Word文档输出json格式

详细设计说明书

亿阳信通股份有限公司

2015年8月18日

# 目录

[目录 2](#_Toc502652462)

[1. 模块设计 3](#_Toc502652463)

[1.1. Word数字化子系统 3](#_Toc502652464)

[1.1.1. 数据库表结构的数字化 3](#_Toc502652465)

[1.1.2. 详细设计的数字化 16](#_Toc502652466)

# 模块设计

## Word数字化子系统

### 数据库表结构的数字化

#### 功能描述

##### 功能说明

实现对数据库表结构数字化文档的读取，将其编写成json格式，并输出为json文件的功能。

##### 调用格式

导入生成的模块，以模块名.方法名（）形式调用

#### 业务文件

##### 业务文件

底层资产

##### **1.1.1.2.2 业务文件格式说明**



#### 外部模块

安装Python-docx包，导入from docx import Document 模块

导入import re正则表达式模块

导入import json模块

#### 处理过程

1. **导入所需模块，创建document对象**
2. **定义一个fun函数，传入document**
3. **对标题格式的判断：**
4. 创建不同标题格式的空列表style\_list，Head2，Head3；
5. 获取文档的段落内容
6. 遍历获取的Word文档内容；
7. 根据模块内提供的方法paras.style.name（），判断输出的内容的标题级数，添加到相对应的列表；
8. **对文档内容的处理：**
9. 建立文档内容的空列表para\_list，遍历传入的Word文档的段落，并将内容添加到列表；
10. 遍历列表的内容，利用正则表达式，提取出固定字段（说明，表空间，主键，索引，字段）；

**5. 对文档表格的处理：**

1） 根据模块内提供的方法，提取表格的固定字段（列名，中文名称，类型，长度，可空，备注）；

2） 建立所有表格内容的空列表table\_list，遍历所有表格内容，将其添加到列表；

1. 建立固定字段以外的表格内容的空列表table\_content，根据列表切片方法，依次取出固定字段下对应的内容，将其添加到列表；
2. 建立表格json对象的空列表Table\_list；
3. 循环遍历table\_content内容，将固定字段和表格内容构造成json键值对形式，并添加到表格json对象的列表里面；

