

1 国内市场

当前，国内已有一个比较成熟的电商网购比价平台，名为“慢慢买”：www.manmanbuy.com。该平台小米应用商店下载量已达 1 亿+（淘宝 136 亿、京东 80 亿、拼多多 213 亿、唯品会 32 亿），虽然用户体量比不上其面向的产品 app，但总体来看其实并不小。此外，“慢慢买”在国内同类竞品市场中基本垄断，无论是广告宣传层面还是 app 下载量层面都基本属于唯一龙头，故选取该 app 作为参考。

“慢慢买”注册为杭州磅礴科技有限公司，旗下几乎只运营“慢慢买”这一个产品（其他都属于附属品），大概有 100 人左右。

慢慢买公司介绍

公司简介

慢慢买成立于2010年7月，是一家电商数据分析技术公司，有多年的商品数据沉淀和分析挖掘经验。公司在海量商品数据采集、存储、清洗、分析挖掘、可视化、历史价格等领域积累了大量核心技术，我们通过消费行为、商品销量、价格趋势、用户口碑等维度，帮助用户更高效的进行决策。

我们针对消费者和企业提供了不同的服务：

消费端产品：慢慢买

慢慢买APP、网站、小程序等平台，为消费者提供全网比价、历史价格查询和最新给力的折扣推荐等服务。多年来我们专注于为用户推荐高性价比的商品，同时开发了全网比价、历史价格查询等购物助手，目标是让用户高效的买到高性价比商品。

企业端产品：比价接口

我们依靠尖端的大数据分析和云计算技术，对海量网购数据进行高效的监控采集和挖掘分析，积累了50多家电商平台，超8000万商品的相关数据，并结合行业需求研发出多款大数据应用产品，一站式解决企业采购、定价、营销及管理等多种问题，助力品牌精准决策。

该公司目前在 boss 直聘上热招的岗位主要有下，基本可以转义为实现比价的技术路线：

（1）审核编辑：人工采集电商平台的热门商品信息；对用户提交的“有需求但还未比价”的商品进行手动补充；每日筛选整理电商平台优惠咨询

（2）C# Web 开发：网页端的建设和维护

（3）爬虫工程师：设计和开发爬虫的监控和报警系统；APP、小程序、h5 的数据抓取；研究爬虫策略和防屏蔽规则，解决封账号、封 IP、验证码、页面跳转等难点，提升网页抓取的效率和质量。

APP 使用体感：可以控制商品不变，对比当前各平台价格；也可以控制商品和平台不变，对比当前和历史价格，预警“假促销”；还可以接受淘宝/京东个人店铺的手动爆料（例如我们的商品承诺是全网最低，则我向慢慢买提供证据以及链接，慢慢买会手动录入我们的商品信息并加以推送）；最近还出了“餐饮外卖”模块，和我们之前的思路很像，也是对比美团、淘宝闪购、京东、饿了么等平台，但是目前只支持大型连锁餐饮，例如奶茶店、KFC、麦当劳、瑞幸库迪、汉堡王等。

总结：整体运作逻辑比较古老，比价数据靠爬虫，比价分析靠人工，对岗位的招聘要求也很低。有意思的是对爬虫工程师的要求还拓展到防屏蔽，怀疑是为了躲避淘宝、京东等公司的防爬虫监控。后续调查发现，除了躲避自己爬别人公司时的监控，还要用 token 保护自己的信息不被别人爬取（原来是有开放 API 免费调用的，前几年给关了，原因未知），但是 2022 年在 github 已经有大神破解了他们的监控 hhhhhhhhhh：

https://github.com/PPsteven/manmanbuy_js_crack。后来又有了新的反爬政策：

今天粗略看了一下，慢慢买当前的反爬措施与我文档中的写的方法并无太大的差别。
只是现在慢慢买更新了，需要登录才能查看信息，这意味着，如果慢慢买记录了用户的信息的话，一旦爬虫频率变高的话可能会使得账户被封。
我可以过两天更新一个新的脚本，并不能保证账户不会被封。

本教程仅作为技术讨论使用，并不鼓励批量爬取慢慢买数据。

至此再无更新。

7、无经验可接受，普通且过度自信不接受

8、学历要求全日制大专及以上

慢慢买部分招聘要求

技术实现的可能性：在有充足时间的情况下，“网购比价”agent 可以完全通过爬虫来获取各平台的价格信息，然后传输给 LLM 做分析，最后再以自然文本或者链接的形式反馈给用户。相当于用 LLM 代替了“慢慢买”公司中的非技术人员，让 LLM 来进行比价分析。在允许的情况下，我觉得爬虫代码都有可能交给 LLM 来写，只是需要我们提前做好提示词工程。

API 调取的可能性：

淘宝：2020 年之后个人用户很难调取了，感觉只能用图像识别或者强行爬虫这样，毕竟 agent 可以逐步完成输入、检索、识别等功能。

京东：能拿到图片下载的 api，但商品详情（例如价格数据）的 api 个人用户暂不支持，应该也是只能图片识别和强行爬虫。

拼多多同理。

主要实现涉及到爬虫、前端开发、php 等技术。