数据清洗工作

1.按照组织均分学者人数进行清洗，在 graphic 目录下实时记录下了已经清洗的人数，并且能从直方图直观的观察到任务进度

2.每次成功提交一条目 bool 清洗开关就会反置，然后该条目就会记录下本次清洗的时间然后从表中移除，每过 30 天开关会重置为未清洗状态，意味着此时需要回溯更新该条目

3.于该组织而言若出现一类字段如头像大面积缺失的情况，可以安排会爬虫解析的同学进行机器清洗，目前原 scholar 爬虫项目的数据清洗功能已十分健壮，直接调用接口即可（TODO: 为了避免语言平台的不同以及部署的不便性我会开始着手将接口以 web server api 的形式部署于云端）

4.目前已经确定必要条目的有

头像 姓名 关键字 email 学科 组织 简介 title

针对无原站链接的情况，可从 email 后缀分析出组织名。google scholar 能以组织划分呈现出清晰的 email，头像，关键字，亦不失为一种快捷的途径。