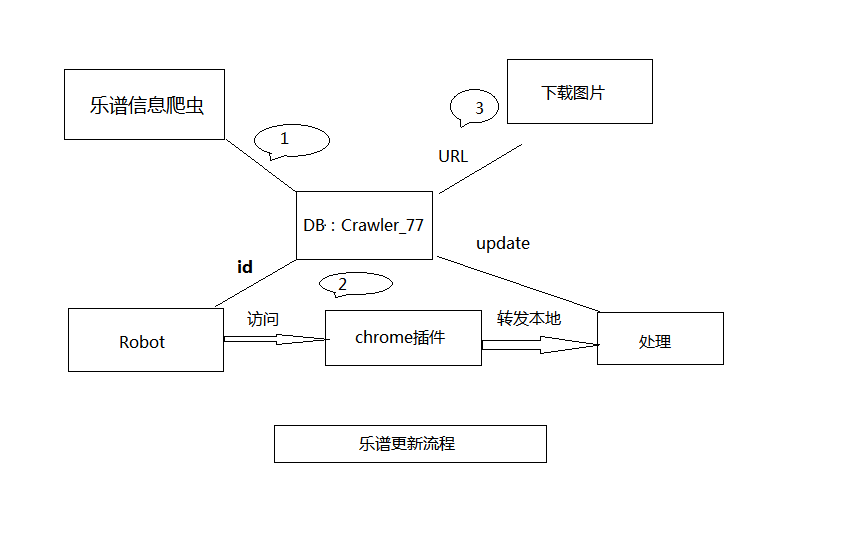
# 乐谱更新

## 概述



分大个大步骤

（爬之前最好更新一下代理Ip ）

1. 爬取乐谱信息：title 之类
2. 根据id 让机器人自动点Flash 然后插件转发，然后把每个乐谱的验证码保存
3. 根据验证码下载图片

三个步骤是顺序分开执行的，依赖于前者。

## 实现

要把项目部署到Tomcat和开启插件才能工作。

第一个步骤以客户端（本地运行F11）方式，第二步骤需要插件和Servlet，第三个需要Servlet。

**爬取乐谱信息**

UpdateClient.java

更新策略：如果蛐蛐网（下面简称77）的最新id 大于本地DB的数据库，下载差值数量的乐谱信息。下载顺序是从小到大，看代码

对HTMl的处理用 htmlParser 这个开源项目

**第二个步骤 下载URL**

衔接在UpdateClient 后面

//下载InfoURL

*executeNewThread*();

"SELECT \* FROM `crawler\_77` where auth\_param is null and isDownloadPic = 0"

Robot 拿到数据 开始自动点击，浏览器位置大小是左遍竖长方形（自己可以调）

没点一下 ，插件转发给Servlet，接受保存

提示：按F3 点进去

**第三个步骤 下载图片**

Downloader.java

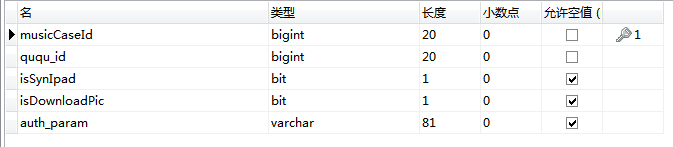
String sql = "SELECT auth\_param FROM `crawler\_77` where auth\_param is not null and isDownloadPic = 0";

拼接URL访问本地，让本地Servlet处理下载；

src/com/ipuzhe/servlet/Download.java

## 数据库

主要关系表：crawler\_77



乐谱id 77网id 是否同步到线上DB 图片下载了没 验证码（flash那个）

其他表跟线上数据库一样，便于同步