**查重 标准流程文**

**文件版本历史**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **文件版本** | **修订日期** | **修订人** | **审核人** | **批准人** | **修订说明** |
| V1.0 | 2022/05/31 | Miang、Chen |  |  | 初版 |
| V2.0 | 2024/10/22 | Nolan |  |  | 2024-10修订 |

目录

[1. 核心词汇注释 1](#_Toc19957)

[1.1. OE号 1](#_Toc26149)

[1.2. SKU 1](#_Toc11160)

[1.3. Order Template 1](#_Toc7974)

[2. 查重 2](#_Toc28406)

[2.1. 查重介绍 2](#_Toc23949)

[2.1.1. 查重目的背景及组成 2](#_Toc8036)

[2.1.2. 查重流程 2](#_Toc31222)

[2.1.3. 查重工具 3](#_Toc28458)

[2.2. 新品目录查重 3](#_Toc4420)

[2.2.1. Excel查重 3](#_Toc13449)

[2.2.2. Python查重 8](#_Toc16878)

[2.3. Order Template查重 11](#_Toc14421)

[2.3.1. Excel查重 11](#_Toc11470)

[2.3.2. Python查重 12](#_Toc28169)

[3. 《已查重》文档规范 12](#_Toc7506)

[3.1. 已查重文档内容规范 13](#_Toc2487)

[3.2. 已查重文档格式规范 13](#_Toc7226)

# 核心词汇注释

## OE号

OE号是指汽车生产厂为方便对配件进行管理而采用的编号。OE在汽车制造业发达的国家，每个汽车生产厂为方便对零部件进行管理，对每种车型的每个零部件都采用不同的编号来区别分类，这个编号就是OE编号。

每个OE厂商都有不同的编号规则。比如：F1DU-10300-AD是FORD福特的编号，0-124-325-003是BOSCH博世的编号，10479923是DELCO德科雷米的编号。

每个OE编号对应唯一的产品，但同一个产品，因为装配在不同的车型上，从而有可能有几个不同的OE编号，通过OE编号可以查询出该产品的类型（发电机、起动机还是电机的零配件），具体的性能和详细参数，也可以知道该电机应用的具体汽车品牌、车型系列、出厂年代，以及对应的发动机的具体型号。

## SKU

SKU全称为Stock Keeping Unit（库存量单位），即库存进出计量的基本单元，可以是以件，盒，托盘等为单位。当下已经被我们引申为产品统一编号的简称，每种产品均对应有唯一的SKU号。

## Order Template

OT的全称为Order Template，实际含义为订单模板；在订单模板中，我司加入风险控制因素，市场分析因素，构成通用的模板；因此我司称OT为可行性分析，可行性分析是要求以全面、系统的分析为主要方法，经济效益为核心，围绕影响项目的各种因素，运用大量的数据资料论证拟建项目是否可行。对整个可行性研究提出综合分析评价，指出优缺点和建议。为了结论的需要，往往还需要加上一些附件，如市场链接、OE信息等。

