# DrugBank数据库处理概要设计

## 数据下载后整体数据处理原则

1、（主控程序会将下载好的更新文件放到相应的文件夹下，主控程序会将下载好的文件进行解压缩放置，即待处理文件夹下直接就是单个文件或多个文件）数据处理的第一步就是进入到自己的数据库待处理文件夹。

2、读取待处理文件夹下的文件，进行文件处理。

3、处理文件时，对数据库中的所有数据进行全部更新。原始数据都有一列入库日期（年月日），当数据更新后，都需要更新入库日期。

4、自己的数据库处理程序需要至少三个数据表，即数据库总表，日志表，删除数据表。

5、处理数据失败时，需要记录数据的唯一标识或出错的日志，写入错误日志，继续处理后续数据。处理单条数据出现异常时，记录日志表，继续处理后续数据。日志表中需要记录处理数据的唯一标识，特殊异常直接插入异常信息即可。

6、数据处理完成后，直接删除待数据库待处理文件夹下的数据。

## 模块说明

1. 目的：文档为DrugBank Database的概要设计介绍，因为是整个项目的一个子系统，所以本文档将整个DrugBank Database作为一个模块处理。本文档将为后续的使用和测试作参考。
2. 模块名称：DrugBankParser
3. 模块介绍：本模块用于处理DrugBank Database的xml文件.
4. 实现环境：本软件在windows 10 64位系统下运行，使用python语言编码，编译器为pycharm专业版。