**6. 构造json对象：**

将全部Word文档内容构造成json键值对格式，把前面的处理结果，放入json字符串里面

**7. 保存json文件：**

将json对象存为json格式的文件

1. **调用fun函数**

#### 异常处理

如果传入文件格式不正确或文件不存在，则不会执行代码功能，需重新传入正确格式文件。

### 详细设计的数字化

#### 功能描述

##### 功能说明

实现对详细设计数字化的文档的读取，将其编写成json格式，并输出为json文件的功能。

##### 调用格式

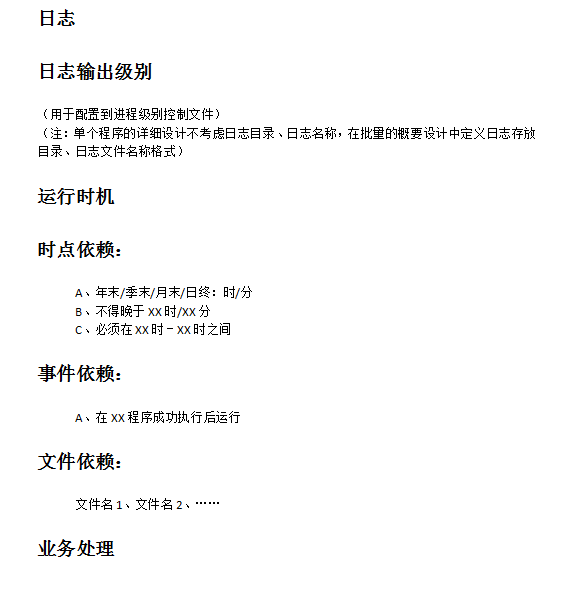
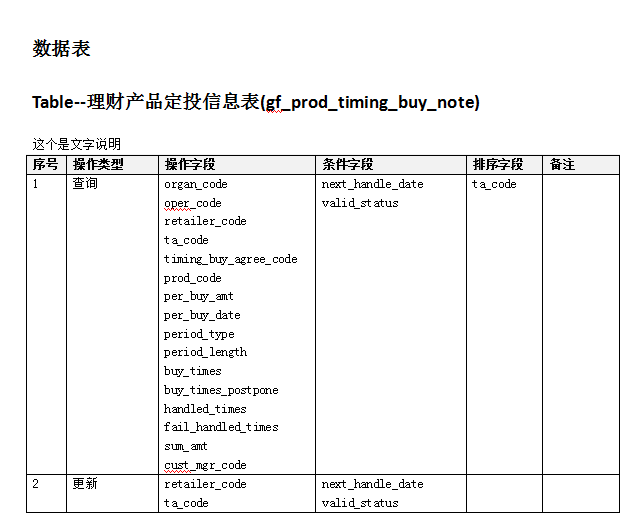
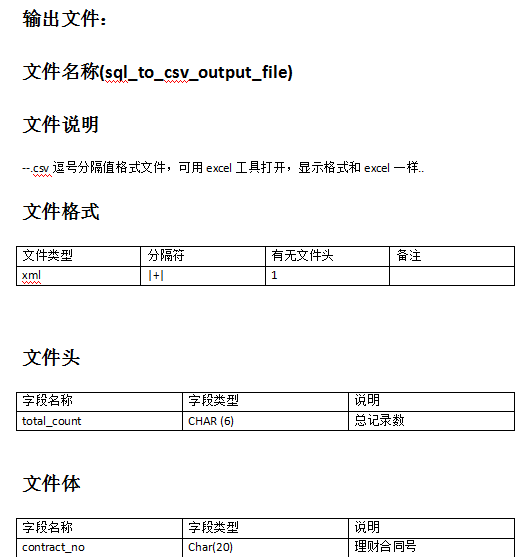
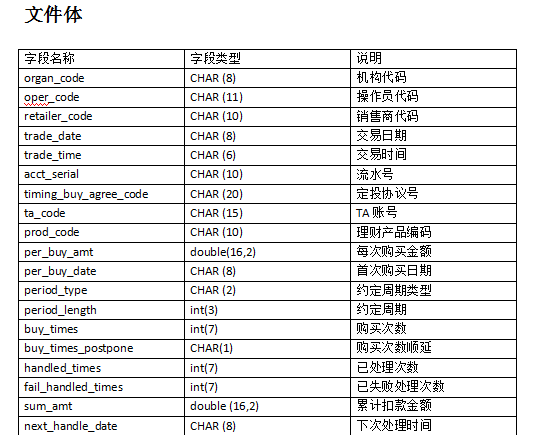
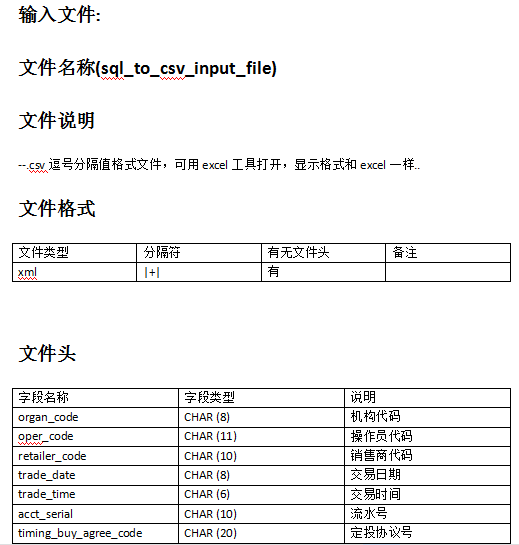
导入生成的模块，以模块名.方法名（）形式调用

#### 业务文件

##### 业务文件

示例详细设计

##### **1.1.1.2.2 业务文件格式说明**



#### 外部模块

安装Python-docx包，导入from docx import Document 模块

导入import re正则表达式模块

导入import json模块

#### 处理过程

**1 导入所需模块，创建document对象**

**2 定义一个fun函数，传入document**

**3 对标题格式的判断：**

1） 创建不同标题格式的空列表Head1，Head2，Head3；

2） 获取文档的段落内容；

3） 遍历获取的Word文档内容；

4） 根据模块内提供的方法paras.style.name（），判断输出的内容的标题级数，添加到相对应的列表；

1. **对文档内容的处理：**
2. 建立文档内容的空列表para\_list，遍历传入的Word文档的段落，并将内容添加到列表；
3. 根据列表切片的方法，依次取出固定字段下对应的内容；

**5. 对文档表格的处理：**

1） 根据模块内提供的方法，提取表格的固定字段；

2） 建立所有表格内容的空列表table\_list，遍历所有表格内容，将其添加到列表；

1. 建立固定字段以外的表格内容的空列表table\_content，根据列表切片方法，依次取出固定字段下对应的内容，将其添加到列表；
2. 建立表格json对象的空列表Table\_list；
3. 循环遍历table\_content内容，将固定字段和表格内容构造成json键值对形式，并添加到表格json对象的列表里面；

**6. 构造json对象：**

将全部Word文档内容构造成json键值对格式，把前面的处理结果，放入json字符串里面

**7. 保存json文件：**

将json对象存为json格式的文件

1. **调用fun函数**

#### 异常处理

如果传入文件格式不正确或文件不存在，则不会执行代码功能，需重新传入正确格式文件。