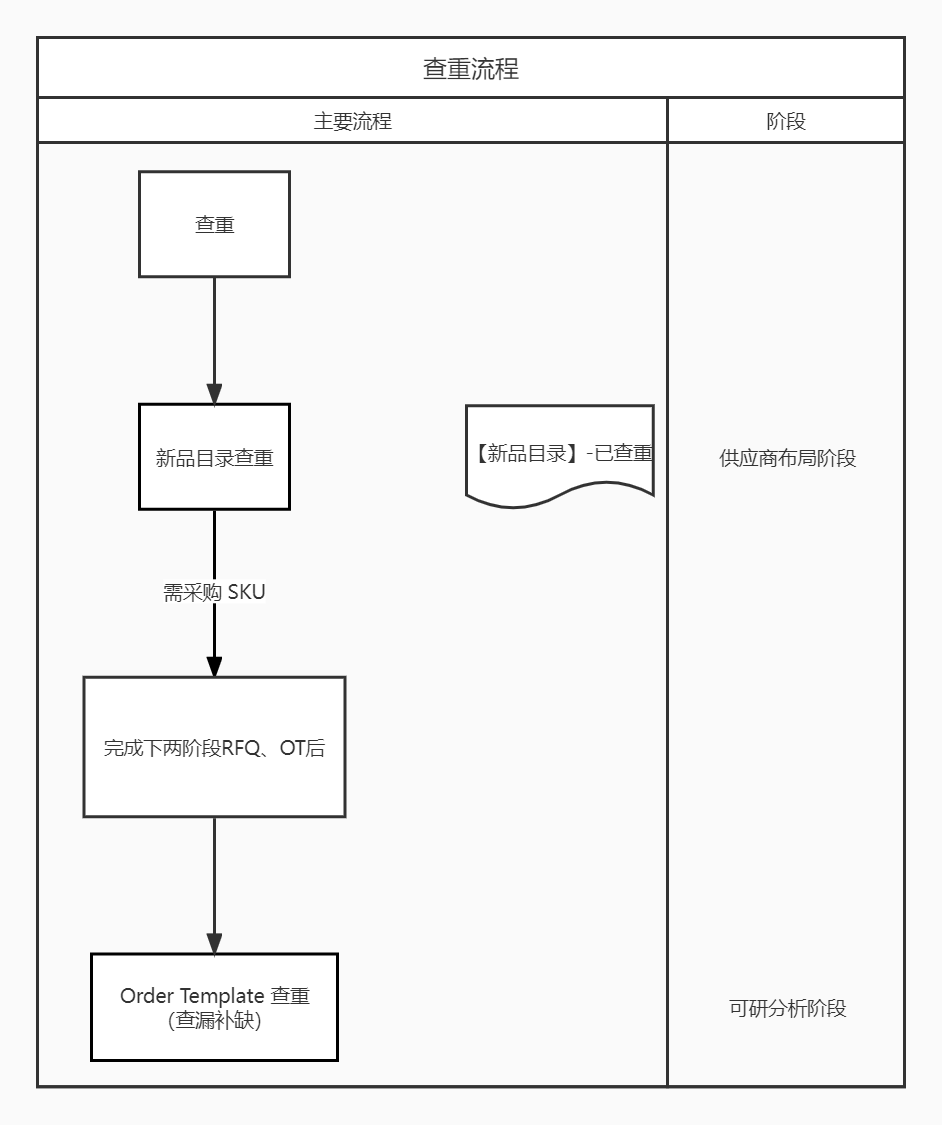
# 查重

## 查重介绍

### 查重目的背景及组成

为避免产品重复开发，需要对产品进行查重检验，过滤掉已开发的产品，从而减少后期不必要的人力及费用产生，目前查重主要分为两个部分：①前期工厂目录初步查重；②完成OT阶段后二次查重。

### 查重流程



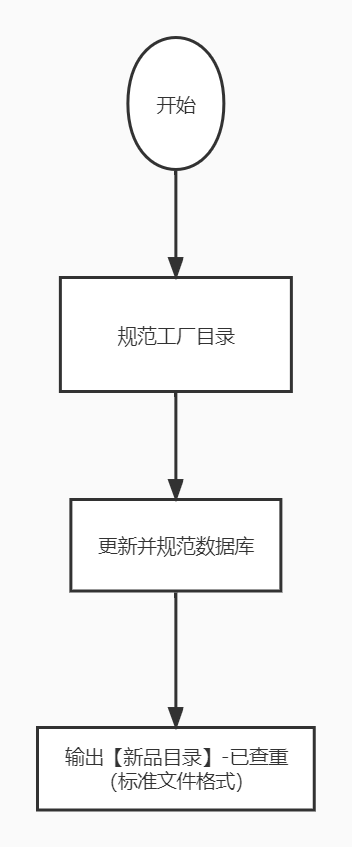
### 查重工具

当前运用的查重工具主要是：Excel、Python、exe查重工具

## 新品目录查重

### **Excel查重**

#### 总体流程



#### 操作流程

##### 规范工厂目录

###### **规范标准**

* OE号间需由英文分号分隔“;”，不得有句号、空格、换行符、逗号、中文分号、斜杠号...等进行分隔。
* OE号中不得带有品牌字样、中文字样，以及各种特殊符号。

###### **目录OE规范步骤**

1. 提取主要字段

系统导出来的数据字段较多，一部分字段我们并不需要用到，为了使数据源看起来更加简洁美观，对我们来说只需要提取如下字段的基本信息，其中OEM.列就是我们需要规范的OE号。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **SKU** | **商品配货名** | **是否采购** | **eBay关键字** | **备注** | **供应商编号** | **OEM.** |
|  |  |  |  |  |  |  |

