

Reddit 内容获取工具链

OpenClaw 中常见的 Reddit 研究/采集流程，按任务复杂度组合搜索、抓取、浏览器和脚本能力

1. 任务输入与判断

先判断目标属于哪一类：单帖查找、多帖汇总、评论深挖、还是批量结构化采集。

2A. 轻量路径：搜索 + 抓取

web_search

搜索关键词，定位 subreddit、帖子 URL、外部讨论入口

web_fetch

抓取帖子正文或页面文本，快速做内容摘要与筛选

2B. 重量路径：浏览器 / 自动化

browser.navigate

打开 Reddit 页面，进入真实渲染环境

browser.snapshot

识别页面结构、按钮、评论区域

browser.act

点击、滚动、展开评论、切排序、深入帖内交互

browser.screenshot / pdf

保存页面证据、导出截图或 PDF

3. 批量化与结构化处理

exec 运行 Python / Node 脚本，对多帖内容进行批量抓取、清洗、去重、关键词提取和结构化汇总。



4. 输出结果

最终输出通常包括：观点摘要、高频关键词、情绪/主题归类、可引用帖子清单，以及可用于汇报的结论文案。

发现层

先用搜索找入口，确认是否值得继续深挖。

获取层

能静态抓取就不动浏览器，拿不到完整内容再升级到浏览器交互。

处理层

当任务需要批量化、结构化、统计分析时，再进入脚本执行阶段。