

JAVA网络爬虫实战

管理学院电子商务研究所

钱洋

- 开发环境的搭建(JDK的安装、环境变量配置、Eclipse安装与操作)
- 基本的数据类型(int/double等)
- 数组操作
- 条件判断与循环
- 集合操作
- 对象和类
- 正则表达式
- Maven工程的创建
- Log4j的使用
- 多线程

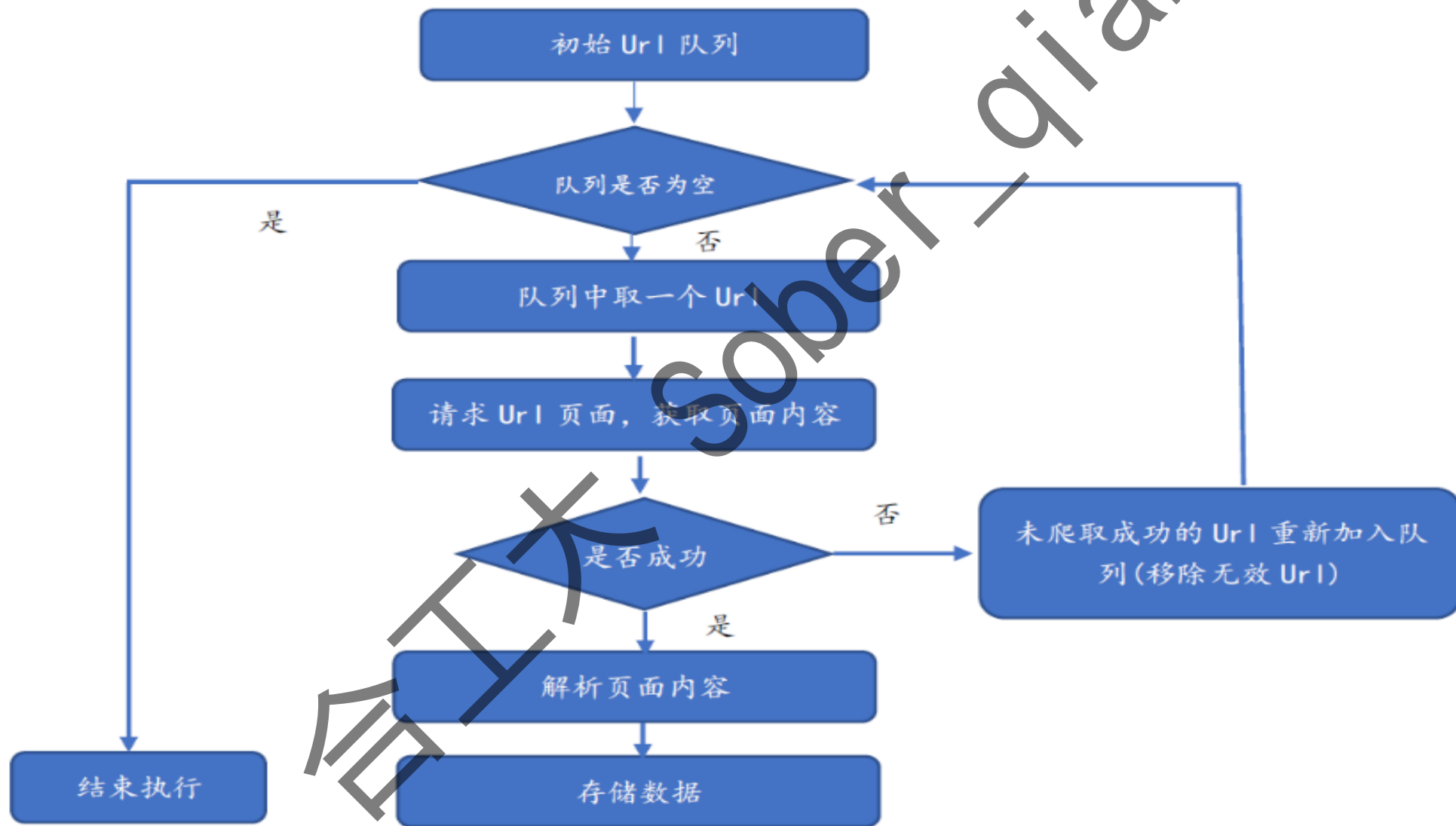
- HTTP协议与网络抓包
- 网页内容获取工具(Jsoup、HttpClient、URLConnection以及OkHttp)
- 网页解析(HTML/XML/JSON)
- 网页内容存储(文件流/Excel/MySQL/NoSQL)—Redis, RocksDB等

```
graph TD; A[网页内容存储] --> B[文本文件/图片、PDF、视频等文件]; B --> C[框架的使用]; C --> D[多线程数据采集、断点数据采集、代理维护等]
```