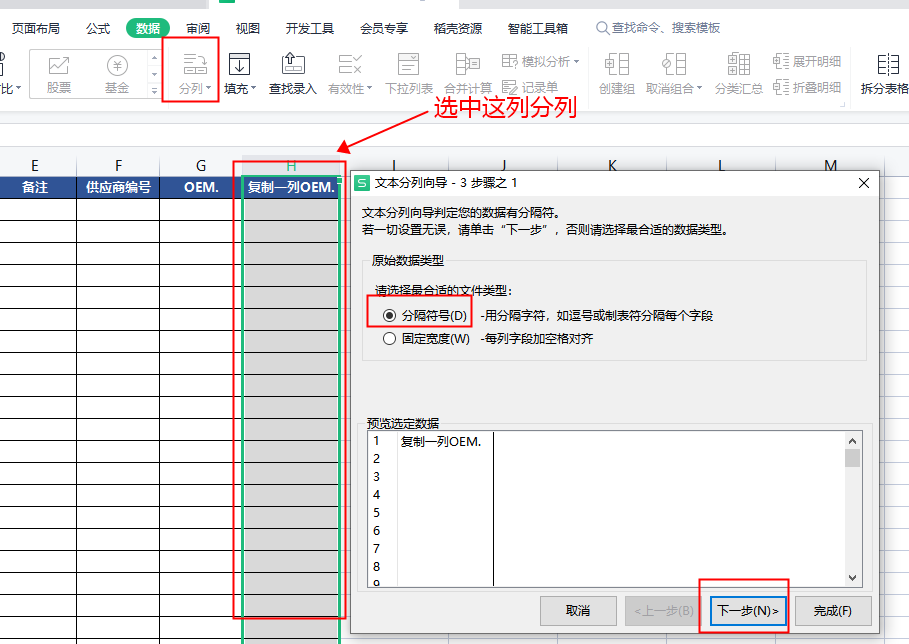
1. 替换

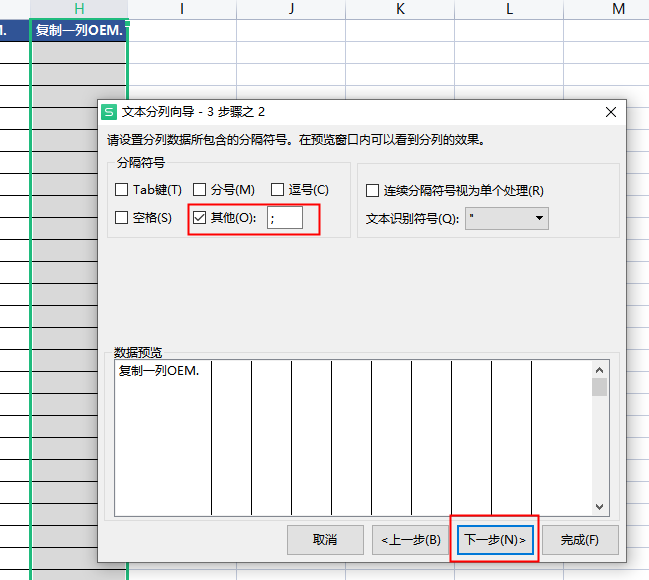
* 若OE中含有换行符、句号、空格、逗号、中文分号、斜杠号等，均替换为“;”。

|  |  |  |  |
| --- | --- | --- | --- |
| **替换前** | **替换后** | **替换操作步骤** | **案例** |
| 换行符（^l） | ; | Ctrl+F，弹出小窗口  ↓  输入被替换的字符串  ↓  输入英文分号";"  ↓  点击全部替换 | 将换行间隔替换为“;”    Tips：换行符可由Ctrl+F→替换→特殊内容→换行符直接找到,如下： |
| 句号（。） |
| 空格（ ） |
| 逗号（，） |
| 中文分号（；） |
| 双分号（;;） |
| 顿号（、） |
| 斜杠号（/） |
| 双斜杠号（//） |
| ... |

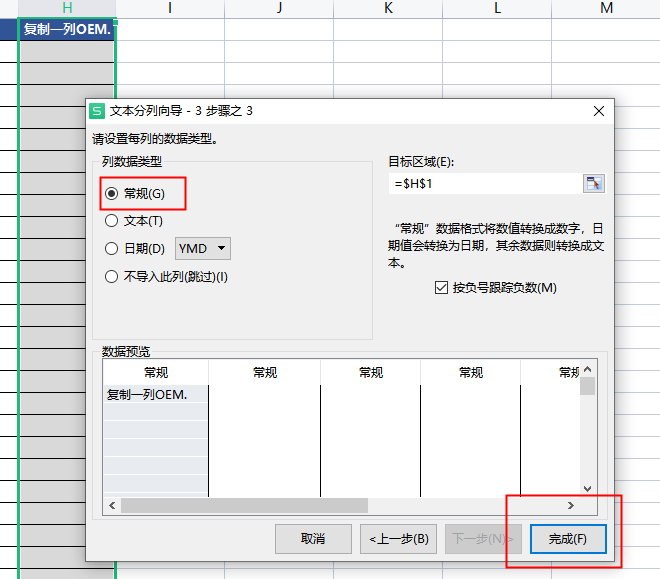
* 若OE号中带有品牌，例如WAI等，或夹杂中文，需要手动删除。
* 若OE号中含有特殊字符，例如星号\*等，需要去市场查一下，是否属于该品牌的OE书写规范，若不属于，需手动删除。
* 若OE号中有带“-”号的，属于不同品牌的OE号书写规范，可以先保留一列，在后面追加一列不带“-”的，也可只保留一列不带“-”号的。

