1.需要pdfminer模块

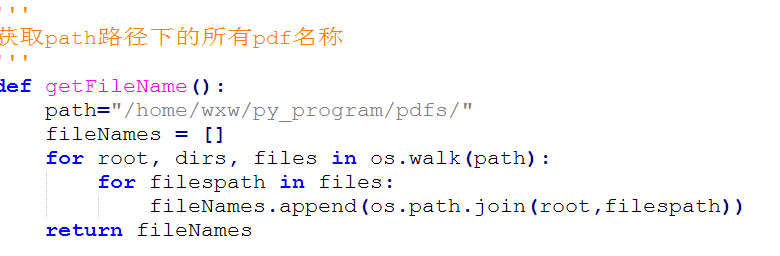
下载源文件包 <http://pypi.python.org/pypi/pdfminer/>

解压

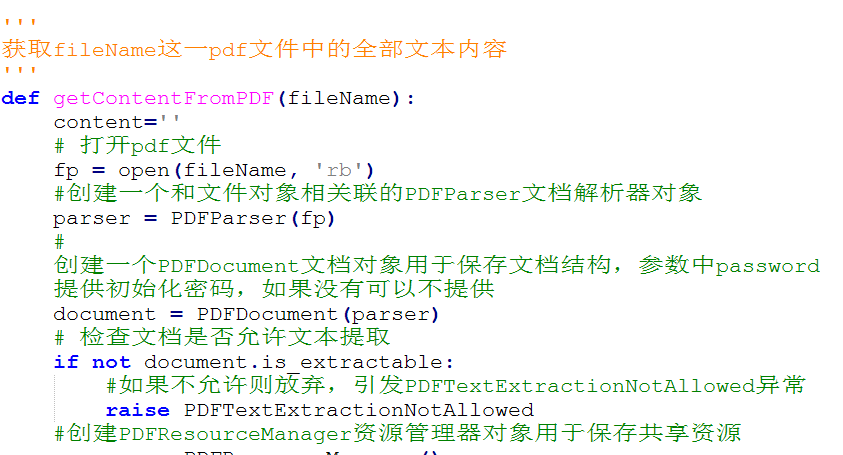
# make cmap（如果需要使用中日韩文字的话）

# python setup.py install

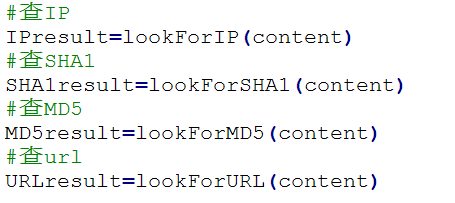
2.要处理的文件都放在这个路径下了，修改一下路径就好



或者是这里，直接给他文件名也可以



3.匹配



所有匹配结果都在这些函数里面处理，函数只返回true or false。。。。

处理结果的数据格式是List，里面每一项都是字符串。

4.所有的都是用的正则去做的，思路非常简单。**欢迎提意见。**