

# DeepKE: A Deep Learning Based Knowledge Extraction Toolkit for Knowledge Base Population

Ningyu Zhang<sup>1</sup>, Xin Xu<sup>1</sup>, Liankuan Tao<sup>1</sup>, Haiyang Yu<sup>1</sup>, Hongbin Ye<sup>1</sup>, Xin Xie<sup>1</sup>, Xiang Chen<sup>1</sup>,  
Zhoubo Li<sup>1</sup>, Lei Li<sup>1</sup>, Xiaozhuan Liang<sup>1</sup>, Yunzhi Yao<sup>1</sup>, Shumin Deng<sup>1</sup>, Wen Zhang<sup>1</sup>,  
Zhenru Zhang<sup>2</sup>, Chuanqi Tan<sup>2</sup>, Fei Huang<sup>2</sup>, Guozhou Zheng<sup>1</sup>, Huajun Chen<sup>1\*</sup>

<sup>1</sup> Zhejiang University & AZFT Joint Lab for Knowledge Engine

<sup>2</sup> Alibaba Group

[www.zjukg.org](http://www.zjukg.org)

## Abstract

We present the first open-source and extensible knowledge extraction toolkit DeepKE, supporting low-resource few-shot and document-level scenarios in knowledge base population. DeepKE implements various information extraction tasks, including named entity recognition, relation extraction and attribute extraction. With a unified framework, DeepKE allows developers and researchers to customize datasets and models to extract information from unstructured texts according to their requirements. Specifically, DeepKE not only provides various functional modules and model implementation for different tasks and scenarios but also organizes all components by consistent frameworks to maintain sufficient modularity and extensibility. Besides, we present an online platform<sup>1</sup> for real-time extraction of various tasks. DeepKE has been equipped with Google Colab tutorials and comprehensive documents<sup>2</sup> for beginners. We release the source code at GitHub<sup>3</sup>, with a demo video<sup>4</sup>.

## 1 Introduction

With the fast development of information extraction techniques, many large-scale knowledge bases (KBs) have been constructed. Those KBs can provide back-end support for knowledge-intensive tasks in real-world applications, such as language understanding (Zhang et al., 2020b; Luo et al., 2021; Zhang et al., 2021b,f), recommendation systems (Wang et al., 2018; Jia et al., 2019), time-series prediction (Deng et al., 2019), and information retrieval (Jin et al., 2020; Zhang et al., 2021c). However, most KBs are far from complete due to

the emerging entities and their relations in real-world applications. Therefore, knowledge base population (KBP), which aims to extract knowledge from the text corpus to complete the missing elements in the KB, has been proposed. For this target, information extraction (IE) is an effective method, which can extract entities, relations from raw texts and link them to KBs (Zheng et al., 2021; Lou et al., 2021; Ye et al., 2021b; Zhang et al., 2021e; Ye et al., 2021a; Zhang et al., 2022).

To date, a few remarkable open-source and long-term maintained information extraction toolkits have been developed, such as Spacy (Vasilev, 2020) for named entity recognition (NER), TagMe (Ferragina and Scaiella, 2010) for entity linking (EL), OpenNRE (Han et al., 2019) for relation extraction, Stanford OpenIE (Martínez-Rodríguez et al., 2018) for open information extraction, RESIN for event extraction (Wen et al., 2021) and so on (Zhang et al., 2020a; Jin et al., 2021). However, there are still several non-trivial issues that hinder the applicability for real-world applications.

Firstly, although information extraction models trained with those tools can achieve promising results, their performance will degrade dramatically when there are only a few training instances (Deng et al., 2020b,a; Zhang et al., 2020c; Deng et al., 2021, 2022). Moreover, it is also non-trivial for these toolkits to be able to extract relations over multiple sentences since many relations are expressed in documents in real-world applications. Therefore, it is necessary to build a knowledge extraction toolkit facilitating knowledge base population that supports **low-resource** and **document-level** scenarios.

In this paper, we share with the community a new open-source knowledge extraction toolkit called **DeepKE**. This toolkit supports knowledge extraction tasks (e.g., named entity recognition, relation extraction, attribute extraction) in the standard supervised setting and two complicated scenarios:

\* Corresponding author: C.Hua (huajunsir@zju.edu.cn)

<sup>1</sup>Project website: <http://deepke.zjukg.cn/>

<sup>2</sup>Docs: <https://zjunlp.github.io/DeepKE/>

<sup>3</sup>GitHub: <https://github.com/zjunlp/DeepKE>

<sup>4</sup>Video: <http://deepke.openkg.cn/demo.mp4>



Figure 1: The examples of tasks with different scenarios in DeepKE.

low-resource and document-level settings. To facilitate flexible usage, we design a unified framework for data processing, data loading, model training and validation. Developers and researchers can quickly customize their datasets and models for various tasks without knowing too many technical details, writing tedious glue code, and conducting hyper-parameter tuning. Moreover, we provide detailed documents and Google Colab tutorials for beginners. An online system is also released to extract structured facts from plain texts with friendly interactive interfaces and fast reaction speed. We will provide maintenance to meet new requests, add new tasks, and fix bugs in the future. In summary, we highlight our contributions as follows:

- We develop and release the first knowledge base population toolkit that supports low-resource and document-level knowledge extraction.
- We provide flexible usage of the toolkit with sufficient modularity as well as automatic hyper-parameter tuning; thus, developers and researchers can implement customized models to extract information from unstructured texts.
- We provide detailed documentation, Google Colab tutorials, an online real-time extraction system and long-term technical support.

## 2 Core Functions

DeepKE is designed for different knowledge extraction tasks, including named entity recognition (NER), relation extraction (RE) and attribute extraction (AE) and so on. As shown in Figure 1, DeepKE supports diverse information extraction tasks in standard supervised, low-resource few-shot, and document-level settings, which makes it flexible to adapt to practical application scenarios.