1. 分列

* 先复制一列最终规范后的OE，再对复制的那列进行分列；
* 点击工具栏数据框，点击分列，弹出窗口；选择分隔符号，点击下一步。
* 选中其他，并输入分号“;”，点击下一步。



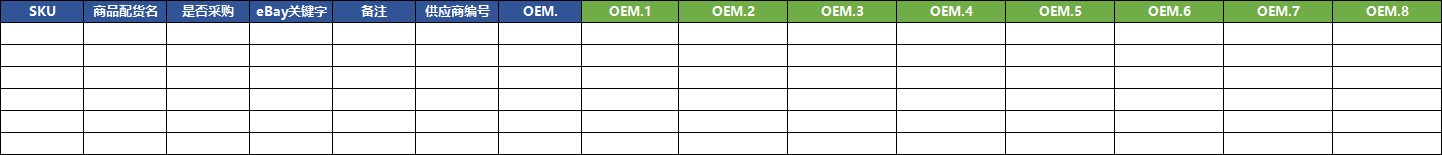
* 选择常规/文本（分列字段格式需和下一步骤数据库分列字段格式一致)，点击完成。



* 分列后，再次检查每一列OE是否都规范完成。

###### **输出规范后OE结果**

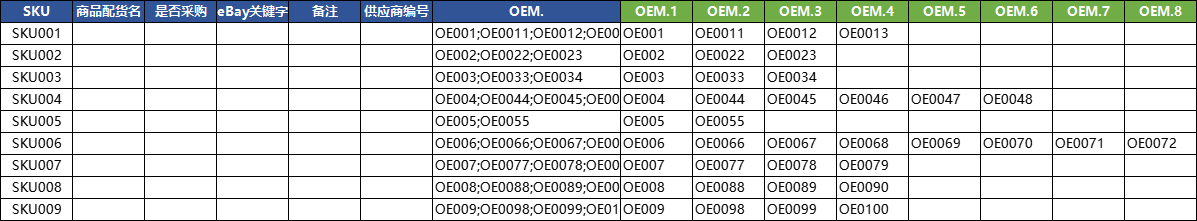
工厂目录OE规范完成后，输出如下图结果：



###### **目录规范后标准格式**

* Title需包含：序号、工厂、工厂号、是否采购、对应OE、查找结果、查找值、OEM.
* 部分工厂可能未提供工厂号情况，可空缺
* 对应OE需填入工厂所提供的完整OE字段
* OEM.为工厂所提供分裂后OE字段，根据OE个数进行调整

文档格式如下图：



##### 更新并规范数据库

###### 规范目的

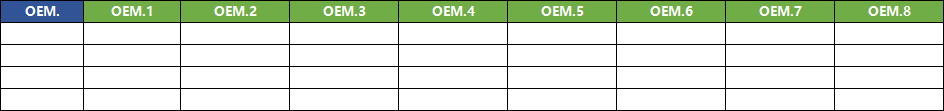
数据库是整个查重的判断基础，完整的数据库对于查重的有效率有着重要作用，能大大降低二次作业以及时间成本。因此，对于系统导出的OE号，也需要进行规范处理，其OE需同工厂一样将OE规范为标准格式。

###### **规范步骤**

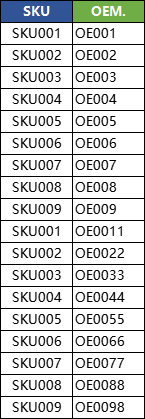
OE处理参考2.2.1.2.1.2.目录OE规范步骤。

###### **最终输出呈现**

系统导出来的OE规范完成后，会输出一份大致如下图的框架：



将上图分列后的横向OE，调整为如下图所示纵向OE格式：



###### **注意事项**

* 分品类建立数据库，防止数据混乱。
* 数据库保持更新，只要前期先处理一次系统导出来的OE，后期只要把新开发的OE更新到后面，不需要每次查重都处理一遍所有的OE。
* OE在进行分列后，部分以“0”开头的OE号可能会被省略“0”，以及OE号中间带“E”的会变成科学计数法，两者均不影响最终的查重结果。

