《文本处理系统》

使用说明书

目录

[使用说明书 1](#_Toc20081)

[一、 概览 2](#_Toc19193)

[二、 功能介绍 3](#_Toc5799)

[三、 注意事项 3](#_Toc11328)

[四、 使用案例 4](#_Toc3133)

# 概览

1. **产品定位**

此产品为一智能化文本编辑软件，算法上采用的是百度文心大模型所提供的开源API，并在其基础上进行的功能拼接与组合，最终实现了手写文本识别、文本纠错、小说续写等功能。并且设计了较好的可视化界面，能够满足普通人的使用需求。

1. **目标人群**

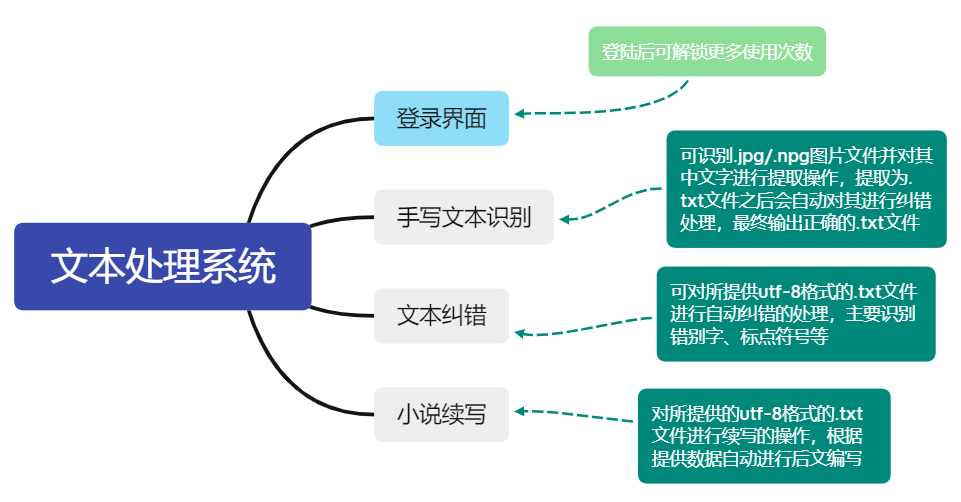
此产品主要面向大学生与中小学生，主要解决论文错字识别、手写文本识别并纠错与小说或文字写作思路提供这三个方面的问题，能够轻松的处理平时所注意不到的文章漏洞，使办公更便捷。

1. **产品优势**

本产品由大学生自主研发，并且提供开源算法，用户可根据实际需求对程序进行改写，来适应不同的需求。

本产品模型架构泛用性较高，用户可根据实际需求去文心大模型官网下载所需API，粘贴进来即可直接调用运行。

# 功能介绍



1. **登陆界面**

用户可选择登录或游客使用，系统默认游客使用，登陆后则可享受更多使用次数，登录API可以从左上角点开链接获取。

1. **手写文本识别**

采用百度飞桨OCR模型，对手写文字进行识别，目前支持短篇幅文本，对于大量手写文本识别时间较长。经过OCR模型后，保存路径为自行建立的.txt，.doc或.docx格式的文件，从而转化为文本模式。该文本后续可以提供给文本纠错模块进行纠错操作，之后会返回正确识别结果在对应的目标文件夹内。

1. **文本纠错**

采用文心大模型构建文本纠错系统，选择所需纠错文件（建议选用.txt）（格式需为utf-8），选择存储位置，即可运行进行纠错，最终返回结果会被保存在目标文件夹下。

1. **小说续写**

采用文心大模型构建小说续写（文本续写）系统，选择需要续写稿件（.txt）（格式需为utf-8），选择存储位置，即可运行进行小说续写操作，最终返回结果会被保存在目标文件夹下。