文本文件/图片、PDF、视频等文件

- 框架的使用Crawler4j、WebCollector以及WebMagic

多线程数据采集、断点数据采集、代理维护等



□ 数据采集目标：中关村在线手机产品相关信息(以OPPO为例)

URL: http://detail.zol.com.cn/cell_phone_index/subcate57_1673_list_1_0_1_2_0_2.html



□ 数据采集目标：中关村在线手机产品相关信息(以OPPO为例)

URL: http://detail.zol.com.cn/cell_phone_index/subcate57_1673_list_1_0_1_2_0_2.html

参考报价: 即将上市 [京东推荐手机](#)

重要参数: 主屏尺寸: 6.2英寸 主屏分辨率: 1520x720像素
前置摄像头: 1600万像素 电池容量: 4230mAh

[+ 加入对比](#)

新闻: 有奖调查:OPPO A7用户购买意向调查
2018-11-14 9

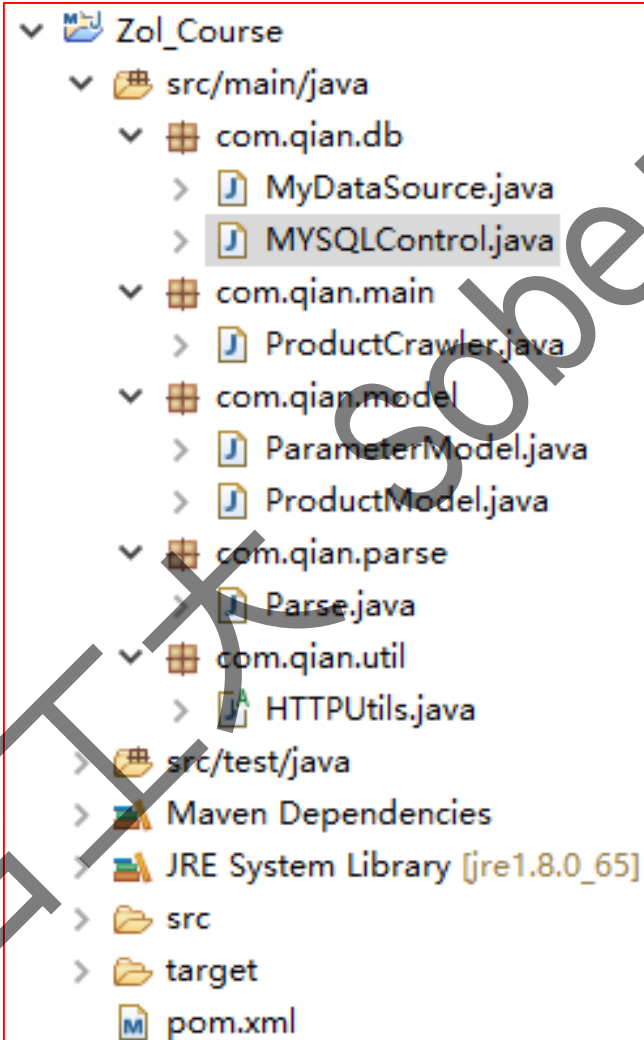
颜色: 共65张图片>>

配置参数 [详细参数](#)

