in tabular data, such as [70], [299], and [300]. Similarly, Muñoz et al. [300] look up the Wikipedia tables for labeling relationships in tabular forms. Krause et al. [301] also expand rule sets for RE via distant supervision.

Rule-based models to perform end-to-end knowledge acquisition are lightweight solutions for specific domains. However, these designs require extra work for maintenance if the domain changes.

D.2 More Embedding-based Models

Embedding-based models lay the foundation for KG completion while providing semantic support for different sub-tasks for knowledge acquisition from semi-structured or unstructured data.

More variants for **translation embedding** (**TranE**) models for KG completion have been developed to search entity-relation feature space via mapping matrices like TransR [302] and TransH [303]. Meanwhile, researchers also consider more tensor-based empirical models for embedding over a completed large graph, such as RESCAL [304] and DistMult [305]. Some knowledge representation models leverage non-linear neural networks to exploit deep knowledge embedding features for KG completion, such as ConvE [306], M-DCN [307], and TransGate [308]. Unstructured entity descriptions are also incorporated for feature enhancement, such as the DKRL model [309] and ConMask model [310]. GCNs are also presented to encode a KG, such as R-GCN [311], W-GCN [312], and COMPGCN [313]. GCNs can also comprehend neighborhood information through semantic diffusion mechanisms. ProjE [314] projects an entity and a relation to distinctive feature spaces through neural operations for capturing another candidate for missing entities. However, when the relation element is missing, a latent vector space of relationship candidates cannot be retrospected. SENN [315] bridges the disparity-distribution-space semantic gaps by multi-task embedding sharing strategy unifying relation, head entity, and tail entity link prediction.

As for ET, novel embedding-based models avail of combing global graph structure features and background knowledge for predicting potential types of entities via representations. Researchers reported that the classical TransE model acts poorly while directly applied to ET tasks. Moon et al.[154] propose the TransE-ET model adjusting the TransE model by optimizing the Euclidean distance between entities and their types representations, limited by insufficient entities types and triples features. New solutions aim at constructing various graphs to share diversified features of entity-related objects for learning embeddings with entity-type features. PTE [17] reduces data noise via a partial-label embedding, which constructs a bipartisan graph between entities and all their types while connecting entities nodes to their related extracted text features. Finally, PTE utilizes the background KG by building a type hierarchy tree with the derived correlation weights. JOIE [316] embeds entity nodes in the ontology-view graph and instance graphs, gathering entity types by top-k ranking between entity and type candidates. Likewise, ConnectE [317] maps entities onto their types and learning knowledge triples embeddings. Practical models improving embeddings on heterogeneous graphs for ET tasks (in Xlore project [42]) also include [318], [319], [320]. We present graph structures for embedding model-based ET in Figure 22.

Embedding-based models are also critical solutions for entity linking via entity embeddings. LIEGE [321] derives distribution context representations to links entities for web pages. Early researchers [322] leverage **Bag-of-word** (**BoW**) for contextual embeddings of entity mentions, then performed clustering to gather linked entity pairs. Later, Lasek et al. [323] extend the BoW model with linguistic embeddings for EL tasks. Researchers also focus on Deep representations for high-performance linking. DSRM [324] employs a deep neural network to exploit semantic relatedness, combining entity descriptions and relationships with types features to obtain deep entity features for linking. EDKate [325] jointly learns low-dimensional embedding of entities and words in the KB and textual data, capturing intrinsic entity-mention features beyond the BoW model. Furthermore, Ganea and Hofmann [18] introduce an attention mechanism for joint embedding and passed semantic interaction for disambiguation. Le and Titov [19] model the latent relations between mentions in the context for embedding, utilizing mention-wise and relation-wise normalization to score pair-wise coherence score function.

Researchers also focus on embedding-based distribution models over multiple semantic structures to handle coreference resolution (CO). Durrett and Klein [326] utilize antecedent

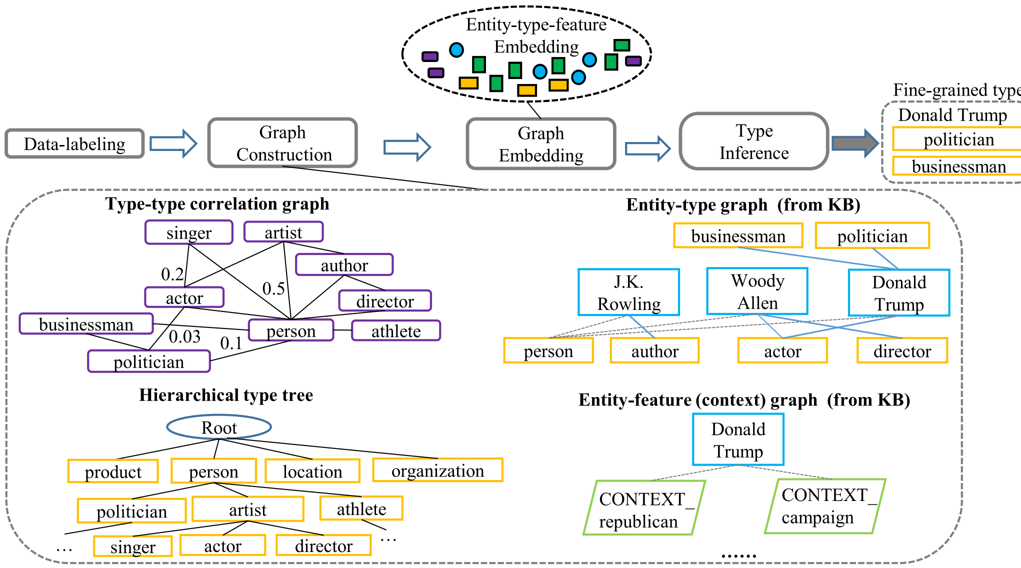

Fig. 22. Illustration of embedding-based ET via heterogeneous graph structures. (PTE [17]).

representations to enable coreference inference through distribution features. Martschat and Strube [327] explore distribution semantics over mention-pairs and tree models to enhance coreference representations, directly picking robust features to optimize the CO task. Chakrabarti et al. [328] further employ the MapReduce framework to cover anaphoric entity names through query context similarity.

As for joint RE, novel distribution embedding-based models are proposed to model the cross-task distributions to bridge the semantic gaps between NER and RC. Ren et al. [329] propose a knowledge-enhanced distribution CoType model for joint extraction tasks. In this model, entity pairs are firstly mapped onto their mentions in the KB, then tagged with entity types and all relation candidates provided by the KB. This model learns embeddings of relation mentions with contextualized lexical and syntax features while training embeddings of the entity mentions with their types, then the contextual relation mention will be derived by its head and tail entities embeddings via TranE [330] model. The CoType model assumes interactive cooccurrence between entities and their relation labels, filling the distribution discrepancy with knowledge from the external domain and extra type features. Noticeably, this model also effectively prevents noises in distant-supervised datasets. However, feature engineering and extra KBs are also needed.

D.3 Rule-Mining Methods for Relation Path Reasoning

Many efforts focus on automatically mining logic rules to pave reasoning paths. There are methods for rule discovering, such as AMIE [177], RLvLR [178], and RuleN [331]. Instead of searching for promising relation path patterns approaching the symbolic essence of knowledge, the rule mining approaches extract and prune logic rules from a reasonable KG, then perform link prediction via the collected rule templates. However, unseen knowledge paths cannot be easily derived by logical rules in incomplete graphs.

Another research direction is to fuel logic rules into neural models to boost path reasoning. KALE [332] jointly embeds first-order logic rules with knowledge embedding to enhance relation inference. RUGE [333] iteratively rectifies KG embeddings via learned soft rules and then performs relation path reasoning. Logic rules are leveraged as the side semantic information into neural models. NeuralLP [179] proposes a neural framework to encode logic rule structures into vectorized embeddings with an attention mechanism. pLogicNet [180] introduces **Markov logic network** (**MLN**) to model uncertain rules. ExpressGNN [181] further employs GCNN to solve neighborhood graphic semantics with logic rules. These rule-based neural models are also regarded as the application of differentiable learning availing for gradient-based optimization algorithms on logic programming.

D.4 Other Advances

Researchers explore more strategies for flexible NER tasks. Transfer Learning shares knowledge between different domains or models. Pan et al. [334] propose **Transfer Joint Embedding** (**TJE**) to jointly embed output labels and input samples from different domains for blending intrinsic entity features. Lin et al. apply [335] a neural network with adaptation layers to transfer parameter features from a model pre-trained on a different domain. **Reinforcement Learning** (**RL**) puts NER models to interact with the environment domain through a behavior agency with a reward policy, such as the **Markov decision process** (**MDP**)-based model [336] and Q-network enhanced model [337]. Noticeably, researchers [338] have also leveraged the RL model for noise reduction in distant-supervised NER data. Adversarial Learning generates counterexamples or perturbations to enforce the robustness of NER models, such as DATNet [339] imposing perturbations on word representations and counterexamples generators ([340] and [341]). Moreover, Active Learning, which queries users to annotate selected samples, has also been applied to NER. Shen et al.[342] incrementally chose the most samples for NER labeling during the training procedures to mitigate the reliance on tagged samples.

Few-shot/zero-shot ET is an intricate challenging issue. Ma et al. [343] model the prototype of entity label embeddings for zero-shot fine-grain ET, naming Proto-HLE, which combines prototypical features with hierarchical type labels for inferring essential features of a new type. Zhang et al. [344] further propose MZET that exploits contextual features and word embeddings with a Memory Network to provide semantic side information for few-shot ET.

More probabilistic-based models are developed for EL tasks. Guo et al. [345] propose a probabilistic model for unstructured data, that leverages the prior probability of an entity, context, and name when performing linking tasks with unstructured data. Han et al. [346] employed a reference graph of entities, assuming that entities co-occurring in the same documents should be semantically related.

Joint models for NER and EL reduce error propagation of the pipeline-based entity recognition tasks. NEREL [347] couples NER and EL by ranking extracted mention-entity pairs to exploit the interaction features between entity mentions and their links. Graphic models are also effective designs to combine **Named Entity Normalization** (**NEN**) labels that convert entity mentions into unambiguous forms, e.g., Washington (Person) and Washington (State). Li et al. [348] incorporated EL with NEN utilizing a factor graph model, forming CRF chains for word entity types and their target nodes. Likewise, MINTREE [349] introduces a tree-based pair-linking model for collective tasks.

Cluster-based solutions handle the CO (Coreference Resolution) task as a pairwise binary classification task (co-referred or not). Early cluster models aim at mention-pair features. Soon et al. [350] propose a single-link clustering strategy to detect anaphoric pairs. Recasens et al. [351] further develop a mention-pair-based cluster to emanate a coreference chain or a singleton leaf. Later, researchers concentrate on entity-based features to exploit complex anaphoric features. Rahman and Ng [352] propose a mention-ranking clustering model to dive into entity characteristics. Stoyanov and Eisner [353] develop agglomerative clustering to merge the best clusters with entity features.

Early researchers concentrate on intriguing statistical-based features for fast end-to-end joint RE, such as **Integer Linear Programming (ILP)**-based algorithm [354] solving entities and relations via conditional probabilistic model, semi-Markov chain model [355] jointly decoding global-level relation features, and MLNs [356] modeling joint logic rules of entity labels and relationships. Early attempts deliver prototypes of entity-relationship interactions. However, statistical patterns are not explicit for intricate contexts.

Few-shot RC designs also consider feature augmentation strategies to mitigate data deficiency with intriguing model designs and background knowledge. Similar to [95], Levy et al. [357] turn zero-shot RC into a reading comprehension problem to comprehend unseen labels by a template converter. Soares et al. [358] compose a compound relation representation for each sentence by the BERT contextualized embeddings of entity pairs and the corresponding sentence. GCNs also deliver extra graph-level features for few-shot learning. Satorras and Estrach [359] propose a novel GCN framework to determine the relation tag of a query sample by calculating the similarity between nodes. Moreover, Qu et al. [360] employ posterior distribution for prototypical vectors. Some designs also avail semi-supervised data augmentation based on metric learning. The previous Neural Snowball [121] (based on RSN) labels the query set via the Siamese network while drawing a similar sample candidate from external distant-supervised sample sets to enrich the support set.

Many early attempts develop random-walk models for relation path reasoning that infer relational logic paths in a latent variable logic graphic model. **Path-Ranking Algorithm (PRA)** [361] generates a feature matrix to sample potential relation paths. However, the feature sparsity in the graph impedes random walk approaches. Semantic enrichment strategies are proposed to mitigate this bottleneck, such as inducing vector space similarity [362] and clustering associated relations [363].

Early attempts aim at the unique attributes of entities for entity matching. Many models leverage distance-based approaches to distributional representations of entity descriptions or definitions. VCU [364] proposes first-order and second-order vector models to embed the description words of an entity pair for comprehensively measuring the conceptual distance. TALN [365] leverages sense-based embedding derived by BabelNet to combine the definitional description of words, which first generates the embedding of each filtered definition word combined with POS-tagger, syntax features via BabelNet, then averages them to obtain a centroid sense to obtain the best matching candidates. String-similarity-based models available for entity matching also include TF-IDF [366] and I-Sub [367].

Graph-based methods achieve feasible performance for entity matching on the medium-scale KG that consists of hierarchical graph structures. ETF [368] learns concept representations through semantic features and graph-based features, including Katz similarity, random walk betweenness centrality, and information propagation score. ParGenFS [369] leverages a graph-based fuzzy cluster algorithm to conceptualize a new entity. This method stimulates the thematic distribution to acquire distinctive concept clusters to search the corresponding location of an entity update in a target KG.

Entity alignment tasks can also be handled by text-similarity-based models that detect surficial similarity between entities when considering the tradeoff between performance and computation cost. Rdf-ai [370] proposes a systematic model to match two entity node graphs, which leverages the string-matching and lexical-feature-similarity comparing algorithms to align available attributes, then calculates the entity similarity for alignment. Similarly, Lime [371] further leverages metric spaces to detect aligned entity pairs, which first generate entity exemplars to filter alignable candidates before similarity computation for entity fusion. Different from small-scale KGs, the shaped large KGs contain meaningful relational paths and enriched concept taxonomy.

HolisticEM [372] employs IDF score to calculate the surficial similarity of entity names for seed generating and utilizes **Personalized PageRank** (**PPR**) to measure distances between entity graphs by traversing their neighbor nodes.

E SEMI-STRCTURED DATA PRE-PROCESSING

Real-world raw data sources include multi-structure contents with irrelevant parts impairing the effect of knowledge extraction. Data preprocessing is necessary for handling a messy data environment. Preprocessing sub-tasks mainly include Content Extraction and Structure Interpretation.

E.1 Content Extraction

Many web pages contain non-content noises such as advertisements. Content extraction tasks aim to erase these irrelevant elements while reserving knowledge content. Users can manually select the main part of a web page (e.g., contents enveloped by "<table>") to achieve this goal by web crawlers such as JSoup, BeautifulSoup, and Web Scraper that retrieve and interpret elements in **Document Object Model** (**DOM**) structures, then users can select the main part of a web page. However, when the data volume is high, manual work will fail to handle them. Mainstream automatic content extraction methods mainly include wrapper-based methods and statistic-based methods.

Wrapper-based methods are the earliest attempts to detect main contents, leveraging matching rules to capture informative content. Off-the-shelf wrapper tools automatically generate rules from semi-structured pages including IEPAD [373] and SoftMealy [374]. Bootstrapping methods iteratively enhance extraction templates with seed examples, such as [375] and [376]. Some toolkits provide user interfaces to optimize extraction templates, such as NoDoSE[377] and DEByE [378]. Template-based wrappers are easy to understand and achieve feasible results where the page structures are well-formed, but fail to grasp the inner contents covered by intricate novel elements or structures.

