



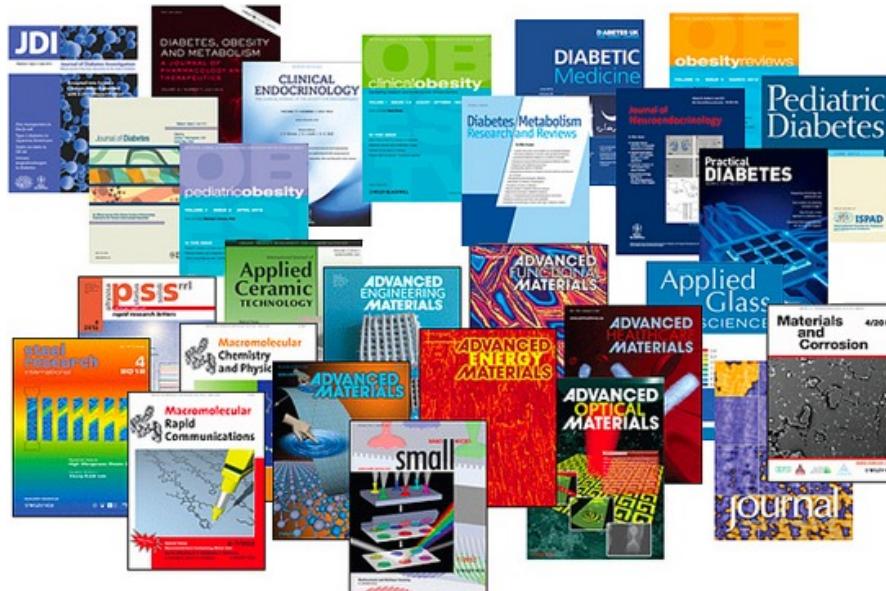
DeepKE: A Deep Learning Based Knowledge Extraction Toolkit for Knowledge Base Population

一个支持低资源、长篇章、多模态的开源知识抽取工具

Ningyu Zhang, Xin Xu, Liankuan Tao, Haiyang Yu, Hongbin Ye, Shuofei Qiao, Xin Xie, Xiang Chen, Zhoubo Li, Lei Li, Xiaozhuan Liang, Yunzhi Yao, Shumin Deng, Peng Wang, Wen Zhang, Zhenru Zhang, Chuanqi Tan, Qiang Chen, Feiyu Xiong, Fei Huang, Guozhou Zheng, Huajun Chen



知识图谱构建



Text2Knowledge

Language Pretrain
BERT、GPT.....

属性：字

张之洞，晚清名臣、清代洋务派代表人物，祖籍直隶南皮，出生于贵州，字孝达。

实体

属性值

isSpouseOf

Jack is married to the microbiologist known as Dr. Germ in the USA.

Entity Pair

- Single Sentence

It was one o'clock when we left **Lauriston Gardens**,
Sherlock Holmes led me meet **Gregson** from **Scotland Yard**.

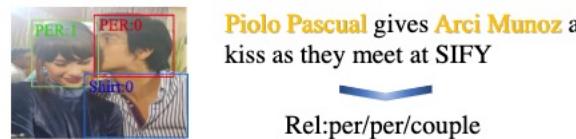
↑
[LOC]
↑
[PER] ↓
[PER] ↓
[ORG]

- Document

... Elias Brown (May 9, 1793 - July 7, 1857) was a **U.S.** Representative from **Maryland**. Born near **Baltimore**, **Maryland**, Brown attended the common schools ... He died near **Baltimore**, **Maryland**, and is interred in a private cemetery near **Eldersburg**, **Maryland** ...