### 2.1 Named Entity Recognition

As an essential task of information extraction, named entity recognition (NER) picks out the entity mentions and classifies them into predefined semantic categories given plain texts (Liu et al., 2021). For instance, given a sentence “It was one o'clock when we left Lauriston Gardens, and Sherlock Holmes led me meet Gregson from Scotland Yard.”, NER models will predict that “Lauriston Gardens” is a location, “Sherlock Holmes” and “Gregson” are persons and “Scotland Yard” is an organization. To achieve supervised entity recognition, DeepKE simply adopts pre-trained language models (Devlin et al., 2019) to encode sentences and make predictions. DeepKE also implements NER models in the few-shot setting (including in-domain and cross-domain) with *LightNER* (Chen et al., 2021a).

## 2.2 Relation Extraction

Relation extraction, a common task in information extraction for knowledge base population, predicts semantic relations between pairs of entities from unstructured texts (Wang et al., 2020b; Wu et al., 2021; Li et al., 2021c,b). To allow users to customize their models, we adopt various models to accomplish standard supervised RE, including CNN (Zeng et al., 2015), RNN (Zhou et al., 2016), Capsule (Zhang et al., 2018a), GCN (Zhang et al., 2018b, 2019), Transformer and BERT (Devlin et al., 2019). Meanwhile, DeepKE provides few-shot and document-level support for RE. For low-resource RE, DeepKE implements *Know-Prompt* (Chen et al., 2021b) which is a recent well-performed few-shot RE method based on prompt-tuning (Zhang et al., 2021d; Li et al., 2021a). Note that few-shot RE is important for real-world applications, which enable users to extract relations with only a few labeled instances (Yu et al., 2020). For document-level RE, DeepKE extracts inter-sentence relational triples within one document via *DocuNet* (Zhang et al., 2021a). Document-level RE is a challenging task that requires integrating information within and across multiple sentences of a document (Nan et al., 2020). To the best of our knowledge, DeepKE is the first toolkit that can handle low-resource and document-level RE.

## 2.3 Attribute Extraction

Attribute extraction plays an important role in the knowledge base population. Given a sentence, entities and queried attribute mentions, AE will infer the corresponding attribute type. For example, given a sentence “诸葛亮，字孔明，三国时期杰出的军事家、文学家、发明家。” (Liang Zhuge, whose courtesy name was Kongming, was an extraordinary strategist, strategist and inventor in the Three Kingdoms period.), an entity “诸葛亮” (Liang Zhuge), and an attribute mention “三国时期” (Three Kingdoms period), DeepKE can predict the corresponding attribute type “朝代” (Dynasty). DeepKE adopts pre-trained language models (Devlin et al., 2019) for AE.

## 3 Toolkit Design and Implementation

The design principle of DeepKE is to achieve named entity recognition, relation extraction and attribute extraction in different scenarios including standard, low-resource and document-level settings. As shown in Figure 2, we build a unified framework

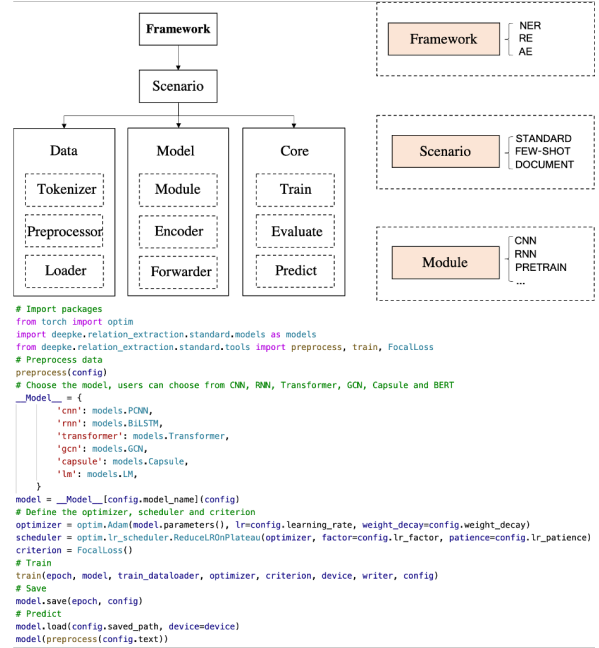


Figure 2: The architecture and example code of DeepKE.

for DeepKE based on PyTorch enabling developers to train models on GPUs for operational efficiency. For every task in each scenario, DeepKE is customized for the specific task objective with respect to *Data*, *Model* and *Core* components. More specifically, we will introduce the detailed design of these three components and the unified framework in the following sections.

### 3.1 Data Module

The data module is designed for preprocessing and loading input data. The tokenizer is able to tokenize input texts into several input tokens. In DeepKE, we implement both word-level tokenization and subword-level tokenization for both English and Chinese. Developers can feed their own datasets into the tokenizer and preprocessor to obtain clean input tokens. Meanwhile, the existing processed input corpus can also be leveraged by models through data loaders.

### 3.2 Model Module

The model module contains main neural networks leveraged to achieve three core tasks. Various neural networks including CNN, RNN, Transformers can be utilized for model implementation which encode texts into specific embedding for corresponding tasks. To adapt to different scenarios, DeepKE utilizes diverse architectures in distinct settings, such as BERT in the standard setting, *LightNER*

and *KnowPrompt* in the low-source setting and *DocuNet* in the document-level setting. We implement the `BasicModel` class with a unified model loader and saver to integrate multifarious neural models.

### 3.3 Core Module

In the core code of DeepKE, `train`, `validate`, and `predict` methods are pivotal components. As for the `train` method, users can simply feed the expected parameters (e.g., the model, data, epoch, optimizer, loss function, .etc.) into it, without writing tedious glue code. The `validate` method is for evaluation. Users can modify the sentences in the configuration for prediction and then utilize the `predict` method to obtain the result.

