**Report 10**

**Part 1: 调查并简要研究三款热门的爬虫软件，在下表中简要介绍这三款爬虫软件，并比较这三款爬虫软件的相似与不同之处，最后简要分析它们的优劣.**

|  |  |  |  |
| --- | --- | --- | --- |
| **软件名称** | **简介** | **比较** | **优劣** |
| **Scraper** | Scraper是一款Chrome扩展程序，适用于在线研究和数据提取。无代码爬虫插件软件 | 操作简单，能够满足基本需求但是可拓展性相较于其他两个软件较为鸡肋。 | 免费易用，适合初学者和专家。  能够将数据导出到Google Sheets。  自动生成XPath，简化数据抓取过程。  可视化工具 |
| **HTTrack** | HTTrack是一款免费且功能强大的网站爬虫软件，它允许用户下载整个网站到本地计算机。 | 免费相较于其他软件，且可以直接将网站内容下载到本地之后，再开始爬虫，不会对于网站服务器造成网络攻击 | 支持多平台，包括Windows、Linux和Unix系统。  能够镜像网站，包括图片、文件、HTML代码等。  用户可以设置下载选项，如并发连接数。  提供代理支持，可通过身份验证提高下载速度。 |
| **亮数据（Bright Data）** | 亮数据是一款强大的数据采集工具，以其全球代理IP网络和强大数据采集技术而闻名。它能够轻松采集各种网页数据，包括产品信息、价格、评论和社交媒体数据等。 | 功能强大且齐全，支持各种各样的服务，住宅代理，抓取自动化，可以定时抓取其中的数据，无需人工操作，且对于采集的数据会给予分析，但是需要收费，并且收费内容较高。 | 全球网络数据采集：提供一站式服务，将全网数据转化为结构化数据库。  商用代理网络：拥有超过7200万个IP，覆盖195个国家，每日更新上百万IP。  高效数据采集：能够达到170000请求/秒，每天处理高达1PT的网络流量。  技术驱动：拥有超3300项授权专利申报，持续引领行业创新。  稳定性：提供99.99%的稳定运行时间，即使在网络高峰期间也能保持稳定. |