**中国商标网开发说明文档**

###### 基本信息

* 开发人：杨羽飞
* 开发日期：2018-08-14~2018-08-15
* 网站名称：中国商标网
* 网站url：<http://sbgg.saic.gov.cn:9080/tmann/annInfoView/homePage.html>
* 网址类型：商标信息

###### **网站抓取接口说明**

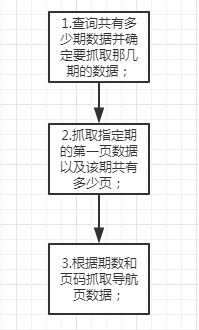
**2.1 获取发布期数数据**

* 抓取需求：获取当前一共有多少期数据，以及最近十期的发布日期；
* 请求接口：http://sbgg.saic.gov.cn:9080/tmann/annInfoView/homePage.html
* 请求方式：GET
* 参数说明：无
* 接口返回数据类型：html
* Headers设置：无
* 其它说明：无

**2.2 获取导航页数据**

* 抓取需求：获取指定期数的指定页码的注册商标数据
* 请求接口：http://sbgg.saic.gov.cn:9080/tmann/annInfoView/annSearchDG.html
* 请求方式：POST
* 参数说明：
  + page：页码，必须；
  + rows：每页多少条数据，当前设置1000；
  + annNum：第几期；
* 返回数据类型：json
* Headers设置：无
* 其它说明：无

###### **网站抓取流程图**



###### **其它说明**

**4.1 去重方式**

由于每期抓取数据中间状态保存在redis中，且每期内数据不重复，因此期内数据不进行去重；

所有抓的期号使用简单去重器进行去重。

**4.2 开发语言**

Nodejs开发，并将进程托管至PM2。

**4.3 测试结果**

无