# 源码部署

## 系统要求

**系统**

Ubuntu x86\_64 系统 >=18.04

python >= 3.8

**安装 python 依赖**

pip3 install -r requirements.txt

**安装torch**

#CPU版本

pip3 install torch --index-url https://download.pytorch.org/whl/cpu

或

#GPU版本

pip3 install torch --index-url https://download.pytorch.org/whl/cu118

**⽀持 CPU 运⾏， 如需使⽤ GPU 运⾏，需硬件配备英伟达显卡**

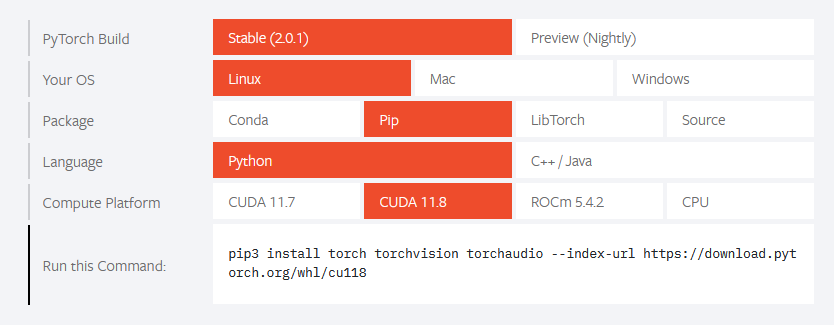
nvidia显卡驱动 > 525.89.02

cuda版本 > 11.8

torch==2.0.0+cu118

或根据具体硬件配置参照torch官方网站建议版本安装

<https://pytorch.org/>



其他硬件版本

<https://pytorch.org/get-started/previous-versions/>

**配置工程是否使用GPU**

打开工程文件中的config/config.py

修改

CONF.GPU= True

或在服务启动时传入device参数，具体见启动服务说明。

## 模型文件

由于依赖的大模型文件较大，建议从官方网站直接下载

### 1 text2vec

安装huggingface下载器

git lfs install

下载文件

git clone <https://huggingface.co/GanymedeNil/text2vec-large-chinese>

或

直接打开页面手动下载

<https://huggingface.co/GanymedeNil/text2vec-large-chinese>

### **2 chatglm2-6b-int4**

安装huggingface下载器

git lfs install

下载文件

git clone <https://huggingface.co/THUDM/chatglm2-6b-int4>

或直接打开页面手动下载

<https://huggingface.co/THUDM/chatglm2-6b-int4>

### 3 配置模型文件

打开工程文件中的config/config.py

修改

CONF.embeding\_model\_path = [text2vec的文件夹地址]

CONF.llm\_model\_path= [chatglm2-6b-int4的文件夹地址]

## 启动服务

进入工程文件夹

**GPU/CPU模式**

通过修改config/config.py文件中的CONF.GPU=True/False 来设置GPU和CPU模式

如果启动时指定device参数，则以device参数传递值为准

运行以下命令启动服务

python3 service.py [--port=] [--device=]

该服务监听本地[port]端口,接收客户端上传的文件并返回分析后的表格格式数据。默认在8000端口监听。如果未指定device参数，以CONF.GPU的配置为准。

启动参数

|  |  |  |  |
| --- | --- | --- | --- |
| **参数名** | **是否必须** | **类型** | **说明** |
| port | 否 | int | 默认在8000端口监听，可根据需要修改端口 |
| device | 否 | string | ‘GPU’/‘CPU’ 选择模型在GPU还是CPU加载  如果不设置该参数，默认使用config文件中的配置项 |

## 调用接口

**客户端脚本**

识别文件中的表格格式：该接口上传[file\_path]指定的本地文件到[ip]:[port]的识别服务，并返回识别后的表格格式数据，默认访问本地8000端口。

进入工程文件夹执行

python3 client.py form\_recognition --file\_path=data/input/支付宝1.pdf [--ip=127.0.0.1] [--port=8000]

接口名**form\_recognition**

调用参数

|  |  |  |  |
| --- | --- | --- | --- |
| **参数名** | **必须** | **类型** | **说明** |
| file\_path | 是 | string | 要识别的文件的本地路径。 |
| ip | 否 | string | 服务所在的ip地址，默认为本机地址127.0.0.1 |
| port | 否 | int | 默认访问8000端口，可根据需要修改 |

返回json文本

**HTTP协议**

**POST**

**接口 <http://[url:port]>/form\_recognition/**

参数

|  |  |  |  |
| --- | --- | --- | --- |
| **参数名** | **必须** | **类型** | **说明** |
| info | 是 | string | 待识别的文件名 |
| data | 是 | string (base64) | 经过base64编码的文件二进制数据 |

返回json文本

**返回值**

返回值为json格式的文本

正确返回：

{

"code":300,

"message":"success",

"data":{

"res1": {

"outside\_infos":

[{"txt": "账户名称： 长沙玉兔信息科技有限公司"},

{"txt": "账号： 155092704"},

{"txt": "开户机构： 中国民生银行股份有限公司长沙分行营业部"}, {"txt": "币种： 人民币元"},

{"txt": "起始日期： 2023-01-01"},

{"txt": "截至日期： 2023-04-09"},

{"txt": "启用高级查询： 是"},

{"txt": "借贷标识： 全部"},

{"txt": "对方账号："},

{"txt": "按发生额查："},

{"txt": "借方累计笔数: 6 借方累计发生额: 123800.00"}, {"txt": "贷方累计笔数: 25 贷方累计发生额: 68169.83"}

]},

"res2": [

{"row\_order": "001\_1",

"header\_row": true,

"1.1": "交易时间",

"1.2": "主机交易流水号",

"1.3": "借方发生额",

"1.4": "贷方发生额",

"1.5": "账户余额",

"1.6": "凭证号",

"1.7": "摘要",

"1.8": "对方账号",

"1.9": "对方账号名称",

"1.10": "对方开户行"},

{"row\_order": "001\_2",

"header\_row": false,

"2.1": "2023-01-05 11:38:56",

"2.2": "50224202301051138563827544649A8B",

"2.3": "0.00",

"2.4": "2237.14",

"2.5": "119291.14",

"2.6": "",

"2.7": "银联入账-0104-0104费7.86元",

"2.8": "991290000023",

"2.9": "银联商务股份有限公司",

"2.10": "中国银联股份有限公司"}],

"doc\_tpye": "excel",

"page\_sum": 1}

}

错误返回：

{

"code":301,

"message":"Type error"

}

**输出格式**

|  |  |
| --- | --- |
| **参数名** | **说明** |
| code | 返回状态码 |
| message | 报错信息 |
| data | 识别数据 |

**code 说明**

|  |  |  |
| --- | --- | --- |
| **代码类型** | **信息** | **说明** |
| 300 | Success | 执行成功 |
| 301 | Type error | 文件格式不符合要求 |
| -1 | unknow | 未知错误 |

**data 参数说明**

|  |  |  |
| --- | --- | --- |
| **名称** | **字段命名** | **内容** |
| 表外数据 | res1 | 第一页的表外数据 |
| 页数 | page\_sum | 总页数  pdf 为pdf数据的实际页数  excel 为以表格分割的页数(每表格一页) |
| 表内数据 | res2 | 1）表内数据，包含每个单元格的行列信息和文字内容，空值则为空；  2）row\_order：行顺序，如 001\_1，代表第一页第一行；  header\_row：true/false，代表该列数据是否属于表头行； |
| 文件类型 | doc\_type | 代表处理的文件类型，分为 excel、epdf，epdf 代表是电子化的 pdf 文件； |