##### 整合数据

1. 新建Excel文档，并命名为【新品目录】-已查重
2. 将2.2.1.2.1.4.目录规范后标准格式数据贴入Sheet1，重命名为“新品目录”
3. 将2.2.1.2.2.3.最终输出呈现数据贴入Sheet2，重命名为“数据源”
4. 在新品目录表“查找结果”列输入公式“=Vlookup（查找值，数据源表，2，0）”并下拉
5. 将“OEM1.”数据复制到“查找值”，筛选“#N/A”项，将“OEM.2”数据复制到“查找值”，粘贴值到可见单元格，若有“OEM3.”、“OEM4.”...等数据，重复上述操作，直至全部“OEMn.”
6. 筛选“查找值”中“#N/A”项，在“是否采购”中填入“否”，然后清除“查重值”中“#N/A”项，将“是否采购”的空白项填为“是”。（过滤已开发的SKU）
7. 全选表格，将表格中的数据全部粘贴为值

### Python查重

#### **现实意义**

当数据体量较大时，excel人工处理的步骤较多，在规范OE的过程中可能会存在一些操作漏洞，而python的逻辑编写可以涵盖目前可能遇到的所有OE规范情况，通过Python程序更好的实现办公自动化，减少人工操作的疏忽，提高查重的准确性。

#### **操作流程**

##### 工厂目录统一模板

不同工厂提供的目录格式可能不唯一，首先要把各式各样的目录转化为我们如下唯一固定的模板，并将文件命名为《工厂新品目录模板》。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **序号** | **工厂** | **工厂号** | **完整OE** | **是否采购** | **QTY** | **含税价格** | **备注** |
|  |  |  |  |  |  |  |  |

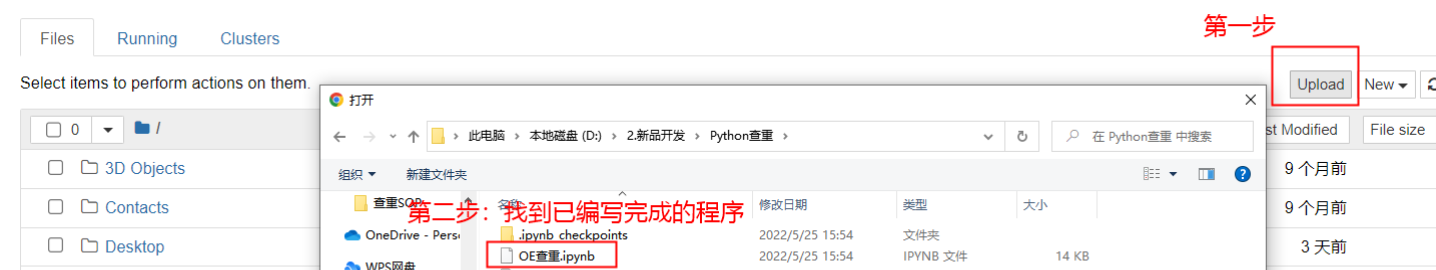
##### 新建输出文件夹

新建一个文件夹，命名“查重输出”，专门用于存放程序输出的文件。（程序将会输出一个文件，名如“【查重输出】\_20240709\_171413”）



##### 上传程序

将写好的python程序上传到Anaconda Navigator上：打开Anaconda Navigator→点击Update，找到已编写完成的程序，点击打开，程序上传完毕，如下图。



##### 更改路径

每台电脑存放文件的路径都不唯一，因此程序里的文件路径不适用于每个人，所以拿到程序后，需要把程序里涉及到的文件路径都改成个人相对应存放的文件路径。

* 打开程序，更改第一个路径：【查重模板】的保存路径；



* 更改第二个路径，即输出文件【查重输出】；

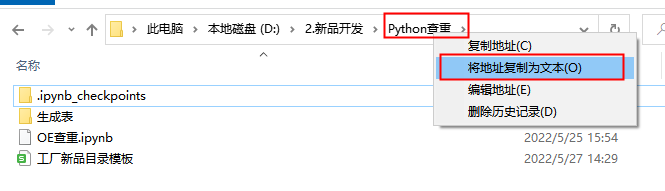


##### 注意事项

* 建议专门建一个文件夹用于存放查重相关的文件，便于文件归类。
* 程序里文件路径的组成：文件地址\文件名称。



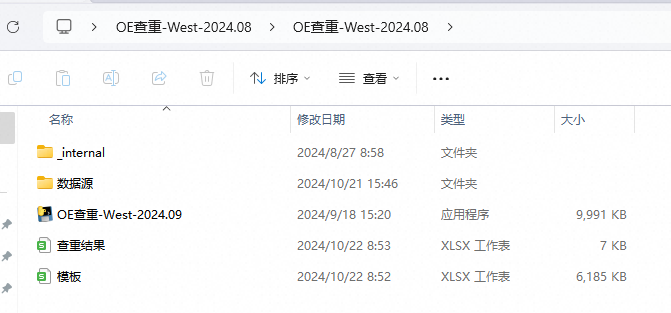
* 复制文件地址的方法：找到需要的文件→点击上方长条框最后一个文件夹→点击鼠标右键→点击“将地址复制为文本”。