# 注意事项

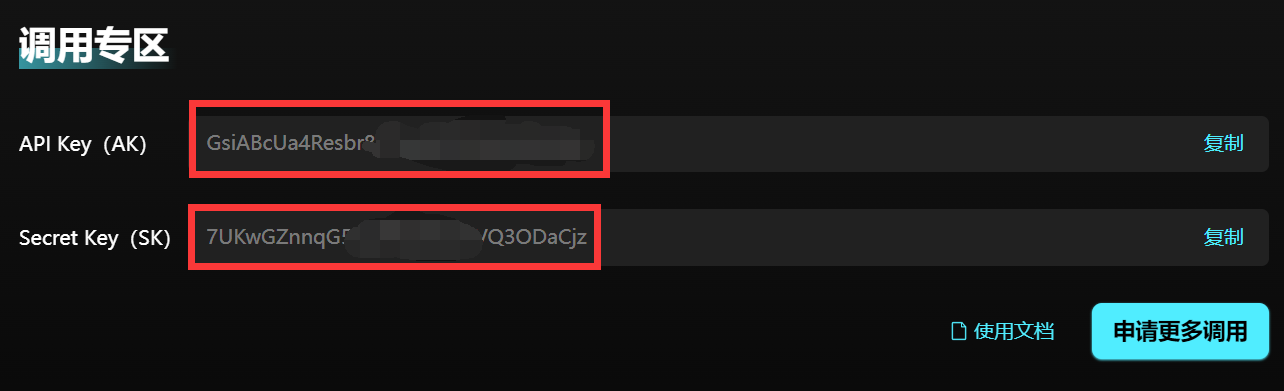
1. 在进行更改API时需要确认账户与密码对应，否则无法运行，需要对其进行覆盖或者重启应用程序。
2. 所有待处理文件等均需限制在1000字以内。
3. 所有待处理文本中不得含有侮辱性词汇，否则程序无法正常运行。
4. 使用OCR时注意上传照片需要清晰可见，行与行直接界限分明，可以兼容.jpg与.npg文件。对于大量手写文字内容，可以拍成多张照片后上传，识别时间随字数量的增大而增大。
5. 使用文本纠错时需上传.txt文件，为防止不必要报错，格式应设置为utf-8格式。
6. 目前文本纠错的输入标准文件为txt文件，若选择doc或docx文件可能会出现异常提示（后续会增添其他文件类型），但是纠错系统仍可以正常运行出正确结果。
7. 有时存在进度条卡顿的情况，请耐心等待一段时间。

# 使用案例

1. 在安装好对应库后，**直接运行texthandle即可打开我们的GUI界面，开始使用**；



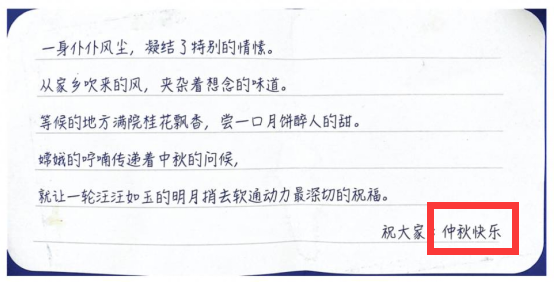
1. 登陆操作（非必要）
   1. 点击界面上方获取API选项即会自动跳转至文心大模型官网，将网页下拉至底部即可看到调用专区，注册后点击申请调用即可得到专属API识别码。



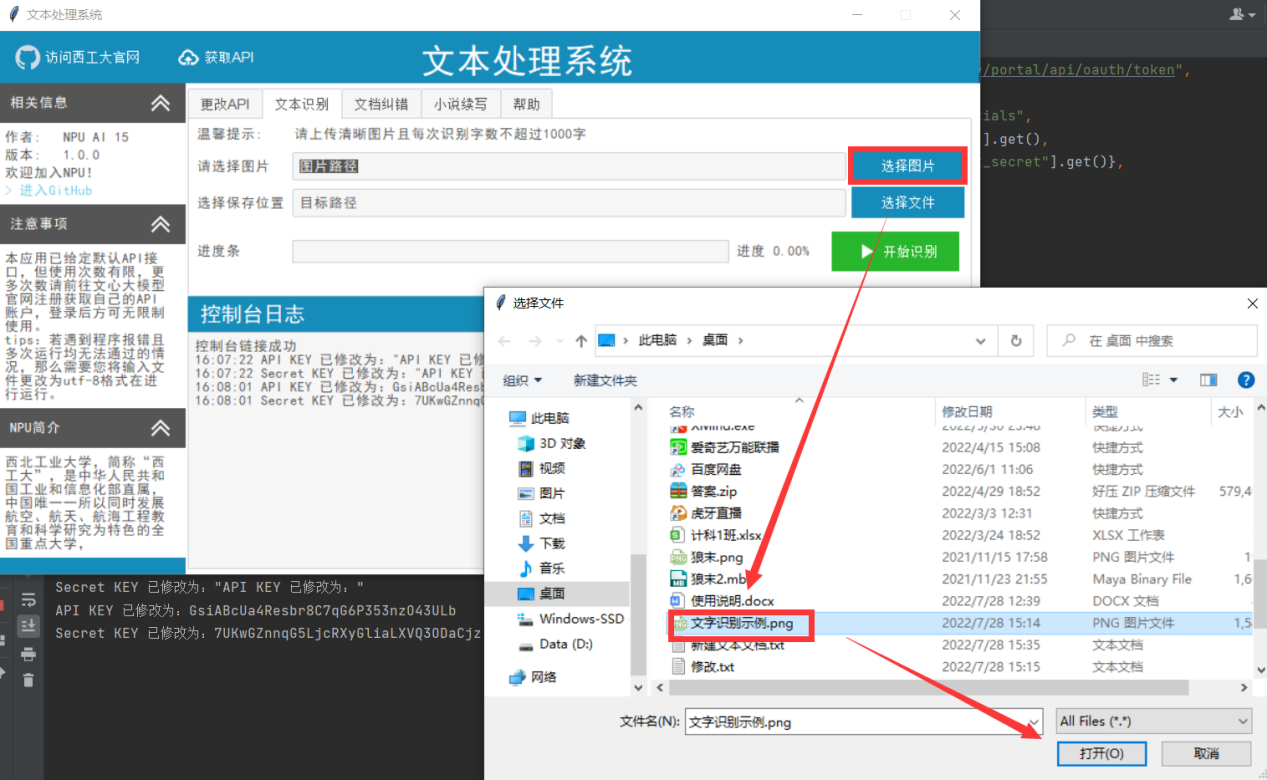
* 1. 将红框处所给API识别码输入进对应位置点击登录即可成功登录。（需注意，这里必须保证所输入API账户密码因对应正确，否则程序无法运行，若误输可重启应用或用正确的对其进行覆盖）