Users can also utilize methods based on statistical features of web pages to obtain informative content. Finn et.al [379] propose an empirical assumption that an informative sub-sequence in a web page contains sufficient enough words with minimal tags. Many models consider statistical features of web contents for extracting informative content, such as CETR [380] (the ratio of text length to tag number), CETD [381] (text density in each sub-tree structure of a DOM tree), and CEPR [382] (path ratio of Web links). Users can utilize WebCollector [47] that integrates the statistic-based models for content extraction. Another heuristic research direction is visual-features-based methods. For example, VIPS [383] utilizes the visual appearances (such as fonts and color types) of a page to build a content structure tree for content extraction.

When content extraction has been performed on a semi-structured page, users will acquire a renewed noise-free semi-structured or unstructured document.

E.2 Structure Interpretation

Many table forms in the web pages function as navigators or style-formatted containers for contents (handled by content extractors), comprising no relational structures. Models shall filter these decorative non-relational web table structures before obtaining relational information.

Relational table interpretation is a binary classification task that determines whether a table is informative. Methods analyze semantic features of table structures for classification. Wang and Hu [384] design a table classifier integrated with **support vector machines** (**SVM**) and decision trees based on the layout and content type features. Similarly, WebTables [385] develops a rule-based classifier based on the table size (number of rows and columns) and tags. Eberius et al. [386] develop a classification system DWTC via the feature of the table matrixes. Many web tables also contain data noises. OCTOPUS [387] further incorporates data cleansing with table classification tasks to filter informative tables.

Developing a table interpretation model includes two steps: first select features in the table forms, then integrate learning models to analyze relational semantics in the data. We recommend that readers refer to [388] for more table syntax features and high-performance model ensembles.

F KNOWLEDGE GRAPH STORAGE

In this section, we provide a brief overview of KG storage tools for different data environments.

Early efforts utilize relational models to perpetuate constructed KGs. Traditional RDBMS provides reliable and swift CRUD operations for table-formed databases. Developers have also employed graph algorithms like depth-first traverse and shortest-path search to enhance relational databases. Ref. [2] includes representative examples like PostgreSQL [389], and filament.[9] However, it can be very costly for a relational database to handle sparse KGs or perform data partition for distribution storage.

Key/value databases are lightweight solutions for saving clusters in large KGs, supporting distributed storage with a simplified flexible data format. Trinity [390] provides a high-performance in-memory Key/Value storage system to manage large KGs with billion nodes, such as Probase. CouchDB [391] utilizes a replication mechanism to maintain dynamic KGs. MapReduce technology automatically transforms data groups into key/value mappings. Hadhoop[10] enables high-throughput parallel computing for KG storage via MapReduce. Pregel [392] develops a superstep mechanism to share messages between vertices for parallel computing.

Another enlightening direction is to design graph databases that fit in knowledge triple structures. Neo4j[393] is a lightweight NoSQL-based graph database supporting embedded dynamic KG storage. SOnes[11] provides object-oriented queries for KG databases. Novel languages are also developed for knowledge storage, such as **resource description framework** (**RDF**) and **Web Ontology Language** (**OWL**).[12] Some graph databases based on RDF optimize the storage of graph structures. For example, gStore [392] improves RDF-structured KG databases via sub-graph matching algorithms.

REFERENCES

[1] Heiko Paulheim. 2017. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web* 8, 3 (2017), 489–508. DOI: http://dx.doi.org/10.3233/SW-160218

[2] Jihong Yan, Chengyu Wang, Wenliang Cheng, Ming Gao, and Aoying Zhou. 2018. A retrospective of knowledge graphs. *Frontiers of Computer Science* 12, 1 (2018), 55–74. DOI: http://dx.doi.org/10.1007/s11704-016-5228-9

[3] Xindong Wu, Jia Wu, Xiaoyi Fu, Jiachen Li, Peng Zhou, and Xu Jiang. 2019. Automatic knowledge graph construction: A report on the 2019 ICDM/ICBK contest. In *Proceedings of the 2019 IEEE International Conference on Data Mining*. 1540–1545. DOI: http://dx.doi.org/10.1109/ICDM.2019.00204

[4] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. 2022. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems* 33, 2 (2022), 494–514. DOI: http://dx.doi.org/10.1109/TNNLS.2021.3070843

[5] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M.

[9] https://filament.sourceforge.net
[10] http://hadoop.apache.org
[11] http://github.com/sones/sones
[12] RDF and OWL are both standards of w3c, see also http://www.w3.org/RDF.

Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. 2021. *Knowledge Graphs*. Morgan & Claypool Publishers. DOI: http://dx.doi.org/10.2200/S01125ED1V01Y202109DSK022

[6] Zhenyu Zhang, Lei Zhang, Dingqi Yang, and Liu Yang. 2022. KRAN: Knowledge refining attention network for recommendation. *ACM Transactions on Knowledge Discovery from Data* 16, 2 (2022), 39:1–39:20. DOI: http://dx.doi.org/10.1145/3470783

[7] Chunquan Chen and Si Li. 2020. Knowledge-based context-aware multi-turn conversational model with hierarchical attention. In *Proceedings of the 2020 International Joint Conference on Neural Networks*. 1–8. DOI: http://dx.doi.org/10.1109/IJCNN48605.2020.9206835

[8] Linmei Hu, Tianchi Yang, Luhao Zhang, Wanjun Zhong, Duyu Tang, Chuan Shi, Nan Duan, and Ming Zhou. 2021. Compare to the knowledge: Graph neural fake news detection with external knowledge. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 754–763. DOI: http://dx.doi.org/10.18653/v1/2021.acl-long.62

[9] Kurt D. Bollacker, Robert P. Cook, and Patrick Tufts. 2007. Freebase: A shared database of structured general human knowledge. In *Proceedings of the AAAI-07, 2007*. 1962–1963. Retrieved from http://www.aaai.org/Library/AAAI/2007/aaai07-355.php

[10] Denny Vrandecic. 2012. Wikidata: A new platform for collaborative data collection. In *Proceedings of the 21st International Conference on World Wide Web*. 1063–1064. DOI: http://dx.doi.org/10.1145/2187980.2188242

[11] Alexander Yates, Michele Banko, Matthew Broadhead, Michael J. Cafarella, Oren Etzioni, and Stephen Soderland. 2007. TextRunner: Open information extraction on the web. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics*. 25–26. Retrieved from https://aclanthology.org/N07-4013/

[12] Oren Etzioni, Michael J. Cafarella, Doug Downey, Stanley Kok, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. 2004. Web-scale information extraction in knowitall: (Preliminary results). In *Proceedings of the 13th international conference on World Wide Web*. 100–110. DOI: http://dx.doi.org/10.1145/988672.988687

[13] Minghui Wu and Xindong Wu. 2019. On big wisdom. *Knowledge and Information Systems* 58, 1 (2019), 1–8. DOI: http://dx.doi.org/10.1007/s10115-018-1282-y

[14] Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. arXiv:1508.01991. Retrieved from http://arxiv.org/abs/1508.01991

[15] Xuezhe Ma and Eduard H. Hovy. 2016. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In *Proceedings of the ACL, 2016, Volume 1: Long Papers*. DOI: http://dx.doi.org/10.18653/v1/p16-1101

[16] Peng Xu and Denilson Barbosa. 2018. Neural fine-grained entity type classification with hierarchy-aware loss. In *Proceedings of the NAACL-HLT, 2018, Volume 1 (Long Papers)*. 16–25. DOI: http://dx.doi.org/10.18653/v1/n18-1002

[17] Xiang Ren, Wenqi He, Meng Qu, Clare R. Voss, Heng Ji, and Jiawei Han. 2016. Label noise reduction in entity typing by heterogeneous partial-label embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1825–1834. DOI: http://dx.doi.org/10.1145/2939672.2939822

[18] Octavian-Eugen Ganea and Thomas Hofmann. 2017. Deep joint entity disambiguation with local neural attention. In *Proceedings of the EMNLP , 2017*. 2619–2629. DOI: http://dx.doi.org/10.18653/v1/d17-1277

[19] Phong Le and Ivan Titov. 2018. Improving entity linking by modeling latent relations between mentions. In *Proceedings of the ACL , 2018, Volume 1: Long Papers*. 1595–1604. DOI: http://dx.doi.org/10.18653/v1/P18-1148

[20] Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. In *Proceedings of the EMNLP, 2017*. 188–197. DOI: http://dx.doi.org/10.18653/v1/d17-1018

[21] Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 1753–1762. DOI: http://dx.doi.org/10.18653/v1/d15-1203

[22] Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short papers)*. DOI: http://dx.doi.org/10.18653/v1/p16-2034

[23] Hongming Zhang, Daniel Khashabi, Yangqiu Song, and Dan Roth. 2020. TransOMCS: From linguistic graphs to commonsense knowledge. In *Proceedings of the IJCAI, 2020*. 4004–4010. DOI: http://dx.doi.org/10.24963/ijcai.2020/554

[24] Hongming Zhang, Xin Liu, Haojie Pan, Haowen Ke, Jiefu Ou, Tianqing Fang, and Yangqiu Song. 2022. ASER: Towards large-scale commonsense knowledge acquisition via higher-order selectional preference over eventualities. *Artif. Intell.* 309 (2022), 103740. DOI: http://dx.doi.org/10.1016/J.ARTINT.2022.103740

[25] Wu Xin-Dong, Sheng Shao-Jing, Jiang Ting-Ting, Bu Chen-Yang, and Wu Ming-Hui. 2020. Huapu-CP: from knowledge graphs to a data central-platform. *Acta Automatica Sinica* 46, 10 (2020), 2045–2059.

[26] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the NAACL-HLT , 2019, Volume 1 (Long and Short Papers)*. 4171–4186. DOI: http://dx.doi.org/10.18653/v1/n19-1423

[27] Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. 2019. DocRED: A large-scale document-level relation extraction dataset. In *Proceedings of the ACL, 2019, Volume 1: Long Papers*. 764–777. DOI: http://dx.doi.org/10.18653/v1/p19-1074

[28] Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2010, Barcelona, Spain, September 20-24, 2010, Proceedings, Part III 21*. 148–163. DOI: http://dx.doi.org/10.1007/978-3-642-15939-8_10

[29] Xiaozhi Wang, Xu Han, Yankai Lin, Zhiyuan Liu, and Maosong Sun. 2018. Adversarial multi-lingual neural relation extraction. In *Proceedings of the 27th International Conference on Computational Linguistics*. 1156–1166. Retrieved from https://aclanthology.org/C18-1099/

[30] Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. 2021. Explainable subgraph reasoning for forecasting on temporal knowledge graphs. In *Proceedings of the International Conference on Learning Representations*. Retrieved from https://openreview.net/forum?id=pGIHq1m7PU

[31] Tianwen Jiang, Tong Zhao, Bing Qin, Ting Liu, Nitesh V. Chawla, and Meng Jiang. 2019. Multi-input multi-output sequence labeling for joint extraction of fact and condition tuples from scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. 302–312. DOI: http://dx.doi.org/10.18653/v1/D19-1029

[32] Abhishek Pradhan, Ketan Kumar Todi, Anbarasan Selvarasu, and Atish Sanyal. 2020. Knowledge graph generation with deep active learning. In *Proceedings of the International Joint Conference on Neural Networks*. 1–8. DOI: http://dx.doi.org/10.1109/IJCNN48605.2020.9207515

[33] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding. 2014. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering* 26, 1 (2014), 97–107. DOI: http://dx.doi.org/10.1109/TKDE.2013.109

[34] Tapas Nayak, Navonil Majumder, Pawan Goyal, and Soujanya Poria. 2021. Deep neural approaches to relation triplets extraction: A comprehensive survey. *Cognitive Computation* 13, 5 (2021), 1215–1232. DOI: http://dx.doi.org/10.1007/s12559-021-09917-7

[35] Sachin Pawar, Pushpak Bhattacharyya, and Girish K. Palshikar. 2021. Techniques for jointly extracting entities and relations: A survey. arXiv:2103.06118. Retrieved from https://arxiv.org/abs/2103.06118

[36] Siddhant Arora. 2020. A survey on graph neural networks for knowledge graph completion. arXiv:2007.12374. Retrieved from https://arxiv.org/abs/2007.12374

[37] Borui Cai, Yong Xiang, Longxiang Gao, He Zhang, Yunfeng Li, and Jianxin Li. 2022. Temporal knowledge graph completion: A survey. arXiv:2201.08236 Retrieved from https://arxiv.org/abs/2201.08236

[38] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering* 29, 12 (2017), 2724–2743. DOI: http://dx.doi.org/10.1109/TKDE.2017.2754499

[39] Lisa Ehrlinger and Wolfram Wöß. 2016. Towards a definition of knowledge graphs. In *Proceedings of the SEMANTiCS, SuCCESS'16, 2016 (CEUR Workshop Proceedings)*. Retrieved from http://ceur-ws.org/Vol-1695/paper4.pdf

[40] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. 2007. DBpedia: A nucleus for a web of open data. In *Proceedings of the International Semantic Web Conference*. 722–735. DOI: http://dx.doi.org/10.1007/978-3-540-76298-0_52

[41] Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Qili Zhu. 2012. Probase: A probabilistic taxonomy for text understanding. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*. 481–492. DOI: http://dx.doi.org/10.1145/2213836.2213891

[42] Zhigang Wang, Juanzi Li, Zhichun Wang, Shuangjie Li, Mingyang Li, Dongsheng Zhang, Yao Shi, Yongbin Liu, Peng Zhang, and Jie Tang. 2013. XLore: A large-scale english-chinese bilingual knowledge graph. In *Proceedings of the ISWC (CEUR Workshop Proceedings)*. 121–124. Retrieved from http://ceur-ws.org/Vol-1035/iswc2013_demo_31.pdf

[43] George A. Miller. 1995. WordNet: A lexical database for english. *Communications of the ACM* 38, 11 (1995), 39–41. DOI: http://dx.doi.org/10.1145/219717.219748

[44] Cynthia Matuszek, John Cabral, Michael J. Witbrock, and John DeOliveira. 2006. An introduction to the syntax and content of cyc. In *Proceedings of the Papers from the 2006 AAAI Spring Symposium, Technical Report SS-06-05, 2006*. 44–49. Retrieved from http://www.aaai.org/Library/Symposia/Spring/2006/ss06-05-007.php

[45] Thomas Steiner, Ruben Verborgh, Raphaël Troncy, Joaquim Gabarró, and Rik Van de Walle. 2012. Adding realtime coverage to the google knowledge graph. In *Proceedings of the 11th International Semantic Web Conference*. Retrieved from http://ceur-ws.org/Vol-914/paper_2.pdf

[46] David S. Wishart, Craig Knox, Anchi Guo, Savita Shrivastava, Murtaza Hassanali, Paul Stothard, Zhan Chang, and Jennifer Woolsey. 2006. DrugBank: A comprehensive resource for *in Silico* Drug Discovery and Exploration. *Nucleic Acids Res.* 34, Database-Issue (2006), 668–672. DOI: http://dx.doi.org/10.1093/nar/gkj067

[47] Wu Gong-Qing, Hu Jun, Li Li, Xu Zhe-Hao, Liu Peng-Cheng, Hu Xue-Gang, and Wu Xin-Dong. 2016. Online web news extraction via tag path feature fusion. ruan jian xue bao. *Journal of Software* 27, 3 (2016), 714–735.

