

算法运行说明

整体基于langchain-chatchat框架

环境 window11 python==3.10.10

```
conda create -n langchain python=3.10.10
conda activate langchain
```

需要pip install如下内容(requirements.txt)

```
torch==2.1.2
torchvision==0.16.2
torchaudio==2.1.2
xformers==0.0.23.post1
transformers==4.37.2
sentence_transformers==2.2.2
langchain==0.0.354
langchain-experimental==0.0.47
pydantic==1.10.13
fschat==0.2.35
openai==1.9.0
fastapi==0.109.0
sse_starlette==1.8.2
nltk==3.8.1
uvicorn>=0.27.0.post1
starlette==0.35.0
unstructured[all-docs] # ==0.11.8
python-magic-bin; sys_platform == 'win32'
SQLAlchemy==2.0.25
faiss-cpu==1.7.4
accelerate==0.24.1
spacy==3.7.2
PyMuPDF==1.23.16
rapidocr_onnxruntime==1.3.8
requests==2.31.0
pathlib==1.0.1
pytest==7.4.3
numexpr==2.8.6
strsimpy==0.2.1
markdownify==0.11.6
tiktoken==0.5.2
tqdm==4.66.1
websockets==12.0
numpy==1.24.4
pandas==2.0.3
einops==0.7.0
transformers_stream_generator==0.0.4
```

```
vllm==0.2.7; sys_platform == "linux"
llama-index==0.9.35
arxiv==2.1.0
youtube-search==2.1.2
duckduckgo-search==3.9.9
metaphor-python==0.1.23

streamlit==1.30.0
streamlit-option-menu==0.3.12
streamlit-antd-components==0.3.1
streamlit-chatbox==1.1.11
streamlit-modal==0.1.0
streamlit-aggrid==0.3.4.post3

httpx==0.26.0
httpx_sse==0.4.0
watchdog==3.0.0
pyjwt==2.8.0
```

接着我们把“代码文件”中的文件都放进Langchain-Chatchat\webui_pages\dialogue文件夹，注意dialogue.py替换原来的dialogue.py文件。

论文pdf原文件放在pdf.py读取的data-set目录对应的文件夹

数据预处理部分

```
cd D:\Langchain-Chatchat\webui_pages\dialogue # 改成你的大模型中的dialogue文件夹中的
路径
python pdf.py
python getquery.py
```

对话部分

写了一个.bat文件可一键运行放在Langchain-Chatchat文件夹中，内容如下

```
:start
cd D:\Langchain-Chatchat
call miniconda3\Scripts\activate.bat
goto :run

:run
python startup.py -a

pause
```

现在可以运行了，最后在Langchain-Chatchat\webui_pages\dialogue文件夹看到result.csv。