

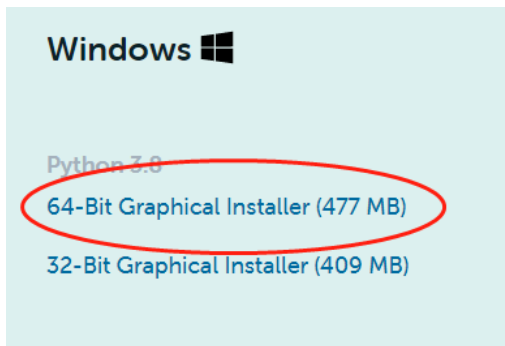
1. 环境配置

依赖 python 环境，建议安装 anaconda 或者使用任意版本 python 在 window 下的安装包。

a) Anaconda 环境安装

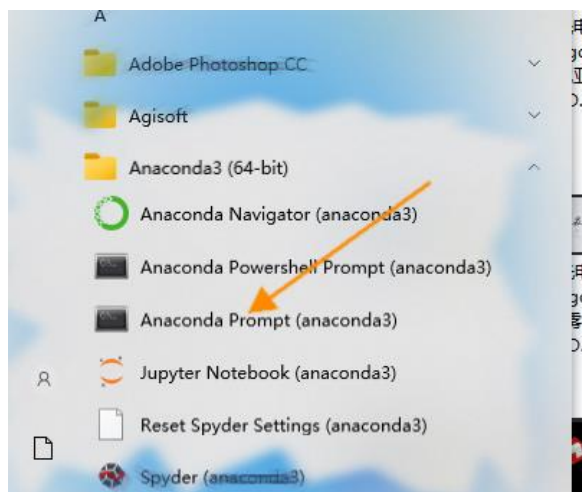
下载链接（当前一般为 64 位操作系统）：

<https://www.anaconda.com/products/individual#Downloads>



下载后双击进行安装，**使用默认推荐即可**。

安装完毕后菜单栏会生成对应的 anaconda 目录，使用 anaconda prompt(anaconda3)



弹出黑色界面，如下所示：



b) Python 环境安装

这时需要新建 python2.7 环境，执行：

```
conda create -n py27 python=2.7
```

输入 Y，等待安装完毕，出现下面结果说明创建成功：

```
Anaconda Prompt (anaconda3)

done
#
# To activate this environment, use
#
#     $ conda activate py27
#
# To deactivate an active environment, use
#
#     $ conda deactivate

(base) C:\Users\ouyangkid>_
```

到这一步，往后每次运行就不需要上面的步骤了。只需要使用

c) 启动运行环境

在 Terminal 中执行如下命令，激活环境

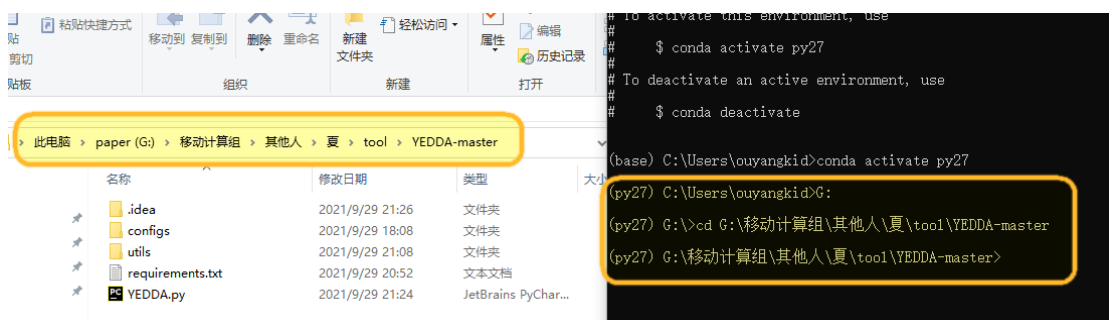
`conda activate py27`

完成对 py27 环境的激活即可使用，激活后会发现最前面括号内的信息已经改变了

```
(base) C:\Users\ouyangkid>conda activate py27
(py27) C:\Users\ouyangkid>
```

进入到程序目录，首先确定存放硬盘，如我的目录位 G 盘，则先使用 G:

再使用 `cd +目录` 进入目标目录



进入成功后，使用 `dir` 命令能够看到 `annotater.py` 文件，如下：

```
(py27) G:\移动计算组\其他人\夏\tool\YEDDA-master>dir
驱动器 G 中的卷是 paper
卷的序列号是 CEF4-414F

G:\移动计算组\其他人\夏\tool\YEDDA-master 的目录
2021/09/29  21:42    <DIR>        .
2021/09/29  21:42    <DIR>        ..
2021/09/29  21:26    <DIR>        .idea
2021/09/29  21:24    41,458 bytes annotater.py
2021/09/29  18:08    <DIR>        config
2021/09/29  20:52    13 requirements.txt
2021/09/29  21:08    <DIR>        utils
                2 个文件        41,458 字节
                5 个目录 161,847,070,720 可用字节

(py27) G:\移动计算组\其他人\夏\tool\YEDDA-master>
```

