Data_scapes:

A study and exploration of data extraction and its possible inferences

Work submitted to Sabine Rosenberg

As part of CART 451 (Networked Media Studio)

By

Sarah Hontoy-Major (40177636)

Concordia University

Faculty of Fine Arts

December 12th, 2022

Technologies used

Node library enabled to query the ISP (Internet Service Provider), its coordinates, and information about the country of origin, and much more (see figure 1). This query can be made whenever a window loads and without client needing to grant permission, meaning the process can be done without the client`s awareness of it.

The IP API`s geolocation information was then sent to the Google Maps API, which used the coordinates to write both a google birds eye view maps and a street view map of the specific point. Using Google Map's complementary Geocode API, I used the former's geolocation information on the latter to infer precise civic addresses.

The project was started on a node framework, thinking it may collect data of some kind from the user, either clicking behaviour or more personal information they were willing to transmit. I however decided against it, as it did not embed itself well with the main topic and thesis of this project. The data collected from the subsequent technologies are simply deleted when the client closes the window. In the context of the current iteration of this project, there is little reason why it is still running on Node. The IP API goes through Node, though there are many other ways of running it that would be potentially more streamlined. However, because this project is an exploration rather experimental than practical, I decided to keep the entirety of the project on Node, so as to be able to easily implement certain libraries in future iterations of the project.

Implemented and possible features

I wanted to implement a lot more things in my project than I ended up doing. The original goal was, in part, to research on what corporations with online presence were using in terms of

collection and tracking technology and parallel those technologies in my project. After thorough investigation of certain websites' privacy policies, namely Shein's, I realized that even though these documents revealed *some* information, it was usually written in ambiguous contexts that did not reveal the full truth (see Figure 2). In the case of Shein, its interface conveniently forbids the client from copying the content of its privacy policy (see Figure 2 in github repository). Different tabs list what the law prescribes and nothing more, each section finishing with a catch-all "we may use your personal information for any other appropriate purpose […] which is reasonably necessary to provide the services […] requested" (Shein privacy policy) which allows them an extensive grey area in which to be able to collect, use, and share its clientele's personal information. The document vaguely mentioned the use of cookies and tracking technologies such as pixel tracking, without elaborating further on which third party platforms they were 'openly' sharing client personal data with (Shein privacy policy).

The research did not allow me to know what technology Shein and similar corporations were using, apart from tracking pixels and other industrial level technologies that I do not have the resource to use in meaningful ways. Instead, I used technologies in the general public's reach and attempted to retrieve as much information as I could from visitors without their knowledge. By without their knowledge, I mean without having permission being asked to the client's browser. Upon login unto the app, the client's isp is retrieved by the IP API, and the following google maps, street view, and geocode are inferred from it. At a micro level, I wanted to parallel how corporations infer information from their client on a much larger scale. Rewards based on what clients were willing to share are also mentioned in the proposal, though I decided against them because they too closely mirroring the mechanics of Shein I was trying to raise awareness on.

I also realized throughout my research that showing the general public the exact technology Shein or others were using would matter very little for them. Emotion and affect is what creates most impact; dry and overwhelming amounts of information will not cut for the majority of people (D'Ignazio and Klein 2020). Furthermore, in the words of Byung-Chul Han, "everything that is time consuming is on the way out. Truth is time consuming" (Han and Steuer 2022, 6). Therefore, I would rather create an impact that affects client and lets them think whatever they want to think about the information on the screen, then to provide context and information for them to read on – which they will likely not. I decided instead to prioritize showing them a low scale version of it, to give them an idea of what was possible to know, in the form of a very short narrative.

Upon loading the website, the client sees a rather blank window with the prompt: 'do you know what websites know about you?' (See Figure 3). After a few seconds, popup messages start appearing on the screen. They are not draggable, but they can be closed -- though they will open again a few seconds after. Popup messages are set to become increasingly invasive and private as they appear. The first ones are information about the client's country and about their specific time zone (see figure 4); following are information about their Wi-Fi provider, in which neighborhood they are situated, and their geolocation (see Figure 5). The subsequent windows are Google maps and street views of that specific location, and then the most precise civic address the geocode API could fetch, followed by hard coded and pointed messages to the client (see Figure 6, Figure 7, Figure 8 on Github repository).

Analysis of ideas and concept

The project had two goals: to research how corporations were collecting information from their online users, and then to showcase the findings accompanied by visitors' personal information inferred from such methods. While I wanted to give more information about privacy policy and what they entail to visitors, I also wanted to show them how much information I could collect from them in similar ways as what Shein and other online giants were doing. However, I realized that the companies and technologies they use to track and collect data are much more cryptic black boxes than their algorithms may be, which we often antagonize because of it. The vague privacy policy left big gaps left for interpretation, and technologies used were named sparsely. Even those that were identified, namely tracking pixels, were not implemented in the context of this project because they require cross platform applications and way more resources than the project allowed.