[48] Yanzeng Li and Lei Zou. 2022. gBuilder: A scalable knowledge graph construction system for unstructured corpus. arXiv:arXiv:2208.09705. Retrieved from https://arxiv.org/abs/2208.09705

[49] Xu Han, Shulin Cao, Xin Lv, Yankai Lin, Zhiyuan Liu, Maosong Sun, and Juanzi Li. 2018. OpenKE: An open toolkit for knowledge embedding. In *Proceedings of the EMNLP, 2018*. 139–144. DOI:http://dx.doi.org/10.18653/v1/d18-2024

[50] Zequn Sun, Qingheng Zhang, Wei Hu, Chengming Wang, Muhao Chen, Farahnaz Akrami, and Chengkai Li. 2020. A benchmarking study of embedding-based entity alignment for knowledge graphs. *Proceedings of the VLDB Endowment* 13, 11 (2020), 2326–2340. Retrieved from http://www.vldb.org/pvldb/vol13/p2326-sun.pdf

[51] Guodong Zhou and Jian Su. 2002. Named entity recognition using an HMM-based chunk tagger. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. 473–480. DOI:http://dx.doi.org/10.3115/1073083.1073163

[52] Jenny Rose Finkel, Trond Grenager, and Christopher D. Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. 363–370. DOI:http://dx.doi.org/10.3115/1219840.1219885

[53] Jun Zhu, Zaiqing Nie, Ji-Rong Wen, Bo Zhang, and Wei-Ying Ma. 2005. 2D conditional random fields for web information extraction. In *Proceedings of the 22nd International Conference on Machine Learning*. 1044–1051. DOI:http://dx.doi.org/10.1145/1102351.1102483

[54] Charles Sutton, Khashayar Rohanimanesh, and Andrew McCallum. 2004. Dynamic conditional random fields: Factorized probabilistic models for labeling and segmenting sequence data. In *Proceedings of the 21st International Conference on Machine Learning*. DOI:http://dx.doi.org/10.1145/1015330.1015422

[55] Jun Zhu, Zaiqing Nie, Ji-Rong Wen, Bo Zhang, and Wei-Ying Ma. 2006. Simultaneous record detection and attribute labeling in web data extraction. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 494–503. DOI:http://dx.doi.org/10.1145/1150402.1150457

[56] Aidan Finn and Nicholas Kushmerick. 2004. Multi-level boundary classification for information extraction. In *Proceedings of the European Conference on Machine Learning*. 111–122. DOI:http://dx.doi.org/10.1007/978-3-540-30115-8_13

[57] Sheng Zhang, Kevin Duh, and Benjamin Van Durme. 2018. Fine-grained entity typing through increased discourse context and adaptive classification thresholds. In *Proceedings of the *SEM@NAACL-HLT, 2018*. 173–179. DOI:http://dx.doi.org/10.18653/v1/s18-2022

[58] Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. 2022. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*. 34, 1 (2022), 50–70. DOI:http://dx.doi.org/10.1109/TKDE.2020.2981314

[59] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research* 12 (2011), 2493–2537. Retrieved from http://dl.acm.org/citation.cfm?id=2078186

[60] Lishuang Li, Liuke Jin, Zhenchao Jiang, Dingxin Song, and Degen Huang. 2015. Biomedical named entity recognition based on extended recurrent neural networks. In *Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine*. 649–652. DOI:http://dx.doi.org/10.1109/BIBM.2015.7359761

[61] Thien Huu Nguyen, Avirup Sil, Georgiana Dinu, and Radu Florian. 2016. Toward mention detection robustness with recurrent neural networks. arXiv:1602.07749. Retrieved from http://arxiv.org/abs/1602.07749

[62] Emma Strubell, Patrick Verga, David Belanger, and Andrew McCallum. 2017. Fast and accurate entity recognition with iterated dilated convolutions. In *Proceedings of the EMNLP, 2017*. 2670–2680. DOI:http://dx.doi.org/10.18653/v1/d17-1283

[63] Ling Luo, Zhihao Yang, Pei Yang, Yin Zhang, Lei Wang, Hongfei Lin, and Jian Wang. 2018. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. *Bioinformatics* 34, 8 (2018), 1381–1388. DOI:http://dx.doi.org/10.1093/bioinformatics/btx761

[64] Andrej Zukov Gregoric, Yoram Bachrach, Pasha Minkovsky, Sam Coope, and Bogdan Maksak. 2017. Neural named entity recognition using a self-attention mechanism. In *Proceedings of the 2017 IEEE 29th International Conference on Tools with Artificial Intelligence*. 652–656. DOI:http://dx.doi.org/10.1109/ICTAI.2017.00104

[65] Alberto Cetoli, Stefano Bragaglia, Andrew D. O'Harney, and Marc Sloan. 2018. Graph convolutional networks for named entity recognition. In *Proceedings of the TLT, 2018*. 37–45. Retrieved from https://aclanthology.org/W17-7607/

[66] Cihan Dogan, Aimore Dutra, Adam Gara, Alfredo Gemma, Lei Shi, Michael Sigamani, and Ella Walters. 2019. Fine-grained named entity recognition using ELMo and wikidata. arXiv:1904.10503. Retrieved from http://arxiv.org/abs/1904.10503

[67] Mingyi Liu, Zhiying Tu, Tong Zhang, Tonghua Su, Xiaofei Xu, and Zhongjie Wang. 2022. LTP: A new active learning strategy for CRF-based named entity recognition. *Neural Processing Letters* (2022), 1–22.

[68] Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. 2020. LUKE: Deep contextualized entity representations with entity-aware self-attention. In *Proceedings of the EMNLP, 2020*. 6442–6454. DOI:http://dx.doi.org/10.18653/v1/2020.emnlp-main.523

[69] Sonse Shimaoka, Pontus Stenetorp, Kentaro Inui, and Sebastian Riedel. 2017. Neural architectures for fine-grained entity type classification. In *Proceedings of the EACL, 2017, Volume 1: Long Papers*. 1271–1280. DOI: http://dx.doi.org/10.18653/v1/e17-1119

[70] Varish Mulwad, Tim Finin, Zareen Syed, and Anupam Joshi. 2010. Using linked data to interpret tables. In *Proceedings of the 1st International Workshop on Consuming Linked Data, 2010 (CEUR Workshop Proceedings)*. Retrieved from http://ceur-ws.org/Vol-665/MulwadEtAl_COLD2010.pdf

[71] Girija Limaye, Sunita Sarawagi, and Soumen Chakrabarti. 2010. Annotating and searching web tables using entities, types and relationships. *Proceedings of the VLDB Endowment* 3, 1 (2010), 1338–1347. DOI: http://dx.doi.org/10.14778/1920841.1921005

[72] Chandra Sekhar Bhagavatula, Thanapon Noraset, and Doug Downey. 2015. TabEL: Entity linking in web tables. In *Proceedings of the International Semantic Web Conference*. 425–441. DOI: http://dx.doi.org/10.1007/978-3-319-25007-6_25

[73] Tianxing Wu, Shengjia Yan, Zhixin Piao, Liang Xu, Ruiming Wang, and Guilin Qi. 2016. Entity linking in web tables with multiple linked knowledge bases. In *Semantic Technology: 6th Joint International Conference, JIST 2016, Singapore, Singapore, November 2-4, 2016, Revised Selected Papers 6*. 239–253. DOI: http://dx.doi.org/10.1007/978-3-319-50112-3_18

[74] Vasilis Efthymiou, Oktie Hassanzadeh, Mariano Rodriguez-Muro, and Vassilis Christophides. 2017. Matching web tables with knowledge base entities: From entity lookups to entity embeddings. In *The Semantic WebISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, October 2125, 2017, Proceedings, Part I 16*. 260–277. DOI: http://dx.doi.org/10.1007/978-3-319-68288-4_16

[75] Nicola De Cao, Gautier Izacard, Sebastian Riedel, and Fabio Petroni. 2021. Autoregressive entity retrieval. In *Proceedings of the ICLR*. Retrieved from https://openreview.net/forum?id=5k8F6UU39V

[76] Jie Cai and Michael Strube. 2010. End-to-end coreference resolution via hypergraph partitioning. In *Proceedings of the COLING, 2010*. 143–151. Retrieved from https://aclanthology.org/C10-1017/

[77] Emili Sapena, Lluís Padró, and Jordi Turmo. 2013. A constraint-based hypergraph partitioning approach to coreference resolution. *Computational Linguistics* 39, 4 (2013), 847–884. DOI: http://dx.doi.org/10.1162/COLI_a_00151

[78] David L. Bean and Ellen Riloff. 2004. Unsupervised learning of contextual role knowledge for coreference resolution. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*. 297–304. Retrieved from https://aclanthology.org/N04-1038/

[79] Eraldo R. Fernandes, Cícero Nogueira dos Santos, and Ruy Luiz Milidiú. 2012. Latent structure perceptron with feature induction for unrestricted coreference resolution. In *Proceedings of the Joint Conference on EMNLP and CoNLL-Shared Task 2012*. 41–48. Retrieved from https://aclanthology.org/W12-4502/

[80] Xue-Feng Xi, Guodong Zhou, Fuyuan Hu, and Baochuan Fu. 2015. A convolutional deep neural network for coreference resolution via modeling hierarchical features. In *Intelligence Science and Big Data Engineering. Big Data and Machine Learning Techniques: 5th International Conference, IScIDE 2015, Suzhou, China, June 14-16, 2015, Revised Selected Papers, Part II 5*. 361–372. DOI: http://dx.doi.org/10.1007/978-3-319-23862-3_36

[81] Jheng-Long Wu and Wei-Yun Ma. 2017. A deep learning framework for coreference resolution based on convolutional neural network. In *Proceedings of the 2017 IEEE 11th International Conference on Semantic Computing*. 61–64. DOI: http://dx.doi.org/10.1109/ICSC.2017.57

[82] Sam Wiseman, Alexander M. Rush, and Stuart M. Shieber. 2016. Learning global features for coreference resolution. In *Proceedings of the NAACL-HLT, 2016*. 994–1004. DOI: http://dx.doi.org/10.18653/v1/n16-1114

[83] Jia-Chen Gu, Zhen-Hua Ling, and Nitin Indurkhya. 2018. A study on improving end-to-end neural coreference resolution. In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data: 17th China National Conference, CCL 2018, and 6th International Symposium, NLP-NABD 2018, Changsha, China, October 1921, 2018, Proceedings 17*. 159–169. DOI: http://dx.doi.org/10.1007/978-3-030-01716-3_14

[84] Rui Zhang, Cícero Nogueira dos Santos, Michihiro Yasunaga, Bing Xiang, and Dragomir R. Radev. 2018. Neural coreference resolution with deep biaffine attention by joint mention detection and mention clustering. In *Proceedings of the ACL*. 102–107. DOI: http://dx.doi.org/10.18653/v1/P18-2017

[85] Jie Ma, Jun Liu, Yufei Li, Xin Hu, Yudai Pan, Shen Sun, and Qika Lin. 2020. Jointly optimized neural coreference resolution with mutual attention. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 402–410. DOI: http://dx.doi.org/10.1145/3336191.3371787

[86] Kevin Clark and Christopher D. Manning. 2016. Deep reinforcement learning for mention-ranking coreference models. In *Proceedings of the EMNLP, 2016*. 2256–2262. DOI: http://dx.doi.org/10.18653/v1/d16-1245

[87] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. 2014. Relation classification via convolutional deep neural network. In *Proceedings of the COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. 2335–2344. Retrieved from https://aclanthology.org/C14-1220/

[88] Thien Huu Nguyen and Ralph Grishman. 2015. Relation extraction: Perspective from convolutional neural networks. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*. 39–48. DOI: http://dx.doi.org/10.3115/v1/w15-1506

[89] Yatian Shen and Xuanjing Huang. 2016. Attention-based convolutional neural network for semantic relation extraction. In *Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2526–2536. Retrieved from https://aclanthology.org/C16-1238/

[90] Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu. 2016. Relation classification via multi-level attention CNNs. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. DOI: http://dx.doi.org/10.18653/v1/p16-1123

[91] Makoto Miwa and Mohit Bansal. 2016. End-to-end relation extraction using LSTMs on sequences and tree structures. In *Proceedings of the ACL, 2016, Volume 1: Long Papers*. DOI: http://dx.doi.org/10.18653/v1/p16-1105

[92] Yi Zhao, Huaiyu Wan, Jianwei Gao, and Youfang Lin. 2019. Improving relation classification by entity pair graph. In *Proceedings of the Asian Conference on Machine Learning*. 1156–1171. Retrieved from http://proceedings.mlr.press/v101/zhao19a.html

[93] Zhijiang Guo, Yan Zhang, and Wei Lu. 2019. Attention guided graph convolutional networks for relation extraction. In *Proceedings of the ACL,2019, Volume 1: Long Papers*. 241–251. DOI: http://dx.doi.org/10.18653/v1/p19-1024

[94] Kang Zhao, Hua Xu, Yue Cheng, Xiaoteng Li, and Kai Gao. 2021. Representation iterative fusion based on heterogeneous graph neural network for joint entity and relation extraction. *Knowledge-Based Systems* 219 (2021), 106888. DOI: http://dx.doi.org/10.1016/j.knosys.2021.106888

[95] Amir D. N. Cohen, Shachar Rosenman, and Yoav Goldberg. 2020. Relation extraction as two-way span-prediction. arXiv:2010.04829. Retrieved from https://arxiv.org/abs/2010.04829

[96] Razvan C. Bunescu and Raymond J. Mooney. 2007. Learning to extract relations from the web using minimal supervision. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Retrieved from https://aclanthology.org/P07-1073/

[97] Frank Reichartz, Hannes Korte, and Gerhard Paass. 2010. Semantic relation extraction with kernels over typed dependency trees. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 773–782. DOI: http://dx.doi.org/10.1145/1835804.1835902

[98] Dmitry Zelenko, Chinatsu Aone, and Anthony Richardella. 2002. Kernel methods for relation extraction. In *Proceedings of the EMNLP*. 71–78. DOI: http://dx.doi.org/10.3115/1118693.1118703

[99] Nicholas FitzGerald, Oscar Täckström, Kuzman Ganchev, and Dipanjan Das. 2015. Semantic role labeling with neural network factors. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 960–970. DOI: http://dx.doi.org/10.18653/v1/d15-1112

[100] Michael Roth and Mirella Lapata. 2016. Neural semantic role labeling with dependency path embeddings. In *Proceedings of the ACL*. DOI: http://dx.doi.org/10.18653/v1/p16-1113

[101] Keshav Kolluru, Samarth Aggarwal, Vipul Rathore, Mausam, and Soumen Chakrabarti. 2020. IMoJIE: Iterative memory-based joint open information extraction. In *Proceedings of the ACL,2020*. 5871–5886. DOI: http://dx.doi.org/10.18653/v1/2020.acl-main.521

[102] Ruidong Wu, Yuan Yao, Xu Han, Ruobing Xie, Zhiyuan Liu, Fen Lin, Leyu Lin, and Maosong Sun. 2019. Open relation extraction: Relational knowledge transfer from supervised data to unsupervised data. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. 219–228. DOI: http://dx.doi.org/10.18653/v1/D19-1021

[103] Varish Mulwad, Tim Finin, and Anupam Joshi. 2013. Semantic message passing for generating linked data from tables. In *The Semantic WebISWC 2013: 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I 12*. 363–378. DOI: http://dx.doi.org/10.1007/978-3-642-41335-3_23

[104] Zhe Chen and Michael J. Cafarella. 2013. Automatic web spreadsheet data extraction. In *Proceedings of the 3RD International Workshop on Semantic Search over the Web*. 1:1–1:8. DOI: http://dx.doi.org/10.1145/2509908.2509909