然后使用清华源更新 pip，复制-粘贴输入即可（该步骤只需要在新电脑执行一次）

```
pip install pip -U
pip config set global.index-url https://pypi.tuna.tsinghua.edu.cn/simple
```

这里即可启动程序，使用 `python annotater.py`



2. 标注工具使用

a) 界面说明



- 首先，通过 open 选择标注文件；之后会在文本显示区域看到文本。
 - *需要注意的是标注过程不要使用中文输入法，会产生字符编码冲突导致文本字符丢失。
 - 接着，通过下方的下拉菜单选择标注字典（这里如果不需要替换，使用默认即可）
- 如果需要定义新的标注类型，或者修改快捷键，需要进入程序目录下的 configs，通过文本编辑器对*.config 文件进行修改。以当前默认文件为例，**当需要进行修改时，建议通过复制粘贴生成一个新文件**，对新文件进行修改。

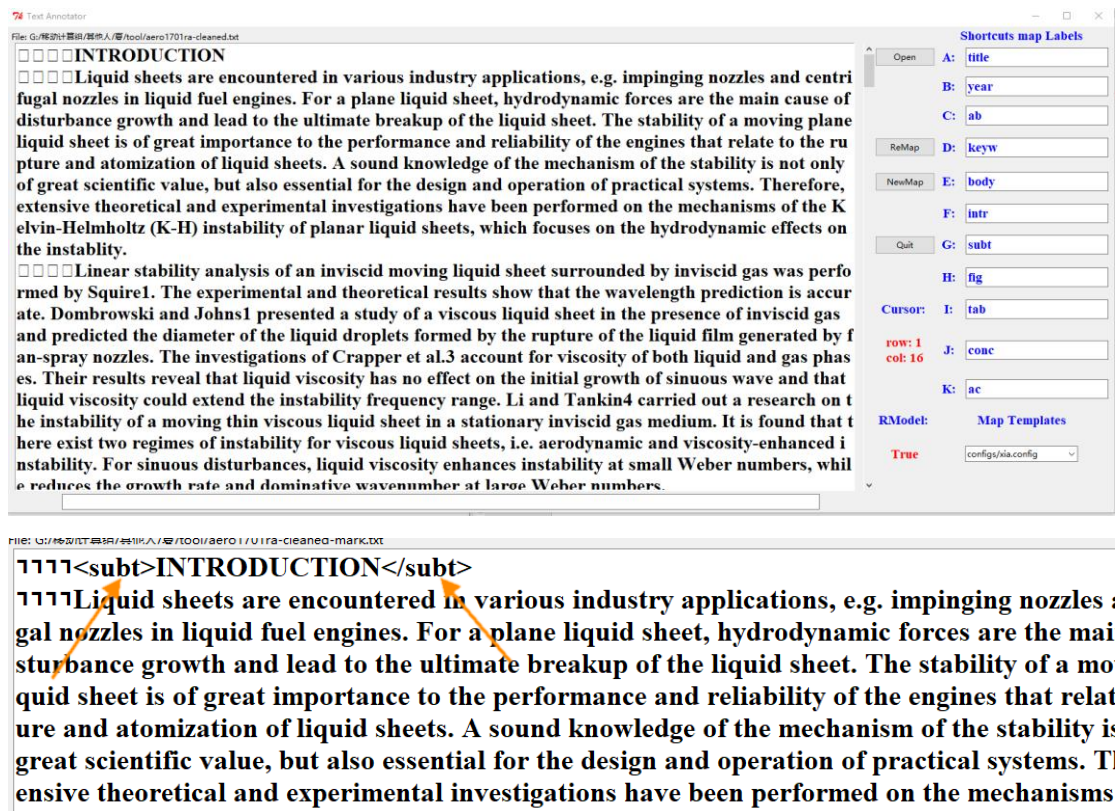
```

1 {"a": "title",
  "b": "ab",
  "c": "year",
  "e": "body",
  "d": "keyw",
  "g": "subt",
  "f": "intr",
  "i": "tab",
  "h": "fig",
  "k": "ac",
  "j": "conc"}

```

文件中的内容格式为：**“快捷键”：“标注类型”**，必须使用双引号+冒号和逗号，半角英文字符。修改完毕后，通过程序界面进行选择即可完成修改，修改完后菜单栏发生对应变化。

- 标注流程，通过鼠标选择需要标注的字符串，选择后会变色，然后单击对应的快捷键一次（注意保证非中文输入法）。




完成操作后，程序会自动保存一个新文件，每次操作程序自动保存一次。

***此外，当标注错误时，通过 `ctrl+z` 进行回滚可取消标注操作。**

***另外，需要当心删除和回车等操作会对文本进行修改。**

***程序增加了对选择目标文本的首、末，以及中间连续空格的检测和删除，但会保留单词之间的单个空格。**

***由于转码问题会出现部分特定符号，该问题由初始文本造成，需要人为修改。**

 <code>aero1701ra-cleaned.txt</code>	2021/9/3 21:14	文本文档	32 KB
 <code>aero1701ra-cleaned-mark.txt</code>	2021/9/29 22:09	文本文档	33 KB

代有-mark 后缀的文件为最终结果。