- | | |
|----------------|---|
| Intra-sentence | (Maryland , country, U.S.)
(Baltimore , located in, Maryland)
(Eldersburg , located in, Maryland) |
| Inter-sentence | (Baltimore , country, U.S.)
(Eldersburg , country, U.S.) |

- MultiModal



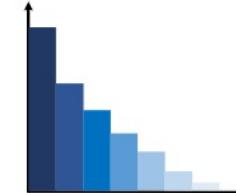
DeepKE

Knowledge Extraction Tool

Diverse Data



Low Resource

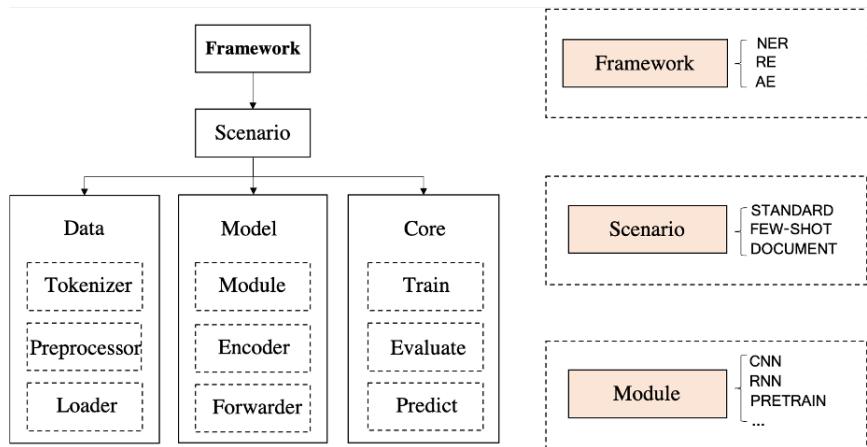


Off-the-shelf Usage Flexible Training

cnSchema
OpenKG 开放的中文知识图谱



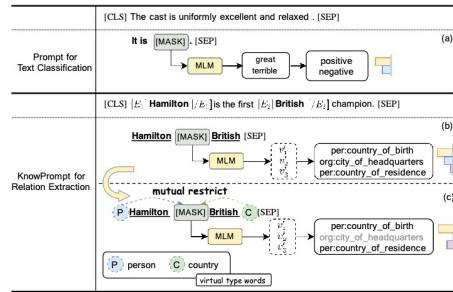
DeepKE: 支持多个经典和前沿模型



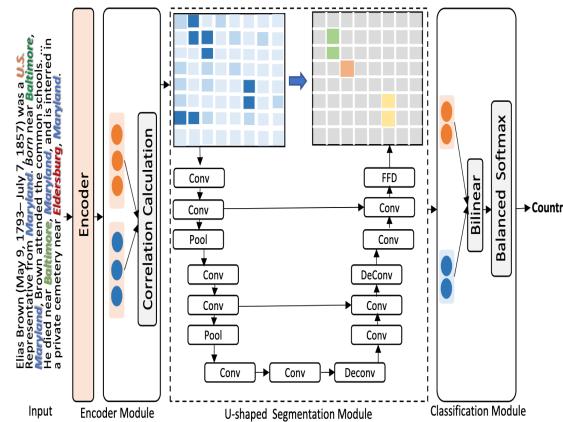
```

# Import packages
from torch import optim
import deepke.relation_extraction.standard.models as models
from deepke.relation_extraction.standard.tools import preprocess, train, FocalLoss
# Preprocess data
preprocess(config)
# Choose the model, users can choose from CNN, RNN, Transformer, GCN, Capsule and BERT
__Model__ = {
    'cnn': models.PCNN,
    'rnn': models.BiLSTM,
    'transformer': models.Transformer,
    'gcn': models.GCN,
    'capsule': models.Capsule,
    'lm': models.LM,
}
model = __Model__[config.model_name](config)
# Define the optimizer, scheduler and criterion
optimizer = optim.Adam(model.parameters(), lr=config.learning_rate, weight_decay=config.weight_decay)
scheduler = optim.lr_scheduler.ReduceLROnPlateau(optimizer, factor=config.lr_factor, patience=config.lr_patience)
criterion = FocalLoss()
# Train
train(epoch, model, train_dataloader, optimizer, criterion, device, writer, config)
# Save
model.save(epoch, config)
# Predict
model.load(config.saved_path, device=device)
model(preprocess(config.text))