### 3.4 Framework Module

The framework module integrates three previously mentioned components and different scenarios. It supports various functions including data processing, model construction and model implementation. Meanwhile, developers and researchers can customize all hyper-parameters by modifying configuration files formatted as “\*.yaml”. We apply *Hydra*<sup>5</sup> to obtain users’ configuration from files. We also offer an off-the-shelf automatic hyperparameter tuning component. In DeepKE, we have implemented frameworks for all application functions mentioned in Section 2. For other future potential application functions, we have reserved interfaces for their implementation.

## 4 Toolkit Usage

In order to adapt to different practical scenarios, DeepKE supports three different settings, including the fully supervised standard setting, low-resource few-shot setting and multi-sentence document-level setting.

### 4.1 Standard Setting

All tasks including NER, RE and AE can be implemented in the fully supervised standard setting using DeepKE. The datasets of these tasks are all annotated with specific information, such as entity mentions, entity categories, entity offsets, relation types and the attributes. DeepKE employs BERT for standard NER and utilizes CNN, RNN, Capsule, GCN, Transformer and BERT for standard RE and AE.

<sup>5</sup><https://hydra.cc/>

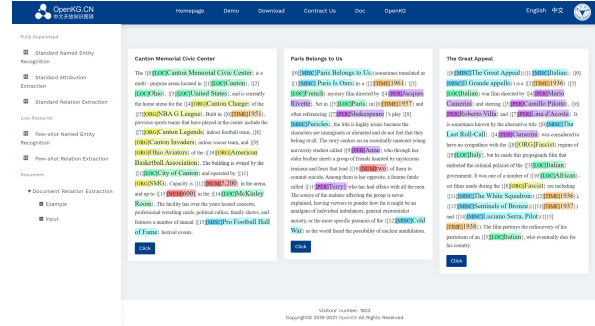


Figure 3: An example of the online system.

### 4.2 Low-resource Setting

In real-world scenarios, labeled data may not be sufficient for deep learning models to make predictions to satisfy users’ specific demands. Therefore, DeepKE provides few-shot support in the low-resource scenario for named entity recognition and relation extraction, which is exceedingly distinctive. We leverage *LightNER*, a generative framework with prompt-guided attention, to achieve both in-domain and cross-domain NER. Meanwhile, we utilize *KnowPrompt* (Chen et al., 2021b) for relation extraction which incorporates knowledge into prompt-tuning with synergistic optimization.

### 4.3 Document-Level Setting

As for relation extraction, relations between two entities not only emerge in one sentence but appear in different sentences within the whole document. Compared to other information extraction toolkits, DeepKE is able to extract relations from documents via *DocuNet* (Zhang et al., 2021a), which predicts an entity-level relation matrix to capture local and global information.

### 4.4 Online System

Besides this toolkit, we release an open-access online system in <http://deepke.zjukg.cn>. As shown in Figure 3, we train our models in different scenarios with multilingual support (English and Chinese) and deploy the model for online access. The online system can be directly applied to recognize named entities, extract relations and classify attributes from plain texts. We also visualize the graph of relational triples extracted, the probabilities of sentence logits and confidence of predicted attribute types for model performance analysis.



## 5 Experiment and Evaluation

### 5.1 Standard Setting

As shown in Table 1, DeepKE achieves named entity recognition, relation extraction and attribute extraction in the standard fully supervised setting.

**Named Entity Recognition** We conduct experiments on two datasets: CoNLL-2003 (Sang and Meulder, 2003) for English and People’s Daily<sup>6</sup> for Chinese. CoNLL-2003 is created for language-independent named entity recognition. Its English part from Reuters Corpus contains 20,864 samples and 35,089 entities including four types of named entities: persons (PER), locations (LOC), organizations (ORG) and miscellaneous (MISC). People’s Daily dataset is a Chinese dataset containing 45,518 entities which is classified into three categories of PER, LOC and ORG. From Table 1, we notice that DeepKE yields comparable performance with various encoders for both English and Chinese.

**Relation Extraction** We conduct RE experiments on the Chinese DuIE dataset<sup>7</sup>. There are 3,000, 1,000 and 1,000 samples for training, validation and test, respectively. Each sample contains one original sentence, the head entity and the tail entity in the sentence, their offsets, and the relation between them. In DuIE, relations are classified into ten categories with their corresponding head and tail entity types. We utilize six different neural networks in DeepKE for evaluation. Users can select models before training by changing only one hyper-parameter. The performance of all different models are shown in Table 1.

**Attribute Extraction** The dataset for attribution extraction is from an online resource<sup>8</sup>. In each sample, one entity is annotated with its attribute type, attribute value and offset. The attributes in the Chinese dataset are classified into six categories. The training set contains 13,815 samples. The validation set contains 3,131 samples and the test set includes 5,921 samples. Like RE, we leverage six neural models to extract attributes from the given sentence to evaluate DeepKE.

<sup>6</sup>[https://github.com/OYE93/Chinese-NLP-Corpus/tree/master/NER/People’s Daily](https://github.com/OYE93/Chinese-NLP-Corpus/tree/master/NER/People's%20Daily)

<sup>7</sup><http://ai.baidu.com/broad/download>

<sup>8</sup>[https://github.com/leefsir/triplet\\_extraction](https://github.com/leefsir/triplet_extraction)

Task	Corpus	Language	Model	F1
NER	CoNLL-2003 People’s Daily	English Chinese	BERT	94.73 95.62
			CNN RNN Capsule GCN Transformer BERT	96.74 94.43 96.23 96.74 96.54 95.79
RE	DuIE	Chinese	CNN RNN Capsule GCN Transformer BERT	94.16 93.06 94.57 94.50 94.15 99.03
			CNN RNN Capsule GCN Transformer BERT	94.16 93.06 94.57 94.50 94.15 99.03

Table 1: F1 scores of various models on named entity recognition (NER), relation extraction (RE), and attribute extraction (AE) tasks in the standard fully supervised setting.

