Longchain-chatchat

1. Create env

```
conda create -n chatchat python=3.10
```

```
conda activate chatchat
```

2. Git clone

```
git clone -b 0.2.final https://github.com/chatchat-space/Langchain-Chatchat.git
```

```
cd Langchain-Chatchat
```

```
pip install -r requirements.txt
```

```
pip install -r requirements_api.txt
```

```
pip install -r requirements_webui.txt
```

3. Pytorch

```
import torch
```

```
torch.cuda.is_available()
```

it should be unsuccessful in most cases

```
pip uninstall torch
```

#CUDA 11.8
```
conda install pytorch==2.1.2 torchvision==0.16.2 torchaudio==2.1.2 pytorch-cuda=11.8 -c pytorch -c nvidia
```
#CUDA 12.1
```
conda install pytorch==2.1.2 torchvision==0.16.2 torchaudio==2.1.2 pytorch-cuda=12.1 -c pytorch -c nvidia
```

4. Model download

```
cd /Users/...../Langchain-Chatchat
```

```
git lfs install
```

```
git clone https://huggingface.co/THUDM/chatglm3-6b
```

```
git clone https://huggingface.co/BAAI/bge-m3
```

5. Initializes the knowledge base

```
python copy_config_example.py
```

```
python init_database.py --recreate-vs
```

6. configuration file needs to be changed a lot. Please use the code I changed directly

If copy my code, you just need to change the

the `MODEL_PATH` section in `model_config.py` , change the path of `bge-m3` and `chatglm3-6b` to your own local placement path,

```python
MODEL_PATH = {
    "embed_model": {
        "ernie-tiny": "nghuyong/ernie-3.0-nano-zh",
        "ernie-base": "nghuyong/ernie-3.0-base-zh",
        "text2vec-base": "shibing624/text2vec-base-chinese",
        "text2vec": "GanymedeNil/text2vec-large-chinese",
        "text2vec-paraphrase": "shibing624/text2vec-base-chinese-paraphrase",
        "text2vec-sentence": "shibing624/text2vec-base-chinese-sentence",
        "text2vec-multilingual": "shibing624/text2vec-base-multilingual",
        "text2vec-bge-large-chinese": "shibing624/text2vec-bge-large-chinese",
        "m3e-small": "moka-ai/m3e-small",
        "m3e-base": "moka-ai/m3e-base",
        "m3e-large": "moka-ai/m3e-large",

        "bge-small-zh": "BAAI/bge-small-zh",
        "bge-base-zh": "BAAI/bge-base-zh",
        #"bge-large-zh": "BAAI/bge-large-zh",
        "bge-large-zh": "/Users/patty.hao/Documents/task5_LLM/Langchain-Chatchat/bge-large-zh",

        "bge-large-zh-noinstruct": "BAAI/bge-large-zh-noinstruct",
        "bge-base-zh-v1.5": "BAAI/bge-base-zh-v1.5",
        "bge-large-zh-v1.5": "BAAI/bge-large-zh-v1.5",

        #"bge-m3": "BAAI/bge-m3",
        "bge-m3": "/Users/patty.hao/Documents/task5_LLM/Langchain-Chatchat/bge-m3",

        "piccolo-base-zh": "sensenova/piccolo-base-zh",
        "piccolo-large-zh": "sensenova/piccolo-large-zh",
        "nlp_gte_sentence-embedding_chinese-large": "damo/nlp_gte_sentence-embedding_chinese-large",
        "text-embedding-ada-002": "your OPENAI_API_KEY",
    },

    "llm_model": {
        "chatglm2-6b": "THUDM/chatglm2-6b",
        "chatglm2-6b-32k": "THUDM/chatglm2-6b-32k",
        #"chatglm3-6b": "THUDM/chatglm3-6b",
        "chatglm3-6b": "/Users/patty.hao/Documents/task5_LLM/Langchain-Chatchat/chatglm3-6b",
        "chatglm3-6b-32k": "THUDM/chatglm3-6b-32k",

        "Orion-14B-Chat": "OrionStarAI/Orion-14B-Chat",
        "Orion-14B-Chat-Plugin": "OrionStarAI/Orion-14B-Chat-Plugin",
        "Orion-14B-LongChat": "OrionStarAI/Orion-14B-LongChat",
```

