

PbIdea 如何导入医保大文件？

国家医保项目以 zip 格式提供药品、材料库下载，zip 文件里面是一个文本文件，文本文件是 utf8 编码，解压出来有好几百 M。这个文件对 PB 来说，比较难处理。PbIdea 有专门功能来处理这个文件。

首先使用 `uo_httpclient` 或 `uo_curl` 对象，根据医保接口文档将 zip 文件下载，存为文件。假设文件存为 `d:\202202187369608330634159091.zip`。接下来介绍如何把这个压缩包里的数据导入数据库。

第一步：解析数据文件，把它分解成若干个小 json 文件。

分解成小块的 json 数据文件的好处是不过份占用内存，可以快速处理完每一个数据导入。

```
uo_yibao yb; yb = create uo_yibao
uo_json jsYB; jsYB = create uo_json
yb.ImportYiBaoZip("d:\
202202187369608330634159091.zip", jsYB)
```

创建 yb 和 json 对象 然后调用函数 `ImportYiBaoZip`。`ImportYiBaoZip` 函数的参数可以是 zip 文件，不需要先解压。也可以是解压后的 txt 文件。

正常情况下，这个过程大概要 10 多秒钟。如果出错，自己察看 jsYB 的内容做相应处理。如果正确的话，jsYB 内容大概是这样：

见下图：

“files” 部分，是一个数组，包含了一个文件列表，可以循环取出来，

用 uo_json 解析后，然后 dwImportJson 导入 DW。

“lines” 部分，是检测到数据中哪些行数据可能有问题，需要人工纠正。
这些行没有被解析到 JSON 里来。

“message” 部分，提示信息，如果出错，这里是错误提示。

“result” 部分，本次总共成功解析了多少行。

```
{
  "files": [
    {
      "file": "D:\\is\\json\\PbJson\\demo\\9\\tmp\\XA001.json"
    },
    {
      "file": "D:\\is\\json\\PbJson\\demo\\9\\tmp\\XA002.json"
    },
    {
      "file": "D:\\is\\json\\PbJson\\demo\\9\\tmp\\XA003.json"
    },
    {
      "file": "D:\\is\\json\\PbJson\\demo\\9\\tmp\\XA004.json"
    },
    .....
  ],
  "lines": [
    {
      "line": 6818
    },
    {
      "line": 176245
    }
  ],
  "message": "parse 179033 lines",
  "result": 179033
}
```

第二步：将解析分割出来的每一个 JSON 导入 DW。

JSON 文件格式是数组，每一行从数据文件解析出来的分字段数据。line 是在数据文件中的行数，f1-fn 是字段名，各自根据业务定义。

```

{
  "line": 1,
  "f1": "XA01ABD075A002010100144",
  "f2": "无",
  "f3": "",
  "f4": "",
  "f5": "",
  "f6": "",
  "f7": "",
  "f8": "地喹氯铵含片",
  "f9": "86900144005704",
  "f10": "片剂",
  "f11": "",
  "f12": "",
  "f13": "",
  "f14": "0.25mg",
  "f15": "",
  "f16": "片剂",
  "f17": "0.25mg",
  "f18": ""
}

```

```
string ls_file
```

```
int i
```

```
i = 0
```

```
do while true
```

```
if jsYB.Get("/files/" + string(i) + "/file",ref ls_file) then
```

```
dw_1.dataobject = "" //这里可以是你自己定义的 DW , 也可以置
```

```
为空 , uo_json 会动态创建一个适合的 DW
```

```
jsData = create uo_json
```

```
if jsData.parseFile(ls_file) then
```

```
if jsData.dwImportJson(dw_1) then
```

```
//这里用你自己的方式更新到数据库
```

```
end if
```

```
end if
```

```
destroy jsData
else
    exit //循环取完所有列表中的文件
end if
i++
loop
```

通过以上代码，循环将所有 JSON 文件导入 DW。

以 300 多 M 的 TXT 文件(ZIP 文件大概 60M 左右)来说 ,预计 20-30 秒能全部导入到 DW。至于多少时间能导入数据库，那看各自业务实现功能了。

当然，数据导入结束后，可以自己再通过循环，取出每一个文件名，并删除之。

大自在 QQ781770213

2022/4/14