APK爬虫工具程序使用说明

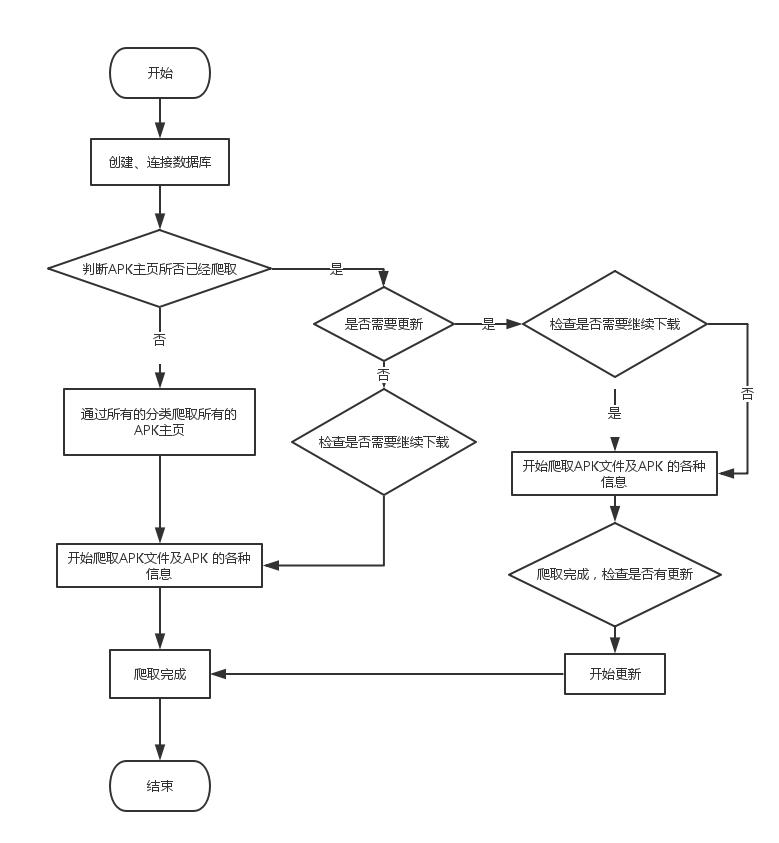
程序说明：本程序爬取的APK来自豌豆荚应用市场。本程序专门对于豌豆荚应用市场，只能爬取豌豆荚应用市场中的APK。

程序思路：

查看豌豆荚的网页版应用市场，发现他组织APK的形式：将APK分成很多类，然后每个类别对应很多APK文件，每个APK文件都有自己的主页，主页中包含APK中的所有基本信息。

所以爬虫的思路就是爬取所有的分类文件，然后爬取所有APK的主页，将APK的所有主页爬取存储到文件中，然后读取文件，再将所有的APK文件及APK的基本所有基本信息爬下来，将APK存储到文件夹，将APK基本信息存储到数据库。

程序框图：



源文件说明：

总共有八个源文件，分别为database.py、main.py、hash\_md5.py、context.py、downjudge.py、crawler.py、check\_update.py、download.py。

下面分别介绍每一个源文件的作用：

database.py:创建数据库、存储数据、执行SQL语句、判断数据库中MD5是否存在。

hash\_md5：提供一个方法对文件进行md5加密，并返回md5值。

context.py:返回APK的各种信息。

downjudge.py:判断是否需要继续下载，返回一个bool类型。

crawler.py:对于所有的分类爬取所有APK主页，并询问是否需要更新。

check\_update.py:检查是否有更新，返回一个bool类型。

download.py:下载APK文件，显示进度条。

main.py:主文件，对另外几个文件的方法进行调用，完成下载。

使用方法：

直接执行main.py即可，然后根据提示操作。