```

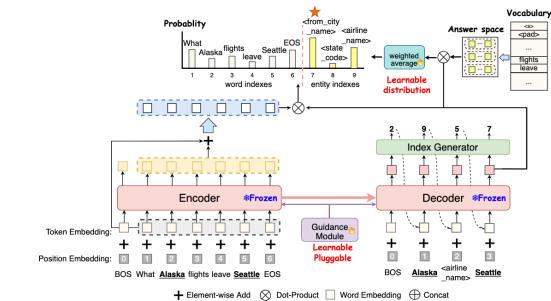


KnowPrompt WWW'22



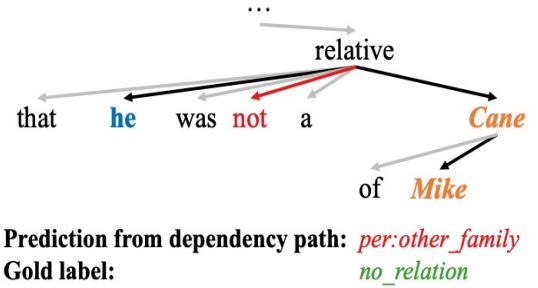
DocuNet IJCAI'21

KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction WWW'22
 LightNER: A Lightweight Generative Framework with Prompt-guided Attention for Low-resource NER Arxiv'21
 Document-level Relation Extraction as Semantic Segmentation IJCAI'21
 Graph Convolution over Pruned Dependency Trees Improves Relation Extraction EMNLP'18



LightNER Arxiv' 21

I had an e-mail exchange with Benjamin Cane of Popular Mechanics which showed that **he** was not a relative of **Mike Cane**.



