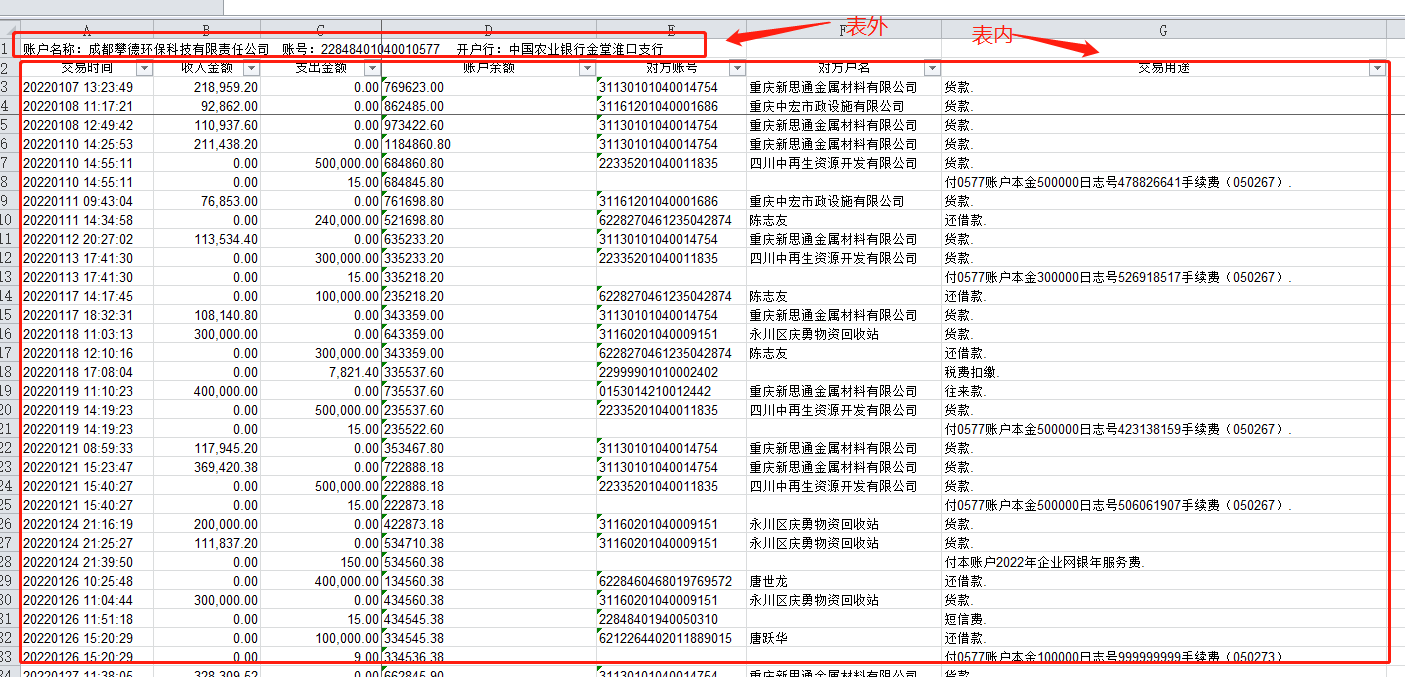
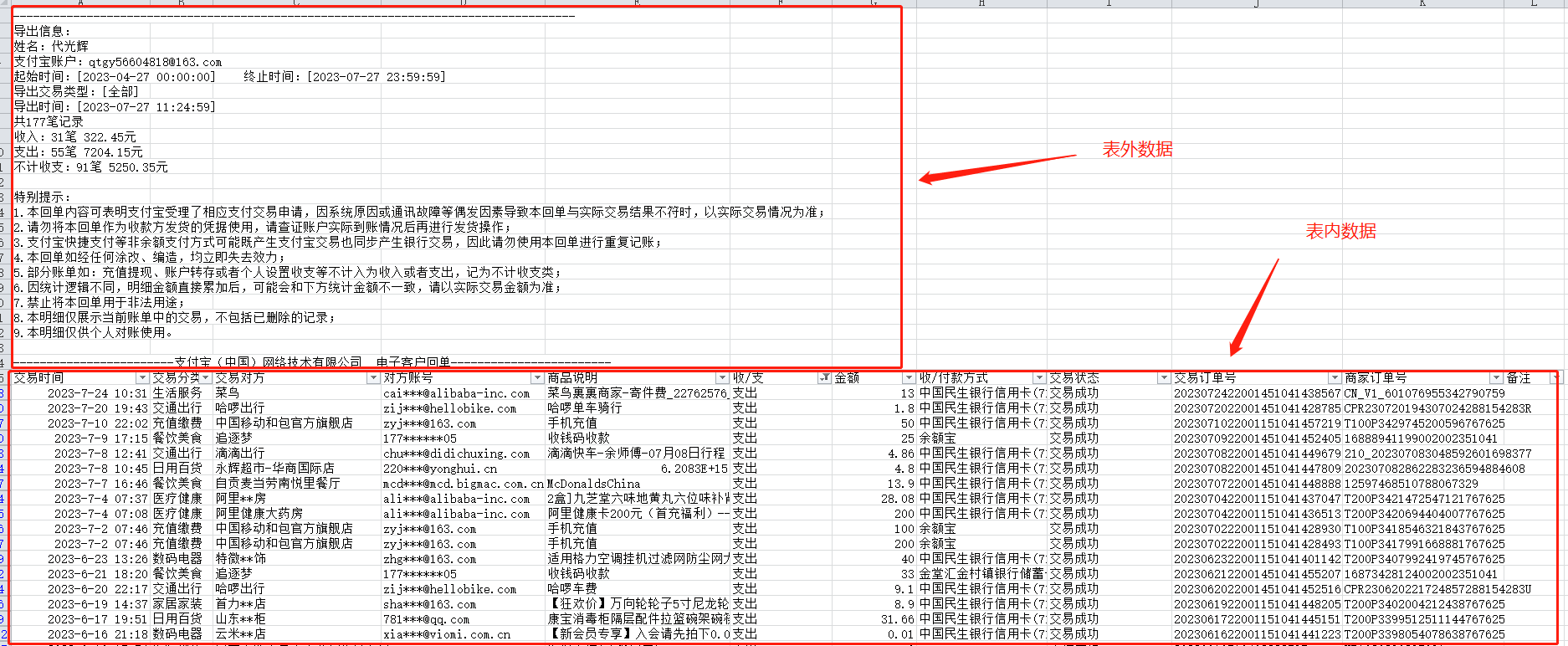
**需求：分析“银行流水数据”，从中筛选并提取指定信息**

