**如何实现增量采集？**

默认情况下，入口URL、列表URL和内容URL这所有的URL都有去重机制，就会对增量采集造成一定的麻烦。  
框架开放了 add\_scan\_url() 接口，让用户可以在一次完整采集过后，添加新的入口URL(比如之前的入口URL、最新列表URL)来进行增量采集。  
通过 add\_scan\_url() 方法添加的URL，不会被框架去重，从而达到增量采集的效果。。。

举个栗子:  
我已经把糗事百科一次性采集完了，而糗百的内容更新都在首页，所以我可以在一次完整采集以后，把首页加入增量采集

$spider->on\_start = function($phpspider)

{

// add\_sacn\_url 没有URL去重机制，可用作增量更新

$phpspider->add\_scan\_url("http://www.qiushibaike.com/");

};