

# Excel-Join 使用说明

## 简介

Excel-Join 是一个将多个 Excel 文件合并成一个 Excel 文件的工具。它可以根据你提供的模板，自动将多个表格的数据合并到一起，并保留模板的格式。

## 适用场景

- 将员工信息表、工资表、考勤表合并成一个完整的员工档案
- 将不同部门的数据表合并成一个汇总表
- 将多个 Excel 文件中的数据按共同列（如员工ID）合并

例如：

### 输入文件1

	A	B	C	D
1	花名册			
2				
3	城市	姓名	工号	职位
4	上海	刘健宁	E0001	客服
5	上海	周梦婷	E0002	客服组长
6	上海	翟菁	E0003	租赁会计
7	上海	刘怡宁	E0004	公司事务副经理
8	上海	沈晗	E0005	客服
9	上海	赵海涛	E0006	工程经理

### 输入文件2

	A	B	C	D	E	F
1	姓名	薪资	入职时间	五险一金起缴时间	劳动合同开始时间	劳动合同结束时间
2	刘健宁	5200	2025-09-08	2025-09-01	2025-09-08	2028-09-07
3	周梦婷	7000	2025-09-15	2025-09-01	2025-09-15	2028-09-14
4	翟菁	8000	2025-09-15	2025-09-01	2025-09-15	2028-09-14
5	刘怡宁	21,420	2024-09-02	2025-10-01	2025-10-01	2027-09-01
6	沈晗	5200	2025-09-22	2025-09-01	2025-09-22	2028-09-21
7	赵海涛	25000	2025-09-22	2025-09-01	2025-09-22	2028-09-21
8	季予峥	8000	2025-10-09	2025-10-01	2025-10-09	2028-10-08
9	汗磊	70000	2025-10-10	2025-10-01	2025-10-10	2028-10-09

### 模板

	A	B	C	D	E	F	G	H	I	J
1										
2	雇员名单									
3										
4	所在城市	账单名称	雇员姓名	职位	岗位性质	Salary	入职时间	五险一金起缴时间	劳动合同开始时间	劳动合同结束时间
5										
6										

### 输出结果

	A	B	C	D	E	F	G	H	I	J
1										
2	雇员名单									
3										
4	所在城市	账单名称	雇员姓名	职位	岗位性质	Salary	入职时间	五险一金起缴时间	劳动合同开始时间	劳动合同结束时间
5	上海		刘健宁	客服		5200	2025-09-08	2025-09-01	2025-09-08	2028-09-07
6	上海		周梦婷	客服组长		7000	2025-09-15	2025-09-01	2025-09-15	2028-09-14
7	上海		翟菁	租赁会计		8000	2025-09-15	2025-09-01	2025-09-15	2028-09-14
8	上海		刘怡宁	公司事务副经理		21420	2024-09-02	2025-10-01	2025-10-01	2027-09-01
9	上海		沈晗	客服		5200	2025-09-22	2025-09-01	2025-09-22	2028-09-21
0	上海		赵海涛	工程经理		25000	2025-09-22	2025-09-01	2025-09-22	2028-09-21
1	上海		季予峥	会计		8000	2025-10-09	2025-10-01	2025-10-09	2028-10-08
2	上海		汪磊	客户洞悉总监		70000	2025-10-10	2025-10-01	2025-10-10	2028-10-09
3	上海		王文怡	运营支持专员		13000	2025-10-15	2025-10-01	2025-10-15	2028-10-14
4	上海		刘一壮	品类供应商质量副经理		20000	2025-10-20	2025-10-01	2025-10-20	2028-10-19
5	上海		孙文莲	人力资源经理		32000	2025-10-20	2025-10-01	2025-10-20	2028-10-19

## 快速开始

### 1. 打开命令行工具:

- Windows: 按 Win + R, 输入 `cmd`, 回车
- Mac: 打开"终端"应用程序
- Linux: 打开终端

### 2. 进入项目目录:

```
cd 项目文件夹路径
```

### 3. 运行合并命令:

```
uv run excel-join 数据文件/ -t 模板.xlsx
```

### 4. 查看结果:

合并完成后, 会在当前目录生成 `joined_output.xlsx` 文件

## 什么是 `uv run`?

- `uv` 是一个 Python 包管理工具, 用于运行项目中的命令
- `uv run` 表示使用项目环境运行后面的命令
- 如果系统提示"找不到 `uv`", 请先安装: `pip install uv`

## 准备工作

### 1. 准备输入文件

将所有需要合并的 Excel 文件放在同一个文件夹中。例如:

```
数据文件/  
├── 员工信息.xlsx  
├── 工资表.xlsx  
└── 考勤记录.xlsx
```

### 2. 准备模板文件

创建一个 Excel 模板文件, 定义你想要的最终输出格式。模板需要包含:

- 表头行: 第一行是列名 (如"姓名"、"工号"、"部门"等)
- 格式: 可以设置字体、颜色、列宽等格式

- **批注**（可选）：在表头单元格添加批注，帮助工具更准确地匹配列

模板示例：

姓名	工号	部门	入职日期	基本工资	考勤天数

添加批注的方法：

1. 在 Excel 中右键点击表头单元格
2. 选择"插入批注"
3. 输入以下格式的批注：

```
同义词：name, 员工姓名, 申请人  
说明：员工的姓名
```

或者简单写：

```
name, 员工姓名, 申请人
```

### 3. 配置 LLM（首次使用需要）

在项目根目录创建 `.env` 文件，填入你的 AI 模型配置：

```
LLM_MODEL_ID=gpt-4  
LLM_API_KEY=你的API密钥  
LLM_BASE_URL=https://api.openai.com/v1
```

