

# 第一组：电影票比价网站

## 1、制作目的

如今大家看电影的购票方式大多都是网络购票，可随着电影票团购网越来越多，在哪个网站进行购票才能取得最大的优惠，这成为一个困扰我们的问题。通常情况下，我们会一个一个的去打开不同的网站，比较他们之间的价格，最后选择一个优惠最多的软件，但无疑，这种方式浪费了大量的时间在比价上面。我们期望作出一个比价站点，通过爬取主流的购票网站实时电影票售价，将不同网站同一部电影的价格呈现给用户，让用户在本网站便可轻松的找到最优惠的价格。

## 2、主要爬取站点及数据使用方式

我们主要爬取的站点有：[猫眼电影](#)、[时光网](#)、[淘票票](#)、[糯米电影](#)、[bilibili](#)、[豆瓣网](#)、[高德地图](#)

抓取内容和数据使用方式见下表

站点	抓取内容	使用方式	已发现反爬措施
<a href="#">猫眼电影</a>	电影院链接、名字、地区、电影播放时间、价格等	比价	字体反爬、链接反爬
<a href="#">时光网</a>	电影院链接、名字、地区、电影播放时间、价格等	比价	动态Ajax加载
<a href="#">淘票票</a>	电影院链接、名字、地区、电影播放时间、价格等	比价	必须登陆
<a href="#">糯米电影</a>	电影院链接、名字、地区、电影播放时间、价格等	比价	
<a href="#">bilibili</a>	视频	播放电影预告片	虚假链接
<a href="#">豆瓣网</a>	评论	爬取关于电影的评论，制作词云	ip限制
<a href="#">高德地图</a>	经纬度	用于计算使用地与电影院之间的距离	

## 3、数据存储方式

对于电影院名字、电影院位置等极少发生改变的数据，我们将其存储在 `SQLite` 数据库中。

对于电影播放时间、电影票价格等会更新的数据，我们进行实时爬取

## 4、预期呈现效果

我们预期制作出一款以比价为主要目标的网站，其中还包含电影预告片、由电影评论生成的词图等信息