I also came to realize through the research that the information I wanted to share with the visitors was dense, nuanced, and incredibly technical; nothing that the general public would really care to read. Instead of making an informational website like I had first planned, I decided to instead prioritize my exploration of data collection and affect my visitors through it. The goal remained to attempt to 'shock' or bring awareness of a problem to the client, but the path taken would be emotional rather than informational; to affect rather than to teach.

The project was presented to a very small pool of people, all students of the same CART-451 course, which are not the target audience, as they all have a somewhat advanced understanding of data collection. Most of them were still shocked to understand how easily one could infer information from public data.

One of the main questions that still remained from my proposal was that, even if people knew how much companies were collecting from them, would it change their behaviour towards them? Would it create any change at all? Although it is only an anecdote which serves no proof, I found it funny and rather telling when a group of people from the class ended up gamifying the app by closing the pop-up messages in a whac-a-mole game mechanics. It made me think of how I believe Shein consumers would react if faced with the reality that my project discusses.

Bibliography

D'Ignazio, Catherine, and Lauren F. Klein. 2020. "On Rational, Scientific, Objective Viewpoints from Mythical, Imaginary, Impossible Standpoints" Data Feminism. Strong Ideas Series. Cambridge, Massachusetts: The MIT Press. P.73-96

Han, Byung-Chul, and Daniel Steuer. 2022. *Non-Things: Upheaval in the Lifeworld*. English edition. Cambridge, UK: Polity.

Shein Privacy Policy. "Shein Privacy Policy". June 1, 2022. Shein, https://ca.shein.com/Privacy-Security-Policy-a-282.html

```
{
  ip: '184.160.24.217',
  continent_code: 'NA',
  continent_name: 'North America',
  country_code2: 'CA',
  country_code3: 'CAN',
  country_name: 'Canada',
  country_capital: 'Ottawa',
  state_prov: 'Quebec',
  district: 'Rosemont-La Petite-Patrie',
  city: 'Montreal',
  zipcode: 'H2V 1C4',
  latitude: '45.52847',
  longitude: '-73.61071',
  is_eu: false,
  calling_code: '+1',
  country_tld: '.ca',
  languages: 'en-CA,fr-CA,iu,en-US',
  country_flag: 'https://ipgeolocation.io/static/flags/ca_64.png',
  geoname_id: '6097081',
  isp: 'Videotron Ltee',
  connection_type: '',
  organization: 'Videotron Telecom Ltee',
  currency: { code: 'CAD', name: 'Canadian Dollar', symbol: 'C$' },
  time_zone: {
    name: 'America/Toronto',
    offset: -5,
    current_time: '2022-12-10 12:38:20.832-0500',
    current_time_unix: 1670693900.832,
    is_dst: false,
    dst_savings: 1
  }
}
```


Figure 1 Information returned by the IP API

Other Purposes.	• We may use your personal information for any other appropriate purpose disclosed to you prior to you providing us your personal information, or which is reasonably necessary to provide the Services or other related products and/or services requested.

Marketing and advertising services	Assistance in reaching potential new customers across multiple communications channels, or sharing with affiliated companies that promote our products on their websites.

D. *Consent*. With your consent, we may disclose your personal information for any purpose that a reasonable person would consider appropriate in the circumstances.

H. *Do Not Track Signal*. Some Internet browsers, such as Internet Explorer, Firefox, and Safari, include the ability to transmit "Do Not Track" or "DNT" signals. Since uniform standards for "DNT" signals have not been adopted, our Site does not currently respond to "DNT" signals.

Figure 2 Shein's Privacy Policy on 1) personal data collection, 2) sharing personal data, 3) consent to share personal data, and 4) on tracking technologies

