

复现课程知识库助手搭建过程（截图）：

下载模型：

```
(InternLM) root@intern-studio: # python /root/data/download.py
2024-01-12 00:19:31.967 - modelscope - INFO - PyTorch version 2.0.1 Found.
2024-01-12 00:19:31.968 - modelscope - INFO - Loading ast index from /root/.cache/modelscope/ast_indexer
2024-01-12 00:19:34.443 - modelscope - INFO - No valid ast index found from /root/.cache/modelscope/ast_indexer, generating ast index from prebuilt!
2024-01-12 00:19:34.617 - modelscope - INFO - Loading done! Current index file version is 1.9.5, with md5 524ee28b48c3c0a502628764d684b10e and a total number of 945 components indexed
2024-01-12 00:19:39.483 - modelscope - INFO - Use user-specified model revision: v1.0.3
```

download_hf:

```
downloading https://hf-mirror.com/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2/resolve/a9c555277f9bcf24f28fa5e092e665fc67c49cd/tf_model.h5 to /root/.cache/huggingface/hub/models--sentence-transformers-paraphrase-multilingual-MiniLM-L12-v2/blobs/22150b6ba00e477c7f816f19884028fff924e2b62e1454088909c72c5add40e.incomplete
downloading https://hf-mirror.com/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2/resolve/a9c555277f9bcf24f28fa5e092e665fc67c49cd/tokenizer.json to /root/.cache/huggingface/hub/models--sentence-transformers-paraphrase-multilingual-MiniLM-L12-v2/blobs/2c3387be76557bd40970cec13153b3bbf80407865484b209e655e4f29076b8.incomplete
downloading https://hf-mirror.com/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2/resolve/a9c555277f9bcf24f28fa5e092e665fc67c49cd/tokenizer_config.json to /root/.cache/huggingface/hub/models--sentence-transformers-paraphrase-multilingual-MiniLM-L12-v2/blobs/0da3507018a1ac625f93179ff60b9b9202cc6c.incomplete
downloading https://hf-mirror.com/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2/resolve/a9c555277f9bcf24f28fa5e092e665fc67c49cd/sentencepiece.bpe.model to /root/.cache/huggingface/hub/models--sentence-transformers-paraphrase-multilingual-MiniLM-L12-v2/blobs/cfc8146ab2a0488e9e2a0c56de79527c11ab599eca145d0a727afcc0db2985.incomplete
downloading https://hf-mirror.com/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2/resolve/a9c555277f9bcf24f28fa5e092e665fc67c49cd/unigram.json to /root/.cache/huggingface/hub/models--sentence-transformers-paraphrase-multilingual-MiniLM-L12-v2/blobs/71b44701d7ef4054205115acfa6ef126c5d2f84bd3aff0c59e48163674d19a6.incomplete
special_tokens_map.json: 239B [00:00, 1.26MB/s]
tokenizer_config.json: 480B [00:00, 2.67MB/s]
sentencepiece.bpe.model: 100%
unigram.json: 14.8MB [00:01, 11.0MB/s]
tf_model.h5: 100%
pytorch_model.bin: 100%
Fetching 14 files: 100%
/root/data/model/sentence-transformer
(InternLM) root@intern-studio: #
```