C-GCN EMNLP'18

第一步 : 进入目标任务文件夹

```
cd example/ner/standard
```

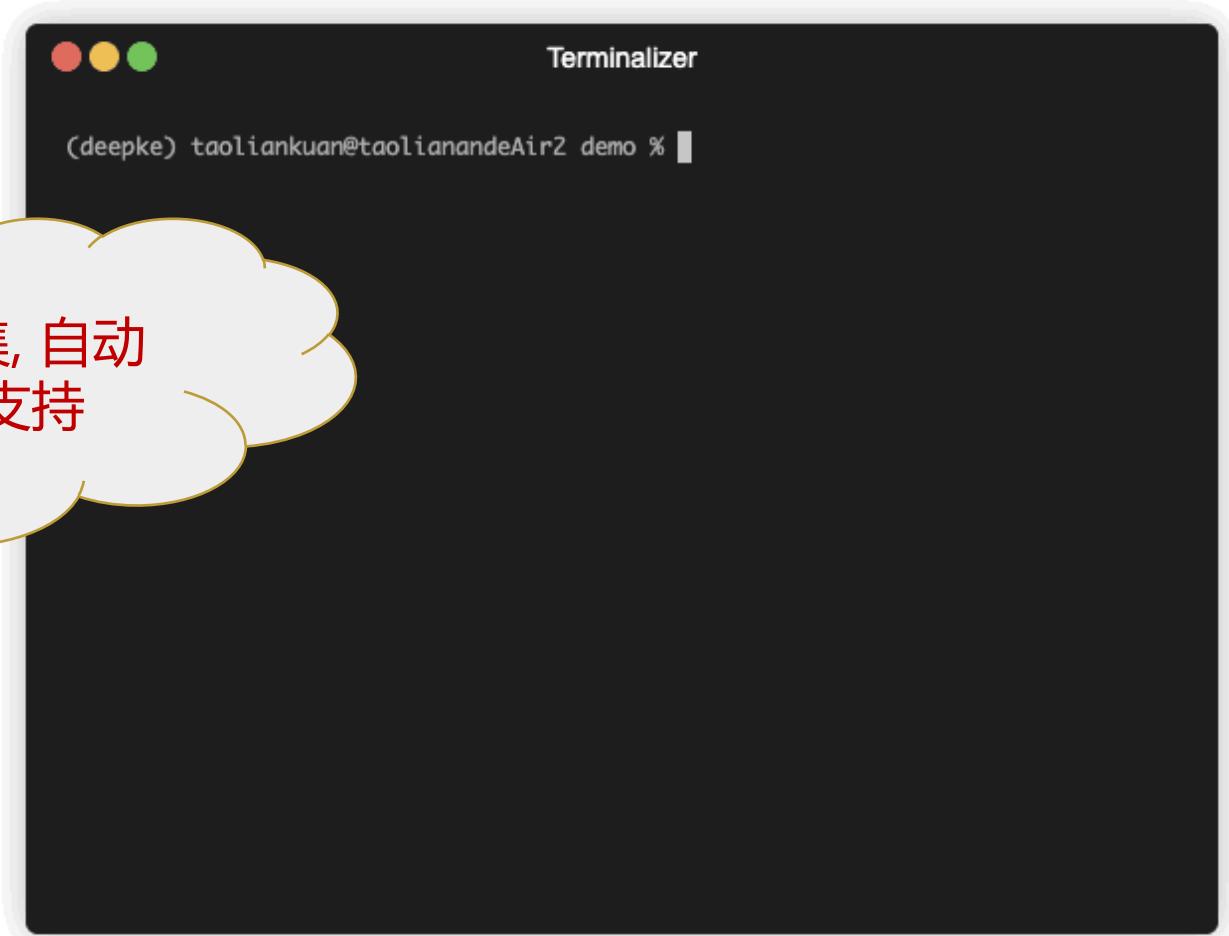
第二步 : 模型训练

```
python run.py
```

自定义数据集, 自动
调参, 容器支持

第三步 : 模型验证

```
python predict.py
```



DeepKE: 支持cnSchema的开箱即用模型



DeepKE 提供了预训练好的支持cnSchem的开箱即用模型：支持抽取**28种实体类型**和**50种关系类型**。

20 Class



50 Relations

Entity Recognition

DeepKE(NER), Roberta-wwm-ext, Chinese

Google dowload

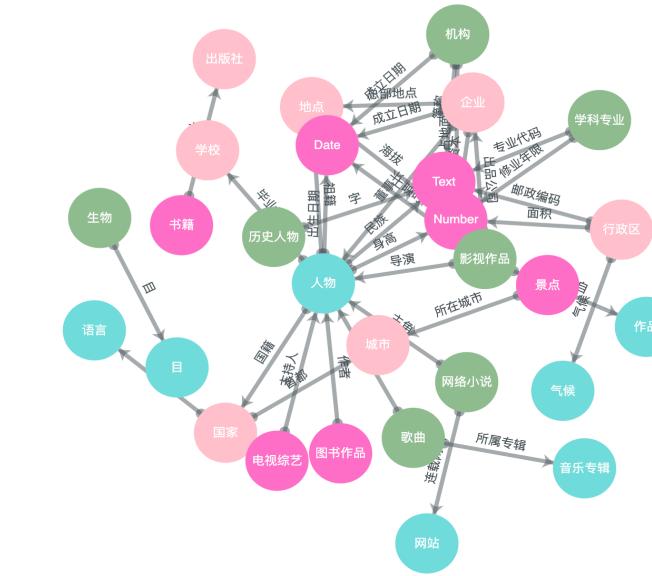
Baidu Netdisk
download(password:u022)