[105] Jun Zhu, Zaiqing Nie, Xiaojiang Liu, Bo Zhang, and Ji-Rong Wen. 2009. StatSnowball: A statistical approach to extracting entity relationships. In *Proceedings of the 18th International Conference on World Wide Web*. 101–110. DOI: http://dx.doi.org/10.1145/1526709.1526724

[106] Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2016. Neural relation extraction with selective attention over instances. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. DOI: http://dx.doi.org/10.18653/v1/p16-1200

[107] Xiaotian Jiang, Quan Wang, Peng Li, and Bin Wang. 2016. Relation extraction with multi-instance multi-label convolutional neural networks. In *Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 1471–1480. Retrieved from https://aclanthology.org/C16-1139/

[108] Bingfeng Luo, Yansong Feng, Zheng Wang, Zhanxing Zhu, Songfang Huang, Rui Yan, and Dongyan Zhao. 2017. Learning with noise: Enhance distantly supervised relation extraction with dynamic transition matrix. In *Proceedings of the ACL, 2017, Volume 1: Long Papers*. 430–439. DOI: http://dx.doi.org/10.18653/v1/P17-1040

[109] Pengda Qin, Weiran Xu, and William Yang Wang. 2018. Robust distant supervision relation extraction via deep reinforcement learning. In *Proceedings of the ACL, 2018, Volume 1: Long Papers*. 2137–2147. DOI: http://dx.doi.org/10.18653/v1/P18-1199

[110] Pengda Qin, Weiran Xu, and William Yang Wang. 2018. DSGAN: Generative adversarial training for distant supervision relation extraction. In *Proceedings of the ACL, 2018, Volume 1: Long Papers*. 496–505. DOI: http://dx.doi.org/10.18653/v1/P18-1046

[111] Yuyun Huang and Jinhua Du. 2019. Self-attention enhanced CNNs and collaborative curriculum learning for distantly supervised relation extraction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. 389–398. DOI: http://dx.doi.org/10.18653/v1/D19-1037

[112] Xu Han, Pengfei Yu, Zhiyuan Liu, Maosong Sun, and Peng Li. 2018. Hierarchical relation extraction with coarse-to-fine grained attention. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2236–2245. DOI: http://dx.doi.org/10.18653/v1/d18-1247

[113] Zhi-Xiu Ye and Zhen-Hua Ling. 2019. Distant supervision relation extraction with intra-bag and inter-bag attentions. In *Proceedings of the NAACL-HLT , 2019, Volume 1 (Long and Short Papers)*. 2810–2819. DOI: http://dx.doi.org/10.18653/v1/n19-1288

[114] Yujin Yuan, Liyuan Liu, Siliang Tang, Zhongfei Zhang, Yueting Zhuang, Shiliang Pu, Fei Wu, and Xiang Ren. 2019. Cross-relation cross-bag attention for distantly-supervised relation extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 419–426. DOI: http://dx.doi.org/10.1609/aaai.v33i01.3301419

[115] Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. 2017. Distant supervision for relation extraction with sentence-level attention and entity descriptions. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 3060–3066. Retrieved from http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14491

[116] Shikhar Vashishth, Rishabh Joshi, Sai Suman Prayaga, Chiranjib Bhattacharyya, and Partha P. Talukdar. 2018. RESIDE: Improving distantly-supervised neural relation extraction using side information. In *Proceedings of the EMNLP, 2018*. 1257–1266. DOI: http://dx.doi.org/10.18653/v1/d18-1157

[117] Ningyu Zhang, Shumin Deng, Zhanlin Sun, Guanying Wang, Xi Chen, Wei Zhang, and Huajun Chen. 2019. Long-tail relation extraction via knowledge graph embeddings and graph convolution networks. In *Proceedings of the NAACL-HLT*. 3016–3025. DOI: http://dx.doi.org/10.18653/v1/n19-1306

[118] Lei Cui, Furu Wei, and Ming Zhou. 2018. Neural open information extraction. In *Proceedings of the ACL*. 407–413. DOI: http://dx.doi.org/10.18F653/v1/P18-2065

[119] Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. 1003–1011. Retrieved from https://aclanthology.org/P09-1113/

[120] Jake Snell, Kevin Swersky, and Richard S. Zemel. 2017. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems* 30, (2017), 4077–4087. Retrieved from https://proceedings.neurips.cc/paper/2017/hash/cb8da6767461f2812ae4290eac7cbc42-Abstract.html

[121] Tianyu Gao, Xu Han, Ruobing Xie, Zhiyuan Liu, Fen Lin, Leyu Lin, and Maosong Sun. 2020. Neural snowball for few-shot relation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 7772–7779. Retrieved from https://aaai.org/ojs/index.php/AAAI/article/view/6281

[122] Oriol Vinyals, Charles Blundell, Tim Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. 2016. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems* 29, (2016), 3630–3638. Retrieved from https://proceedings.neurips.cc/paper/2016/hash/90e1357833654983612fb05e3ec9148c-Abstract.html

[123] Wenhan Xiong, Mo Yu, Shiyu Chang, Xiaoxiao Guo, and William Yang Wang. 2018. One-shot relational learning for knowledge graphs. In *Proceedings of the EMNLP, 2018*. 1980–1990. DOI: http://dx.doi.org/10.18653/v1/d18-1223

[124] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the International Conference on Machine Learning*. 1126–1135. Retrieved from http://proceedings.mlr.press/v70/finn17a.html

[125] Tsendsuren Munkhdalai and Hong Yu. 2017. Meta networks. In *Proceedings of the International Conference on Machine Learning*. 2554–2563. Retrieved from http://proceedings.mlr.press/v70/munkhdalai17a.html

[126] Tongtong Wu, Xuekai Li, Yuan-Fang Li, Gholamreza Haffari, Guilin Qi, Yujin Zhu, and Guoqiang Xu. 2021. Curriculum-meta learning for order-robust continual relation extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 10363–10369. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/17241

[127] Miao Fan, Yeqi Bai, Mingming Sun, and Ping Li. 2019. Large margin prototypical network for few-shot relation classification with fine-grained features. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2353–2356. DOI: http://dx.doi.org/10.1145/3357384.3358100

[128] Tianyu Gao, Xu Han, Zhiyuan Liu, and Maosong Sun. 2019. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 6407–6414. DOI: http://dx.doi.org/10.1609/aaai.v33i01.33016407

[129] Zhi-Xiu Ye and Zhen-Hua Ling. 2019. Multi-level matching and aggregation network for few-shot relation classification. In *Proceedings of the ACL, 2019, Volume 1: Long Papers*. 2872–2881. DOI: http://dx.doi.org/10.18653/v1/p19-1277

[130] Tianyu Gao, Xu Han, Hao Zhu, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2019. FewRel 2.0: Towards more challenging few-shot relation classification. In *Proceedings of the EMNLP-IJCNLP, 2019*. 6249–6254. DOI: http://dx.doi.org/10.18653/v1/D19-1649

[131] Suncong Zheng, Yuexing Hao, Dongyuan Lu, Hongyun Bao, Jiaming Xu, Hongwei Hao, and Bo Xu. 2017. Joint entity and relation extraction based on a hybrid neural network. *Neurocomputing* 257 (2017), 59–66. DOI: http://dx.doi.org/10.1016/j.neucom.2016.12.075

[132] Tsu-Jui Fu, Peng-Hsuan Li, and Wei-Yun Ma. 2019. GraphRel: Modeling text as relational graphs for joint entity and relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 1409–1418. DOI: http://dx.doi.org/10.18653/v1/p19-1136

[133] Zhepei Wei, Jianlin Su, Yue Wang, Yuan Tian, and Yi Chang. 2020. A novel cascade binary tagging framework for relational triple extraction. In *Proceedings of the ACL, 2020*. 1476–1488. DOI: http://dx.doi.org/10.18653/v1/2020.acl-main.136

[134] Suncong Zheng, Feng Wang, Hongyun Bao, Yuexing Hao, Peng Zhou, and Bo Xu. 2017. Joint extraction of entities and relations based on a novel tagging scheme. In *Proceedings of the ACL 2017, Volume 1: Long Papers*. 1227–1236. DOI: http://dx.doi.org/10.18653/v1/P17-1113

[135] Yucheng Wang, Bowen Yu, Yueyang Zhang, Tingwen Liu, Hongsong Zhu, and Limin Sun. 2020. TPLinker: Single-stage joint extraction of entities and relations through token pair linking. In *Proceedings of the COLING, 2020*. 1572–1582. DOI: http://dx.doi.org/10.18653/v1/2020.coling-main.138

[136] Giannis Bekoulis, Johannes Deleu, Thomas Demeester, and Chris Develder. 2018. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Systems with Applications* 114 (2018), 34–45. DOI: http://dx.doi.org/10.1016/j.eswa.2018.07.032

[137] Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. 2019. Entity-relation extraction as multi-turn question answering. In *Proceedings of the ACL, 2019, Volume 1: Long Papers*. 1340–1350. DOI: http://dx.doi.org/10.18653/v1/p19-1129

[138] Hao Chen, Chenwei Zhang, Jun Li, Philip S. Yu, and Ning Jing. 2022. KGGen: A generative approach for incipient knowledge graph population. *IEEE Transactions on Knowledge and Data Engineering* 34, 5 (2022), 2254–2267. DOI: http://dx.doi.org/10.1109/TKDE.2020.3014166

[139] Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen-tau Yih. 2017. Cross-sentence n-ary relation extraction with graph LSTMs. *Transactions of the Association for Computational Linguistics* 5 (2017), 101–115. Retrieved from https://transacl.org/ojs/index.php/tacl/article/view/1028

[140] Linfeng Song, Yue Zhang, Zhiguo Wang, and Daniel Gildea. 2018. N-ary relation extraction using graph-state LSTM. In *Proceedings of the EMNLP, 2018*. 2226–2235. DOI: http://dx.doi.org/10.18653/v1/d18-1246

[141] Sunil Kumar Sahu, Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Inter-sentence relation extraction with document-level graph convolutional neural network. In *Proceedings of the ACL, 2019, Volume 1: Long Papers*. 4309–4316. DOI: http://dx.doi.org/10.18653/v1/p19-1423

[142] Hao Zhu, Yankai Lin, Zhiyuan Liu, Jie Fu, Tat-Seng Chua, and Maosong Sun. 2019. Graph neural networks with generated parameters for relation extraction. In *Proceedings of the ACL, 2019, Volume 1: Long Papers*. 1331–1339. DOI: http://dx.doi.org/10.18653/v1/p19-1128

[143] Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs. In *Proceedings of the EMNLP-IJCNLP , 2019*. 4924–4935. DOI: http://dx.doi.org/10.18653/v1/D19-1498

[144] Yi Luan, Dave Wadden, Luheng He, Amy Shah, Mari Ostendorf, and Hannaneh Hajishirzi. 2019. A general framework for information extraction using dynamic span graphs. In *Proceedings of the NAACL-HLT, 2019, Volume 1 (Long and Short Papers)*. 3036–3046. DOI: http://dx.doi.org/10.18653/v1/n19-1308

[145] Guoshun Nan, Zhijiang Guo, Ivan Sekulic, and Wei Lu. 2020. Reasoning with latent structure refinement for document-level relation extraction. In *Proceedings of the ACL, 2020*. 1546–1557. DOI: http://dx.doi.org/10.18653/v1/2020.acl-main.141

[146] Shuang Zeng, Runxin Xu, Baobao Chang, and Lei Li. 2020. Double graph based reasoning for document-level relation extraction. In *Proceedings of the EMNLP, 2020*. 1630–1640. DOI: http://dx.doi.org/10.18653/v1/2020.emnlp-main.127

[147] Wang Xu, Kehai Chen, and Tiejun Zhao. 2021. Document-level relation extraction with reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 14167–14175. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/17667

[148] Wang Xu, Kehai Chen, and Tiejun Zhao. 2021. Discriminative reasoning for document-level relation extraction. In *Proceedings of the ACL/IJCNLP , 2021.* 1653–1663. DOI: http://dx.doi.org/10.18653/v1/2021.findings-acl.144

[149] Zhenyu Zhang, Bowen Yu, Xiaobo Shu, Mengge Xue, Tingwen Liu, and Li Guo. 2021. From what to why: Improving relation extraction with rationale graph. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP.*

[150] Ningyu Zhang, Xiang Chen, Xin Xie, Shumin Deng, Chuanqi Tan, Mosha Chen, Fei Huang, Luo Si, and Huajun Chen. 2021. Document-level relation extraction as semantic segmentation. In *Proceedings of the IJCAI ,2021.* 3999–4006. DOI: http://dx.doi.org/10.24963/ijcai.2021/551

[151] Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. 2021. Document-level relation extraction with adaptive thresholding and localized context pooling. In *Proceedings of the AAAI Conference on Artificial Intelligence.* 14612–14620. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/17717

[152] Qingyu Tan, Ruidan He, Lidong Bing, and Hwee Tou Ng. 2022. Document-level relation extraction with adaptive focal loss and knowledge distillation. In *Proceedings of the ACL.* 1672–1681. DOI: http://dx.doi.org/10.18653/v1/2022.findings-acl.132

[153] Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. In *Proceedings of the ACL, 2019 , Volume 1: Long Papers.* 4762–4779. DOI: http://dx.doi.org/10.18653/v1/p19-1470

[154] Changsung Moon, Paul Jones, and Nagiza F. Samatova. 2017. Learning entity type embeddings for knowledge graph completion. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.* 2215–2218. DOI: http://dx.doi.org/10.1145/3132847.3133095

[155] Guoliang Ji, Kang Liu, Shizhu He, and Jun Zhao. 2016. Knowledge graph completion with adaptive sparse transfer matrix. In *Proceedings of the AAAI Conference on Artificial Intelligence.* 985–991. Retrieved from http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11982

[156] Ivana Balazevic, Carl Allen, and Timothy M. Hospedales. 2019. TuckER: Tensor factorization for knowledge graph completion. In *Proceedings of the EMNLP-IJCNLP, 2019.* 5184–5193. DOI: http://dx.doi.org/10.18653/v1/D19-1522

[157] Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Proceedings of the Advances in Neural Information Processing Systems.* 926–934. Retrieved from https://proceedings.neurips.cc/paper/2013/hash/b337e84de8752b27eda3a12363109e80-Abstract.html

[158] Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. 2020. Learning hierarchy-aware knowledge graph embeddings for link prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence.* 3065–3072. Retrieved from https://aaai.org/ojs/index.php/AAAI/article/view/5701

[159] Guanglin Niu, Bo Li, Yongfei Zhang, and Shiliang Pu. 2022. CAKE: A scalable commonsense-aware framework for multi-view knowledge graph completion. In arXiv:2202.13785. Retrieved from https://arxiv.org/abs/2202.13785

[160] Liang Wang, Wei Zhao, Zhuoyu Wei, and Jingming Liu. 2022. SimKGC: Simple contrastive knowledge graph completion with pre-trained language models. In *Proc. ACL.* DOI: http://dx.doi.org/10.48550/arXiv.2203.02167

[161] Kai Wang, Yu Liu, and Quan Z. Sheng. 2022. Swift and sure: Hardness-aware contrastive learning for low-dimensional knowledge graph embeddings. In *Proceedings of the ACM Web Conference 2022.* 838–849. DOI: http://dx.doi.org/10.1145/3485447.3511927

[162] Agustín Borrego, Daniel Ayala, Inma Hernández, Carlos R. Rivero, and David Ruiz. 2021. CAFE: Knowledge graph completion using neighborhood-aware features. *Engineering Applications of Artificial Intelligence* 103 (2021), 104302. DOI: http://dx.doi.org/10.1016/j.engappai.2021.104302

