**总结:**

- 尽量使用find_elements_by_xpath--获取不到不会报错
- 不能准确定位，类名中有空格的，一律使用xpath
- xpath分组后，在使用xpath时候，需要加（.//） 点表示当前节点

```
# coding=utf-8
from selenium import  webdriver
import time

class DouyuSpider:
    def __init__(self):
        self.start_url = "https://www.douyu.com/directory/all"
        self.driver = webdriver.Chrome()

    def get_content_list(self):
        li_list = self.driver.find_elements_by_xpath("//ul[@id='live-list-contentbox']/li
        content_list = []
        for li in li_list:
            item = {}
            item["room_img"]=li.find_element_by_xpath(".//span[@class='imgbox']/img").ge
            item["room_title"] = li.find_element_by_xpath("./a").get_attribute("title")
            item["room_cate"] = li.find_element_by_xpath(".//span[@class='tag ellipsis']
            item["anchor_name"] = li.find_element_by_xpath(".//span[@class='dy-name elli
            item["watch_num"] = li.find_element_by_xpath(".//span[@class='dy-num fr']").
            print(item)
            content_list.append(item)
        #获取下一页的元素
        next_url = self.driver.find_elements_by_xpath("//a[@class='shark-pager-next']")
        next_url = next_url[0] if len(next_url)>0 else None
        return content_list,next_url

    def save_content_list(self,content_list):
        pass


    def run(self):#实现主要逻辑
        #1.start_url
        #2.发送请求，获取响应
```

```python
        self.driver.get(self.start_url)
        #3.提取数据，提取下一页的元素
        content_list,next_url = self.get_content_list()
        #4.保存数据
        self.save_content_list(content_list)
        #5.点击下一页元素，循环
        while next_url is not None:
            next_url.click()
            time.sleep(3)
            content_list,next_url = self.get_content_list()
            self.save_content_list(content_list)


if __name__ == '__main__':
    douyu = DouyuSpider()
    douyu.run()
```