

6.1 网页爬虫——软件爬虫

2023 年 9 月 8 日

6.1 网页爬虫——软件爬虫



这里推荐使用八爪鱼采集器，它是模拟人的采集方式，比如打开网页、点击网页的按钮等对网页上的信息进行全自动提取，提供了一种无需编程就可以爬取数据的方法。

安装方法如下：1. 访问网址：<https://www.bazhuayu.com> 2. 安装完成后，注册账号。3. 查看八爪鱼教程：<https://www.bazhuayu.com/tutorial8/81zdypz>

练习

本文通过应用数据采集软件（八爪鱼），从 JD 平台上采集和巧克力相关的数据。

采集方法如下：1. 搜索 JD 平台，使用关键词“巧克力”，获取到链接地址 2. 应用八爪鱼上智能爬取功能，采集第 1 步的链接地址 3. 采集样本数为 1000+ 个样本，并保存为 xlsx 格式 4. 转换 xlsx 格式文件为 csv 格式文件