[163] Junkang Wu, Wentao Shi, Xuezhi Cao, Jiawei Chen, Wenqiang Lei, Fuzheng Zhang, Wei Wu, and Xiangnan He. 2021. DisenKGAT: Knowledge graph embedding with disentangled graph attention network. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management.* 2140–2149. DOI: http://dx.doi.org/10.1145/3459637.3482424

[164] Arvind Neelakantan, Benjamin Roth, and Andrew McCallum. 2015. Compositional vector space models for knowledge base completion. In *ACL, 2015, Volume 1: Long Papers.* 156–166. DOI: http://dx.doi.org/10.3115/v1/p15-1016

[165] Rajarshi Das, Arvind Neelakantan, David Belanger, and Andrew McCallum. 2017. Chains of reasoning over entities, relations, and text using recurrent neural networks. In *Proc. EACL, 2017, Volume 1: Long Papers.* 132–141. DOI: http://dx.doi.org/10.18653/v1/e17-1013

[166] Wenhan Xiong, Thien Hoang, and William Yang Wang. 2017. DeepPath: A reinforcement learning method for knowledge graph reasoning. In *Proc. EMNLP, 2017.* 564–573. DOI: http://dx.doi.org/10.18653/v1/d17-1060

[167] Xi Victoria Lin, Richard Socher, and Caiming Xiong. 2018. Multi-hop knowledge graph reasoning with reward shaping. In *Proc. EMNLP, 2018.* 3243–3253. DOI: http://dx.doi.org/10.18653/v1/d18-1362

[168] Zixuan Li, Xiaolong Jin, Saiping Guan, Yuanzhuo Wang, and Xueqi Cheng. 2018. Path reasoning over knowledge graph: A multi-agent and reinforcement learning based method. In *Proceedings of the 2018 IEEE International Conference on Data Mining Workshops.* 929–936. DOI: http://dx.doi.org/10.1109/ICDMW.2018.00135

[169] Yelong Shen, Jianshu Chen, Po-Sen Huang, Yuqing Guo, and Jianfeng Gao. 2018. M-Walk: Learning to walk over graphs using monte carlo tree search. In *Advances in Neural Information Processing Systems* (2018), 6787–6798. Retrieved from https://proceedings.neurips.cc/paper/2018/hash/c6f798b844366ccd65d99bc7f31e0e02-Abstract.html

[170] Yuyu Zhang, Hanjun Dai, Zornitsa Kozareva, Alexander J. Smola, and Le Song. 2018. Variational reasoning for question answering with knowledge graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 6069–6076. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16983

[171] Sara Kardani-Moghaddam, Rajkumar Buyya, and Kotagiri Ramamohanarao. 2021. ADRL: A hybrid anomaly-aware deep reinforcement learning-based resource scaling in clouds. *IEEE Transactions on Parallel and Distributed Systems* 32, 3 (2021), 514–526. DOI: http://dx.doi.org/10.1109/TPDS.2020.3025914

[172] Heng Wang, Shuangyin Li, Rong Pan, and Mingzhi Mao. 2019. Incorporating graph attention mechanism into knowledge graph reasoning based on deep reinforcement learning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. 2623–2631. DOI: http://dx.doi.org/10.18653/v1/D19-1264

[173] Mingming Zheng, Yanquan Zhou, and Qingyao Cui. 2021. Hierarchical policy network with multi-agent for knowledge graph reasoning based on reinforcement learning. In *Proceedings of the International Conference on Knowledge Science, Engineering and Management*. 445–457. DOI: http://dx.doi.org/10.1007/978-3-030-82136-4_36

[174] Prayag Tiwari, Hongyin Zhu, and Hari Mohan Pandey. 2021. DAPath: Distance-aware knowledge graph reasoning based on deep reinforcement learning. *Neural Networks* 135 (2021), 1–12. DOI: http://dx.doi.org/10.1016/j.neunet.2020.11.012

[175] Shuangyin Li, Heng Wang, Rong Pan, and Mingzhi Mao. 2021. MemoryPath: A deep reinforcement learning framework for incorporating memory component into knowledge graph reasoning. *Neurocomputing* 419 (2021), 273–286. DOI: http://dx.doi.org/10.1016/j.neucom.2020.08.032

[176] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38. DOI: http://dx.doi.org/10.1016/j.artint.2018.07.007

[177] Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. 2013. AMIE: Association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd international conference on World Wide Web*. 413–422. DOI: http://dx.doi.org/10.1145/2488388.2488425

[178] Pouya Ghiasnezhad Omran, Kewen Wang, and Zhe Wang. 2018. Scalable rule learning via learning representation. In *Proceedings of the IJCAI, 2018*. 2149–2155. DOI: http://dx.doi.org/10.24963/ijcai.2018/297

[179] Fan Yang, Zhilin Yang, and William W. Cohen. 2017. Differentiable learning of logical rules for knowledge base reasoning. In *Proceedings of the Advances in Neural Information Processing Systems*. 2319–2328. Retrieved from https://proceedings.neurips.cc/paper/2017/hash/0e55666a4ad822e0e34299df3591d979-Abstract.html

[180] Meng Qu and Jian Tang. 2019. Probabilistic logic neural networks for reasoning. In *Proceedings of the Advances in Neural Information Processing Systems*. 7710–7720. Retrieved from https://proceedings.neurips.cc/paper/2019/hash/13e5ebb0fa112fe1b31a1067962d74a7-Abstract.html

[181] Yuyu Zhang, Xinshi Chen, Yuan Yang, Arun Ramamurthy, Bo Li, Yuan Qi, and Le Song. 2020. Efficient probabilistic logic reasoning with graph neural networks. In *ICLR*. Retrieved from https://openreview.net/forum?id=rJg76kStwH

[182] Vicente Iván Sánchez Carmona, Tim Rocktäschel, Sebastian Riedel, and Sameer Singh. 2015. Towards extracting faithful and descriptive representations of latent variable models. In *Proceedings of the AAAI Spring Symposia Series*. Retrieved from http://www.aaai.org/ocs/index.php/SSS/SSS15/paper/view/10304

[183] Yatin Nandwani, Ankesh Gupta, Aman Agrawal, Mayank Singh Chauhan, Parag Singla, and Mausam. 2020. OxKBC: Outcome explanation for factorization based knowledge base completion. In *Proceedings of the Automated Knowledge Base Construction*. DOI: http://dx.doi.org/10.24432/C5V88F

[184] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. 2019. GNNExplainer: Generating explanations for graph neural networks. In *Proceedings of the Advances in Neural Information Processing Systems*. 9240–9251. Retrieved from https://proceedings.neurips.cc/paper/2019/hash/d80b7040b773199015de6d3b4293c8ff-Abstract.html

[185] Pouya Pezeshkpour, Yifan Tian, and Sameer Singh. 2019. Investigating robustness and interpretability of link prediction via adversarial modifications. In *Proceedings of the NAACL-HLT, 2019, Volume 1 (Long and Short Papers)*. 3336–3347. DOI: http://dx.doi.org/10.18653/v1/n19-1337

[186] Ruobing Xie, Zhiyuan Liu, Fen Lin, and Leyu Lin. 2018. Does william shakespeare REALLY write hamlet? Knowledge representation learning with confidence. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4954–4961. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16577

[187] Tiansi Dong, Zhigang Wang, Juanzi Li, Christian Bauckhage, and Armin B. Cremers. 2019. Triple classification using regions and fine-grained entity typing. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 77–85. DOI: http://dx.doi.org/10.1609/aaai.v33i01.330177

[188] Elvira Amador-Domínguez, Emilio Serrano, Daniel Manrique, Patrick Hohenecker, and Thomas Lukasiewicz. 2021. An ontology-based deep learning approach for triple classification with out-of-knowledge-base entities. *Information Sciences* 564 (2021), 85–102. DOI: http://dx.doi.org/10.1016/j.ins.2021.02.018

[189] Dai Quoc Nguyen, Tuan Nguyen, and Dinh Phung. 2020. A relational memory-based embedding model for triple classification and search personalization. In *Proceedings of the ACL, 2020*. 3429–3435. DOI: http://dx.doi.org/10.18653/v1/2020.acl-main.313

[190] Tao Sun, Jiaojiao Zhai, and Qi Wang. 2020. NovEA: A novel model of entity alignment using attribute triples and relation triples. In *Knowledge Science, Engineering and Management: 13th International Conference, KSEM 2020, Hangzhou, China, August 2830, 2020, Proceedings, Part I 13*. 161–173. DOI: http://dx.doi.org/10.1007/978-3-030-55130-8_14

[191] Fuzhen He, Zhixu Li, Qiang Yang, An Liu, Guanfeng Liu, Pengpeng Zhao, Lei Zhao, Min Zhang, and Zhigang Chen. 2019. Unsupervised entity alignment using attribute triples and relation triples. In *Database Systems for Advanced Applications: 24th International Conference, DASFAA 2019, Chiang Mai, Thailand, April 2225, 2019, Proceedings, Part I 24*. 367–382. DOI: http://dx.doi.org/10.1007/978-3-030-18576-3_22

[192] Hsiu-Wei Yang, Yanyan Zou, Peng Shi, Wei Lu, Jimmy Lin, and Xu Sun. 2019. Aligning cross-lingual entities with multi-aspect information. In *Proceedings of the EMNLP-IJCNLP , 2019*. 4430–4440. DOI: http://dx.doi.org/10.18653/v1/D19-1451

[193] Zequn Sun, Wei Hu, and Chengkai Li. 2017. Cross-lingual entity alignment via joint attribute-preserving embedding. In *The Semantic WebISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, October 2125, 2017, Proceedings, Part I 16*. 628–644. DOI: http://dx.doi.org/10.1007/978-3-319-68288-4_37

[194] Bayu Distiawan Trisedya, Jianzhong Qi, and Rui Zhang. 2019. Entity alignment between knowledge graphs using attribute embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 297–304. DOI: http://dx.doi.org/10.1609/aaai.v33i01.3301297

[195] Wang-Chiew Tan. 2020. Technical perspective: Entity matching with magellan. *Communications of the ACM* 63, 8 (2020), 82. DOI: http://dx.doi.org/10.1145/3405464

[196] Michael Sejr Schlichtkrull and Héctor Martínez Alonso. 2016. MSejrKu at SemEval-2016 Task 14: Taxonomy enrichment by evidence ranking. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*. 1337–1341. DOI: http://dx.doi.org/10.18653/v1/s16-1209

[197] Sidharth Mudgal, Han Li, Theodoros Rekatsinas, AnHai Doan, Youngchoon Park, Ganesh Krishnan, Rohit Deep, Esteban Arcaute, and Vijay Raghavendra. 2018. Deep learning for entity matching: A design space exploration. In *Proceedings of the 2018 International Conference on Management of Data*. 19–34. DOI: http://dx.doi.org/10.1145/3183713.3196926

[198] David Jurgens and Mohammad Taher Pilehvar. 2016. SemEval-2016 task 14: Semantic taxonomy enrichment. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. 1092–1102. DOI: http://dx.doi.org/10.18653/v1/s16-1169

[199] Max Berrendorf, Evgeniy Faerman, and Volker Tresp. 2021. Active learning for entity alignment. In *Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28April 1, 2021, Proceedings, Part I 43*. 48–62. DOI: http://dx.doi.org/10.1007/978-3-030-72113-8_4

[200] Hao Zhu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. 2017. Iterative entity alignment via joint knowledge embeddings. In *Proceedings of the IJCAI , 2017*. 4258–4264. DOI: http://dx.doi.org/10.24963/ijcai.2017/595

[201] Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. 2018. Bootstrapping entity alignment with knowledge graph embedding. In *Proceedings of the IJCAI, 2018*. 4396–4402. DOI: http://dx.doi.org/10.24963/ijcai.2018/611

[202] Qingheng Zhang, Zequn Sun, Wei Hu, Muhao Chen, Lingbing Guo, and Yuzhong Qu. 2019. Multi-view knowledge graph embedding for entity alignment. In *Proceedings of the IJCAI, 2019*. 5429–5435. DOI: http://dx.doi.org/10.24963/ijcai.2019/754

[203] Muhao Chen, Yingtao Tian, Mohan Yang, and Carlo Zaniolo. 2017. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In *Proceedings of the IJCAI, 2017*. 1511–1517. DOI: http://dx.doi.org/10.24963/ijcai.2017/209

[204] Zijie Huang, Zheng Li, Haoming Jiang, Tianyu Cao, Hanqing Lu, Bing Yin, Karthik Subbian, Yizhou Sun, and Wei Wang. 2022. Multilingual knowledge graph completion with self-supervised adaptive graph alignment. In *Proceedings of the ACL*. DOI: http://dx.doi.org/10.48550/arXiv.2203.14987

[205] Xiao Liu, Haoyun Hong, Xinghao Wang, Zeyi Chen, Evgeny Kharlamov, Yuxiao Dong, and Jie Tang. 2022. SelfKG: Self-supervised entity alignment in knowledge graphs. In *Proceedings of the ACM Web Conference 2022*. 860–870. DOI: http://dx.doi.org/10.1145/3485447.3511945

[206] Muhao Chen, Yingtao Tian, Kai-Wei Chang, Steven Skiena, and Carlo Zaniolo. 2018. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. In *Proceedings of the IJCAI, 2018*. 3998–4004. DOI: http://dx.doi.org/10.24963/ijcai.2018/556

[207] Bo Chen, Jing Zhang, Xiaobin Tang, Hong Chen, and Cuiping Li. 2020. JarKA: Modeling attribute interactions for cross-lingual knowledge alignment. In *Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 1114, 2020, Proceedings, Part I 24*. 845–856. DOI: http://dx.doi.org/10.1007/978-3-030-47426-3_65

[208] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. 2018. Cross-lingual knowledge graph alignment via graph convolutional networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 349–357. DOI: http://dx.doi.org/10.18653/v1/d18-1032

[209] Yao Zhu, Hongzhi Liu, Zhonghai Wu, and Yingpeng Du. 2021. Relation-aware neighborhood matching model for entity alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4749–4756. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/16606

[210] Yuting Wu, Xiao Liu, Yansong Feng, Zheng Wang, Rui Yan, and Dongyan Zhao. 2019. Relation-aware entity alignment for heterogeneous knowledge graphs. In *Proceedings of the IJCAI, 2019*. 5278–5284. DOI: http://dx.doi.org/10.24963/ijcai.2019/733

[211] Kun Xu, Liwei Wang, Mo Yu, Yansong Feng, Yan Song, Zhiguo Wang, and Dong Yu. 2019. Cross-lingual knowledge graph alignment via graph matching neural network. In *Proceeding of the ACL, 2019, Volume 1: Long Papers*. 3156–3161. DOI: http://dx.doi.org/10.18653/v1/p19-1304

[212] Zhiyuan Liu, Yixin Cao, Liangming Pan, Juanzi Li, and Tat-Seng Chua. 2020. Exploring and evaluating attributes, values, and structures for entity alignment. In *Proceedings of the EMNLP, 2020*. 6355–6364. DOI: http://dx.doi.org/10.18653/v1/2020.emnlp-main.515

[213] Yixin Cao, Zhiyuan Liu, Chengjiang Li, Zhiyuan Liu, Juanzi Li, and Tat-Seng Chua. 2019. Multi-channel graph neural network for entity alignment. In *Proceedings of the ACL, 2019, Volume 1: Long Papers*. 1452–1461. DOI: http://dx.doi.org/10.18653/v1/p19-1140

[214] Shixian Jiang, Tiezheng Nie, Derong Shen, Yue Kou, and Ge Yu. 2021. Entity alignment of knowledge graph by joint graph attention and translation representation. In *Web Information Systems and Applications: 18th International Conference, WISA 2021, Kaifeng, China, September 2426, 2021, Proceedings 18*. 347–358. DOI: http://dx.doi.org/10.1007/978-3-030-87571-8_30