Relation Extraction

DeepKE(RE), Roberta-wwm-ext, Chinese

Google download

Baidu Netdisk
download(password:29oe)



```
1 {
2     "text": "查尔斯·阿兰基斯 (Charles Aránguiz)，1989年4月17日出生于智利圣地亚哥，智利职业足球运动员",
3     "spo_list": [
4         {
5             "predicate": "出生地",
6             "object_type": "地点",
7             "subject_type": "人物",
8             "object": "圣地亚哥",
9             "subject": "查尔斯·阿兰基斯"
10        },
11        {
12            "predicate": "出生日期",
13            "object_type": "Date",
14            "subject_type": "人物",
15            "object": "1989年4月17日",
16            "subject": "查尔斯·阿兰基斯"
17        }
18    ]
19 }
```

DeepKE

[Search docs](#)

GETTING STARTED

- Start
- Install
- Example
- FAQ

PACKAGE

- DeepKE
 - Attribution Extraction
 - Name Entity Recognition
 - Few Shot
 - Models
 - Module
 - Utils
 - Standard
- Relation Extraction

Models

deepke.name_entity_recognition.few_shot.models.model module

```
class deepke.name_entity_recognition.few_shot.models.model.PromptBartEncoder(encoder) [source]
```

Bases: [torch.nn.modules.module.Module](#)

```
forward(src_tokens, attention_mask=None, past_key_values=None) [source]
```

Defines the computation performed at every call.

Should be overridden by all subclasses.

Note

Although the recipe for forward pass needs to be defined within this function, one should call the [Module](#) instance afterwards instead of this since the former takes care of running the registered hooks while the latter silently ignores them.

```
class deepke.name_entity_recognition.few_shot.models.model.PromptBartDecoder(decoder, pad_token_id, label_ids, use_prompt=False, prompt_len=10, learn_weights=False) [source]
```

Bases: [torch.nn.modules.module.Module](#)

```
forward(tgt_tokens, prompt_state) [source]
```

Defines the computation performed at every call.

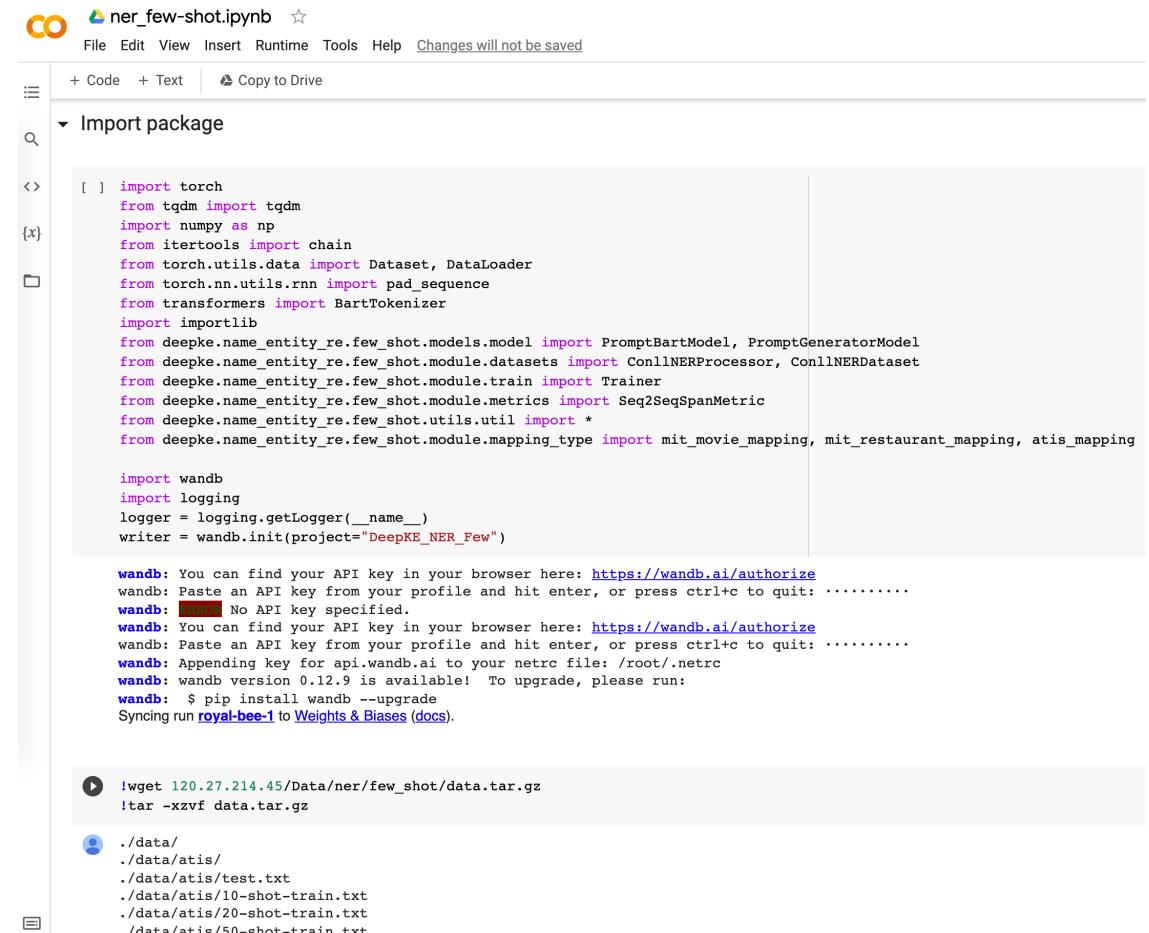
Should be overridden by all subclasses.