Model	Entity Category				
	PER	ORG	LOC*	MISC*	Overall
LC-BERT	76.25	75.32	61.55	59.35	68.12
LC-BART	75.70	73.59	58.70	57.30	66.82
Template.	84.49	72.61	71.98	73.37	75.59
<b>DeepKE (LightNER)</b>	<b>90.96</b>	<b>76.88</b>	<b>81.57</b>	<b>82.08</b>	<b>78.97</b>

Table 2: F1 scores of in-domain low-resource NER on CoNLL-2003. \* indicates low-resource entity types (100-shot).

### 5.2 Low-resource Setting

In the low-resource setting, we conduct experiments with few-shot named entity recognition and relation extraction.

**Named Entity Recognition** We conduct experiments in both in-domain and cross-domain few-shot settings. According to (Cui et al., 2021), to construct the in-domain few-shot scenario, we reduce the number of training samples for certain entity categories by downsampling one dataset. Specifically, from CoNLL-2003, we choose **100** “LOC” and **100** “MISC” as the low-resource entities and 2,496 “PER” and 3,763 “ORG” as the rich-resource entities. We leverage *LightNER* in

Model	Dataset		
	MIT Movie	MIT Restaurant	ATIS
Neigh.Tag.	1.4	3.6	3.4
Example.	29.6	26.1	16.5
MP-NSP	36.8	48.2	74.8
LC-BERT	45.2	40.9	78.5
LC-BART	30.4	11.1	74.4
Template.	54.2	60.3	88.9
<b>DeepKE (LightNER)</b>	<b>75.6</b>	<b>67.4</b>	<b>89.4</b>

Table 3: F1 scores of cross-domain few-shot NER (20-shot).

Method	Split		
	K=8	K=16	K=32
Fine-Tuning	41.3	65.2	80.1
GDPNet	42.0	67.5	81.2
PTR	70.5	81.3	84.2
<b>DeepKE (KnowPrompt)</b>	<b>74.3</b>	<b>82.9</b>	<b>84.8</b>

Table 4: F1 scores of few-shot relation extraction

DeepKE to carry out the few-shot experiments and adopt BERT and BART (Lewis et al., 2020) with label-specific classifier layers as baselines denoted as *LC-BERT* and *LC-BART*. We also use template-based BART (*Template.*) (Cui et al., 2021) as competitive baseline. From Table 2, *LightNER* outperforms other methods for both rich- and low-resource entity types, which illustrates that DeepKE has an outstanding performance in in-domain few-shot NER. In the cross-domain setting where the target entity categories and textual style are different from the source domain with restricted labeled data available for training, we adopt the CoNLL-2003 dataset as an ordinary domain and MIT Movie Review (Liu et al., 2013), MIT Restaurant Review (Liu et al., 2013) and Airline Travel Information Systems (ATIS) (Hakkani-Tür et al., 2016) datasets as target domains. We first train *LightNER* on CoNLL-2003 and fine-tune it on 20-shot target domain datasets (randomly sampled per entity category). We employ prototype-based *Neigh.Tag.* (Wiseman and Stratos, 2019), *Example.* (example-based NER) (Ziyadi et al., 2020), *MP-NSP* (Multi-prototype+NSP) (Huang et al., 2020), *LC-BERT*, *LC-BART* and *Template.* as competitive baselines. From Table 3, we notice that our model achieve the most excellent few-shot performance.

**Relation Extraction** For few-shot relation extraction, we use SemEval 2010 Task-8 (Hendrickx et al., 2010), which is a conventional dataset of relation classification with nine bidirectional relations and one relation OTHER. The dataset is split into 6,507 instances for training, 1,493 instances for validation, and 2,717 for test. We utilize *KnowPrompt* to conduct 8-, 16-, and 32-shot experiments. From Table 4, we observe that DeepKE outperforms the baseline methods *Fine-tuning* (fine-tuning vanilla PLMs), *GDPNet* (Xue et al., 2021) and *PTR* (Han et al., 2021).

### 5.3 Document-level Setting

DeepKE can extract intra- and inter- sentence relations among multiple entities within one document.

Model	F1
GEDA-BERT <sub>base</sub>	55.74
LSR-BERT <sub>base</sub>	59.05
GLRE-BERT <sub>base</sub>	57.40
GAIN-BERT <sub>base</sub>	61.24
HeterGSAN-BERT <sub>base</sub>	59.45
BERT <sub>base</sub>	53.20
BERT-TS <sub>base</sub>	53.92
HIN-BERT <sub>base</sub>	55.60
CorefBERT <sub>base</sub>	56.96
ATLOP-BERT <sub>base</sub>	61.30
<b>DeepKE (DocuNet-BERT<sub>base</sub>)</b>	<b>61.86</b>
BERT <sub>large</sub>	58.69
CorefBERT <sub>large</sub>	58.83
RoBERTa <sub>large</sub>	59.62
CorefRoBERTa <sub>large</sub>	60.25
ATLOP-RoBERTa <sub>large</sub>	63.40
<b>DeepKE (DocuNet-RoBERTa<sub>large</sub>)</b>	<b>64.55</b>

Table 5: Results of DeepKE on document-level RE.

**Relation Extraction** We leverage a large-scale document-level RE dataset, DocRED (Ye et al., 2020), containing 3,053/1,000/1,000 instances for training, validation and testing, respectively. DeepKE employs *DocuNet* (Zhang et al., 2021a) to extract relations of multiple entity pairs in a long chapter at once. We use cased BERT-base and RoBERTa-large as the encoder. Compared with graph-based models, including *GEDA* (Li et al., 2020), *LSR* (Nan et al., 2020), *GLRE* (Wang et al., 2020a) and *GAIN* (Zeng et al., 2020), *HeterGSAN* (Xu et al., 2021); and transformer-based models, including *BERT<sub>base</sub>* (Wang et al., 2019), *BERT-TS<sub>base</sub>* (Wang et al., 2019), *HIN-BERT<sub>base</sub>* (Tang et al., 2020), *CorefBERT<sub>base</sub>* (Ye et al., 2020), and *ATLOP<sub>base</sub>* (Zhou et al., 2021), our model appears the best performance as shown in Table 5.

