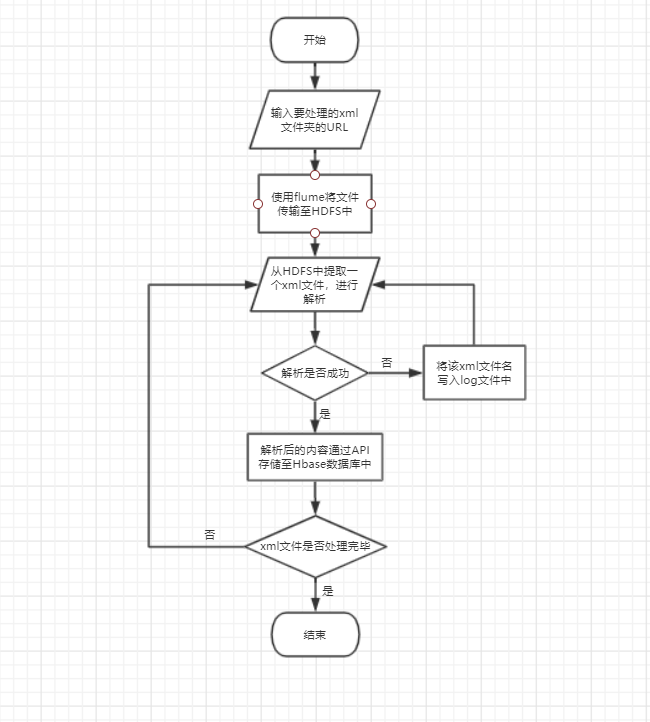
**现有工作基础**：秦龙的爬虫所爬的知网论文xml格式文件

**前置工作问题**：在用旧的flume程序将xml文件放入Hbase数据库时，由于文件数据格式的变化而使操作不超过。且旧的flume程序不可维护

**本周工作内容**：编写程序将HDFS中的每个xml文件进行解析，并转存至Hbase数据库中，总的程序流程图如下：

**下周工作内容**：解决BUG，完成程序中flume部分，完成程序总体调试，编写相应Shell脚本实现自动化。