主屏尺寸: 6.2英寸 主屏分辨率: 1520x720像素	后置摄像头: 1300万像素+200万像素 前置摄像头: 1600万像素
电池容量: 4230mAh 电池类型: 不可拆卸式电池	核心数: 八核 内存: 3GB

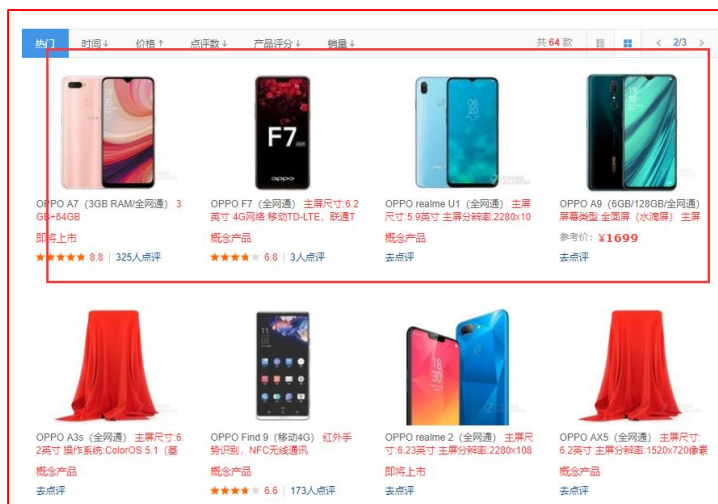
□ 程序编写步骤

(1) 框架内容



□ 程序编写步骤

(2) 确定要采集的字段



产品id、产品名称、
产品URL、产品价格



产品id、产品配置

□ 程序编写步骤

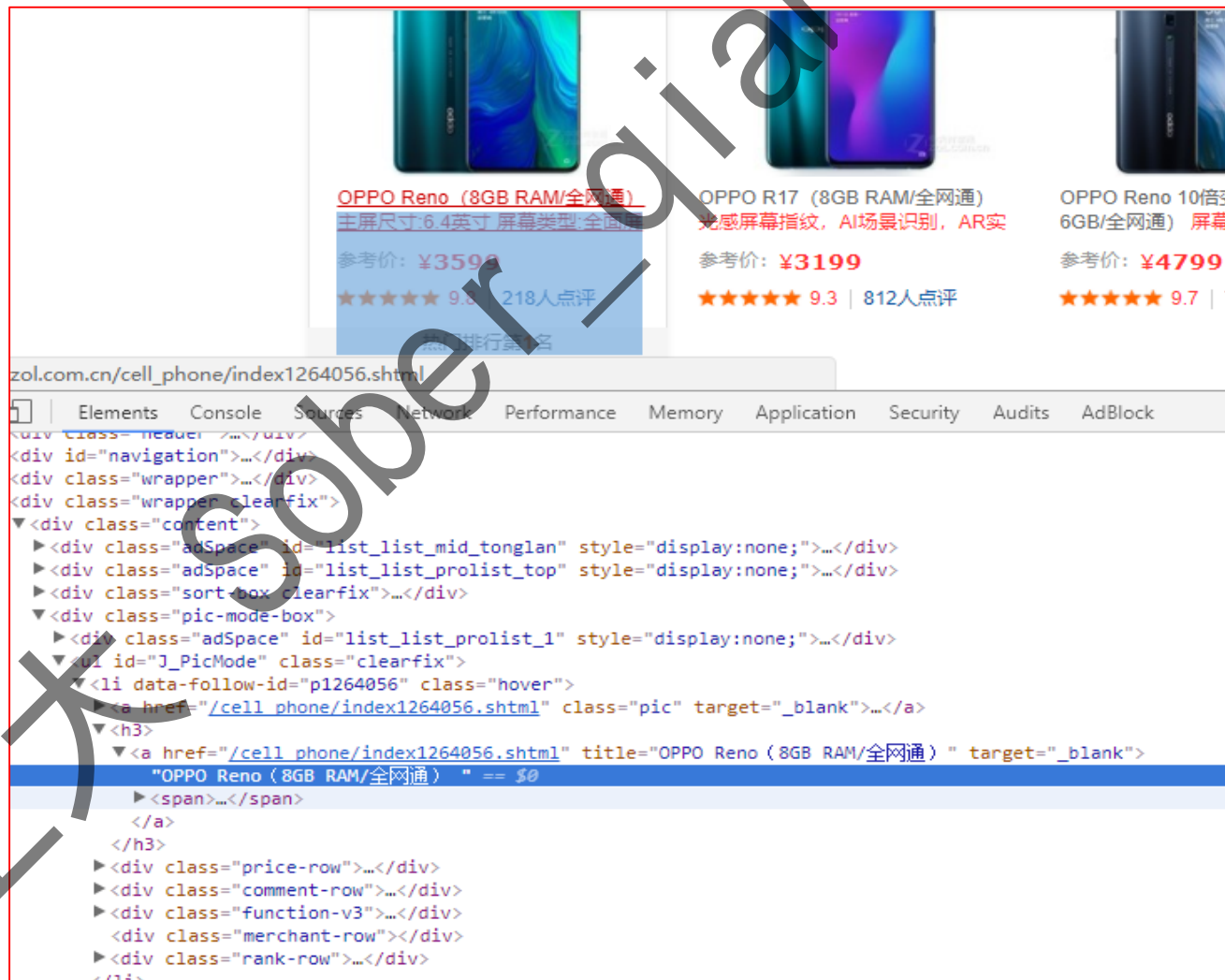
(3) 编写model

(4) 编写HTTPTUtils

(5) 编写Parse

(6) 编写db

(7) 编写main



□ 数据采集结果

product_id	product_name	product_url	product_price
p1115953	OPPO R9 (全网通) 美颜4.0, VOOC快充, 金属机身, 超窄边框	http://detail.zol.com.cn/cell_phone/index1115953.shtml	2499
p1166362	OPPO 5X Zoom (全网通) 主屏尺寸:0英寸 4G网络:移动TD-LTE, 联通TD-L	http://detail.zol.com.cn/cell_phone/index1166362.shtml	概念产品
p1167114	OPPO F3 (全网通) 主屏尺寸:0英寸 4G网络:移动TD-LTE, 联通TD-LTE, 联	http://detail.zol.com.cn/cell_phone/index1167114.shtml	停产
p1170998	OPPO R9sk (全网通) 微缝天线, 固态指纹按钮, VOOC快充, 双核对焦	http://detail.zol.com.cn/cell_phone/index1170998.shtml	2599
p1183683	OPPO MIX (全网通) 主屏尺寸:0英寸 主屏材质:AMOLED 4G网络:移动TD-L	http://detail.zol.com.cn/cell_phone/index1183683.shtml	概念产品
p1184068	OPPO R11s (4GB RAM/全网通) 全面屏一体全金属外观, 超清双摄像头模式	http://detail.zol.com.cn/cell_phone/index1184068.shtml	2299
p1206990	OPPO R15 (6GB RAM/全网通) 19:9超视野全面屏, 双面玻璃设计, AI 智能	http://detail.zol.com.cn/cell_phone/index1206990.shtml	2399
p1207284	OPPO 可折叠手机 (全网通) 主屏尺寸:0英寸 4G网络:移动TD-LTE, 联通TD-L	http://detail.zol.com.cn/cell_phone/index1207284.shtml	概念产品
