

**提问:【迫切希望您能讲一下!! 我们爬取豆瓣的时候网站是有25页显示电影信息。但是有的网站是一页显示全部信息,但是需要在底部点击加载更多,这个时候网页的网址并没有发生任何变化。这个时候我怎么爬取一个网页上的全部信息呢?(比如说政府留言板)】**

---

## 什么是异步加载的数据?

### 表现形式: 6个例子

1. NHK新闻:<https://www3.nhk.or.jp/news/special/coronavirus/latest-news/>
2. 人民网:<http://liuyan.people.com.cn/threads/list?fid=5062&position=1>
3. B站评论区:<https://www.bilibili.com/video/BV1Mf4y197cj>
4. 百度地图:<https://map.baidu.com/@11585725,3557560,15.71z>
5. 网易邮箱注册:<https://mail.163.com/register/index.htm>
6. 花瓣网:<https://huaban.com/search/?q=%E7%BE%8E%E9%A3%9F>

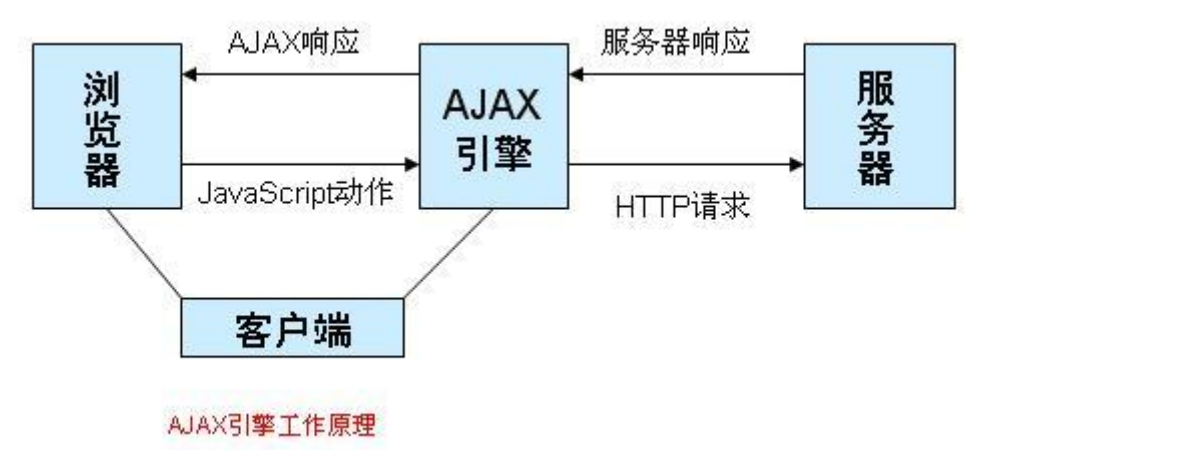
辅助工具:

- Chrome浏览器调试工具
- Postman

### 应用场景

- 优化用户体验
  - 用户行为分析
  - 提升开发效率
-

实现原理:



概念	一句话简述	参考资料
AJAX	通过Javascript和XML异步获取数据的技术	<a href="https://www.runoob.com/ajax/ajax-tutorial.html">https://www.runoob.com/ajax/ajax-tutorial.html</a>
异步	url路径不变,部分刷新页面,动态加载内容	
JavaScript	一个操作浏览器元素的脚本语言	<a href="https://www.runoob.com/js/js-tutorial.html">https://www.runoob.com/js/js-tutorial.html</a>
XML	可以自定义标签的标记语言	<a href="https://www.runoob.com/xml/xml-tutorial.html">https://www.runoob.com/xml/xml-tutorial.html</a>
JSON	文本格式,本质是Javascript对象	<a href="https://www.runoob.com/json/json-tutorial.html">https://www.runoob.com/json/json-tutorial.html</a>
	python解析json	<a href="https://www.runoob.com/python3/python3-json.html">https://www.runoob.com/python3/python3-json.html</a>

辅助工具:

JSON在线解析: <https://www.sojson.com/>

相同内容,JSON和XML比较

XML:

```
<?xml version="1.0" encoding="UTF-8"?>
<note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
</note>
```

JSON:

```
{
  "note": [
    {"to": "Tove"},
    {"from": "Jani"},
    {"heading": "Reminder"},
    {"body": "Don't forget me this weekend!"}
  ]
}
```

---

## 这样的数据如何获取？

---

### 0. 判断是否为异步加载

### 1. 找到异步请求的连接并分析规律

- 参数
- 规律

### 2. 获取返回的JSON数据并解析

- 转换
- 解析