On the Viability of Manual Scraping

Created on: 09-26-2023 10:10

WHAT

- process of manually retrieving data from websites without automated tools

PROS

control and flexibility

- high degree of precision when scraping
- human scraper can easily adapt to any website structures

lightweight

- potentially quicker than automated scraping
- ideal for small static databases

CONS

error-prone

- humans have larger overhead when switching between scraping targets

time-consuming

- does not scale well with data size
- not feasible for realtime requirements

HOW

inspect element

- copy paste the data from the browser's inspection panel

copy paste

- copy paste the data directly from the webpage

screenshots and OCR

- apply OCR on captured screenshots and convert to text data

paid services

- pay freelancers to do the scraping

CONCLUSION

In the context of our project, manual scraping is great for compiling a small collection of data for initial testing.

It is also useful as a backup solution for any websites that are incompatible with automated scraping methods.

Beyond that, it is less than ideal for populating the production database.