**如何爬取列表页中的数据?**

一般情况下, 我们只需爬取内容页的数据即可, 不过有时候列表网页中也会有需要爬取的数据, 那想要爬取这部分数据, $phpspider->add\_url($url, $options)函数

举个栗子:

在爬取[爱游网](http://www.ai.tt/)的时候, 除了基本的内容页信息外, 还需要爬取浏览次数(或阅读量), 但是这些数据在列表页中, 这就需要在on\_list\_page回调函数中做处理

$configs = array(

// configs的其他成员

...

'fields' => array(

array(

'name' => "question\_view\_count",

// 在内容页中通过XPath提取浏览次数(或阅读量)

'selector' => "//a[contains(@class,'page-view')]",

'required' => true,

),

),

);

$spider->on\_list\_page = function($page, $content, $phpspider)

{

// 在列表页中通过XPath提取到内容页URL

$content\_url = selector::select($content, "//a[contains(@class,'s xst')]/@href");

// 在列表页中通过XPath提取到浏览次数(或阅读量)

$page\_views = selector::select($content, "//td[contains(@class,'num')]/em");

// 拼出包含浏览次数(或阅读量)的HTML代码

$page\_views = '<div><a class="page-view">' . $page\_views + '</a></div>';

$options = array(

'method' => 'get',

'context\_data' => $page\_views,

);

$phpspider->add\_url($content\_url, $options);

// 返回true继续提取其他列表页URL

return true;

};