**一. 银行流水数据样例：**





**二. 需要处理的数据：**

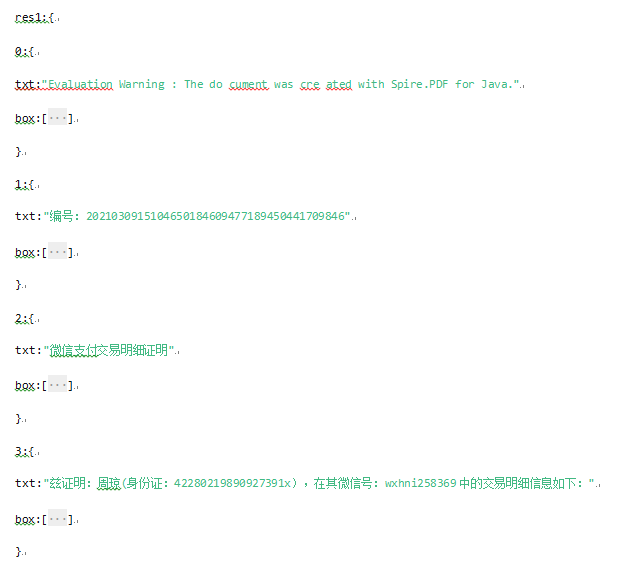
给到的数据data分为四部分，表外数据（res1）、表内数据（res2）和page\_sum、doc\_type四部分；



**1）doc\_type为文件类型，分为excel、epdf、image三种类型；不同类型体现在res1中是否包含坐标信息；**

**2）res1数据是指（表外数据），包含内容和坐标信息；（doc\_type值为image的是包含坐标信息的，excel和epdf的没有坐标信息；）0值是指res2前面的表外数据，1值是指res2后面的表外数据；**

**3）以下是res2数据（表内数据）的输出格式，包含每个单元格的行列信息和内容：**

**** 

**三. 需求说明**

**需求1. 分析res1数据，识别机构类型（agent\_type：bank/alipay/wechat），同时提取以下关键字段信息；输出格式参考第四部分的res1部分；**

1）agent\_type类型为bank的，输出以下字段信息

|  |  |  |
| --- | --- | --- |
| **字段** | **说明** | **输出内容** |
| agent\_type | 机构类型 | **bank** |
| bank\_name | 参考“金融机构信息对照表” | GY0001 |
| accout\_num | 银行账号号码 | 6228480329572520271 |
| account\_name | 账号的户名 | 李数 |
| date\_start | 起始日期 | 2022/05/22 |
| date\_end | 结束日期 | 2022/11/22 |
| account\_type | 账户类型 | 对私 |