[215] Tianwen Jiang, Tong Zhao, Bing Qin, Ting Liu, Nitesh V. Chawla, and Meng Jiang. 2019. The role of c̈ondition̈: A novel scientific knowledge graph representation and construction model. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1634–1642. DOI: http://dx.doi.org/10.1145/3292500.3330942

[216] Tingyue Zheng, Ziqiang Xu, Yufan Li, Yuan Zhao, Bin Wang, and Xiaochun Yang. 2021. A novel conditional knowledge graph representation and construction. In *Proceedings of the CAAI International Conference on Artificial Intelligence*. 383–394. DOI: http://dx.doi.org/10.1007/978-3-030-93049-3_32

[217] Fei Cheng and Yusuke Miyao. 2017. Classifying temporal relations by bidirectional LSTM over dependency paths. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 1–6. DOI: http://dx.doi.org/10.18653/v1/P17-2001

[218] Yuanliang Meng, Anna Rumshisky, and Alexey Romanov. 2017. Temporal information extraction for question answering using syntactic dependencies in an LSTM-based architecture. In *Proceedings of the EMNLP, 2017*. 887–896. DOI: http://dx.doi.org/10.18653/v1/d17-1092

[219] Siddharth Vashishtha, Benjamin Van Durme, and Aaron Steven White. 2019. Fine-grained temporal relation extraction. In *Proceedings of the ACL, 2019, Volume 1: Long Papers*. 2906–2919. DOI: http://dx.doi.org/10.18653/v1/p19-1280

[220] Puneet Mathur, Rajiv Jain, Franck Dernoncourt, Vlad I. Morariu, Quan Hung Tran, and Dinesh Manocha. 2021. TIMERS: Document-level temporal relation extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 524–533. DOI: http://dx.doi.org/10.18653/v1/2021.acl-short.67

[221] Julien Leblay and Melisachew Wudage Chekol. 2018. Deriving validity time in knowledge graph. In *Companion Proceedings of the The Web Conference 2018*. 1771–1776. DOI: http://dx.doi.org/10.1145/3184558.3191639

[222] Shib Sankar Dasgupta, Swayambhu Nath Ray, and Partha P. Talukdar. 2018. HyTE: Hyperplane-based temporally aware knowledge graph embedding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2001–2011. DOI: http://dx.doi.org/10.18653/v1/d18-1225

[223] Alberto García-Durán, Sebastijan Dumancic, and Mathias Niepert. 2018. Learning sequence encoders for temporal knowledge graph completion. In *Proceedings of the EMNLP, 2018*. 4816–4821. https://aclanthology.org/D18-1516/

[224] Yu Liu, Wen Hua, Kexuan Xin, and Xiaofang Zhou. 2019. Context-aware temporal knowledge graph embedding. In *Web Information Systems EngineeringWISE 2019: 20th International Conference, Hong Kong, China, November 2630, 2019, Proceedings 20*. 583–598. DOI: http://dx.doi.org/10.1007/978-3-030-34223-4_37

[225] Lifan Lin and Kun She. 2020. Tensor decomposition-based temporal knowledge graph embedding. In *Proceedings of the 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence*. 969–975. DOI: http://dx.doi.org/10.1109/ICTAI50040.2020.00151

[226] Timothée Lacroix, Guillaume Obozinski, and Nicolas Usunier. 2020. Tensor decompositions for temporal knowledge base completion. In *ICLR*. Retrieved from https://openreview.net/forum?id=rke2P1BFwS

[227] Pengpeng Shao, Dawei Zhang, Guohua Yang, Jianhua Tao, Feihu Che, and Tong Liu. 2022. Tucker decomposition-based temporal knowledge graph completion. *Knowledge-Based Systems* 238 (2022), 107841. DOI: http://dx.doi.org/10.1016/j.knosys.2021.107841

[228] Wessel Radstok, Mel Chekol, and Yannis Velegrakis. 2021. Leveraging static models for link prediction in temporal knowledge graphs. In *Proceedings of the 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence*. 1034–1041. DOI: http://dx.doi.org/10.1109/ICTAI52525.2021.00165

[229] Jaehun Jung, Jinhong Jung, and U Kang. 2021. Learning to walk across time for interpretable temporal knowledge graph completion. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 786–795. DOI: http://dx.doi.org/10.1145/3447548.3467292

[230] Hao Liu, Shu-wang Zhou, Changfang Chen, Tianlei Gao, Jiyong Xu, and Minglei Shu. 2022. Dynamic knowledge graph reasoning based on deep reinforcement learning. *Knowledge-Based Systems* 241 (2022), 108235. DOI: http://dx.doi.org/10.1016/j.knosys.2022.108235

[231] Tingsong Jiang, Tianyu Liu, Tao Ge, Lei Sha, Sujian Li, Baobao Chang, and Zhifang Sui. 2016. Encoding temporal information for time-aware link prediction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2350–2354. DOI: http://dx.doi.org/10.18653/v1/d16-1260

[232] Kui Yu, Xianjie Guo, Lin Liu, Jiuyong Li, Hao Wang, Zhaolong Ling, and Xindong Wu. 2020. Causality-based feature selection: Methods and evaluations. *ACM Computing Surveys* 53, 5 (2020), 111:1–111:36. DOI: http://dx.doi.org/10.1145/3409382

[233] Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. 2017. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *Proceedings of the International Conference on Machine Learning*. 3462–3471. Retrieved from http://proceedings.mlr.press/v70/trivedi17a.html

[234] Woojeong Jin, Changlin Zhang, Pedro A. Szekely, and Xiang Ren. 2019. Recurrent event network for reasoning over temporal knowledge graphs. arXiv:1904.05530. Retrieved from http://arxiv.org/abs/1904.05530

[235] Xuhui Li, Liuyan Liu, Xiaoguang Wang, Yiwen Li, Qingfeng Wu, and Tieyun Qian. 2021. Towards evolutionary knowledge representation under the big data circumstance. *The Electronic Library* 39, 3 (2021), 392–410. DOI: http://dx.doi.org/10.1108/EL-11-2020-0318

[236] Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. 2020. DyERNIE: Dynamic evolution of riemannian manifold embeddings for temporal knowledge graph completion. In *EMNLP, 2020*. 7301–7316. DOI: http://dx.doi.org/10.18653/v1/2020.emnlp-main.593

[237] Tony Gracious, Shubham Gupta, Arun Kanthali, Rui M. Castro, and Ambedkar Dukkipati. 2021. Neural latent space model for dynamic networks and temporal knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4054–4062. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/16526

[238] Yuchen Yan, Lihui Liu, Yikun Ban, Baoyu Jing, and Hanghang Tong. 2021. Dynamic knowledge graph alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4564–4572. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/16585

[239] Kristinn Andersen, Helgi Thorbergsson, Saemundur E Thorsteinsson, and Karl S Gudmundsson. 2020. The small languages in the large world of technology. In *Proceedings of the 2020 IEEE International Professional Communication Conference*. IEEE, 30–33.

[240] Tristan Miller, Christian Hempelmann, and Iryna Gurevych. 2017. SemEval-2017 Task 7: Detection and interpretation of english puns. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. 58–68. DOI: http://dx.doi.org/10.18653/v1/S17-2005

[241] Cunxiang Wang, Shuailong Liang, Yili Jin, Yilong Wang, Xiaodan Zhu, and Yue Zhang. 2020. SemEval-2020 Task 4: Commonsense validation and explanation. In *Proceedings of the SemEval@COLING, 2020*. 307–321. DOI: http://dx.doi.org/10.18653/v1/2020.semeval-1.39

[242] Hongbin Ye, Ningyu Zhang, Hui Chen, and Huajun Chen. 2022. Generative knowledge graph construction: A review. In *EMNLP*. 1–17.

[243] Ye Liu, Hui Li, Alberto García-Durán, Mathias Niepert, Daniel Oñoro-Rubio, and David S. Rosenblum. 2019. MMKG: Multi-modal knowledge graphs. In *The Semantic Web: 16th International Conference, ESWC 2019, Portoro, Slovenia, June 26, 2019, Proceedings 16*. 459–474. DOI: http://dx.doi.org/10.1007/978-3-030-21348-0_30

[244] Shahi Dost, Luciano Serafini, Marco Rospocher, Lamberto Ballan, and Alessandro Sperduti. 2022. Aligning and linking entity mentions in image, text, and knowledge base. *Data and Knowledge Engineering* 138 (2022), 101975. DOI: http://dx.doi.org/10.1016/j.datak.2021.101975

[245] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He. 2022. A survey on knowledge graph-based recommender systems. *IEEE Trans. Knowl. Data Eng.* 34, 8 (2022), 3549–3568. DOI: http://dx.doi.org/10.1109/TKDE.2020.3028705

[246] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. DKN: Deep knowledge-aware network for news recommendation. In *WWW*. Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis (Eds.), 1835–1844. DOI: http://dx.doi.org/10.1145/3178876.3186175

[247] Cairong Yan, Shuai Liu, Yanting Zhang, Zijian Wang, and Pengwei Wang. 2021. A multi-task learning approach for recommendation based on knowledge graph. In *International Joint Conference on Neural Networks, IJCNN 2021*. IEEE, 1–8. DOI: http://dx.doi.org/10.1109/IJCNN52387.2021.9533556

[248] Jeff Z. Pan, Siyana Pavlova, Chenxi Li, Ningxi Li, Yangmei Li, and Jinshuo Liu. 2018. Content based fake news detection using knowledge graphs. In *ISWC*. 669–683. DOI: http://dx.doi.org/10.1007/978-3-030-00671-6_39

[249] Yaqian Dun, Kefei Tu, Chen Chen, Chunyan Hou, and Xiaojie Yuan. 2021. KAN: Knowledge-aware attention network for fake news detection. In *AAAI*. 81–89. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/16080

[250] Shengsheng Qian, Jun Hu, Quan Fang, and Changsheng Xu. 2021. Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection. *ACM Trans. Multim. Comput. Commun. Appl.* 17, 3 (2021), 98:1–98:23. DOI: http://dx.doi.org/10.1145/3451215

[251] Seungwhan Moon, Pararth Shah, Anuj Kumar, and Rajen Subba. 2019. OpenDialKG: Explainable conversational reasoning with attention-based walks over knowledge graphs. In *ACL*. 845–854. DOI: http://dx.doi.org/10.18653/v1/p19-1081

[252] Jinghui Qin, Zheng Ye, Jianheng Tang, and Xiaodan Liang. 2020. Dynamic knowledge routing network for target-guided open-domain conversation. In *AAAI, IAAI, EAAI 2020*. 8657–8664. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/6390

[253] Bo Xu, Yong Xu, Jiaqing Liang, Chenhao Xie, Bin Liang, Wanyun Cui, and Yanghua Xiao. 2017. CN-DBpedia: A never-ending chinese knowledge extraction system. In *IEA/AIE, 2017, Proceedings, Part II*. 428–438. DOI: http://dx.doi.org/10.1007/978-3-319-60045-1_44

[254] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: A core of semantic knowledge. In *Proc. WWW, 2007*. 697–706. DOI: http://dx.doi.org/10.1145/1242572.1242667

[255] Jindong Chen, Ao Wang, Jiangjie Chen, Yanghua Xiao, Zhendong Chu, Jingping Liu, Jiaqing Liang, and Wei Wang. 2019. CN-probase: A data-driven approach for large-scale chinese taxonomy construction. In *Proc. ICDE, 2019*. 1706–1709. DOI: http://dx.doi.org/10.1109/ICDE.2019.00178

[256] Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In *Proc. ACL, 2010*. 216–225. Retrieved from https://aclanthology.org/P10-1023/

[257] Hugo Liu and Push Singh. 2004. ConceptNeta practical commonsense reasoning tool-kit. *BT Technology Journal* 22, 4 (2004), 211–226.

[258] Zhendong Dong and Qiang Dong. 2003. HowNet - a hybrid language and knowledge resource. In *ICNLP, 2003, Proceedings*. 820–824. DOI: http://dx.doi.org/10.1109/NLPKE.2003.1276017

[259] Shiyi Han, Yuhui Zhang, Yunshan Ma, Cunchao Tu, Zhipeng Guo, Zhiyuan Liu, and Maosong Sun. 2016. THUOCL: Tsinghua Open Chinese Lexicon Tsinghua University (2016). Retrieved from http://thuocl.thunlp.org/

[260] Jie Tang, Duo Zhang, and Limin Yao. 2007. Social network extraction of academic researchers. In *Proc. ICDM, 2007*. 292–301. DOI: http://dx.doi.org/10.1109/ICDM.2007.30

[261] Fanjin Zhang, Xiao Liu, Jie Tang, Yuxiao Dong, Peiran Yao, Jie Zhang, Xiaotao Gu, Yan Wang, Bin Shao, Rui Li, and Kuansan Wang. 2019. OAG: Toward linking large-scale heterogeneous entity graphs. In *KDD, 2019*. 2585–2595. DOI: http://dx.doi.org/10.1145/3292500.3330785

[262] Huajun Chen, Ning Hu, Guilin Qi, Haofen Wang, Zhen Bi, Jie Li, and Fan Yang. 2021. OpenKG chain: A blockchain infrastructure for open knowledge graphs. *Data Intell.* 3, 2 (2021), 205–227. DOI: http://dx.doi.org/10.1162/dint_a_00095

[263] Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proc. EMNLP*.