### exe工具查重

利用exe查重工具进行便捷查重。exe工具获取路径：

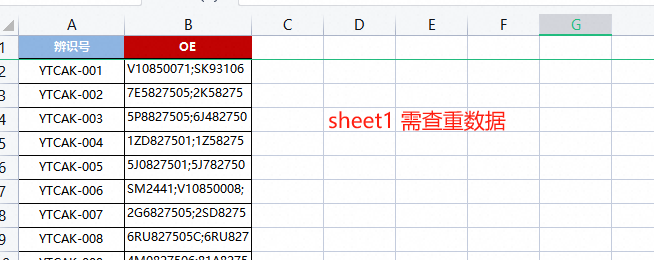
#### **操作流程**

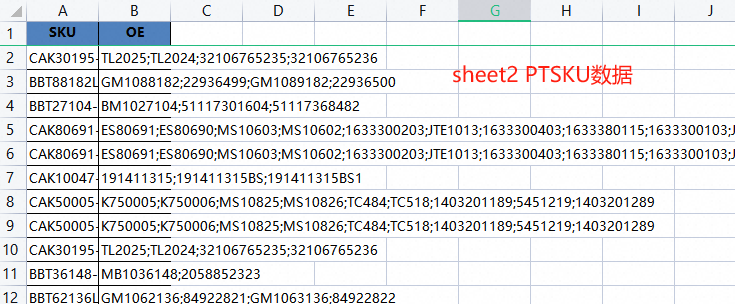


* OE查重工具文件夹内共有五部分，其中“数据源”文件夹内保存全量“PTSKU数据”或“查重用OE”数据，相关数据每周更新。

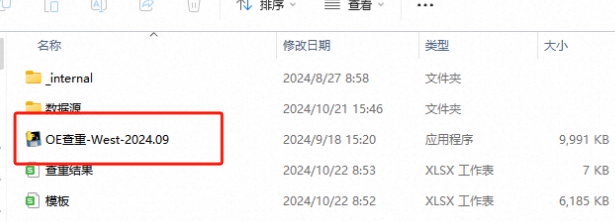
##### 数据输入和输出

* “模板”和“查重结果”两个Excel文件分别对应“需查重数据”和已查重的“数据输出”，打开“模板”文件，共有两张sheet表。如下图。
  + 第一张表为需查重数据，将需要查重的数据贴入对应位置，第一列“辨识号”不能为重复数据，方便区分；第二列“OE”数据为需要查重的数据，如已命名SKU，可将SKU作为辨识号输入，如还未命名SKU，则可自主取名编号作为辨识号如“YTDCR-001”、“YTDCR-002”…下拉。
  + 第二张表为PTSKU，如已确定查重的品类，将对应品类的全量PTSKU数据放入“PTSKU”的sheet中；如为新品查重或同时有多个品类查重，则可将全量PTSKU或多个已知品类的SKU和OE数据一并放入“PTSKU”的sheet中。