Otherwise:

Under Langchain-Chatchat\configs

There are ....

a. basic_config.py : for some basic settings, do not need change at all

b. kb_config : config for knowledge base

At the end of code, change `TEXT_SPLITTER_NAME` into `"SpacyTextSplitter"`

```
1  #TEXT_SPLITTER_NAME = "ChineseRecursiveTextSplitter"
```

```
2  TEXT_SPLITTER_NAME = "SpacyTextSplitter"
```

c. model_config.py : config for choosing different models

Change:

```
1  #EMBEDDING_MODEL = "bge-large-zh"
2  EMBEDDING_MODEL = "bge-m3"
3
4  ....
5
6  #LLM_MODELS = ["chatglm3-6b", "zhipu-api", "openai-api"]
7  LLM_MODELS = ["chatglm3-6b"]
```

In addition to the `MODEL_PATH` section in `model_config.py`, change the path of `bge-m3` and `chatglm3-6b` to your own local placement path,

```
MODEL_PATH = {
    "embed_model": {
        "ernie-tiny": "nghuyong/ernie-3.0-nano-zh",
        "ernie-base": "nghuyong/ernie-3.0-base-zh",
        "text2vec-base": "shibing624/text2vec-base-chinese",
        "text2vec": "GanymedeNil/text2vec-large-chinese",
        "text2vec-paraphrase": "shibing624/text2vec-base-chinese-paraphrase",
        "text2vec-sentence": "shibing624/text2vec-base-chinese-sentence",
        "text2vec-multilingual": "shibing624/text2vec-base-multilingual",
        "text2vec-bge-large-chinese": "shibing624/text2vec-bge-large-chinese",
        "m3e-small": "moka-ai/m3e-small",
        "m3e-base": "moka-ai/m3e-base",
        "m3e-large": "moka-ai/m3e-large",

        "bge-small-zh": "BAAI/bge-small-zh",
        "bge-base-zh": "BAAI/bge-base-zh",
        #"bge-large-zh": "BAAI/bge-large-zh",
        "bge-large-zh": "/Users/patty.hao/Documents/task5_LLM/Langchain-Chatchat/bge-large-zh",

        "bge-large-zh-noinstruct": "BAAI/bge-large-zh-noinstruct",
        "bge-base-zh-v1.5": "BAAI/bge-base-zh-v1.5",
        "bge-large-zh-v1.5": "BAAI/bge-large-zh-v1.5",

        #"bge-m3": "BAAI/bge-m3",
        "bge-m3": "/Users/patty.hao/Documents/task5_LLM/Langchain-Chatchat/bge-m3",

        "piccolo-base-zh": "sensenova/piccolo-base-zh",
        "piccolo-large-zh": "sensenova/piccolo-large-zh",
        "nlp_gte_sentence-embedding_chinese-large": "damo/nlp_gte_sentence-embedding_chinese-large",
        "text-embedding-ada-002": "your OPENAI_API_KEY",
    },

    "llm_model": {
        "chatglm2-6b": "THUDM/chatglm2-6b",
        "chatglm2-6b-32k": "THUDM/chatglm2-6b-32k",
        #"chatglm3-6b": "THUDM/chatglm3-6b",
        "chatglm3-6b": "/Users/patty.hao/Documents/task5_LLM/Langchain-Chatchat/chatglm3-6b",
        "chatglm3-6b-32k": "THUDM/chatglm3-6b-32k",

        "Orion-14B-Chat": "OrionStarAI/Orion-14B-Chat",
        "Orion-14B-Chat-Plugin": "OrionStarAI/Orion-14B-Chat-Plugin",
        "Orion-14B-LongChat": "OrionStarAI/Orion-14B-LongChat",
```

d. prompt_config.py : give some default prompt to the chat model , can change the response language , just delete some sentences saying "the response should be chinese"

e. server_config.py : about some webUI page port or API port

7. start

```
python startup.py -a
```

When ...showing `Email:` message in blue color press `Enter` otherwise it will not auto open the Web Page