1. 将需处理数据放进FU文件夹，存储文件类型为csv，数据格式为A列：需处理的基金名组合；B列：论文的唯一标志符（本次案例中为WOS号）（注意：第一行不需要属性名FU、WOS标记，文件内全部为待处理数据）。因已加载多进程，在处理文件前，如果文件内数据过多（超过一百万），可以上网搜索csv文件分割器，将其分割成若干较小文件（六万到十万条数据左右即可），可提高处理速度。
2. 检查文件内数据是否有问题（可通过筛选文件C列数据非空查看），若是出现异常数据分列现象（出错数据多是因为csv文件逗号存储法的问题），若数据量较小可手动修改；若数据量较大，可将出错数据条目剪切至“错误数据.xlsx”文件中，并运行faultFile.py,可将出错数据一次性转化为正常数据correctFiles.csv，将其放进FU文件夹待处理。若无问题，进行下一步。
3. 确保FU文件夹内待处理的文件数据结构全都正常后，运行fuMerge.py，运行过程较长，请耐心等待，可通过任务管理器查看python程序是否结束判断其是否还在运行。产生FuDealt文件夹，其中为处理后数据。
4. 产生fauleFile.csv文件，其内若有记载有问题的文件名，找到相应文件，查看其中数据是否有分列异常数据问题，进行步骤2修改后重新处理出错文件（注意：把FU文件夹中已处理完成的文件移除，否则会出现重复处理现象）。
5. 运行deDuplication.py文件，可生成dpubFunds.csv文件，存储去重后基金名。