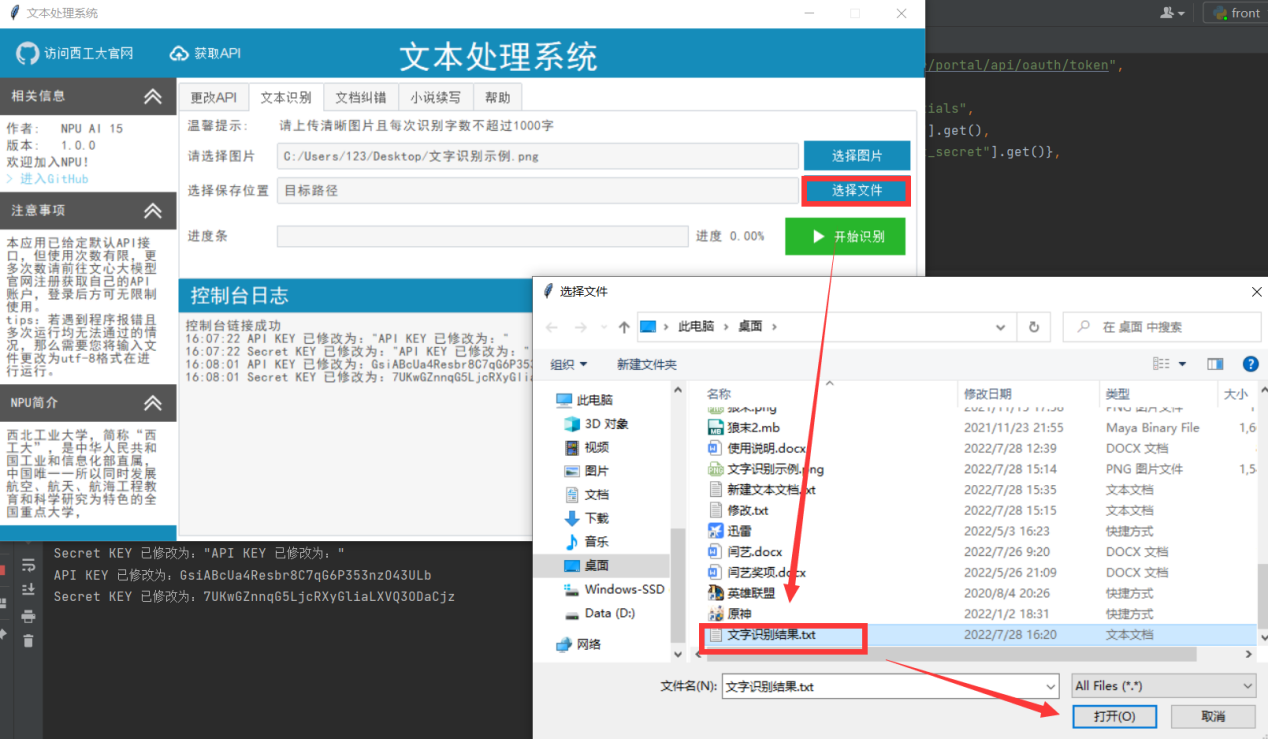
1. 由于其他功能操作类似，此处仅以文字识别为例介绍使用方法，所识别文本如下所示：



* 1. 如上图所示，是所识别图片中分为多行数据且其中有错别字，此选项的作用就是将所有文字识别为电子文件并将其中语法、字句错误更改。
  2. 首先准备好图片文件与保存目标文件；
  3. 然后点击选择图片填入所需识别的图片；



* 1. 点击选择文件选择保存位置



* 1. 之后点击开始识别按钮即可开始运行程序，待运行结束后结果便被保存在目标文件下了。



* 1. 如上图所示，手写文字已被识别，并且以正常保存至目标文件内。