

笔试题:

- 1、现有 2017, 2018, 2019 年参加 CSS 大会的人员名单 (1 年 1 份, 一共 3 份), 字段分别为《姓名》, 《年龄》, 《身份证》, 《手机号》, 《公司名》
 - 2、统计连续 3 年都参加了安全大会的人, 并且以年龄高->低排序输出
 - 3、数据手动/自动构造均可, 比如每年构造 1 万个示例, 程序设计请考虑实际大数据的情况
- 温馨提示:
- 1、以自己最拿手的语言编写, 优选 c++, c, go, python, java
 - 2、注意 《存储空间》, 《性能》, 《代码风格》

语言: python, 所用库: xlrd: 用于读取数据, collection: 用于存储的数据结构

设计思路: 由于参赛者需要三年都参加才会被考虑, 因此从前两张表中只提取 ID 信息, 并将结果进行交集 $O(n\log n)$ 。遍历第三章表, 判断 ID 是否在前两张表的交集中 $O(M*N)$ (M, N 分别为两张表大小), 将存在着交集的 ID 的所有信息取出并存入年龄作为 key 的 defaultdict(list) 中, 在输出打印前对 list 的 key 进行从小到大排序 $O(n\log n)$ 。

性能: 整体时间复杂度是 $O(M*N+n\log n)$, 所需存储的内容: 前两年的 ID, 三年均参加的人的信息, $O(M1+M2+N)$, 在所测试的 10000 个用例 (九千多人三年均参加), 笔记本用时 1s

备注: 附件 test1.py 为主程序, peopleGenerate.py 为了生成测试所用人员信息。People1, 2, 3 三张 excel 表格分别为 2017, 2018, 2019 年人员信息

部分运行结果:

```
['祝介健', 22.0, '231124199812091843', 123.0, 'seu']
['倪怕线', 22.0, '450421199802132012', 123.0, 'seu']
['杨伙合', 22.0, '632321199803162395', 123.0, 'seu']
['孙受都', 22.0, '430801199812121133', 123.0, 'seu']
['褚林误', 22.0, '440514199812262822', 123.0, 'seu']
['郝初圣', 22.0, '41092619981112266X', 123.0, 'seu']
['韦则南', 22.0, '430412199806172513', 123.0, 'seu']

['里斯', 19.0, 'FJKLSKDJFDLSK', 123.0, 'seu']
['张三', 18.0, 'fsfjklS1123', 123.0, 'seu']
['梅洛瑜', 17.0, '36050220021111041X', 12312.0, 'seu']
```