p1208720	OPPO R15梦镜版 (梦境红/全网通) 19:9超视野全面屏, 双面玻璃设计, AI 智	http://detail.zol.com.cn/cell_phone/index1208720.shtml	2999
p1209152	OPPO F7 (全网通) 主屏尺寸:6.2英寸 4G网络:移动TD-LTE, 联通TD-LTE, 联	http://detail.zol.com.cn/cell_phone/index1209152.shtml	概念产品
p1209829	OPPO A1 (3GB RAM/全网通) AI智慧美颜, 面部识别, 全面屏, 快速充电	http://detail.zol.com.cn/cell_phone/index1209829.shtml	1000
p1210656	OPPO R15梦镜版 (陶瓷黑/梦境紫/全网通) 19:9超视野全面屏, 双面玻璃设计	http://detail.zol.com.cn/cell_phone/index1210656.shtml	3299
p1211599	OPPO A3 (全网通) 128G大内存, 钻石纹理设计, 超视野全面屏, AI智慧美颜	http://detail.zol.com.cn/cell_phone/index1211599.shtml	1599
p1212068	OPPO R15 (4GB RAM/全网通) 4GB+128GB	http://detail.zol.com.cn/cell_phone/index1212068.shtml	2399
p1212408	OPPO A1 (4GB RAM/全网通) AI智慧美颜, 面部识别, 全面屏	http://detail.zol.com.cn/cell_phone/index1212408.shtml	1399
p1213467	OPPO Find X (标准版/全网通) 双曲面柔性屏, 全隐藏式3D摄像头, VOOC快	http://detail.zol.com.cn/cell_phone/index1213467.shtml	4999
p1213525	OPPO realme 1 (全网通) 主屏尺寸:0英寸 4G网络:移动TD-LTE, 联通TD-LT	http://detail.zol.com.cn/cell_phone/index1213525.shtml	停产
p1213828	OPPO R15星云特别版 (全网通) 19:9超视野全面屏, 双面玻璃设计, AI 智能	http://detail.zol.com.cn/cell_phone/index1213828.shtml	2699
p1214344	OPPO R15s (全网通) 主屏尺寸:6.2英寸 操作系统:Android 8.1 主屏材质:TFT	http://detail.zol.com.cn/cell_phone/index1214344.shtml	概念产品
p1220645	OPPO Find X兰博基尼版 (全网通) 双曲面柔性屏, 全隐藏式3D摄像头, VOC	http://detail.zol.com.cn/cell_phone/index1220645.shtml	9999
