# 系统说明

## 爬虫识别

### 引入

在node.js中，通过以下的代码使用爬虫识别模块。使用的参数为用户的请求。

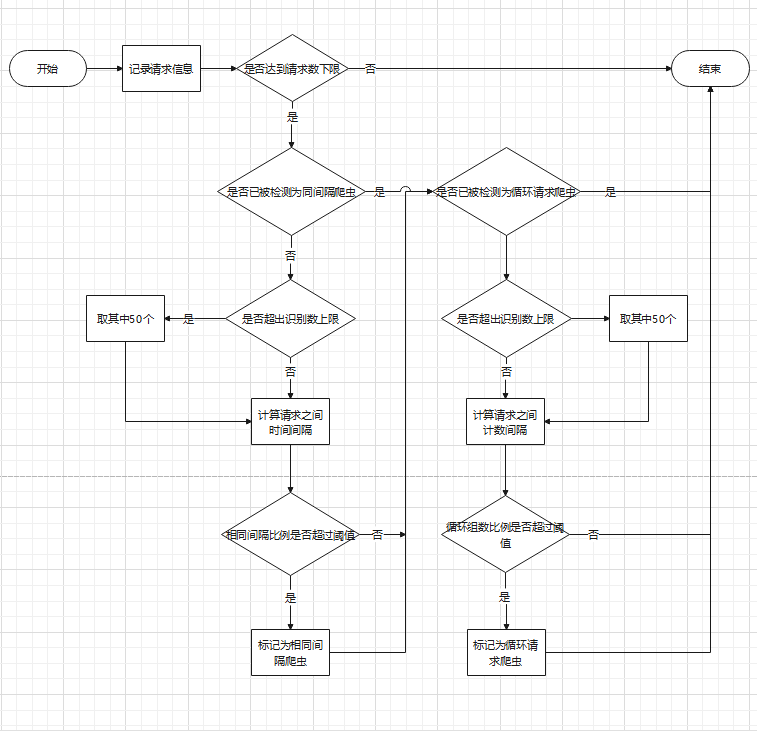
const recCrawler = require("./crawlers");

recCrawler(req);

### 程序流程

**数据库：**连接ip-request表”reqs”、已识别的爬虫表”craws”。

流程图如下图所示。



**请求间隔：**对于请求的时间间隔进行计算，取至少20个、至多50个最新请求。当请求间隔的时间（毫秒）相等的个数超过所取样本的50%时，标记为爬虫，标记理由设置为相同间隔sameGap。

**循环请求：**计算请求的计数间隔，取至少20个、至多50个最新请求。当请求间隔的计数相等的个数（也就是成循环组的请求个数）超过所取样本的50%时，标记为爬虫，标记理由设置为循环请求loopApi。

以上两个部分的识别仅当请求个数大于20个，且客户端的爬虫标记理由没有对应的内容时才会启用。