

知乎—爬虫

2017 年 8 月 3 日

1 模拟登录

1. 知乎爬取时必须带上header
2. `_xref` 可以从登录网页的源码中提取，但是提取和登陆的网页必须是同一个。`urllib2.urlopen` 是打开不同的，可以使用requests库建立session。

2 爬取follow , 问题等内容

知乎网页是动态加载问题，怎么解决？