p1220675	OPPO Find X (国际版/全网通) VOOC快充, 3D面部识别, 全隐藏式3D摄像	http://detail.zol.com.cn/cell_phone/index1220675.shtml	即将上市
p1221126	OPPO A5 (4GB RAM/全网通) 面部识别, 智能相册, OPPO互传, 全面屏	http://detail.zol.com.cn/cell_phone/index1221126.shtml	1500
p1221131	OPPO A5 (移动定制版/全网通) 面部识别, 智能相册, OPPO互传	http://detail.zol.com.cn/cell_phone/index1221131.shtml	即将上市
p1221830	OPPO Find X (超级快充版/全网通) 双曲面柔性屏, 全隐藏式3D摄像头, VO	http://detail.zol.com.cn/cell_phone/index1221830.shtml	5499
p1222890	OPPO A3s (全网通) 主屏尺寸:6.2英寸 操作系统:ColorOS 5.1 (基于Androi	http://detail.zol.com.cn/cell_phone/index1222890.shtml	概念产品
p1224911	OPPO F9 (全网通) 主屏尺寸:0英寸 CPU型号:联发科 Helio P60 (MT6771)	http://detail.zol.com.cn/cell_phone/index1224911.shtml	即将上市
p1225806	OPPO R17 (8GB RAM/全网通) 光感屏幕指纹, AI场景识别, AR实景导航, ;	http://detail.zol.com.cn/cell_phone/index1225806.shtml	3199
p1225826	OPPO R17 Pro (6GB RAM/全网通) Super VOOC快充, TOF 3D立体摄像头	http://detail.zol.com.cn/cell_phone/index1225826.shtml	3299
p1228346	OPPO F9 Pro (全网通) 主屏尺寸:6.3英寸 操作系统:Color OS 5.2 (基于And	http://detail.zol.com.cn/cell_phone/index1228346.shtml	概念产品
p1228566	OPPO AX5 (全网通) 主屏尺寸:6.2英寸 主屏分辨率:1520x720像素 屏幕像素	http://detail.zol.com.cn/cell_phone/index1228566.shtml	概念产品
p1228622	OPPO realme 2 (全网通) 主屏尺寸:6.23英寸 主屏分辨率:2280x1080像素 屏	http://detail.zol.com.cn/cell_phone/index1228622.shtml	即将上市