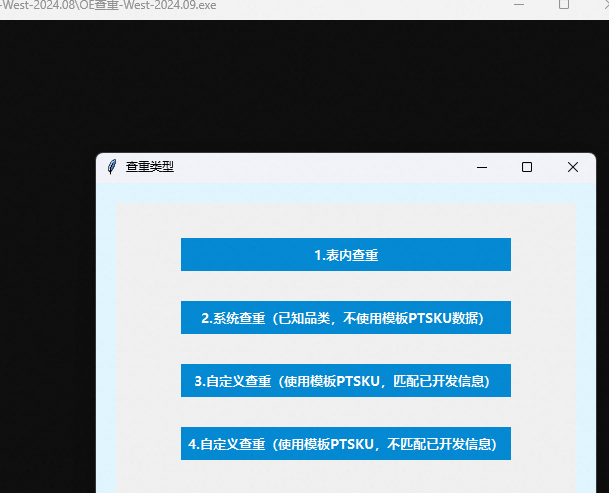




##### exe工具使用介绍

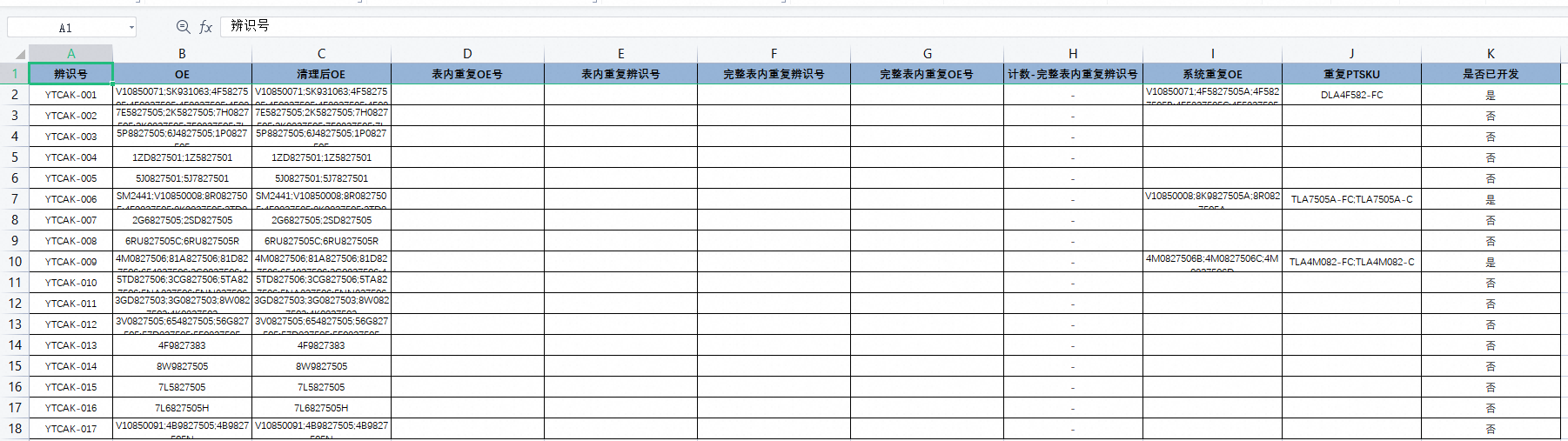
* 在模板中贴入“需查重数据”和“PTSKU”数据以后，保存excel文件，打开OE查重的exe工具。
* 在工具界面中，对应查重需求进行选择。（一般情况下，使用4.自定义查重）

1. 、如选择1.表内查重。那么仅会进行sheet1需查重数据的表内查重，判断是否存在表内重复的情况；
2. 、如选择2.系统查重。即会对PTSKU内已有品类进行查重，此时会用到“数据源”文件夹中的数据，进行已知品类的系统查重；
3. 、如选择3.自定义查重（匹配已开发信息）。那么会与PTSKU数据进行查重，且会匹配已经开发的具体信息；（耗时较久）
4. 、如选择4.自定义查重（不匹配已开发信息）。那么会使用sheet2中的PTSKU数据进行查重且不匹配开发的具体信息。（常用）