Note

Although the recipe for forward pass needs to be defined within this function, one should call the [Module](#) instance afterwards instead of this since the former takes care of running the registered hooks while the latter silently ignores them.

```
decode(tokens, state) [source]
```

详细在线文档



The screenshot shows a Google Colab notebook titled "ner_few-shot.ipynb". The code cell contains imports for torch, tqdm, numpy, and various PyTorch and DeepKE modules. It also includes logging setup and a wandb initialization command. Below the code cell, there is a terminal window showing the execution of commands to download data and extract tar files.

```
File Edit View Insert Runtime Tools Help Changes will not be saved
+ Code + Text Copy to Drive
Import package
[ ] import torch
from tqdm import tqdm
import numpy as np
from itertools import chain
from torch.utils.data import Dataset, DataLoader
from torch.nn.utils.rnn import pad_sequence
from transformers import BartTokenizer
import importlib
from deepke.name_entity_re.few_shot.models.model import PromptBartModel, PromptGeneratorModel
from deepke.name_entity_re.few_shot.module.datasets import ConllNERProcessor, ConllNERDataset
from deepke.name_entity_re.few_shot.module.train import Trainer
from deepke.name_entity_re.few_shot.module.metrics import Seq2SeqSpanMetric
from deepke.name_entity_re.few_shot.utils.util import *
from deepke.name_entity_re.few_shot.module.mapping_type import mit_movie_mapping, mit_restaurant_mapping, atis_mapping

import wandb
import logging
logger = logging.getLogger(__name__)
writer = wandb.init(project="DeepKE_NER_Few")

wandb: You can find your API key in your browser here: https://wandb.ai/authorize
wandb: Paste an API key from your profile and hit enter, or press ctrl+c to quit: .....
wandb: No API key specified.
wandb: You can find your API key in your browser here: https://wandb.ai/authorize
wandb: Paste an API key from your profile and hit enter, or press ctrl+c to quit: .....
wandb: Appending key for api.wandb.ai to your netrc file: /root/.netrc
wandb: wandb version 0.12.9 is available! To upgrade, please run:
wandb: $ pip install wandb --upgrade
Syncing run royal-bee-1 to Weights & Biases (docs).

!wget 120.27.214.45/Data/ner/few_shot/data.tar.gz
!tar -xzvf data.tar.gz

./data/
./data/atis/
./data/atis/test.txt
./data/atis/10-shot-train.txt
./data/atis/20-shot-train.txt
./data/atis/50-shot-train.txt
```