□ 数据采集结果

product_id	parameter
p1115953	主屏尺寸: 5.5英寸主屏分辨率: 1920x1080像素后置摄像头: 1300万像素前置摄像头: 1600万像素电池容量:
p1166362	主屏尺寸: 暂无数据主屏分辨率: 暂无数据后置摄像头: 暂无数据前置摄像头: 暂无数据电池容量: 暂无数据
p1167114	主屏尺寸: 暂无数据主屏分辨率: 暂无数据后置摄像头: 暂无数据前置摄像头: 暂无数据电池容量: 暂无数据
p1170998	主屏尺寸: 5.5英寸主屏分辨率: 1920x1080像素后置摄像头: 1600万像素前置摄像头: 1600万像素电池容量:
p1183683	主屏尺寸: 暂无数据主屏分辨率: 暂无数据后置摄像头: 暂无数据前置摄像头: 暂无数据电池容量: 暂无数据
p1184068	主屏尺寸: 6.01英寸主屏分辨率: 2160x1080像素后置摄像头: 2000万像素+1600万像素前置摄像头: 2000万
p1206990	主屏尺寸: 6.28英寸主屏分辨率: 2280x1080像素后置摄像头: 1600万像素+500万像素前置摄像头: 2000万
p1207284	主屏尺寸: 暂无数据主屏分辨率: 暂无数据后置摄像头: 暂无数据前置摄像头: 暂无数据电池容量: 暂无数据
p1208720	主屏尺寸: 6.28英寸主屏分辨率: 2280x1080像素后置摄像头: 2000万像素+1600万像素前置摄像头: 2000万
p1209152	主屏尺寸: 6.2英寸主屏分辨率: 暂无数据前置摄像头: 2500万像素后置摄像头: 暂无数据电池容量: 暂无数据
p1209829	主屏尺寸: 5.7英寸主屏分辨率: 1440x720像素后置摄像头: 1300万像素前置摄像头: 800万像素电池容量: 3
p1210656	主屏尺寸: 6.28英寸主屏分辨率: 2280x1080像素后置摄像头: 2000万像素+1600万像素前置摄像头: 2000万
p1211599	主屏尺寸: 6.2英寸主屏分辨率: 2280x1080像素后置摄像头: 1600万像素前置摄像头: 800万像素电池容量:
p1212068	主屏尺寸: 6.28英寸主屏分辨率: 2280x1080像素后置摄像头: 1600万像素+500万像素前置摄像头: 2000万
p1212408	主屏尺寸: 5.7英寸主屏分辨率: 1440x720像素后置摄像头: 1300万像素前置摄像头: 800万像素电池容量: 3
p1213467	主屏尺寸: 6.42英寸主屏分辨率: 2340x1080像素后置摄像头: 2000万像素+1600万像素前置摄像头: 2500万
p1213525	主屏尺寸: 暂无数据主屏分辨率: 暂无数据后置摄像头: 暂无数据前置摄像头: 暂无数据电池容量: 暂无数据
p1213828	主屏尺寸: 6.28英寸主屏分辨率: 2280x1080像素后置摄像头: 1600万像素+500万像素前置摄像头: 2000万
p1214344	主屏尺寸: 6.2英寸主屏分辨率: 1520x720像素后置摄像头: 1300万像素+200万像素前置摄像头: 800万像素
p1220645	主屏尺寸: 6.42英寸主屏分辨率: 2340x1080像素后置摄像头: 2000万像素+1600万像素前置摄像头: 2500万
p1220675	主屏尺寸: 6.42英寸主屏分辨率: 暂无数据后置摄像头: 1600万像素+2000万像素前置摄像头: 2500万像素