2）type类型为alipay的，输出以下字段信息

|  |  |  |
| --- | --- | --- |
| **字段** | **说明** | **输出内容** |
| agent\_type | 机构类型 | alipay |
| accout\_num | 支付宝账号号码 | caiyingbiao@hotmail.com |
| account\_name | 支付宝账号姓名 | 李数 |
| idcard\_num | 用户身份证号码 | 330724197805240011 |
| date\_start | 起始日期 | 2022/05/22 |
| date\_end | 结束日期 | 2022/11/22 |
| account\_type | 账户类型 | 对私 |

3）type类型为wechat的，输出以下字段信息

|  |  |  |
| --- | --- | --- |
| **字段** | **说明** | **输出内容** |
| agent\_type | 机构类型 | wechat |
| accout\_num | 微信账号号码 | xshaitt |
| account\_name | 微信账号姓名 | 李数 |
| idcard\_num | 身份证号码 | 330724197805240011 |
| date\_start | 起始日期 | 2022/05/22 |
| date\_end | 结束日期 | 2022/11/22 |
| account\_type | 账户类型 | 对私 |

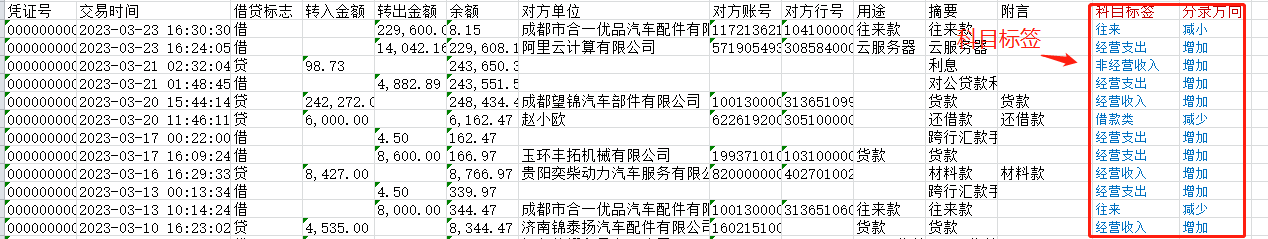
**需求2：分析res2数据，每行判断是否表头行，给每行非表头行的数据添加“科目标签”和“分录方向”；**

**输出格式参考第四部分的res2；**

**1）检测每行数据是否是表头行/** **header\_row，显示”true”或”false”**

**2）检测每行数据（表头行除外）相应的“科目标签/accounting\_item”和“分录方向/accouting\_entry”， 五类科目下面一共划分为18个科目标签，如下：**

**3）检测每行数据（表头行除外）的交易对象属于什么机构属性/trader\_nature；**





**A、对私账户的费用的科目标签：**

查看附件《科目标签&居民消费支出分类》

**B、对公账户的费用的科目标签：**

查看附件《科目标签&居民消费支出分类》

**需求3：分析res2数据，固定以下字段数据，以规范格式输出；输出格式参考第四部分的res3；**

**1）agent\_type为bank，输出以下字段数据；（备注如果有多个，就输出多个）**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **交易日期** | **交易金额** | **收支类型** | **余额** | **备注（摘要）** | **对方账号** | **对方户名** |
|  | 2022/03/15 | 3571.42 | 收入 | 3571.42 | 跨行转账（文本） | 622841……8978（文本） | 张\*\*（文本） |
|  | 2022/03/16 | 100.00 | 支出 | 3471.42 | 支付宝（文本） | 2155\*\*690（文本） | 支付宝网络技术有限公司（文本） |

**2）agent\_type为wechat，输出以下字段数据（wechat的格式也不一定是不变的，根据原数据显示列，针对日期和金额格式做规范化处理）**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **交易日期** | **交易类型** | **收/支/其他** | **交易方式** | **交易金额** | **交易对方** |
|  | 2022/03/15（yyyy/mm/dd） | 商户消费（文本） | 支出（文本） | 中国银行（6633）（文本） | 3571.42（数值） | 王\*\* |
|  | 2022/03/16（yyyy/mm/dd） | 微信红包（文本） | 支出（文本） | 零钱（文本） | 3471.42（数值） | 发出群红包 |
|  | 2022/03/16（yyyy/mm/dd） | 二维码收款（文本） | 收入（文本） | / | 1500.00（数值） | 曾\*\* |

