爬**https://www.vmgirls.com/心得**

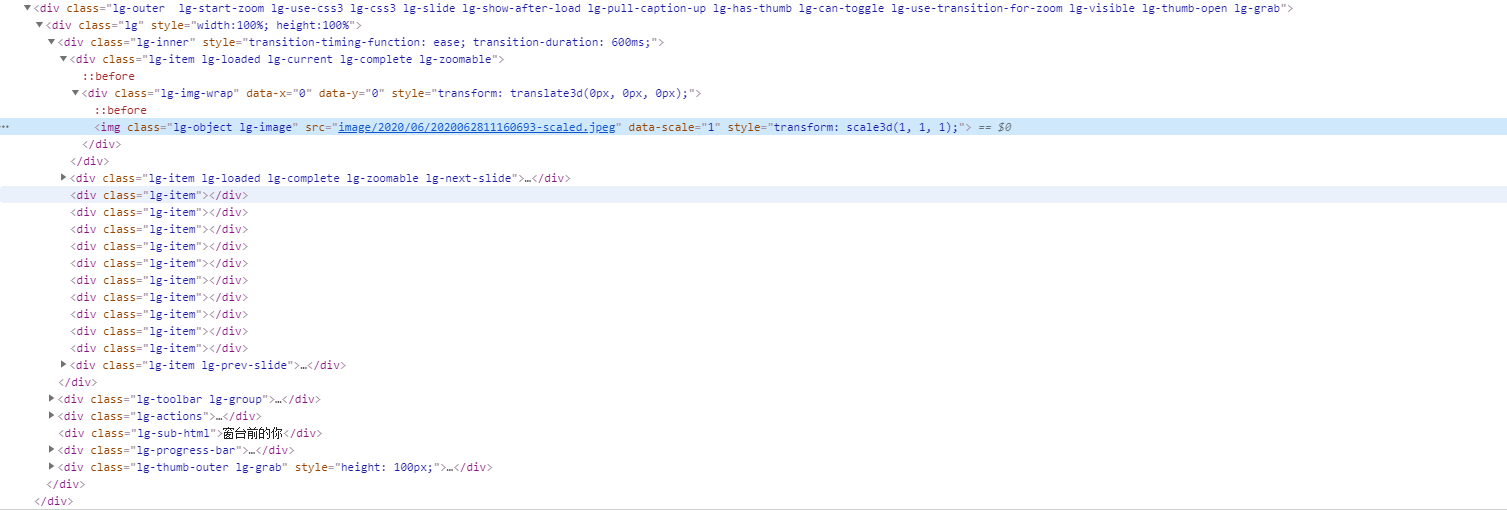
一、分析单个图片页面

<https://www.vmgirls.com/13992.html>本页面的13张图片，双击某张图片后，会进入另一个页面，[https://www.vmgirls.com/13992.html#lg=1&slide=0](https://www.vmgirls.com/13992.html" \l "lg=1&slide=0)，滑动会显示下一张图片，url也会变动，[https://www.vmgirls.com/14244.html#lg=1&slide=1](https://www.vmgirls.com/14244.html" \l "lg=1&slide=0)、......、[https://www.vmgirls.com/14244.html#lg=1&slide=1](https://www.vmgirls.com/14244.html" \l "lg=1&slide=0)2，即slide从0 到12 之间循环。

通过页面分析，13992.html得到url



和slide=0 得到url



相同，故只需对13992.html解析，得到13张图片url下载即可

通过对爬取页面内容解析发现，<**h1 class="post-title h3"**>窗台前的你</**h1**> ，就是图片的标题，每一张图片链接的title都是此名称，





故，title=html.xpath(**"//h1[@class='post-title h3']/text()"**)[0]

urllist=html.xpath(**f"//a[@title='{title}']/@href"**)

就可得到标题和url列表

1. 分析首页

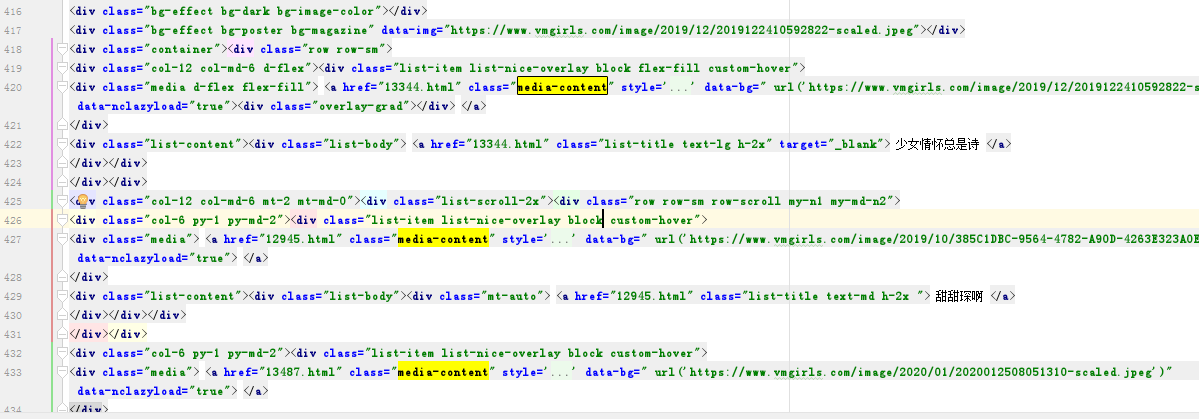
此网站没有对页面进行分页，虽然有菜单、分类，但实际下，部分类容重复，所有主题内容都在首页上。

通过对页面分析发现，每个主题的url都在



补充完整就可得到二级页面的地址。

通过对爬取页面分析，



仅 urllist=html.xpath(**"//a[@class='media-content']/@href"**) 就可得到 二级页面列表

1. 多线程下载

将所有下载url找到后，用title 作文件夹目录，下载图片url中带年月日，取做文件名，放到下载列表中，