## 6 Conclusion

In practical application, the knowledge base population struggles for low-resource and document-level scenarios. To this end, we propose DeepKE, an open-source and extensible knowledge extraction toolkit, with the functions of named entity recognition, relation extraction and attribute extraction. We conduct extensive experiments that demonstrate the models implemented by DeepKE are effective, efficient and can achieve comparable performance compared to some state-of-the-art methods. Besides, an online system supporting real-time extraction without training and deploying is provided. We will offer long-term maintenance to fix bugs, solve issues and meet new requests.

## References

- Xiang Chen, Ningyu Zhang, Lei Li, Xin Xie, Shumin Deng, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2021a. [Lightner: A lightweight generative framework with prompt-guided attention for low-resource NER](#). *CoRR*, abs/2109.00720.
- Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2021b. [Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction](#). *CoRR*, abs/2104.07650.
- Leyang Cui, Yu Wu, Jian Liu, Sen Yang, and Yue Zhang. 2021. [Template-based named entity recognition using BART](#). In *Proceedings of ACL/IJCNLP*, volume ACL/IJCNLP 2021 of *Findings of ACL*.
- Shumin Deng, Ningyu Zhang, Hui Chen, Chuanqi Tan, Fei Huang, Changliang Xu, and Huajun Chen. 2022. Low-resource extraction with knowledge-aware pairwise prototype learning. *Knowledge-Based Systems*, 235:107584.
- Shumin Deng, Ningyu Zhang, Jiaojian Kang, Yichi Zhang, Wei Zhang, and Huajun Chen. 2020a. [Meta-learning with dynamic-memory-based prototypical network for few-shot event detection](#). In *Proceedings of WSDM*, pages 151–159. ACM.
- Shumin Deng, Ningyu Zhang, Luoqiu Li, Hui Chen, Huaixiao Tou, Mosha Chen, Fei Huang, and Huajun Chen. 2021. [Ontoed: Low-resource event detection with ontology embedding](#). *CoRR*, abs/2105.10922.
- Shumin Deng, Ningyu Zhang, Zhanlin Sun, Jiaoyan Chen, and Huajun Chen. 2020b. [When low resource NLP meets unsupervised language model: Meta-pretraining then meta-learning for few-shot text classification \(student abstract\)](#). In *Proceedings of AAAI*, pages 13773–13774. AAAI Press.
- Shumin Deng, Ningyu Zhang, Wen Zhang, Jiaoyan Chen, Jeff Z. Pan, and Huajun Chen. 2019. [Knowledge-driven stock trend prediction and explanation via temporal convolutional network](#). In *Proceedings of WWW*, pages 678–685. ACM.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of NAACL-HLT*, pages 4171–4186. Association for Computational Linguistics.
- Paolo Ferragina and Ugo Scaiella. 2010. [TAGME: on-the-fly annotation of short text fragments \(by wikipedia entities\)](#). In *Proceedings of CIKM*, pages 1625–1628. ACM.
- Dilek Hakkani-Tür, Gökhan Tür, Asli Celikyilmaz, Yun-Nung Chen, Jianfeng Gao, Li Deng, and Ye-Yi Wang. 2016. [Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM](#). In *Proceedings of INTERSPEECH*, pages 715–719. ISCA.
- 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*.
- Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. 2021. [PTR: prompt tuning with rules for text classification](#). *CoRR*, abs/2105.11259.
- Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. 2010. [Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals](#). In *Proceedings of SemEval@ACL Workshop*.
- Jiaxin Huang, Chunyuan Li, Krishan Subudhi, Damien Jose, Shobana Balakrishnan, Weizhu Chen, Baolin Peng, Jianfeng Gao, and Jiawei Han. 2020. [Few-shot named entity recognition: A comprehensive study](#).
- Qianghuai Jia, Ningyu Zhang, and Nengwei Hua. 2019. [Context-aware deep model for entity recommendation in search engine at alibaba](#). *CoRR*, abs/1909.04493.
- Qiao Jin, Chuanqi Tan, Mosha Chen, Ming Yan, Songfang Huang, Ningyu Zhang, and Xiaozhong Liu. 2020. [Aliababa DAMO academy at TREC precision medicine 2020: State-of-the-art evidence retriever for precision medicine with expert-in-the-loop active learning](#). In *Proceedings of TREC*, volume 1266 of *NIST Special Publication*. National Institute of Standards and Technology (NIST).
- Zhuoran Jin, Yubo Chen, Dianbo Sui, Chenhao Wang, Zhipeng Xue, and Jun Zhao. 2021. [CogIE: An information extraction toolkit for bridging texts and CogNet](#). In *Proceedings of ACL-IJCNLP*, pages 92–98.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. [BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension](#). In *Proceedings of ACL*, pages 7871–7880. Association for Computational Linguistics.
- Bo Li, Wei Ye, Zhonghao Sheng, Rui Xie, Xiangyu Xi, and Shikun Zhang. 2020. [Graph enhanced dual attention network for document-level relation extraction](#). In *Proceedings of COLING*, pages 1551–1560. International Committee on Computational Linguistics.
- Chengxi Li, Feiyu Gao, Jiajun Bu, Lu Xu, Xiang Chen, Yu Gu, Zirui Shao, Qi Zheng, Ningyu Zhang, Yongpan Wang, and Zhi Yu. 2021a. [Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis](#). *CoRR*, abs/2109.08306.

