

# 二十、破解有反爬虫的网站

不学这节， 可以抓取95%的网页  
学了这节， 可以抓取99%的网页



网页用web scraper选不中?

没事，咱们自己来~



# 例子 | 知乎文章标题

<https://www.zhihu.com/people/zhang-jia-wei/posts>

<h2 class="ContentItem-title">

<a href="www.baidu.com">不用写代码的爬虫课 </a>

</h2>

web scraper: `h2.ContentItem-title a`

自己写: `h2.ContentItem-title`



# 例子 | 知乎文章标题

<https://www.zhihu.com/people/zhang-jia-wei/posts>

他的文章

按时间排序

h2.ContentItem-title | 654 × 28

可以像 C 罗这样永远冲刺向前，本身就是种伟大天赋吧？

 张佳玮 

文学 话题的优秀回答者

2,526 人赞了该文章



约稿。2016 年的夏天，在波尔多，德科这么跟我说 C 罗。2004 年，C 罗在国家队还是个边路球员，喜欢那些漂亮的花式。这些年，他逐渐重视效率，变成了一个更纯粹的得分手，更加强壮了。那，有一点，C 罗是个很努... [阅读全文](#)

2.5K

214 条评论

分享

收藏

举报

Elements

Console

Sources

Network

Performance

Memory

Application

Security

Audits

Web Scraper

<div class="List-item">

<div class="ContentItem ArticleItem" data-zop="{\"authorName\": \"张佳玮\", \"itemId\": 38345088, \"title\": \"可以像 C 罗这样永远冲刺向前，本身就是种伟大天赋吧？\", \"itemprop\": \"zhihu:article\" itemtype=\"http://schema.org/SocialMediaPosting\" itemscope data-za-detail-view-path-module=\"Post\" {\"has\_image\": false, \"has\_video\": false, \"content\": {\"type\": \"Post\", \"token\": \"38345088\", \"upvote\_num\": 2526, \"comment\_num\": 214, \"pu... 1529587566000, \"author\_member\_hash\_id\": \"f9de84865e3e8455a09af78bfe4d1da5\"}}}>

<h2 class="ContentItem-title">

<a href=\"//zhuanlan.zhihu.com/p/38345088\" target=\"\_blank\" data-za-detail-view-element\_name=\"Title\">可以像 C 罗这样永

</h2>

<div class="ContentItem-meta">...</div>



# 例子 | 知乎标题区域

<https://www.zhihu.com/people/zhang-jia-wei/posts>

```
<div class="List-item">
```

```
.....
```

```
</div>
```

web scraper: `div.List-item`

自己写: `div.ContentItem` or  
`div.List-item` or  
`div.ContentItem.ArticleItem`



<https://www.zhihu.com/people/zhang-jia-wei/posts>





# 例子 | 知乎

<https://www.zhihu.com/people/zhang-jia-wei/posts>

<button class="Button LikeButton ContentItem-action"

.....

2.5K

.....

</button>

web scraper: `button.Button.LikeButton`

自己写: `button.Button.LikeButton.ContentItem-action` or

`button.Button.LikeButton` or

`button.LikeButton`



# 例子 | 知乎

<https://www.zhihu.com/people/zhang-jia-wei/posts>



button.Button.LikeButton.ContentItem-action.-sitemap-select-item-selected | 81.25 x 34

2.5K 214 条评论 分享 收藏 举报

### 邓肯对波波维奇与马刺说过的，最冷的一个笑话

张佳玮    
文学话题的优秀回答者

```
<meta itemprop="url" content="https://www.zhihu.com/p/38345088">
<meta itemprop="datePublished" content="2018-06-21T13:26:06.000Z">
<meta itemprop="dateModified" content="2018-06-21T13:26:06.000Z">
<meta itemprop="commentCount" content="214">
<meta itemprop="image" content="https://pic4.zhimg.com/v2-02927ffa3cf39d527438ebb0624a58c9_r.jpg">
<div class="RichContent is-collapsed">
  <div class="RichContent-cover">...</div>
  <div class="RichContent-inner RichContent-inner--collapsed">...</div>
  <div class="ContentItem-actions">
    <button type="button" class="Button LikeButton ContentItem-action -sitemap-select-item-selected">...</button> == $0
    <button type="button" class="Button ContentItem-action Button--plain Button--withIcon Button--withLabel">...</button>
    <div class="Popover ShareMenu ContentItem-action">...</div>
    <button type="button" class="Button ContentItem-action Button--plain Button--withIcon Button--withLabel">...</button>
    <button type="button" class="Button ContentItem-action Button--plain">...</button>
  </div>
</div>
```



# 牛刀小试

[https://www.xiaohongshu.com/user/profile/  
55593a11484fb660cc777dd0?tab=notes](https://www.xiaohongshu.com/user/profile/55593a11484fb660cc777dd0?tab=notes)



# 牛刀小试

[http://tieba.baidu.com/f?](http://tieba.baidu.com/f?kw=%E5%B0%8F%E7%B1%B3&red_tag=q2525413622)

[kw=%E5%B0%8F%E7%B1%B3&red\\_tag=q2525413622](http://tieba.baidu.com/f?kw=%E5%B0%8F%E7%B1%B3&red_tag=q2525413622)



web scraper不是万能