

**基于阿里云的会计发票识别智能录入系统**

**基于阿里云的会计发票识别智能录入系统**

### 作品简介

本系统主要实现了对发票的分类、信息识别和整理导出，识别模块借助了阿里云文字识别中的发票识别云服务。用户使用时登录网址上传发票图片文件，系统进行自动识别，识别后的数据分类存储在Excel表中，用户可根据需求自行下载。

### 技术路线

**1.研究背景**

现实场景中，很多发票是机打纸质发票，在财务处理过程中往往需要将发票中显示的信息电子化。当前会计发票自动识别并录入的应用并不广泛，有的还停留在扫描成图片形式存档的阶段，所以有必要开发发票识别的工具。

需求分析：

人们需要将发票信息手动输入系统或识别成图片才能显示成电子化的形式，但是手动输入较麻烦，同时扫描不方便，所以，对于这一方面的需求是很大的。

核心需求：直接将发票识别出数据查看使用。

**2.技术手段**

利用阿里云发票识别接口，采用PHP页面传图片，使用Python解析数据。

**3.设计过程的具体步骤**

1. **了解发票的相关信息**

发票主要包含销售方企业信息、购方企业信息以及销售的商品信息、金额、税额、开票日期等数据。

**B.设计思路**

用户注册时会生成一个用户唯一的文件夹，登录时会生成一个session用于校验。上传需要识别的发票后，程序会根据发票提供的数据用Python解析识别后生成包含数据的Excel表再返给用户查看或下载使用。主要方法是首先用Python定义类接口，接受PHP发送的文件路径，便于后期存储数据；用户提供图片后会自动存储在生成的文件中，之后向服务器发送请求，将图片存储在文件中后对图片进行简单的类型解析，将不同的发票类型归类；最后就是转码的过程，将图片的格式由API转换识别后存储在列表中，同时调用API接口返回数据，解析返回的数据，转换为我们通常常见的数据格式。最后将解析好的数据写入Excel中，再将Excel表格存到用户对应的文件夹中，方便用户查看。

### 模块设计

* **Python部分**

定义一个类，Get\_Img\_Data：

1. 接收用户上传文件，定义函数get\_img\_file，用来接收用户上传的文件并读取。
2. 对图片格式进行压缩，定义函数decode\_img，把图片格式转化为api能够识别的格式。
3. 将格式化的图片传给api进行识别，定义函数posturl，读取数据保存到网页，并且拿到网页数据。
4. 解析网页数据，定义函数get\_data，对api解析出来的数据进行分析，拿取用户需要的数据。
5. 将数据保存到excel，定义函数to\_excel，将存在列表的数据传入excel
6. 定义run函数，调用上述函数完成一个完整的程序。
7. 在main函数里执行run函数并返回值让php接收。