[264] Mausam, Michael Schmitz, Stephen Soderland, Robert Bart, and Oren Etzioni. 2012. Open language learning for information extraction. In *Proc. EMNLP-CoNLL, ACL 2012*. 523–534. Retrieved from https://aclanthology.org/D12-1048/

[265] Steven Bird. 2006. NLTK: The natural language toolkit. In *ACL, 2006*. Nicoletta Calzolari, Claire Cardie, and Pierre Isabelle (Eds.). DOI: http://dx.doi.org/10.3115/1225403.1225421

[266] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford CoreNLP natural language processing toolkit. In *Proc. ACL, 2014*. 55–60. DOI: http://dx.doi.org/10.3115/v1/p14-5010

[267] Ziqi Zhang. 2017. Effective and efficient semantic table interpretation using TableMiner$^+$. *Semantic Web* 8, 6 (2017), 921–957. DOI: http://dx.doi.org/10.3233/SW-160242

[268] Marco Cremaschi, Anisa Rula, Alessandra Siano, and Flavio De Paoli. 2019. MantisTable: A tool for creating semantic annotations on tabular data. In *ESWC 2019 Satellite Events, 2019, Revised Selected Papers*. 18–23. DOI: http://dx.doi.org/10.1007/978-3-030-32327-1_4

[269] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python. (2020). DOI: http://dx.doi.org/10.5281/zenodo.1212303
[270] Xu Han, Tianyu Gao, Yuan Yao, Deming Ye, Zhiyuan Liu, and Maosong Sun. 2019. OpenNRE: An open and extensible toolkit for neural relation extraction. In *Proceedings of EMNLP-IJCNLP: System Demonstrations.* 169–174. DOI: http://dx.doi.org/10.18653/v1/D19-3029
[271] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of KDD.* 855–864.
[272] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of WWW.* 1067–1077.
[273] Wei Hu and Yuzhong Qu. 2008. Falcon-AO: A practical ontology matching system. *J. Web Semant.* 6, 3 (2008), 237–239. DOI: http://dx.doi.org/10.1016/j.websem.2008.02.006
[274] Hao Peng, Haoran Li, Yangqiu Song, Vincent W. Zheng, and Jianxin Li. 2021. Differentially private federated knowledge graphs embedding. In *CIKM, 2021.* 1416–1425. DOI: http://dx.doi.org/10.1145/3459637.3482252
[275] Mingyang Chen, Wen Zhang, Zonggang Yuan, Yantao Jia, and Huajun Chen. 2021. FedE: Embedding knowledge graphs in federated setting. In *IJCKG, 2021.* 80–88. DOI: http://dx.doi.org/10.1145/3502223.3502233
[276] Cynthia Dwork. 2008. Differential privacy: A survey of results. In *TAMC 2008.* Manindra Agrawal, Ding-Zhu Du, Zhenhua Duan, and Angsheng Li (Eds.), Vol. 4978, 1–19. DOI: http://dx.doi.org/10.1007/978-3-540-79228-4_1
[277] NhatHai Phan, Minh N. Vu, Yang Liu, Ruoming Jin, Dejing Dou, Xintao Wu, and My T. Thai. 2019. Heterogeneous gaussian mechanism: Preserving differential privacy in deep learning with provable robustness. In *IJCAI 2019.* 4753–4759. DOI: http://dx.doi.org/10.24963/ijcai.2019/660
[278] Ferdinando Fioretto, Cuong Tran, Pascal Van Hentenryck, and Keyu Zhu. 2022. Differential privacy and fairness in decisions and learning tasks: A survey. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence.* ijcai.org, 5470–5477. DOI: http://dx.doi.org/10.24963/ijcai.2022/766
[279] Mohammad Mahdi Khalili, Xueru Zhang, Mahed Abroshan, and Somayeh Sojoudi. 2021. Improving fairness and privacy in selection problems. In *AAAI 2021, IAAI 2021, EAAI 2021.* 8092–8100. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/16986
[280] Aisling McGlinchey and Oliver Mason. 2021. Observations on the bias of nonnegative mechanisms for differential privacy. arXiv:2101.02957. Retrieved from https://arxiv.org/abs/2101.02957
[281] Ken Mishima and Hayato Yamana. 2022. A survey on explainable fake news detection. *IEICE Trans. Inf. Syst.* 105-D, 7 (2022), 1249–1257.
[282] Rubén Tolosana, Rubén Vera-Rodríguez, Julian Fiérrez, Aythami Morales, and Javier Ortega-Garcia. 2020. Deepfakes and beyond: A survey of face manipulation and fake detection. *Inf. Fusion* 64 (2020), 131–148.
[283] Luciano Floridi and Massimo Chiriatti. 2020. GPT-3: Its nature, scope, limits, and consequences. *Minds Mach.* 30, 4 (2020), 681–694. DOI: http://dx.doi.org/10.1007/s11023-020-09548-1
[284] Eric Mitchell, Yoonho Lee, Alexander Khazatsky, Christopher D. Manning, and Chelsea Finn. 2023. DetectGPT: Zero-Shot Machine-Generated Text Detection using Probability Curvature. In *ICML 2023 (Proceedings of Machine Learning Research, Vol. 202).* PMLR, 24950–24962. Retrieved from https://proceedings.mlr.press/v202/mitchell23a.html
[285] Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. KG-BERT: BERT for knowledge graph completion. arXiv:1909.03193. Retrieved from https://arxiv.org/abs/1909.03193
[286] Michihiro Yasunaga, Antoine Bosselut, Hongyu Ren, Xikun Zhang, Christopher D. Manning, Percy Liang, and Jure Leskovec. 2022. Deep bidirectional language-knowledge graph pretraining. In *NeurIPS.* Retrieved from http://papers.nips.cc/paper_files/paper/2022/hash/f224f056694bcfe465c5d84579785761-Abstract-Conference.html
[287] Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick S. H. Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander H. Miller. 2019. Language models as knowledge bases?. In *EMNLP-IJCNLP.* 2463–2473. DOI: http://dx.doi.org/10.18653/v1/D19-1250
[288] Ye Liu, Yao Wan, Lifang He, Hao Peng, and Philip S. Yu. 2021. KG-BART: Knowledge graph-augmented BART for generative commonsense reasoning. In *AAAI.* 6418–6425. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/16796
[289] Henry Weld, Xiaoqi Huang, Siqu Long, Josiah Poon, and Soyeon Caren Han. 2023. A Survey of Joint Intent Detection and Slot Filling Models in Natural Language Understanding. *ACM Comput. Surv.* 55, 8 (2023), 156:1–156:38. DOI: http://dx.doi.org/10.1145/3547138
[290] Henry Weld, Xiaoqi Huang, Siqi Long, Josiah Poon, and Soyeon Caren Han. 2021. A survey of joint intent detection and slot-filling models in natural language understanding. arXiv:2101.08091. Retrieved from https://arxiv.org/abs/2101.08091
[291] Nicholas Kushmerick. 2000. Wrapper induction: Efficiency and expressiveness. *Artificial Intelligence* 118, 1-2 (2000), 15–68.

[292] David Buttler, Ling Liu, and Calton Pu. 2001. A fully automated object extraction system for the world wide web. In *Proc. ICDCS, 2001*. 361–370. DOI: http://dx.doi.org/10.1109/ICDSC.2001.918966

[293] Beth M. Sundheim. 1996. The message understanding conferences. In *TIPSTER TEXT PROGRAM PHASE II: Proceedings of a Workshop held at Vienna, 1996*. 35–37. DOI: http://dx.doi.org/10.3115/1119018.1119025

[294] S. Thenmalar, Balaji Jagan, and T. V. Geetha. 2015. Semi-supervised bootstrapping approach for named entity recognition. arXiv:1511.06833. Retrieved from https://arxiv.org/abs/1511.06833

[295] Sergey Brin. 1998. Extracting patterns and relations from the world wide web. In *The World Wide Web and Databases, International Workshop WebDB'98, 1998, Selected Papers*. 172–183. DOI: http://dx.doi.org/10.1007/10704656_11

[296] Eugene Agichtein and Luis Gravano. 2000. *Snowball*: Extracting Relations from Large Plain-text Collections. In *Proc. ACM, 2000*. 85–94. DOI: http://dx.doi.org/10.1145/336597.336644

[297] Meng Jiang, Jingbo Shang, Taylor Cassidy, Xiang Ren, Lance M. Kaplan, Timothy P. Hanratty, and Jiawei Han. 2017. MetaPAD: Meta pattern discovery from massive text corpora. In *KDD, 2017*. 877–886. DOI: http://dx.doi.org/10.1145/3097983.3098105

[298] Yanif Ahmad, Tudor Antoniu, Sharon Goldwater, and Shriram Krishnamurthi. 2003. A type system for statically detecting spreadsheet errors. In *ASE, 2003*. 174–183. DOI: http://dx.doi.org/10.1109/ASE.2003.1240305

[299] Yoones A. Sekhavat, Francesco Di Paolo, Denilson Barbosa, and Paolo Merialdo. 2014. Knowledge base augmentation using tabular data. In *Proc. WWW, 2014 (CEUR Workshop Proceedings)*. Retrieved from http://ceur-ws.org/Vol-1184/ldow2014_paper_02.pdf

[300] Emir Muñoz, Aidan Hogan, and Alessandra Mileo. 2014. Using linked data to mine RDF from wikipedia's tables. In *WSDM, 2014*. 533–542. DOI: http://dx.doi.org/10.1145/2556195.2556266

[301] Sebastian Krause, Hong Li, Hans Uszkoreit, and Feiyu Xu. 2012. Large-scale learning of relation-extraction rules with distant supervision from the web. In *ISWC, 2012, Proceedings, Part I*, Vol. 7649. 263–278. DOI: http://dx.doi.org/10.1007/978-3-642-35176-1_17

[302] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proc. AAAI-15, 2015*. 2181–2187. Retrieved from http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9571

[303] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proc. AAAI-14, 2014*. 1112–1119. Retrieved from http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8531

[304] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In *Proc. ICML, 2011*. 809–816. Retrieved from https://icml.cc/2011/papers/438_icmlpaper.pdf

[305] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR, 2015, Conference Track Proceedings*. Retrieved from http://arxiv.org/abs/1412.6575

[306] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2D knowledge graph embeddings. In *AAAI, IAAI, EAAI, 2018*. 1811–1818. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17366

[307] Zhaoli Zhang, Zhifei Li, Hai Liu, and Neal N. Xiong. 2022. Multi-scale dynamic convolutional network for knowledge graph embedding. *IEEE Trans. Knowl. Data Eng.* 34, 5 (2022), 2335–2347. DOI: http://dx.doi.org/10.1109/TKDE.2020.3005952

[308] Jun Yuan, Neng Gao, and Ji Xiang. 2019. TransGate: Knowledge graph embedding with shared gate structure. In *AAAI, IAAI, EAAI, 2019*. 3100–3107. DOI: http://dx.doi.org/10.1609/aaai.v33i01.33013100

[309] Yukun Zuo, Quan Fang, Shengsheng Qian, Xiaorui Zhang, and Changsheng Xu. 2018. Representation learning of knowledge graphs with entity attributes and multimedia descriptions. In *BigMM, 2018*. 1–5. DOI: http://dx.doi.org/10.1109/BigMM.2018.8499179

[310] Baoxu Shi and Tim Weninger. 2018. Open-world knowledge graph completion. In *AAAI, IAAI, EAAI, 2018*. 1957–1964. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16055

[311] Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *ESWC, 2018, Proceedings*. 593–607. DOI: http://dx.doi.org/10.1007/978-3-319-93417-4_38

[312] Chao Shang, Yun Tang, Jing Huang, Jinbo Bi, Xiaodong He, and Bowen Zhou. 2019. End-to-end structure-aware convolutional networks for knowledge base completion. In *AAAI, IAAI, EAAI, 2019*. 3060–3067. DOI: http://dx.doi.org/10.1609/aaai.v33i01.33013060

[313] Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha P. Talukdar. 2020. Composition-based multi-relational graph convolutional networks. In *ICLR*. Retrieved from https://openreview.net/forum?id=BylA_C4tPr

[314] Baoxu Shi and Tim Weninger. 2017. ProjE: Embedding projection for knowledge graph completion. In *Proc. AAAI-17, 2017*. 1236–1242. Retrieved from http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14279

[315] Saiping Guan, Xiaolong Jin, Yuanzhuo Wang, and Xueqi Cheng. 2018. Shared embedding based neural networks for knowledge graph completion. In *CIKM, 2018*. 247–256. DOI: http://dx.doi.org/10.1145/3269206.3271704

[316] Junheng Hao, Muhao Chen, Wenchao Yu, Yizhou Sun, and Wei Wang. 2019. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts. In *KDD, 2019*. 1709–1719. DOI: http://dx.doi.org/10.1145/3292500.3330838

[317] Yu Zhao, Anxiang Zhang, Ruobing Xie, Kang Liu, and Xiaojie Wang. 2020. Connecting embeddings for knowledge graph entity typing. In *Proc. ACL, 2020*. 6419–6428. DOI: http://dx.doi.org/10.18653/v1/2020.acl-main.572

[318] Hailong Jin, Lei Hou, and Juanzi Li. 2018. Type hierarchy enhanced heterogeneous network embedding for fine-grained entity typing in knowledge bases. In *NLP-NABD, 2018, Proceedings*. 170–182. DOI: http://dx.doi.org/10.1007/978-3-030-01716-3_15

[319] Hailong Jin, Lei Hou, Juanzi Li, and Tiansi Dong. 2018. Attributed and predictive entity embedding for fine-grained entity typing in knowledge bases. In *Proc. COLING, 2018*. 282–292. Retrieved from https://aclanthology.org/C18-1024/

[320] Yixin Cao, Lei Hou, Juanzi Li, Zhiyuan Liu, Chengjiang Li, Xu Chen, and Tiansi Dong. 2018. Joint representation learning of cross-lingual words and entities via attentive distant supervision. In *Proc. EMNLP, 2018*. 227–237. DOI: http://dx.doi.org/10.18653/v1/d18-1021

[321] Wei Shen, Jianyong Wang, Ping Luo, and Min Wang. 2012. LIEGE: : Link entities in web lists with knowledge base. In *KDD, 2012*. 1424–1432. DOI: http://dx.doi.org/10.1145/2339530.2339753

[322] Amit Bagga and Breck Baldwin. 1998. Entity-based cross-document coreferencing using the vector space model. In *COLING-ACL, 1998, Proceedings of the Conference*. 79–85. DOI: http://dx.doi.org/10.3115/980845.980859

[323] Ivo Lasek and Peter Vojtáš. 2012. Various approaches to text representation for named entity disambiguation. In *IIWAS, 2012*. 256–262. DOI: http://dx.doi.org/10.1145/2428736.2428776

[324] Hongzhao Huang, Larry P. Heck, and Heng Ji. 2015. Leveraging deep neural networks and knowledge graphs for entity disambiguation. arXiv:1504.07678. Retrieved from https://arxiv.org/abs/1504.07678

[325] Wei Fang, Jianwen Zhang, Dilin Wang, Zheng Chen, and Ming Li. 2016. Entity disambiguation by knowledge and text jointly embedding. In *SIGNLL, CoNLL, ACL, 2016*. 260–269. DOI: http://dx.doi.org/10.18653/v1/k16-1026

[326] Greg Durrett and Dan Klein. 2013. Easy victories and uphill battles in coreference resolution. In *Proc. EMNLP, 2013, A meeting of SIGDAT, a Special Interest Group of the ACL*. 1971–1982. Retrieved from https://aclanthology.org/D13-1203/

[327] Sebastian Martschat and Michael Strube. 2015. Latent structures for coreference resolution. *Trans. Assoc. Comput. Linguistics* 3 (2015), 405–418. Retrieved from https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/604

[328] Kaushik Chakrabarti, Surajit Chaudhuri, Tao Cheng, and Dong Xin. 2012. A framework for robust discovery of entity synonyms. In *KDD, 2012*. 1384–1392. DOI: http://dx.doi.org/10.1145/2339530.2339743

[329] Xiang Ren, Zeqiu Wu, Wenqi He, Meng Qu, Clare R. Voss, Heng Ji, Tarek F. Abdelzaher, and Jiawei Han. 2017. CoType: Joint extraction of typed entities and relations with knowledge bases. In *Proc. WWW, 2017*. 1015–1024. DOI: http://dx.doi.org/10.1145/3038912.3052708

[330] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NeurIPS, Proceedings, 2013*. 2787–2795. Retrieved from https://proceedings.neurips.cc/paper/2013/hash/1cecc7a77928ca8133fa24680a88d2f9-Abstract.html

[331] Christian Meilicke, Manuel Fink, Yanjie Wang, Daniel Ruffinelli, Rainer Gemulla, and Heiner Stuckenschmidt. 2018. Fine-grained evaluation of rule- and embedding-based systems for knowledge graph completion. In *ISWC, 2018, Proceedings, Part I*. 3–20. DOI: http://dx.doi.org/10.1007/978-3-030-00671-6_1

[332] Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. 2016. Jointly embedding knowledge graphs and logical rules. In *Proc. EMNLP, 2016*. 192–202. DOI: http://dx.doi.org/10.18653/v1/d16-1019

[333] Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. 2018. Knowledge graph embedding with iterative guidance from soft rules. In *AAAI, IAAI, EAAI, 2018*. 4816–4823. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16369

[334] Sinno Jialin Pan, Zhiqiang Toh, and Jian Su. 2013. Transfer joint embedding for cross-domain named entity recognition. *ACM Trans. Inf. Syst.* 31, 2 (2013), 7. DOI: http://dx.doi.org/10.1145/2457465.2457467

[335] Bill Yuchen Lin and Wei Lu. 2018. Neural adaptation layers for cross-domain named entity recognition. In *Proc. EMNLP, 2018*. 2012–2022. DOI: http://dx.doi.org/10.18653/v1/d18-1226