- Luoqiu Li, Xiang Chen, Hongbin Ye, Zhen Bi, Shumin Deng, Ningyu Zhang, and Huajun Chen. 2021b. On robustness and bias analysis of bert-based relation extraction. In *CKKS*, pages 43–59. Springer.
- Luoqiu Li, Xiang Chen, Ningyu Zhang, Shumin Deng, Xin Xie, Chuanqi Tan, Mosha Chen, Fei Huang, and Huajun Chen. 2021c. [Normal vs. adversarial: Saliency-based analysis of adversarial samples for relation extraction](#). *CoRR*, abs/2104.00312.
- Jingjing Liu, Panupong Pasupat, Scott Cyphers, and James R. Glass. 2013. [Asgard: A portable architecture for multilingual dialogue systems](#). In *Proceedings of ICASSP*, pages 8386–8390. IEEE.
- Kun Liu, Yao Fu, Chuanqi Tan, Mosha Chen, Ningyu Zhang, Songfang Huang, and Sheng Gao. 2021. [Noisy-labeled NER with confidence estimation](#). In *Proceedings of NAACL-HLT*, pages 3437–3445. Association for Computational Linguistics.
- Dongfang Lou, Zhilin Liao, Shumin Deng, Ningyu Zhang, and Huajun Chen. 2021. [Mlbinet: A cross-sentence collective event detection network](#). In *Proceedings of ACL/IJCNLP*, pages 4829–4839. Association for Computational Linguistics.
- Junjie Luo, Xi Chen, Jichao Sun, Yuejia Xiang, Ningyu Zhang, and Xiang Wan. 2021. [Field embedding: A unified grain-based framework for word representation](#). In *In Proceedings of NAACL*.
- José-Lázaro Martínez-Rodríguez, Ivan López-Arévalo, and Ana B. Ríos-Alvarado. 2018. [Openie-based approach for knowledge graph construction from text](#). *Expert Syst. Appl.*, 113:339–355.
- Guoshun Nan, Zhijiang Guo, Ivan Sekulic, and Wei Lu. 2020. [Reasoning with latent structure refinement for document-level relation extraction](#). In *Proceedings of ACL*, pages 1546–1557. Association for Computational Linguistics.
- Erik F. Tjong Kim Sang and Fien De Meulder. 2003. [Introduction to the conll-2003 shared task: Language-independent named entity recognition](#). In *Proceedings of HLT-NAACL*, pages 142–147. ACL.
- Hengzhu Tang, Yanan Cao, Zhenyu Zhang, Jiangxia Cao, Fang Fang, Shi Wang, and Pengfei Yin. 2020. [HIN: hierarchical inference network for document-level relation extraction](#). In *Proceedings of PAKDD*, volume 12084 of *Lecture Notes in Computer Science*, pages 197–209. Springer.
- Yuli Vasiliev. 2020. *Natural Language Processing with Python and SpaCy: A Practical Introduction*. No Starch Press.
- Difeng Wang, Wei Hu, Ermei Cao, and Weijian Sun. 2020a. [Global-to-local neural networks for document-level relation extraction](#). In *Proceedings of EMNLP*, pages 3711–3721. Association for Computational Linguistics.
- Hong Wang, Christfried Focke, Rob Sylvester, Nilesh Mishra, and William Yang Wang. 2019. [Fine-tune bert for docred with two-step process](#). *CoRR*, abs/1909.11898.
- Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. [DKN: deep knowledge-aware network for news recommendation](#). In *Proceedings of WWW*, pages 1835–1844. ACM.
- Zifeng Wang, Rui Wen, Xi Chen, Shao-Lun Huang, Ningyu Zhang, and Yefeng Zheng. 2020b. [Finding influential instances for distantly supervised relation extraction](#). *CoRR*, abs/2009.09841.
- Haoyang Wen, Ying Lin, Tuan Lai, Xiaoman Pan, Sha Li, Xudong Lin, Ben Zhou, Manling Li, Haoyu Wang, Hongming Zhang, Xiaodong Yu, Alexander Dong, Zhenhailong Wang, Yi Fung, Piyush Mishra, Qing Lyu, Dídac Surís, Brian Chen, Susan Windisch Brown, Martha Palmer, Chris Callison-Burch, Carl Vondrick, Jiawei Han, Dan Roth, Shih-Fu Chang, and Heng Ji. 2021. [RESIN: A dockerized schema-guided cross-document cross-lingual cross-media information extraction and event tracking system](#). In *In Proceedings of NAACL-HLT*.
- Sam Wiseman and Karl Stratos. 2019. [Label-agnostic sequence labeling by copying nearest neighbors](#). In *Proceedings of ACL*, pages 5363–5369. Association for Computational Linguistics.
- 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 AAAI*.
- Wang Xu, Kehai Chen, and Tiejun Zhao. 2021. [Document-level relation extraction with reconstruction](#). In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 14167–14175. AAAI Press.
- Fuzhao Xue, Aixin Sun, Hao Zhang, and Eng Siong Chng. 2021. [Gdpnet: Refining latent multi-view graph for relation extraction](#). In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 14194–14202. AAAI Press.
- Deming Ye, Yankai Lin, Jiaju Du, Zhenghao Liu, Peng Li, Maosong Sun, and Zhiyuan Liu. 2020. [Coreferential reasoning learning for language representation](#). In *Proceedings of EMNLP*, pages 7170–7186. Association for Computational Linguistics.



