

Data Patterns:

an exploration of big data and (mis)information collection

Work submitted to Sabine Rosenberg

As part of CART 451 (Interactive Media)

By

Sarah Hontoy-Major (40177636)

Concordia University

Faculty of Fine Arts

November 9th, 2022

The goal of this project was originally to reflect and bring a greater sense of awareness on the extent to which companies with an online presence have access to high levels of personal information from their clients and platform users. Taking the giant online retailer Shein as a case study, I wanted to find ways to incorporate similar technologies the corporation uses on my own online platform, while making my users aware of the data that is unknowingly harvested from them. Implementing similar technologies of an omnipresent giant corporations like Shein was however far too unrealistic within the resources of this project. The Shein inspired aesthetic I had envisioned for the project's online interface also failed to translate the original goal of battling against opaque infrastructures by bringing transparency and awareness to the process of personal data collection.

At this point in the project, research on Shein's own tracking and collection methods revealed there was little relevance in implementing those systems unto my project's platform if my goal is to affect the public by showing them how much websites can know about them. Instead, I have turned towards public access APIs and solutions that are easier to implement. Without the intention to do so originally, my project has slightly shifted to understanding how 'simple' it is to get a maximum of personal information about virtually anyone (willing to access your website). However, this remains a research and exploration project, as I understand increasingly better what it takes to infer information from provided data.

As of now, I was able to gather information about a client's IP address upon entering on the website in addition to its geolocation coordinates, by using an IP lookup API. From there, I was able to infer precise civic addresses by using google maps' geocode API and posting back street view and map view of the location with google maps API. All of that was done without the client needing to accept or decline browsers' usual alert about sharing one's geolocation

information. By inferring all information from the client's IP address, a public domain information, I was able to bypass the client's browser security wall and collect the data without the client's knowledge. This, in a much more simplistic way, parallels what giant companies like Shein infer from information collected from their users, at a much greater scale.

Though I have used up most of what can be gathered publicly without knowledge or warning on the client side, I still wish to understand how far people are willing to go to know how much websites can collect from them. The next implementations on the website will therefore consist of social media login APIs to get a better understanding of the information that transit from social media platforms to other online platforms that require those logins. Preliminary research hints towards different platforms collecting and sharing different sets of data, which will be interesting to research, explore, and document (auth0 API). Another avenue of exploration could be the usage of VPNs and how much we are able to know from its specific IP address' download history. For that, the 'I Know what you Download' API will be explored and implemented if time and relevancy allows for it (IKnowWhatYouDownloadAPI). To infer statistics from the project's online platform usage, such as the percentage of people willing to share their social media accounts on the website, creating unique IDs for all users may be helpful to generate useful information. Again, if time and relevancy allow for it, a database, likely on MongooseDB, will have to be implemented to keep track of those stored information. For privacy reasons, no personal and/or sensitive information will get stored in that database in regard to clients visiting the website; these will get deleted whenever the window closes. The only information stored in the database would help infer conclusions from clients' behaviours on the website, such as buttons clicked.

Lastly, as mentioned previously, the UI presented in the prototype acts as a placeholder and a general idea of the original intention of the project. However, experimenting with this aesthetic made me realized it was not conveying the original goal of the project as I intended at first. Instead, a commercial looking website slowly fragmenting, or 'breaking', as users visit and input more personal information into it will be preferred and explored. The anxiety inducing and overwhelming pop-ups mechanics will still be experimented with, as the idea of having 'information box' with personal information 'surprise' the user can be an interesting avenue to explore as well.

Bibliography

Auth0 Social Media login API, <https://auth0.com/learn/social-login>

IKnowWhatYouDownload API, <https://iknowwhatyoudownload.com/en/api/>