[336] Karthik Narasimhan, Adam Yala, and Regina Barzilay. 2016. Improving information extraction by acquiring external evidence with reinforcement learning. In *Proc. EMNLP, 2016*. 2355–2365. DOI: http://dx.doi.org/10.18653/v1/d16-1261

[337] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin A. Riedmiller, Andreas Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. 2015. Human-level control through deep reinforcement learning. *Nat.* 518, 7540 (2015), 529–533. DOI: http://dx.doi.org/10.1038/nature14236

[338] YaoSheng Yang, Wenliang Chen, Zhenghua Li, Zhengqiu He, and Min Zhang. 2018. Distantly supervised NER with partial annotation learning and reinforcement learning. In *Proc. COLING, 2018*. 2159–2169. Retrieved from https://aclanthology.org/C18-1183/

[339] Joey Tianyi Zhou, Hao Zhang, Di Jin, Hongyuan Zhu, Meng Fang, Rick Siow Mong Goh, and Kenneth Kwok. 2019. Dual adversarial neural transfer for low-resource named entity recognition. In *Proc. ACL , 2019, Volume 1: Long Papers*. 3461–3471. DOI: http://dx.doi.org/10.18653/v1/p19-1336

[340] Pengfei Cao, Yubo Chen, Kang Liu, Jun Zhao, and Shengping Liu. 2018. Adversarial transfer learning for chinese named entity recognition with self-attention mechanism. In *Proc. EMNLP, 2018*. 182–192. DOI: http://dx.doi.org/10.18653/v1/d18-1017

[341] Jing Li, Deheng Ye, and Shuo Shang. 2019. Adversarial transfer for named entity boundary detection with pointer networks. In *Proc. IJCAI, 2019*. 5053–5059. DOI: http://dx.doi.org/10.24963/ijcai.2019/702

[342] Yanyao Shen, Hyokun Yun, Zachary C. Lipton, Yakov Kronrod, and Animashree Anandkumar. 2018. Deep active learning for named entity recognition. In *ICLR, 2018, Conference Track Proceedings*. Retrieved from https://openreview.net/forum?id=ry018WZAZ

[343] Yukun Ma, Erik Cambria, and Sa Gao. 2016. Label embedding for zero-shot fine-grained named entity typing. In *COLING, 2016, Proceedings of the Conference: Technical Papers*. 171–180. Retrieved from https://aclanthology.org/C16-1017/

[344] Tao Zhang, Congying Xia, Chun-Ta Lu, and Philip S. Yu. 2020. MZET: Memory augmented zero-shot fine-grained named entity typing. In *Proc. COLING, 2020*. 77–87. DOI: http://dx.doi.org/10.18653/v1/2020.coling-main.7

[345] Yuhang Guo, Wanxiang Che, Ting Liu, and Sheng Li. 2011. A graph-based method for entity linking. In *IJCNLP, 2011*. 1010–1018. Retrieved from https://aclanthology.org/I11-1113/

[346] Xianpei Han, Le Sun, and Jun Zhao. 2011. Collective entity linking in web text: A graph-based method. In *SIGIR*. 765–774. DOI: http://dx.doi.org/10.1145/2009916.2010019

[347] Avirup Sil and Alexander Yates. 2013. Re-ranking for joint named-entity recognition and linking. In *CIKM, 2013*. 2369–2374. DOI: http://dx.doi.org/10.1145/2505515.2505601

[348] Xiaohua Liu, Ming Zhou, Xiangyang Zhou, Zhongyang Fu, and Furu Wei. 2012. Joint inference of named entity recognition and normalization for tweets. In *Proc. ACL,2012, Volume 1: Long Papers*. 526–535. Retrieved from https://aclanthology.org/P12-1055/

[349] Minh C. Phan, Aixin Sun, Yi Tay, Jialong Han, and Chenliang Li. 2019. Pair-linking for collective entity disambiguation: Two could be better than all. *IEEE Trans. Knowl. Data Eng.* 31, 7 (2019), 1383–1396. DOI: http://dx.doi.org/10.1109/TKDE.2018.2857493

[350] Wee Meng Soon, Hwee Tou Ng, and Chung Yong Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Comput. Linguistics* 27, 4 (2001), 521–544. DOI: http://dx.doi.org/10.1162/089120101753342653

[351] Marta Recasens, Marie-Catherine de Marneffe, and Christopher Potts. 2013. The life and death of discourse entities: Identifying singleton mentions. In *NAACL-HLT, Proceedings, 2013*. 627–633. Retrieved from https://aclanthology.org/N13-1071/

[352] Md. Altaf ur Rahman and Vincent Ng. 2009. Supervised models for coreference resolution. In *Proc. EMNLP, 2009, A meeting of SIGDAT, a Special Interest Group of the ACL*. 968–977. Retrieved from https://aclanthology.org/D09-1101/

[353] Veselin Stoyanov and Jason Eisner. 2012. Easy-first coreference resolution. In *Proc. COLING, 2012*. 2519–2534. Retrieved from https://aclanthology.org/C12-1154/

[354] Dan Roth and Wen-tau Yih. 2004. A linear programming formulation for global inference in natural language tasks. In *HLT-NAACL, 2004*. 1–8. Retrieved from https://aclanthology.org/W04-2401/

[355] Qi Li and Heng Ji. 2014. Incremental joint extraction of entity mentions and relations. In *ACL, 2014, Volume 1: Long Papers*. 402–412. DOI: http://dx.doi.org/10.3115/v1/p14-1038

[356] Sachin Pawar, Pushpak Bhattacharyya, and Girish Keshav Palshikar. 2017. End-to-end relation extraction using neural networks and markov logic networks. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017*. Mirella Lapata, Phil Blunsom, and Alexander Koller (Eds.), Association for Computational Linguistics, 818–827. DOI: http://dx.doi.org/10.18653/v1/e17-1077

[357] Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. 2017. Zero-shot relation extraction via reading comprehension. In *Proc. CoNLL, ACL, 2017*. 333–342. DOI: http://dx.doi.org/10.18653/v1/K17-1034

[358] Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the blanks: Distributional similarity for relation learning. In *Proc. ACL, 2019, Volume 1: Long Papers*. 2895–2905. DOI: http://dx.doi.org/10.18653/v1/p19-1279

[359] Victor Garcia Satorras and Joan Bruna Estrach. 2018. Few-shot learning with graph neural networks. In *ICLR, 2018, Conference Track Proceedings*. Retrieved from https://openreview.net/forum?id=BJj6qGbRW

[360] Meng Qu, Tianyu Gao, Louis-Pascal A. C. Xhonneux, and Jian Tang. 2020. Few-shot relation extraction via bayesian meta-learning on relation graphs. In *Proc. ICML, 2020*. 7867–7876. Retrieved from http://proceedings.mlr.press/v119/qu20a.html

[361] Ni Lao and William W. Cohen. 2010. Relational retrieval using a combination of path-constrained random walks. *Mach. Learn.* 81, 1 (2010), 53–67. DOI: http://dx.doi.org/10.1007/s10994-010-5205-8

[362] Matt Gardner, Partha Pratim Talukdar, Jayant Krishnamurthy, and Tom M. Mitchell. 2014. Incorporating vector space similarity in random walk inference over knowledge bases. In *Proc. EMNLP, 2014, A meeting of SIGDAT, a Special Interest Group of the ACL*. 397–406. DOI: http://dx.doi.org/10.3115/v1/d14-1044

[363] Quan Wang, Jing Liu, Yuanfei Luo, Bin Wang, and Chin-Yew Lin. 2016. Knowledge base completion via coupled path ranking. In *Proc. ACL, 2016, Volume 1: Long Papers*. DOI: http://dx.doi.org/10.18653/v1/p16-1124

[364] Bridget T. McInnes. 2016. VCU at semeval-2016 task 14: Evaluating definitional-based similarity measure for semantic taxonomy enrichment. In *Proc. SemEval@NAACL-HLT, 2016*. 1351–1355. DOI: http://dx.doi.org/10.18653/v1/s16-1212

[365] Luis Espinosa Anke, Francesco Ronzano, and Horacio Saggion. 2016. TALN at SemEval-2016 task 14: Semantic taxonomy enrichment via sense-based embeddings. In *Proc. SemEval@NAACL-HLT, 2016*. 1332–1336. DOI: http://dx.doi.org/10.18653/v1/s16-1208

[366] Karen Spärck Jones. 2004. A statistical interpretation of term specificity and its application in retrieval. *J. Documentation* 60, 5 (2004), 493–502. DOI: http://dx.doi.org/10.1108/00220410410560573

[367] Giorgos Stoilos, Giorgos B. Stamou, and Stefanos D. Kollias. 2005. A string metric for ontology alignment. In *ISWC , 2005, Proceedings*. 624–637. DOI: http://dx.doi.org/10.1007/11574620_45

[368] Nikhita Vedula, Patrick K. Nicholson, Deepak Ajwani, Sourav Dutta, Alessandra Sala, and Srinivasan Parthasarathy. 2018. Enriching taxonomies with functional domain knowledge. In *SIGIR, 2018*. 745–754. DOI: http://dx.doi.org/10.1145/3209978.3210000

[369] Dmitry Frolov, Susana Nascimento, Trevor I. Fenner, and Boris G. Mirkin. 2019. Using taxonomy tree to generalize a fuzzy thematic cluster. In *FUZZ-IEEE, 2019*. 1–6. DOI: http://dx.doi.org/10.1109/FUZZ-IEEE.2019.8859015

[370] François Scharffe, Yanbin Liu, and Chuguang Zhou. 2009. Rdf-ai: An architecture for rdf datasets matching, fusion and interlink. In *Proc. IJCAI workshop on IR-KR, 2019*. 23.

[371] Axel-Cyrille Ngonga Ngomo and Sören Auer. 2011. LIMES - a time-efficient approach for large-scale link discovery on the web of data. In *Proc. IJCAI, 2011*. 2312–2317. DOI: http://dx.doi.org/10.5591/978-1-57735-516-8/IJCAI11-385

[372] Maria Pershina, Mohamed Yakout, and Kaushik Chakrabarti. 2015. Holistic entity matching across knowledge graphs. In *IEEE BigData, 2015*. 1585–1590. DOI: http://dx.doi.org/10.1109/BigData.2015.7363924

[373] Chia-Hui Chang, Chun-Nan Hsu, and Shao-Chen Lui. 2003. Automatic information extraction from semi-structured web pages by pattern discovery. *Decis. Support Syst.* 35, 1 (2003), 129–147. DOI: http://dx.doi.org/10.1016/S0167-9236(02)00100-8

[374] Chun-Nan Hsu and Ming-Tzung Dung. 1998. Generating finite-state transducers for semi-structured data extraction from the web. *Inf. Syst.* 23, 8 (1998), 521–538. DOI: http://dx.doi.org/10.1016/S0306-4379(98)00027-1

[375] Paulo Braz Golgher, Altigran Soares da Silva, Alberto H. F. Laender, and Berthier A. Ribeiro-Neto. 2001. Bootstrapping for example-based data extraction. In *Proc. CIKM, 2001*. 371–378. DOI: http://dx.doi.org/10.1145/502585.502648

[376] Andrew Carlson and Charles Schafer. 2008. Bootstrapping information extraction from semi-structured web pages. In *ECML/PKDD, 2008, Proceedings, Part I*. 195–210. DOI: http://dx.doi.org/10.1007/978-3-540-87479-9_31

[377] Brad Adelberg. 1998. NoDoSE - A tool for semi-automatically extracting semi-structured data from text documents. In *SIGMOD, 1998*. Laura M. Haas and Ashutosh Tiwary (Eds.), 283–294. DOI: http://dx.doi.org/10.1145/276304.276330

[378] Alberto H. F. Laender, Berthier A. Ribeiro-Neto, and Altigran Soares da Silva. 2002. DEByE - data extraction by example. *Data Knowl. Eng.* 40, 2 (2002), 121–154. DOI: http://dx.doi.org/10.1016/S0169-023X(01)00047-7

[379] Aidan Finn, Nicholas Kushmerick, and Barry Smyth. 2001. Fact or fiction: Content classification for digital libraries. In *DELOS, 2001 (ERCIM Workshop Proceedings)*. Retrieved from http://www.ercim.org/publication/ws-proceedings/DelNoe02/AidanFinn.pdf

[380] Tim Weninger, William H. Hsu, and Jiawei Han. 2010. CETR: Content extraction via tag ratios. In *Proc. WWW, 2010*. 971–980. DOI: http://dx.doi.org/10.1145/1772690.1772789

[381] Fei Sun, Dandan Song, and Lejian Liao. 2011. DOM based content extraction via text density. In *SIGIR, 2011*. 245–254. DOI: http://dx.doi.org/10.1145/2009916.2009952

[382] Gong-Qing Wu, Li Li, Xuegang Hu, and Xindong Wu. 2013. Web news extraction via path ratios. In *CIKM, 2013*. 2059–2068. DOI: http://dx.doi.org/10.1145/2505515.2505558

[383] Deng Cai, Shipeng Yu, Ji-Rong Wen, and Wei-Ying Ma. 2003. Extracting content structure for web pages based on visual representation. In *APWeb, 2003, Proceedings*. 406–417. DOI: http://dx.doi.org/10.1007/3-540-36901-5_42

[384] Yalin Wang and Jianying Hu. 2002. A machine learning based approach for table detection on the web. In *WWW, 2002*. 242–250. DOI: http://dx.doi.org/10.1145/511446.511478

[385] Michael J. Cafarella, Alon Y. Halevy, Yang Zhang, Daisy Zhe Wang, and Eugene Wu. 2008. Uncovering the relational web. In *11th International Workshop on the Web and Databases, WebDB*. Retrieved from http://webdb2008.como.polimi.it/images/stories/WebDB2008/paper30.pdf

[386] Julian Eberius, Katrin Braunschweig, Markus Hentsch, Maik Thiele, Ahmad Ahmadov, and Wolfgang Lehner. 2015. Building the dresden web table corpus: A classification approach. In *BDC, 2015*. 41–50. DOI: http://dx.doi.org/10.1109/BDC.2015.30
[387] Michael J. Cafarella, Alon Y. Halevy, and Nodira Khoussainova. 2009. Data integration for the relational web. *Proc. VLDB Endow.* 2, 1 (2009), 1090–1101. DOI: http://dx.doi.org/10.14778/1687627.1687750
[388] Shuo Zhang and Krisztian Balog. 2020. Web table extraction, retrieval, and augmentation: A survey. *ACM Trans. Intell. Syst. Technol.* 11, 2 (2020), 13:1–13:35. DOI: http://dx.doi.org/10.1145/3372117
[389] Bruce Momjian. 2001. *PostgreSQL: Introduction and Concepts*. Vol. 192. Addison-Wesley New York.
[390] Bin Shao, Haixun Wang, and Yatao Li. 2013. Trinity: A distributed graph engine on a memory cloud. In *SIGMOD, 2013*. 505–516. DOI: http://dx.doi.org/10.1145/2463676.2467799
[391] J. Chris Anderson, Jan Lehnardt, and Noah Slater. 2010. CouchDB - The Definitive Guide: Time to Relax. O'Reilly.
[392] Lei Zou, Jinghui Mo, Lei Chen, M. Tamer Özsu, and Dongyan Zhao. 2011. gStore: Answering SPARQL queries via subgraph matching. *Proc. VLDB Endow.* 4, 8 (2011), 482–493. DOI: http://dx.doi.org/10.14778/2002974.2002976
[393] Jim Webber. 2012. A programmatic introduction to Neo4j. In *SPLASH, Proceedings, 2012*. 217–218. DOI: http://dx.doi.org/10.1145/2384716.2384777

Received 30 June 2022; revised 29 May 2023; accepted 21 August 2023