- Hongbin Ye, Ningyu Zhang, Zhen Bi, Shumin Deng, Chuanqi Tan, Hui Chen, Fei Huang, and Huajun Chen. 2021a. [Learning to ask for data-efficient event argument extraction](#). *CoRR*, abs/2110.00479.
- Hongbin Ye, Ningyu Zhang, Shumin Deng, Mosha Chen, Chuanqi Tan, Fei Huang, and Huajun Chen. 2021b. [Contrastive triple extraction with generative transformer](#). In *Proceedings of AAAI*, pages 14257–14265. AAAI Press.
- Haiyang Yu, Ningyu Zhang, Shumin Deng, Hongbin Ye, Wei Zhang, and Huajun Chen. 2020. [Bridging text and knowledge with multi-prototype embedding for few-shot relational triple extraction](#). In *Proceedings of COLING*.
- Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. [Distant supervision for relation extraction via piecewise convolutional neural networks](#). In *Proceedings of EMNLP*, pages 1753–1762. The Association for Computational Linguistics.
- Shuang Zeng, Runxin Xu, Baobao Chang, and Lei Li. 2020. [Double graph based reasoning for document-level relation extraction](#). In *Proceedings of EMNLP*, pages 1630–1640.
- Ningyu Zhang, Xiang Chen, Xin Xie, Shumin Deng, Chuanqi Tan, Mosha Chen, Fei Huang, Luo Si, and Huajun Chen. 2021a. [Document-level relation extraction as semantic segmentation](#). In *Proceedings of IJCAI*, pages 3999–4006. ijcai.org.
- Ningyu Zhang, Shumin Deng, Zhen Bi, Haiyang Yu, Jiacheng Yang, Mosha Chen, Fei Huang, Wei Zhang, and Huajun Chen. 2020a. [Openue: An open toolkit of universal extraction from text](#). In *Proceedings of EMNLP*, pages 1–8. Association for Computational Linguistics.
- Ningyu Zhang, Shumin Deng, Xu Cheng, Xi Chen, Yichi Zhang, Wei Zhang, and Huajun Chen. 2021b. [Drop redundant, shrink irrelevant: Selective knowledge injection for language pretraining](#). In *Proceedings of IJCAI*, pages 4007–4014. ijcai.org.
- Ningyu Zhang, Shumin Deng, Juan Li, Xi Chen, Wei Zhang, and Huajun Chen. 2020b. [Summarizing chinese medical answer with graph convolution networks and question-focused dual attention](#). In *Findings of EMNLP*.
- Ningyu Zhang, Shumin Deng, Zhanlin Sun, Jiaoyan Chen, Wei Zhang, and Huajun Chen. 2020c. [Relation adversarial network for low resource knowledge graph completion](#). In *Proceedings of WWW*, pages 1–12. ACM / IW3C2.
- 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 NAACL-HLT*, pages 3016–3025. Association for Computational Linguistics.
- Ningyu Zhang, Shumin Deng, Zhanling Sun, Xi Chen, Wei Zhang, and Huajun Chen. 2018a. [Attention-based capsule network with dynamic routing for relation extraction](#). In *In Proceedings of EMNLP*.
- Ningyu Zhang, Shumin Deng, Hongbin Ye, Wei Zhang, and Huajun Chen. 2022. [Robust triple extraction with cascade bidirectional capsule network](#). *Expert Syst. Appl.*, 187:115806.
- Ningyu Zhang, Qianghuai Jia, Shumin Deng, Xiang Chen, Hongbin Ye, Hui Chen, Huaixiao Tou, Gang Huang, Zhao Wang, Nengwei Hua, and Huajun Chen. 2021c. [Alicg: Fine-grained and evolvable conceptual graph construction for semantic search at alibaba](#). In *Proceedings of KDD*, pages 3895–3905. ACM.
- Ningyu Zhang, Luoqiu Li, Xiang Chen, Shumin Deng, Zhen Bi, Chuanqi Tan, Fei Huang, and Huajun Chen. 2021d. [Differentiable prompt makes pre-trained language models better few-shot learners](#). *CoRR*, abs/2108.13161.
- Ningyu Zhang, Hongbin Ye, Shumin Deng, Chuanqi Tan, Mosha Chen, Songfang Huang, Fei Huang, and Huajun Chen. 2021e. [Contrastive information extraction with generative transformer](#). *IEEE ACM Trans. Audio Speech Lang. Process.*, 29:3077–3088.
- Ningyu Zhang, Hongbin Ye, Jiacheng Yang, Shumin Deng, Chuanqi Tan, Mosha Chen, Songfang Huang, Fei Huang, and Huajun Chen. 2021f. [LOGEN: few-shot logical knowledge-conditioned text generation with self-training](#). *CoRR*, abs/2112.01404.
- Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018b. [Graph convolution over pruned dependency trees improves relation extraction](#). In *Proceedings of EMNLP*, pages 2205–2215. Association for Computational Linguistics.
- Hengyi Zheng, Rui Wen, Xi Chen, Yifan Yang, Yunyan Zhang, Ziheng Zhang, Ningyu Zhang, Bin Qin, Xu Ming, and Yefeng Zheng. 2021. [PRGC: potential relation and global correspondence based joint relational triple extraction](#). In *Proceedings of ACL/IJCNLP*, pages 6225–6235. Association for Computational Linguistics.
- 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 ACL*, pages 207–212, Berlin, Germany. Association for Computational Linguistics.
- Wenxuan Zhou, Kevin Huang, Tengyu Ma, and Jing Huang. 2021. [Document-level relation extraction with adaptive thresholding and localized context pooling](#). In *In proceedings of AAAI*, pages 14612–14620. AAAI Press.
- Morteza Ziyadi, Yuting Sun, Abhishek Goswami, Jade Huang, and Weizhu Chen. 2020. [Example-based named entity recognition](#).

## A Toolkit Usage Details

In this section, we introduce how to use DeepKE exhaustively.

### A.1 Build a Model From Scratch

**Prepare the Runtime Environment** Users can clone the source code from the DeepKE GitHub repository and create a runtime environment. There are two convenient methods to create the environment. Users can choose to either leverage *Anaconda* or run the docker file provided in the repository. Besides, all dependencies can be installed by running `pip install deepke` directly. If developers would like to modify the source code of DeepKE, the following commands should be executed: running `python setup.py install`, modifying code and then running `python setup.py develop`. To train specific information extraction models, corresponding datasets (e.g., default or customized datasets) should be leveraged. All datasets need to be downloaded or uploaded in the folder named *data*.

**Named Entity Recognition** As shown in Table 6, the input data files with BIO tags for standard and few-shot NER contains two columns separated by a single space. Each word has been put on a separate line, and there is an empty line after each sentence. The two columns represent two items: the word and the named entity tag. Before training, all datasets with the formats mentioned above should be fed into NER models through the data loader. Developers can implement training and evaluating by running example code *run.py* directly to obtain a fine-tuned NER language model, which will be used in the prediction period. For inference, users can run *predict.py* with a single sentence input and obtain the output recognized entity mentions and types.