下载 NLTK 相关资源：

```
chive: punkt.zip
creating: punkt/
inflating: punkt/greek.pickle
inflating: punkt/estonian.pickle
inflating: punkt/turkish.pickle
inflating: punkt/polish.pickle
creating: punkt/PY3/
inflating: punkt/PY3/greek.pickle
inflating: punkt/PY3/estonian.pickle
inflating: punkt/PY3/turkish.pickle
inflating: punkt/PY3/polish.pickle
inflating: punkt/PY3/russian.pickle
inflating: punkt/PY3/czech.pickle
inflating: punkt/PY3/portuguese.pickle
inflating: punkt/PY3/README
inflating: punkt/PY3/dutch.pickle
inflating: punkt/PY3/norwegian.pickle
inflating: punkt/PY3/slovene.pickle
inflating: punkt/PY3/english.pickle
inflating: punkt/PY3/danish.pickle
inflating: punkt/PY3/finnish.pickle
inflating: punkt/PY3/swedish.pickle
inflating: punkt/PY3/spanish.pickle
inflating: punkt/PY3/german.pickle
inflating: punkt/PY3/italian.pickle
inflating: punkt/PY3/french.pickle
inflating: punkt/russian.pickle
inflating: punkt/czech.pickle
inflating: punkt/portuguese.pickle
inflating: punkt/README
inflating: punkt/dutch.pickle
inflating: punkt/norwegian.pickle
inflating: punkt/slovene.pickle
inflating: punkt/english.pickle
inflating: punkt/danish.pickle
inflating: punkt/finnish.pickle
inflating: punkt/swedish.pickle
inflating: punkt/spanish.pickle
inflating: punkt/german.pickle
inflating: punkt/italian.pickle
inflating: punkt/french.pickle
inflating: punkt/.DS_Store
inflating: punkt/PY3/malayalam.pickle
inflating: punkt/malayalam.pickle
(InternLM) root@intern-studio: /nltk_data/tokenizers# cd ../taggers
(InternLM) root@intern-studio: /nltk_data/taggers# unzip averaged_perceptron_tagger.zip
chive: averaged_perceptron_tagger.zip
creating: averaged_perceptron_tagger/
inflating: averaged_perceptron_tagger/averaged_perceptron_tagger.pickle
(InternLM) root@intern-studio: /nltk_data/taggers#
```

下载本项目代码：

```
(InternLM) root@intern-studio:~/nltk_data/taggers# cd /root/data
(InternLM) root@intern-studio:~/data# git clone https://github.com/InternLM/tutorial
Cloning into 'tutorial'...
remote: Enumerating objects: 446, done.
remote: Counting objects: 100% (208/208), done.
remote: Compressing objects: 100% (119/119), done.
remote: Total 446 (delta 163), reused 105 (delta 89), pack-reused 238
Receiving objects: 100% (446/446), 11.38 MiB | 11.64 MiB/s, done.
Resolving deltas: 100% (207/207), done.
```

