

爬虫步骤

获取数据

- `urllib.request.Request()`---改变用户代理为浏览器访问
- `urllib.request.urlopen(req)`---获取网页源码
- `response.read().decode('utf-8')`---读出并规格化源码

解析数据

- `BeautifulSoup(html,"html.parser")`--创建bsDOM树
- `soup.find_all('div',class_="item")`---找到网页中最小所需单元
- `re.compile(r'')` ---设置匹配规则
- `re.findall(findLink,item)`----使用re库进行模式匹配出需要的数据

保存数据

- `xlwt.Workbook(encoding="utf-8",style_compression=0)`---创建workbook对象
- `book.add_sheet("豆瓣电影Top250",cell_overwrite_ok=True)`---创建excel表
- `sheet.write(0,i,col[i])`----写入数据
- `book.save(savepath)`---保存数据