

FDDC2018 金融算法挑战赛 02-A 股上市上市公司公告信息抽取

致 FDDC2018 参赛选手，

恭喜各位顺利进入复赛阶段，在这里需要提醒各位，8 月 8 日 22:00 之后，选手不可再修改程序，该程序在平台上运行生成的预测结果，将作为复赛的主要评分依据；因此，请选手抓紧 8 月 8 日由官方运行程序的机会并认真根据反馈修改，确保程序可以顺利运行。

为了提高选手优化模型的效率，组委会特梳理了关于答案提交的常见问题，请选手务必认真阅读。

预祝各位取得理想成绩~

FDDC2018 组委会

赛题二提交格式注意事项：

*请参考 “submit_sample.zip”

- 1、以 utf-8(without BOM) 为编码格式
- 2、以 tab 分隔的三个 txt 文件，不要改变文件名称
- 3、所有字符串类型字段均忽略大小写和空格
- 4、需将所有标点符号、英文字符及数字转化成半角

一、“资产重组” 类型注意事项

1、针对“资产重组-标的公司”、“资产重组-交易对方” 字段，公告原文中可能同时披露公司简称和公司全称；选手在提交答案时，可同时提供全称及简称 2 个答案，答案间用竖线 “|” 分隔，其一和标准答案相同即为正确；

“资产重组-交易对方” 答案用竖线分隔两个集合示例：a、b、c | aa、bb、cc

2、针对“资产重组-交易对方”、“资产重组-评估法”、“重大合同-联合体成员”字段，类型为多个值的集合，答案次序没有统一标准；选手在提交该字段答案时，使用顿号（、）进行分隔，每个值得类型为字符串，评测时将忽略值得顺序；

3、针对“资产重组-标的名称”字段，公告中存在“xx.xx%股份/股权/债权”、“xx.xx00%（的/之）股份/股权/债权”等不同的表述，这些表述在评测时 将由程序统一归一化为“xx.xx%股份/股权/债权”，选手在提交结果时无需做改动；

二、“股东增减持”类型注意事项

- 1、存在部分增减持价格以区间表示，选手按规则统一取**最高价**即可；
- 2、公告原文中，有具体变动日期的，根据具体日期处理；没有公布具体变动日期的，根据公告日处理，对应的公告日请参考公告日期数据文件；

三、“重大合同”类型注意事项（红色标注为新增内容）

- 1、**重大合同中，合同金额同时有使用人民币和美元单位的金融，优先使用“人民币”；**
- 2、重大合同中，如果有多位联合体成员，使用**顿号（、）**分隔；**“和”、“及”、“以及”等连词在测试集的答案中也将统一替换成顿号；**
- 2、重大合同中，子公司中标的情况下，“乙方”就是子公司，即输出的是子公司，项目名称就是子公司中标的名称，为——对应的关系；
- 3、重大合同中，逗号需保留，全角逗号需要转半角；
- 4、重大合同中，“乙方”和“联合体成员”字段的抽取结果需去掉括号及括号中的内容；

注，仅针对这两个字段；

5、重大合同中，当“项目名称”在公告原文中出现多次且不一致时，根据如下标准进行判断：如果在项目情况介绍中提出了“项目名称：xxx”的，优先以词为准；

6、重大合同中，“项目名称”字段补充强调一个处理规则：对于项目名称为“xxxxx”项目时，需要**去掉引号**，标准化为 xxxxx 项目——举例：我公司成功中标“某某某污水处理”项目，抽取出来的项目名称应为：某某某污水处理项目；

四、其他无法分类的规则...

1、如果主键中有值缺失，还需要把这一项抽取出来吗？

需要。主键中的某些字段是可以为空的，例如“重大合同-甲方”，这种情况仍需要把这条记录抽取出来。哪些字段为空在“样例说明”文档中有详细介绍。

2、关于“联合体”和“交易对方”字段，只有多个值全部正确才算字段正确；