**3）agent\_type为alipay，输出以下字段数据（alipay的格式也不一定是不变的，根据原数据显示列，针对日期和金额格式做规范化处理）**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **收/支** | **交易对方** | **商品说明** | **收/付款方式** | **金额** | **交易日期** |
| 其他 | 天弘基金管理有限公司（文本） | 转账收款到余额宝（文本） | 余额宝（文本） | 1.77（数值） | 2022/03/16（yyyy/mm/dd） |
| 收入 | \*平（文本） | 转账（文本） | 余额宝（文本） | 500.77（数值） | 2022/03/16（yyyy/mm/dd） |
| 支出 | \*明（文本） | 收款（文本） | 中国建设银行(3955)（文本） | 1500.00（数值） | 2022/03/16（yyyy/mm/dd） |
| 支出 | 滴滴出行（文本） | 滴滴快车-王师傅（文本） | 花呗（文本） | 12.8（数值） | 2022/03/16（yyyy/mm/dd） |

**四、输出格式字段说明**

|  |  |
| --- | --- |
| **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** |
| 400 | Success | 执行成功 |
| 401 | Type error | 文件格式不符合要求 |
| -1 | unknow | 未知错误 |

**data说明**

|  |  |  |
| --- | --- | --- |
| **参数名** | **说明** | **内容** |
| page\_sum | 总共页数 | 001 |
| doc\_type | 数据来源的文件格式 | Image、excel、epdf |
| res1 | 表外信息的数据提取 |  |
| res2 | 表内每行数据的以及科目标签 |  |
| res3 | 表内数据的指定字段信息提取 |  |

{“code”：“200”,

“message”:”success”,

“data”：{

page\_sum:”001”

doc\_type：”image”

**res1:**[

agent\_type：”bank”,

bank\_name：” GY0001”,

accout\_num：” 6228480329572520271”,

account\_name：” 李数”,

account\_type：” 对私”,

]

**res2:** [

0:{

row\_order:"001\_001"

header\_row:”true”

accounting\_item:””

accouting\_entry: ""

trader\_nature:””

1.1:"交易单号"

1.2:"交易时间"

1.3:"交易类型"

1.4:"收/支/其他"

1.5:"交易方式"

1.6:"金额(元)"

1.7:"交易对方"

1.8:"商户单号"

}

1:{

row\_order:"001\_002"

header\_row:false

accounting\_item: "往来"

accouting\_entry: "增加"

trader\_nature:”个人”

2.1:"1000050001202103090508p29889237"

2.2:"p021-03-0g15:46:30"

2.3:"转账"

2.4:"收入"

2.5:"7"

2.6:"15000.00"

2.7:"胡美树"

2.8:""

}

2:{

row\_order:"001\_003"

header\_row:false

accounting\_item: "往来"

accouting\_entry: "减少"

trader\_nature:”personality”

3.1:"000050001210309000543 2083786447849"

3.2:"2021-03-0g 15:44:21"

3.3:"转账"

3.4:"支出"

3.5:"湖北农信（(73"

3.6:"15000.00"

3.7:"胡美树"

3.8:"1000050001202103 p91503090780922"

}

3:{

row\_order:"001\_004"

header\_row:false

accounting\_item: "消费支出"

accouting\_entry: "增加"

trader\_nature:”其他商家”

4.1:"000107101210309000603"

4.2:"2021-03-09"

4.3:"扫二维码付"

4.4:"支出"

4.5:"54) 零钱"

4.6:"40.00"

4.7:"从头再来"

4.8:"1000107101202103"

}

]

**res3:[**

0:{

row\_order:"001\_001"

header\_row:true

1.1:"交易日期"

1.2:"交易金额"

1.3:"收支类型"

1.4:"余额"

1.5:"备注（摘要）"

1.6:"对方账号"

1.7:"对方户名"

}

1:{

row\_order:"001\_002"

header\_row:false

2.1:" 2021/03/15"

2.2:"3.00"

2.3:"收入"

2.4:"300.00"

2.5:"充值话费"

2.6:"3326451512"

2.7:"胡美树"

}

2:{

row\_order:"001\_003"

header\_row:false

3.1:"000050001210309000543 2083786447849"

3.2:"2021-03-0g 15:44:21"

3.3:"转账"

3.4:"支出"

3.5:"湖北农信（(73"

3.6:"15000.00"

3.7:"胡美树"

}

3:{

row\_order:"001\_004"

header\_row:false

4.1:"000107101210309000603"

4.2:"2021-03-09"

4.3:"扫二维码付"

4.4:"支出"

4.5:"54) 零钱"

4.6:"40.00"

4.7:"从头再来"

}

]

}

**五、交付内容**

1、源代码

2、镜像文件

3、部署文档（源码、镜像的部署文档&接口文档）