* **网页部分**

定义getFiles，根据上传的是单张还是多张图片来构建上传文件信息。

1.用户注册，用户登录信息校验成功之后使用mkdir函数来为用户生成用户私人的文件夹。

2.登录时，用户登录数据库数据判断成功以后，使用session存储用户的基本信息。

3.进入请求进入invoice页面以后使用session中的信息和用户数据库中的进行比对，如果比对成功就进入invoice页面，反之就回到登录页。

4.图片判断，定义uploadFile用来判断单张图片的格式，大小，上传方式是否符合条件。

5.定义getUniName得到唯一的字符串，作为图片上传的文件的路径。

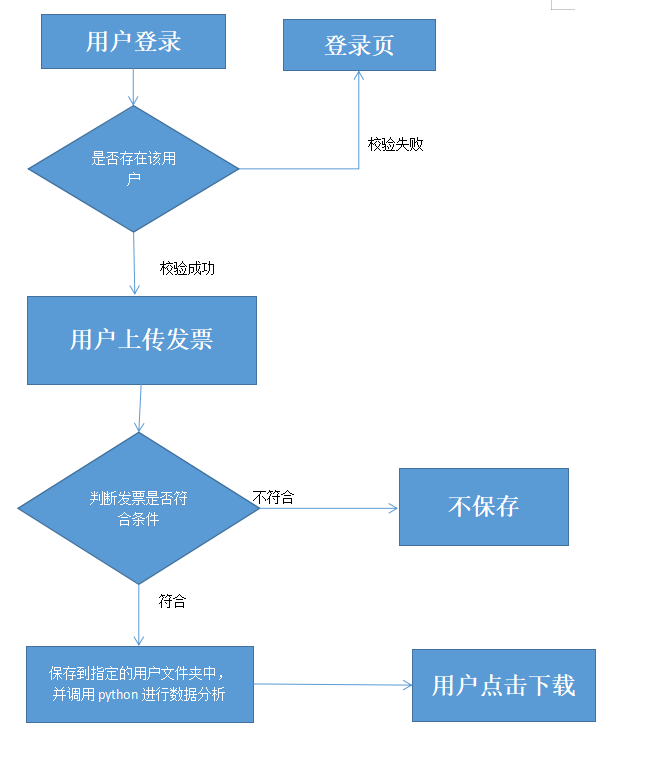
6.定义getExt函数，利用strtolower，pathinfo等函数得到文件的扩展名。

* **核心模块：**

Python中主要是图片压缩模块，第三个api解析模块和第四个数据分析模块。

网页部分主要是先调用getFiles函数获得文件信息，然后利用uploadFile来判断图片是否符合上传的标准，最后调用getExt，getUniName来参与上传完成模块。使用fileShow来获得用户之前上传发票进行识别生成的表格，并为用户提供下载。

* **程序流程**



### 实验结果或应用效果

实现了不同的用户自己选择图片上传，发票的识别和数据自动生成对应的Excel数据表格且可下载使用的一个功能。具体如下：

1.通过网址进入页面有选择文件，下载和上传文件界面则登录成功。

2.选择要识别的图片进行上传，当出现Excel表格时表示发票数据解析完成用户可以查看和下载，否则不能。

3.用户可以根据需要选择下载，下载完成后该系统功能使用结束。

该结果在实际应用中更加方便快捷，省去注册登录的环节，用户直接输入网址就可以使用，同时数据生成时会自动分类存放在对应的Excel表格中，方便用户查看数据和下载使用。

### 总结和展望

**总结：**

我们在做这个系统的时候遇到很多问题，如PHP传值Python时页面传值成功，但是启用失败；上传图片的遍历问题（一次只能上传一张，遍历上传图片时对图片格式大小进行判断，超过后就不能上传图片）等问题，但是最终通过查阅资料和向老师请教成功的解决了此类问题。

通过本系统的设计与实现，团队成员意识到，一个系统即便功能很少，但是编写程序时也要考虑很多实际问题，如果不能从用户角度去考虑是否可行，即使将程序做出来也没有实用价值。

同时，当考虑做某一系统时应该先构建大致框架，确定使用的语言及要实现的主要功能，进行适当的项目管理，避免进度拖沓。这次经历也使我们明白一个系统的完成大致需要做些什么，怎么做，组内成员又该怎么去合理分工才能高效率、高质量完成程序。

这是一次非常有趣、有益的学习经历。

**展望：**

在时间有限、技术有限的情况下，我们还是尽力完成了系统核心模块。这个系统只实现了发票数据识别解析的一个过程，虽然使用更加方便了，但还需要增加一些人性化的模块，如：

1.当用户上传的图片不能够满足条件时则返回相应的条件以及图片名称。

2.当Python文件在识别发票图片的时候页面提示等候，当识别成功的时候提示下载。把数据上传到对应的系统里面。

我们希望之后做类似的系统时，能够更加完善诸如此类的交互细节问题。同时希望可以再创新一些，结合系统测试结果，不断收集用户反馈，调整、完善、优化系统模块，做出让用户满意的产品。