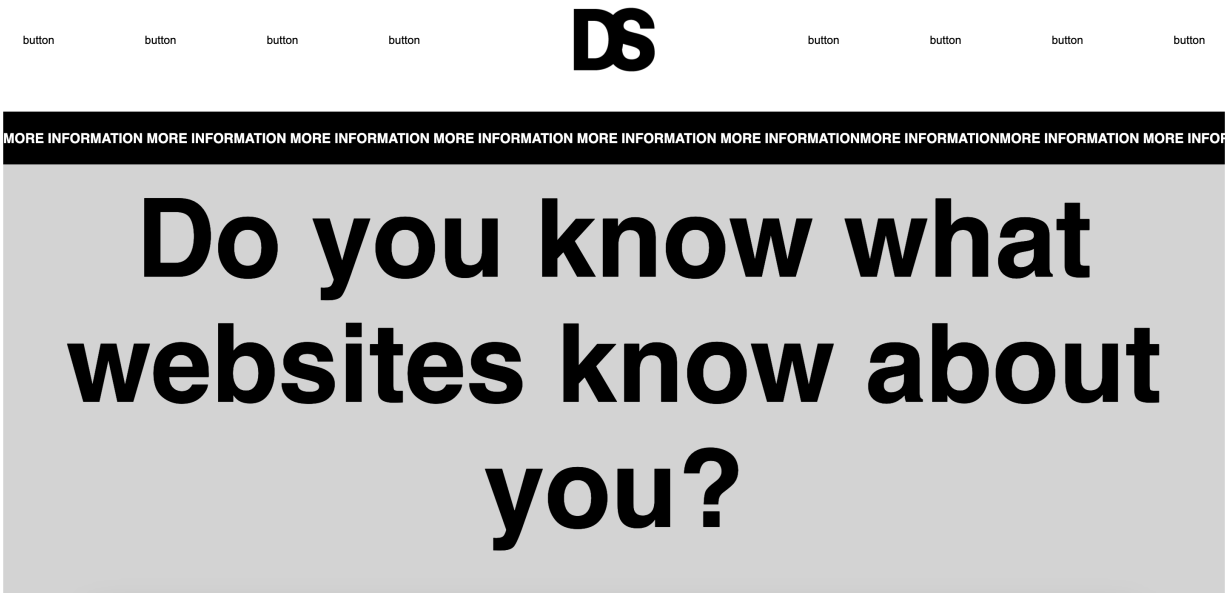

Figure 4 The first pop-up messages refer to the client's country and information about its flag, currency, capital, specific time zone of the client, ...

Figure 5 The first pop-up messages refer to the client's country and information about its flag, currency, capital, specific time zone of the client, ...

Figure 6 The second set of pop-up messages give information about the client's wifi provider, where it is located, the client's IP address and to what geolocation coordinates the address points to

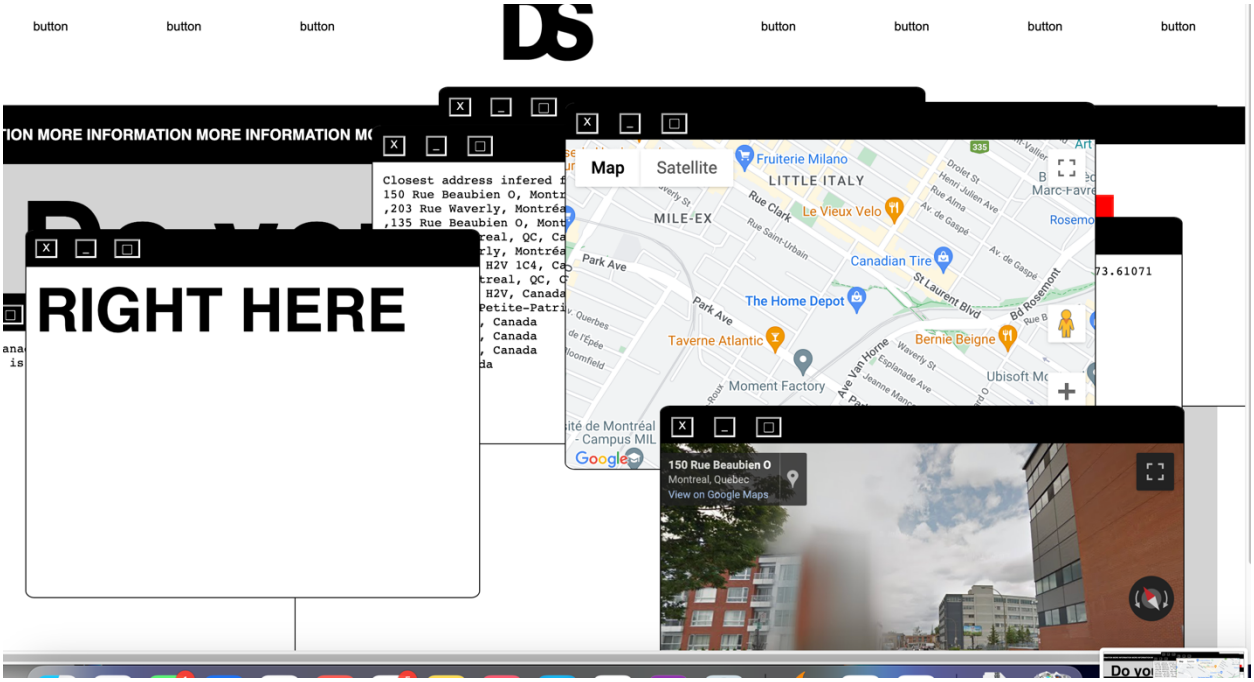

Figure 7 The next pop-up messages post different maps (birds eye view and street view) based on the geolocation provided in previous pop-up messages

Figure 8 The last pop-up messages are hard-coded strings to stress the information provided by previous pop-up messages