数据收集：

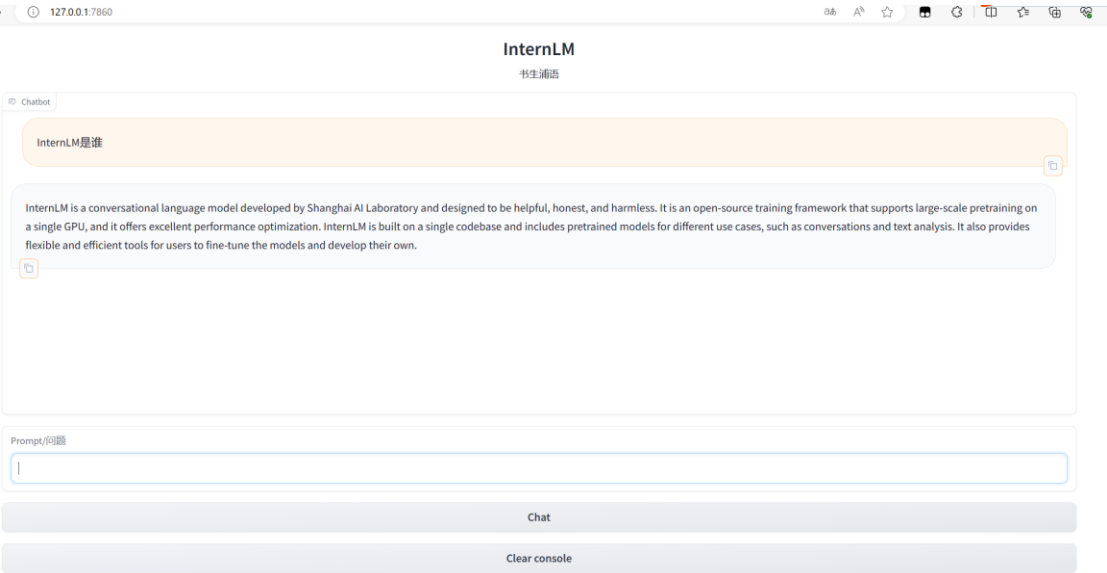
```
Resolving deltas: 100% (207/207), done.
(InternLM) root@intern-studio: /data# # 进入到数据库盘
(InternLM) root@intern-studio: /data# cd /root/data
(InternLM) root@intern-studio: /data# # clone 上述开源仓库
(InternLM) root@intern-studio: /data# git clone https://gitee.com/open-compass/opencompass.git
Cloning into 'opencompass'...
remote: Enumerating objects: 4843, done.
remote: Total 4843 (delta 0), reused 0 (delta 0), pack-reused 4843
Receiving objects: 100% (4843/4843), 1.48 MiB | 1.80 MiB/s, done.
Resolving deltas: 100% (2941/2941), done.
(InternLM) root@intern-studio: /data# git clone https://gitee.com/InternLM/lmdeploy.git
Cloning into 'lmdeploy'...
remote: Enumerating objects: 4485, done.
remote: Counting objects: 100% (4485/4485), done.
remote: Compressing objects: 100% (1494/1494), done.
remote: Total 4485 (delta 2914), reused 4485 (delta 2914), pack-reused 0
Receiving objects: 100% (4485/4485), 2.23 MiB | 1.94 MiB/s, done.
Resolving deltas: 100% (2914/2914), done.
(InternLM) root@intern-studio: /data# git clone https://gitee.com/InternLM/xtuner.git
Cloning into 'xtuner'...
remote: Enumerating objects: 3735, done.
remote: Counting objects: 100% (1150/1150), done.
remote: Compressing objects: 100% (252/252), done.
remote: Total 3735 (delta 920), reused 1106 (delta 895), pack-reused 2585
Receiving objects: 100% (3735/3735), 742.80 KiB | 1.21 MiB/s, done.
Resolving deltas: 100% (2741/2741), done.
(InternLM) root@intern-studio: /data# git clone https://gitee.com/InternLM/InternLM-XComposer.git
Cloning into 'InternLM-XComposer'...
remote: Enumerating objects: 680, done.
remote: Counting objects: 100% (680/680), done.
remote: Compressing objects: 100% (273/273), done.
remote: Total 680 (delta 361), reused 680 (delta 361), pack-reused 0
Receiving objects: 100% (680/680), 10.74 MiB | 3.94 MiB/s, done.
Resolving deltas: 100% (361/361), done.
(InternLM) root@intern-studio: /data# git clone https://gitee.com/InternLM/lagent.git
Cloning into 'lagent'...
remote: Enumerating objects: 414, done.
remote: Counting objects: 100% (414/414), done.
remote: Compressing objects: 100% (188/188), done.
remote: Total 414 (delta 197), reused 414 (delta 197), pack-reused 0
Receiving objects: 100% (414/414), 214.97 KiB | 1000.00 KiB/s, done.
Resolving deltas: 100% (197/197), done.
```

知识库搭建：

```
(InternLM) root@intern-studio: /data/demo# python create_db.py
0% | 0/25 [00:00:00, 211t/s]
/root/.conda/envs/InternLM/lib/python3.10/site-packages/unstructured/documents/html.py:498: FutureWarning: The behavior of this method will change in future versions. Use specific 'len(elem)' or 'elem is n
ot None' test instead.
  rows = body.findall("tr") if body else []
20% | 5/25 [00:19:01:00, 3.03s/it]
/root/.conda/envs/InternLM/lib/python3.10/site-packages/unstructured/documents/html.py:498: FutureWarning: The behavior of this method will change in future versions. Use specific 'len(elem)' or 'elem is n
ot None' test instead.
  rows = body.findall("tr") if body else []
100% | 25/25 [00:19:00:00, 1.25t/s]
100% | 9/9 [00:00:00:00, 22.87t/s]
100% | 18/18 [00:00:00:00, 44.92t/s]
100% | 72/72 [00:02:00:00, 28.31t/s]
100% | 113/113 [00:04:00:00, 23.68t/s]
100% | 26/26 [00:01:00:00, 19.60t/s]
```

InternLM 接入 LangChain：

部署 Web Demo：



- 提醒：
- 1. 初始化数据库时间可能较长，请耐心等待。
 - 2. 使用中如果出现异常，将会在文本输入框进行展示，请不要惊慌。