##### 具体输出

在程序运行完毕之后，查重信息会在“查重结果”表格中输出。在这里以常用的4.自定义查重做具体输出示例。



其中包含了辨识号、OE、重复OE、重复SKU、是否开发等等信息，最后三列包含了已开发的具体SKU和OE信息。

由于工具中“查重结果”表格会覆盖上一次的查重数据，如需保存本次查重信息，可以将“查重结果”另存到新的路径中，避免重复操作。

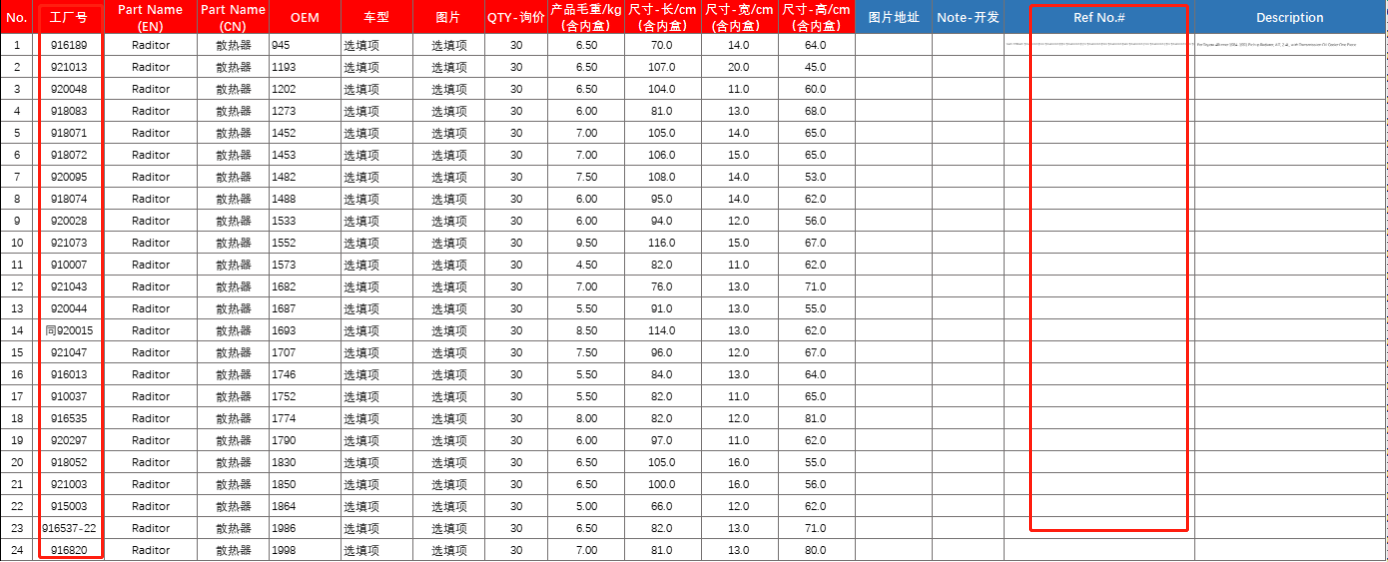
## Order Template查重

工厂所提供的OE会存在遗漏或者有误情况，因此，在做OT时，会进行产品OE号的补充。新品目录查重，只能排除大部分重复的概率，可能会漏掉某些OE。所以在完成OT后需要进行二次查重，减少重复的概率。

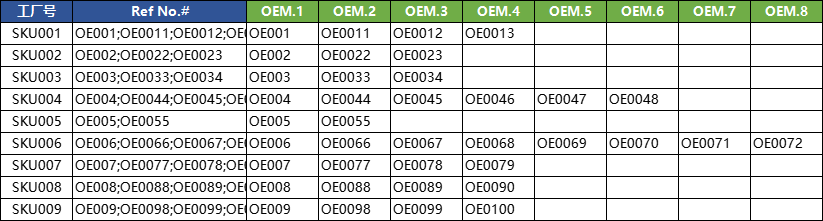
### Excel查重

#### **操作步骤**

* 新建EXCEL文档，命名为：【Order Template】-工厂品名-年月日-已查重。根据Order Template文件：【Order Template】-工厂品名-年月日，提取工厂号、Ref No.#，如下图所示。



* 根据2.2.1.2.1.2.目录OE规范步骤，将Order Template数据进行处理，得到如下图格式数据。



* 进行2.2.1.2.3.整合数据

### Python查重

#### **操作步骤**

根据Order Template文件：【Order Template】-工厂品名-年月日，参照2.2.3.2.2.新建输出文件夹、2.2.3.2.3.上传程序、2.2.3.2.4.更改路径、2.2.3.2.5.注意事项四个步骤，进行相对应的修改调整

# 《已查重》文档规范

根据查重结果需要以规范形式输出《已查重》文档，供新品开发以及Order Template流程的数据确认。文档的规范主要包含内容规范与格式规范。

## 已查重文档内容规范

已查重文档一般需包含以下信息：NO（即序号）、工厂、工厂号、完整OE、是否采购、重复OE、备注。具体内容规范如下：

|  |  |  |
| --- | --- | --- |
| **内容** | **规范要求** | **备注** |
| NO（即序号） | 标题、内容居中 |  |
| 工厂 | 标题、内容居中 |  |
| 工厂号 | 标题、内容居中 |  |
| 完整OE | 标题居中、内容左对齐 |  |
| 是否采购 | 标题、内容居中 | 是/否 |
| 重复OE | 标题居中、内容左对齐 | 可选 |
| 备注 | 标题居中、内容左对齐 |  |

## 已查重文档格式规范

已查重文档格式规范包括所用字体，行高、文档命名等要求，具体如下：

|  |  |  |
| --- | --- | --- |
| **内容** | **规范要求** | **例子** |
| 字体 | 1.首行：微软雅黑，10号加粗  2.其他行：微软雅黑，9号 |  |
| 字体颜色 | 1.首行：白色  2.其他行：黑色 |  |
| 单元格颜色填充 | 1. 首行：红色填充 2. 其他行：无填充色 |  |
| 行高 | 1.首行：20磅  2.其他行：16.8磅 |  |
| 边框 | 1. 外边框加粗 2. 添加内边框 |  |
| 网格线 | 网格线-不显示 |  |
| 文档命名 | 【新品目录】-已查重  【Order Template】-工厂品名-年月日-已查重 | 【Order Template】-福州扬腾撑杆-20220501-已查重 |