**Relation Extraction** The training input with the CSV format of standard RE is shown in Table 7. There are five components in the format, including a sentence, a relation, the head and tail entity of the relation, the head entity offset and the tail entity offset. For few-shot RE, one input sample, as shown in Figure 4, contains sentence tokens including words and punctuation, the head entity and tail entities with their mention names and position spans, and the relation between them. For example, an input of few-shot relation extraction

Word	Named Entity Tag
U.N.	B-ORG
official	O
Ekeus	B-PER
heads	O
for	O
Baghdad	B-LOC
.	O
Israel	B-LOC
approves	O
Arafat	B-PER
s	O
flight	O
to	O
West	B-LOC
Bank	I-LOC
.	O

Table 6: Examples of the input format for NER.

Sentence	Relation	Head	HO	Tail	TO
When it comes to beautiful sceneries in Hangzhou, West Lake first emerges in mind.	city: located in	West Lake	50	Hangzhou	40
Harry Potter, a wizard, graduated from Hogwarts School of Witchcraft and Wizardry.	school: graduated from	Harry Potter	0	Hogwarts School of Witchcraft and Wiz- ardry	39

Table 7: Examples of the input format for standard RE. HO: Head Offset, TO: Tail Offset.

instance is the format of `{"token": ["the", "dol-phin", "uses", "its", "flukes", "for", "swimming", "and", "its", "flippers", "for", "steering", "."], "h": {"name": "dolphin", "pos": [1, 2]}, "t": {"name": "flukes", "pos": [4, 5]}, "relation": "Component-Whole(e2,e1)"}` (h: head entity, t: tail entity, pos: position). The document-level RE training format is shown in Figure 5. One sample consists of a sample title, sentences separated into words and punctuation in one document, an entity set (including entity mentions, sentence IDs the entities are located in, entity position spans and entity types in the document) and a relation label set (including the head and tail entity IDs, relations and evidence sentence IDs). After training and validation, users can run the `predict` function given an input sentence with head and tail entity to obtain corresponding relations.

**Attribution Extraction** The input CSV files formatted as Table 8 should be given to train the attribution extraction (AE) model. One sample contains six components: a raw sentence, a queried attribute type, an entity and its offset, the entity's

```

Data Format:
{
  'token': [tokens in a sentence],
  "h": {
    "name": mention_name,
    "pos": [postion of mention in a sentence]
  },
  "t": {
    "name": mention_name,
    "pos": [postion of mention in a sentence]
  },
  "relation": relation
}

```

Figure 4: The input format of few-shot RE.

```

Data Format:
{
  'title',
  'sents': [
    [word in sent 0],
    [word in sent 1]
  ]
  'vertexSet': [
    [
      { 'name': mention_name,
        'sent_id': mention in which sentence,
        'pos': postion of mention in a sentence,
        'type': NER_type}
      {author mention}
    ],
    [anthoer entity]
  ]
  'labels': [
    {
      'h': idx of head entity in vertexSet,
      't': idx of tail entity in vertexSet,
      'r': relation,
      'evidence': evidence sentences' id
    }
  ]
}

```

Figure 5: The input format of document-level RE.

corresponding attribute value and the attribute mention offset. After training, users will obtain a fine-tuned AE model, which can be leveraged to infer attributes. Given a sentence with an entity and a candidate attribute mention, the AE model will predict the attribute type with confidence. Note that all operations mentioned above are guided in the example code file *run.py* and *predict.py*.

## A.2 Auto-Hyperparameter Tuning

To achieve automatic hyper-parameters fine-tuning, DeepKE adopts *Weight & Biases*, a machine learning toolkit for developers to reduce label-intensive

Sentence	Attribute	Entity	EO	AV	AVO
1903年，亨利·福特创建福特汽车公司	创始人	福特	9	亨利·福特	6
吴会期，字行可，号子官，明朝工部郎中	朝代	吴会期	0	明朝	12

Table 8: Examples of the input format AE. EO: Entity Offset, AV: Attribute Value, AVO: Attribute Value Offset.

hyper-parameter tuning. With DeepKE, users can visualize results and tune hyper-parameters automatically. Note that all metrics and hyper-parameter configurations can be customized to meet diverse settings for different tasks. For more details of automatic hyper-parameter tuning, please refer to the official document<sup>9</sup>.

## A.3 Notebook Tutorials

We provide Google Colab tutorials and jupyter notebooks in the GitHub repository as an exemplary implementation of every task in different scenarios. These tutorials can be run directly, thus, leading developers and researchers to have a whole picture of DeepKE’s powerful functions.

## B Contributions

**Ningyu Zhang** from Zhejiang University, AZFT Joint Lab for Knowledge Engine, Hangzhou Innovation Center conducted the whole development of DeepKE and wrote the paper.

**Xin Xu** from Zhejiang University, AZFT Joint Lab for Knowledge Engine developed the standard NER and wrote the paper.

**Liankuan Tao, Haiyang Yu** from Zhejiang University, AZFT Joint Lab for Knowledge Engine developed the standard RE and AE, the deepke python package, and documents and provides consistent maintenance.

**Hongbin Ye** from Zhejiang University, AZFT Joint Lab for Knowledge Engine constructed the online demo.

**Xin Xie, Xiang Chen** from Zhejiang University, AZFT Joint Lab for Knowledge Engine developed the few-shot relation extraction model Know-Prompt and the document-level relation extraction model DocuNet.

**Zhoubo Li, Lei Li, Xiaozhuan Liang, Yunzhi Yao, Shumin Deng, Wen Zhang** from Zhejiang University, AZFT Joint Lab for Knowledge Engine developed the Google Colab and proofread the paper.

**Zhenru Zhang, Chuanqi Tan, Fei Huang** from Alibaba Group, proofread the paper, advised the project.

**Guozhou Zheng, Huajun Chen** from Zhejiang University, AZFT Joint Lab for Knowledge Engine advised the project, suggested tasks, and led the research.

<sup>9</sup><https://docs.wandb.ai>