□JDK下载

□MySQL

□Navicat

<http://114.213.252.89/>

Filename

[Anaconda3-4.2.0-Windows-x86_64.exe](#)

[Anaconda3-5.2.0-Windows-x86_64.exe](#)

[Chrome_46.0.2490.80.rar](#)

[Chrome_55.0.2883.75%5B.rar](#)

[CrawlerJarUpdate.zip](#)

[CrawlerStudy/](#)

[CrawlerTeaching.zip](#)

[DG2015Setup_1185E.exe](#)

[Data/](#)

[EditPlus4.0.0.626.exe](#)

[Endnote_8/](#)

[Feige5.1.161207_for_Windows.exe](#)

[Git-2.11.0-64-bit.exe](#)

[HP_LaserJet_M1130-M1210_Series_I.zip](#)

[IBM_SPSS_Statistics_21.0_x86.exe](#)

[Linux_OS/](#)

[MathType6.9b_Cracked_CHS.exe](#)

[MemTest64.exe](#)

[Microsoft_Toolkit.exe](#)

[Office_2013_64.rar](#)

[Office_2013_X86.rar](#)

□ 采集每款产品的评论信息

<http://detail.zol.com.cn/1240/1239500/review.shtml>

OPPO A7 (3GB RAM/全网通) 点评

综述介绍 报价 图片 (65) 参数 点评 (325) 竞品对比 评测行情 论坛 问答 视频 维修

快速点评 高级点评

8.8

★★★★★

8.5

性价比

7.0

性能

9.3

续航

9.0

外观

8.8

拍照

325人评分

4-5星	94.2%
3-4星	2.5%
2-3星	0%
1-2星	0.3%

大家都说:

外观漂亮(40)

续航出色(40)

性价比高(33)

系统流畅(27)

电池容量大(23)

手感舒适(19)

机身轻薄(14)

外观不是很满意(6)

机身内存小(3)

续航一般(3)

最有用 有图 最新 长篇 说不好

好·说 带上它去生活 为它写一篇好说吧

□ 采集每款产品的评论信息

<http://detail.zol.com.cn/1240/1239500/review.shtml>

最有用

有图

最新

长篇

说不好



sina 70503533bw

优质点评人

购买信息
价格: 3199元
时间: 2018-12
地点: 北京

★长篇

静谧而美好, 将内涵深藏的OPPO R17

★★★★★ 9.6

续航: 10 拍照: 10 性能: 8 外观: 10 性价比: 10

优点: 炫酷的光感屏幕指纹解锁 好看的水滴屏及91.5%的屏占比 让工作与生活更高效的ColorOS5.2系统

缺点: 没有NFC功能 没有无线充电功能 虽然搭载了快充, 但是如果搭载超级闪充就更好了

前些天不小心摔碎了手机的屏幕, 去维修, 换个屏幕的价格太贵了, 还是决定直接换新手机。在B站偶然中看到了关于R17的测评, 感觉各方面都很实用, 挺适合我使用的。仔细想了一下, 第二天就去官网下单了。下单后闲来无事

[查看全文>](#)



发表于: 2018年12月21日 01:47

3条评论 218赞



原小欧

★长篇

内外兼修, OPPO A7上手体验

★★★★★ 9.2

续航: 10 拍照: 8 性能: 8 外观: 10 性价比: 10



谢谢