

Excel Compare

Project Overview

针对 Excel 文件内容对比的工具：

- 可以在配置文件中指定多列作为联合唯一业务标识，以此关联两表的行
- 可以在配置文件中指定多列进行数据对比
- 单独对比每一列的值，并标识出差异（黄色背景及批注）
- 对比结果中会新增两列：联合唯一业务标识，整行的对比结果
- 对比结果同时包含统计信息：总行数、一致数、不一致数、缺失数及主键重复数
- 可以指定运行时使用的配置文件

Requirements

- Python >= 3.7

Installation

```
pip install pandas openpyxl
```

Config

Main config

- `main_compare_file_path`
主文件路径（生成的对比结果文件会以此文件为基础）
- `sub_compare_file_path`
副文件路径
- `output_path`
默认值： ""
对比结果文件的生成路径（无需指定文件名）（未配置时会在当前命令执行目录生成）
- `skip_both`
默认值： True
在跳过时，副文件是否同时跳过

Sheets config

- `index`
sheet 的 index，从 0 开始
- `name`
sheet 的名称，当和 `index` 同时存在时，以 `index` 为准

- `skip_lines`

默认值: 0

对比时跳过的行数（第一行默认为表头默认跳过），例如需要跳过第二行时，配置此参数为 1

- `unique_columns`

作为联合唯一业务标识的列

Note

- 1、定义栏位的结构可以为 `Columns config` 或 字符串
- 2、为 字符串 时直接使用对应单元格的原始值进行处理
- 3、`Columns config` 可以和普通字符串同时出现在数组中

- `compare_columns`

数据对比的列

Note

说明同 `unique_columns`

Columns config

- `name`

列名

- `sub`

非必需，截取规则为：

- 当数组长度为 1 时，截取前 n 位，例如截取前 2 位则设置为 [2]
- 当数组长度为 2 时，按照指定指定区间进行截取，例如截取 2 至 5 位则设置为 [1, 5]
- 当数组长度为 2 且第一个元素为 null 时，可截取后 n 位，例如截取后 2 位则设置为 [null, -2]

```
{
  "main_compare_file_path": "your path/e1.xlsx",
  "sub_compare_file_path": "your path/e2.xlsx",
  "output_path": "your path/",
  "skip_both": true,
  "sheets": [
    {
      "index": 1,
      "name": "Sheet1",
      "skip_lines": 0,
      "unique_columns": [
        {
          "name": "部门",
          "sub": [
            null,
            -2
          ]
        },
        "系统"
      ]
    }
  ],
}
```

```

        "compare_columns": [
            {
                "name": "状态1",
                "sub": [
                    2
                ]
            },
            {
                "name": "状态2",
                "sub": [
                    1
                ]
            }
        ]
    },
    {
        "index": 0,
        "skip_lines": 3,
        "unique_columns": [
            "部门",
            "系统名称",
            "系统名称2"
        ],
        "compare_columns": [
            "阶段1",
            "阶段2",
            "阶段3"
        ]
    },
    {
        "name": "Sheet1",
        "unique_columns": [
            "部门",
            "系统名称",
            "系统名称2"
        ],
        "compare_columns": [
            "阶段1",
            "阶段2",
            "阶段3"
        ]
    },
    {
        "unique_columns": [
            "部门",
            "系统名称",
            "系统名称2"
        ],
        "compare_columns": [
            "阶段1",
            "阶段2",
            "阶段3"
        ]
    }
]

```

```
    },
    {
      "index": -1,
      "unique_columns": [
        "部门",
        "系统名称",
        "系统名称2"
      ],
      "compare_columns": [
        "阶段1",
        "阶段2",
        "阶段3"
      ]
    },
    {
      "index": -1,
      "name": "S1",
      "unique_columns": [
        "部门",
        "系统名称",
        "系统名称2"
      ],
      "compare_columns": [
        "阶段1",
        "阶段2",
        "阶段3"
      ]
    }
  ]
}
```

Usage

Run the project with the following command:

Note

`配置文件路径` 非必须，未指定时会自动读取当前命令执行目录下的 `config.json` 文件

```
py .\main.py <配置文件路径>
```