

Java 爬虫之 当当图书畅销榜近 24 小时 TOP500

作者: [ab](#)

原文链接: <https://ld246.com/article/1526979263322>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

```

package com.example.demo.web;
import org.jsoup.Jsoup;
import org.jsoup.nodes.Document;
import org.jsoup.nodes.Element;
import org.jsoup.select.Elements;

import java.io.IOException;

public class aa {
    public static void main(String\[\] args) {
        StringBuilder sb = new StringBuilder();
        //当当数据是一页20条, 共25页
        for (int i = 1;i <= 25;i++){
            Document document;
            String url = "http://bang.dangdang.com/books/bestsellers/01.00.00.00.00.00-24hours
0-0-1-"+i;
            try {
                document = Jsoup.connect(url).timeout(50000).get();
                Elements es = null;
                if(document != null){
                    es = document.getElementsByAttributeValue("class","bang\_list\_box").first().getE
lementsByTag("li");
                }
                for (Element element : es) {
                    //获取排名
                    Element ele = element.getElementsByAttributeValue("class","list\_num").first();
                    if (ele == null){
                        ele = element.getElementsByAttributeValue("class","list\_num red").first();
                    }
                    try {
                        String id = ele.text();
                        sb.append(id + " ");
                    }catch (Exception e){
                        //System.err.println("ele id:"+element);
                    }
                    //获取封面
                    Element link = element.getElementsByAttributeValue("class","pic").first();
                    try {
                        Element ing = link.getElementsByTag("img").first();
                        String href = ing.attr("src");
                        sb.append(href+ ", ");
                    }catch (Exception e){
                        //System.err.println("ele href:"+element);
                    }
                    //获取书名
                    Element name = element.getElementsByAttributeValue("class","name").first();
                    try {
                        Element nam = name.getElementsByTag("a").first();
                        String na = nam.text();
                        sb.append(na + ", ");
                    }catch (Exception e){
                        // System.err.println("ele nam:"+element);
                    }
                    //获取作者, 出版社, 出版日期

```

```

Elements publisher_info = element.getElementsByAttributeValue("class","publisher_info");
    for (Element a : publisher_info){
        try {
            Elements publisher = a.getElementsByTag("a");
            for (Element b : publisher){
                String publishe = b.text();
                sb.append(publishe + "|");
            }
        }catch (Exception e){
            //System.err.println("ele publisher:" + element);
        }

        try {
            Element c = a.getElementsByTag("span").first();
            String publi = c.text();
            sb.append(publi + ", ");
        }catch (Exception e){}
    }
    //获取价格
    Element price = element.getElementsByAttributeValue("class","price").first();
    try {
        Element price_r = price.getElementsByAttributeValue("class","price_r").first();
        String pric = price_r.text();
        sb.append(pric + " ").append("\n");
    }catch (Exception e){
        //System.err.println("ele price_r:" + element);
    }
    }
    } catch (IOException e) {
        e.printStackTrace();
    }
    }
    System.out.println(sb.toString());
}
}
}

```

不会贴图，效果图不贴了...