Google Colab教程

DeepKE是一个支持cnSchema、**低资源、长篇章、多模态**的知识图谱抽取开源工具



DeepKE

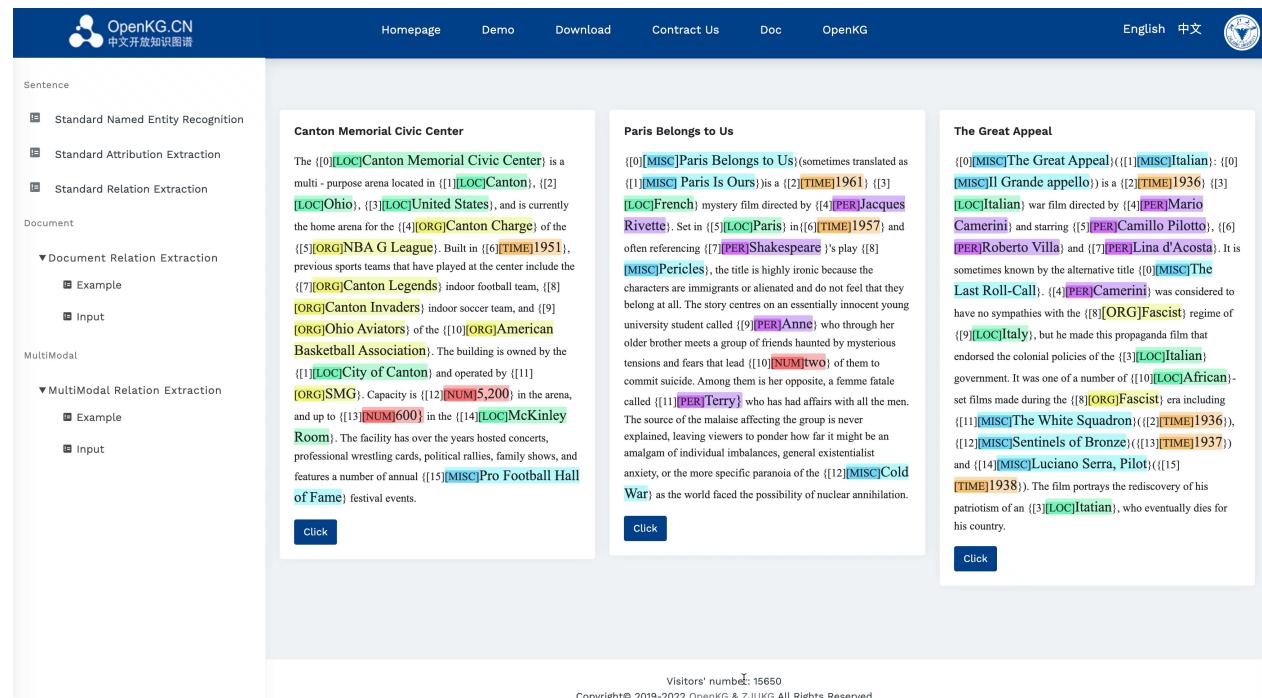
Knowledge Extraction Tool

DeepKE Public

An Open Toolkit for Knowledge Graph Extraction and Construction

Python ⭐ 1.2k 📂 355

<https://github.com/zjunlp/DeepKE>



The screenshot shows the DeepKE online demo interface. At the top, there's a navigation bar with links to Homepage, Demo, Download, Contact Us, Doc, and OpenKG. Below the navigation, there are several cards displaying extracted knowledge from sentences. One card for 'Canton Memorial Civic Center' shows its location in Canton, Ohio, and its role as the home arena for the Canton Charge. Another card for 'Paris Belongs to Us' discusses its capacity and history. A third card for 'The Great Appeal' provides details about the film's plot and cast. Each card includes a 'Click' button at the bottom.

Visitors' number: 15650
Copyright© 2019-2022 OpenKG & ZJUKG All Rights Reserved





谢谢!

A C C E P T M Y E N D L E S S G R A T I T U D E