**CPA抓取数据程序部署说明**

**一、部署目录文件结构**

* knowtion-spider
  + knowtion-spider-0.0.1.jar应用程序
  + application.properties启动配置文件
  + jsonTempjson模板文件夹
    - clinicalTrial.json
    - drug.json
    - evidence.json
    - gene.json
    - mutationStatistics.json
    - protein.json
    - regimen.json
    - variant.json
  + merge数据合并文件夹
  + doc文档
    - CPA与老库合并返回的结果示例.xlsx
    - CPA一次运行流程图.pdf
    - CPA泳道图.pdf
    - 部署说明.md
  + logs日志文件夹
  + atomikos分布式事务相关，自动生成

**二、启动**

1. 指定内部配置文件启动： 启动命令：java -jar knowtion-spider-0.0.1.jar --spring.profiles.active=dev。 该命令启动使用的是knowtion-spider-0.0.1.jar包内部的src\main\resources\application-dev.properties配置文件。 需要在代码打包前修改好要用的配置文件，application-dev.properties会覆盖application.properties的配置，前者没有的配置会使用后者的即默认配置。
2. 指定外部配置文件启动： 启动命令：java -jar knowtion-spider-0.0.1.jar --spring.config.location=application.properties。 该命令使用的是文件目录中的application.properties作为配置文件，该配置文件必须包含全部的必须的配置。修改配置后，重新以此命令启动即可。

* 程序会在每天早上9点定时发送前一天的错误日志。

**三、JSON结构变更的处理**

1. 当需要修改JSON结构时，关闭程序，然后修改相应的json文件，修改完成确认无误后重启即可。
2. json值的类型说明：
   * "String":字符串类型
   * 0:数字类型
   * true:布尔类型
   * null:表示无论待对比的JOSN有没有该键都视为没有改变。

**四、关于数据合并的处理**

1. 当发送待审核数据给相关人员失败时会在merge文件夹下生成CoincideData-20171122.xlsx的文件（可以在错误日志中查看关于文件位置的描述）。 ***注意：***相同的待审核数据在一次程序运行周期（启动应用程序到关闭应用程序）内只会发送一次邮件，如果发送待审核文件失败(也只会生成一次文件)，需要手动获取上面所说的文件,该文件不会自动删除。
2. 返回的审核结果手动保存在merge文件夹下，返回的格式参照doc/merge/CPA与老库合并返回的结果示例.xlsx文件，程序会定时扫描这些文件并根据审核结果进行处理。 文件必须以merge+数字.xlsx的规则命名,如merge20171122.xlsx；merge.xlsx是保留文件，用来存储合并失败的，不能人工更改。程序会自动删除已处理的文件，不能人工删除。