## 基本使用方法

### 最简单的命令

```
uv run excel-join 数据文件/ -t 模板.xlsx
```

这个命令会：

1. 将 `数据文件/` 文件夹中的所有 Excel 文件合并
2. 按照 `模板.xlsx` 的格式输出
3. 生成 `joined_output.xlsx` 文件

### 指定输出文件名

```
uv run excel-join 数据文件/ -t 模板.xlsx -o 员工完整档案.xlsx
```

### 合并指定的文件

```
uv run excel-join 员工信息.xlsx 工资表.xlsx 考勤记录.xlsx -t 模板.xlsx -o 合并结果.xlsx
```

## 参数说明

## 必需参数

- 输入文件: 要合并的 Excel 文件或文件夹 (可以多个)
- `-t 模板文件` 或 `--template 模板文件`: 模板 Excel 文件的路径

## 可选参数

- `-o 输出文件` 或 `--output 输出文件`: 指定输出文件名 (默认: `joined_output.xlsx`)
- `-p 主表名` 或 `--primary 主表名`: 指定主表名称 (默认使用第一个文件的第一个工作表)
  - 什么是主表? 主表是合并时的"基准表", 所有数据都会保留, 其他表的数据会"挂"到主表上
  - 主表名格式: 文件名\_工作表名, 例如 `员工信息_Sheet1`
- `-k 列名` 或 `--key-column 列名`: 指定用于过滤空行的列名
  - 什么是关键列? 选择一个列 (如"工号"), 如果该列为空, 则整行数据会被过滤掉
- `-c 配置目录` 或 `--config-dir 配置目录`: 配置文件目录 (默认: `./config`)
- `--rebuild-cache`: 强制重新生成 SQL, 忽略缓存
- `-v` 或 `--verbose`: 显示详细处理过程
- `-q` 或 `--quiet`: 只显示错误信息

## 使用示例

### 示例1: 合并员工数据

假设有以下文件:

- `员工信息.xlsx` (包含: 工号、姓名、部门、入职日期)
- `工资表.xlsx` (包含: 工号、基本工资、奖金)
- `考勤记录.xlsx` (包含: 工号、出勤天数、请假天数)

模板文件 `员工档案模板.xlsx`:

工号	姓名	部门	入职日期	基本工资	奖金	出勤天数	请假天数

执行命令:

```
uv run excel-join 员工信息.xlsx 工资表.xlsx 考勤记录.xlsx -t 员工档案模板.xlsx -o 员工完整档案.xlsx
```

### 示例2: 合并整个文件夹的文件

```
uv run excel-join 数据文件/ -t 模板.xlsx -o 汇总结果.xlsx
```

### 示例3: 指定主表

如果希望以某个特定表格为基准进行合并:

```
uv run excel-join 数据文件/ -t 模板.xlsx --primary 员工信息_Sheet1 -o 结果.xlsx
```

## 示例4：过滤空行

如果数据中有很多空行，可以指定一个关键列来过滤：

```
uv run excel-join 数据文件/ -t 模板.xlsx -k 工号 -o 结果.xlsx
```

## 工作原理

Excel-Join 的工作流程：

1. **读取数据**：将所有 Excel 文件读入内存
2. **智能匹配**：使用 AI 分析每个表格的列，找到可以关联的列（如"工号"、"员工ID"）
3. **合并数据**：根据关联列将多个表格合并成一个表格
4. **格式输出**：按照模板的格式（字体、颜色、列宽等）输出结果

什么是"共同列"？

- 共同列是多个表格中都存在的列，用于关联数据
- 例如：员工信息表有"工号"，工资表也有"工号"，通过"工号"将两个表关联起来
- 如果没有共同列，AI 会尝试根据列名相似度进行匹配

## 常见问题

### 1. 提示"Template file not found"

- 检查模板文件路径是否正确
- 确保模板文件存在且是 `.xlsx` 格式

### 2. 提示"No Excel files found"

- 检查输入路径是否正确
- 确保文件夹中包含 `.xlsx` 或 `.xls` 文件

### 3. 合并结果不正确

- 检查模板中的列名是否与数据文件中的列名匹配
- 在模板中添加批注，提供同义词帮助 AI 更好地匹配列
- 使用 `--rebuild-cache` 参数强制重新生成 SQL

### 4. 提示 LLM 配置错误

- 检查 `config/.env` 文件是否存在
- 确保 API 密钥和 URL 配置正确
- 确保网络可以访问 AI 服务

### 5. 处理速度慢

- 首次运行需要调用 AI 生成 SQL，会比较慢
- 后续运行会使用缓存，速度会快很多
- 如果修改了数据结构，使用 `--rebuild-cache` 重新生成

## 注意事项

1. **文件格式**：输入文件必须是 Excel 格式（`.xlsx` 或 `.xls`）
2. **模板格式**：模板文件必须是 `.xlsx` 格式

3. **列名匹配**: 模板中的列名应尽量与数据文件中的列名一致
4. **共同列**: 合并是基于共同列进行的, 确保各表格有可以关联的列 (如工号、ID等)
5. **隐藏工作表**: 隐藏的工作表会被自动跳过
6. **缓存机制**: 相同的数据结构会使用缓存, 加快处理速度

## 获取帮助

---

如果遇到问题, 可以:

1. 使用 `-v` 参数查看详细处理过程
2. 检查 `config/` 目录下的配置文件是否正确
3. 确保所有输入文件没有被其他程序占用