Final Report Of
CSIT 6910A Independent Project

Collect, Analyze and Visualize Campus Big Data

Name: YAN Ziqi

Student Number: 20543632

Supervisor: QU Huamin

The Hong Kong University of Science and Technology

1 Abstract

This project builds a heatmap by using wifi data. This wifi data includes the number of devices that connect to wifi access points and location of wifi access points.

The implementation of this project includes cleaning the row data and visualize these processed data. Row data is gotten from ITSC system periodically. The visualized result is a heatmap of the library. To visualize the data, javascript library D3.js and heatmap.js are used. Students can find the crowd level of each part of library by referring to this heatmap.

2 Content

3 Introduction

Nowadays, computer technologies are widely used in our life, among these technologies, data visualization technology is always used to create a smart campus, which provide teachers and students, moreover, everyone on campus, with a great deal of convenience.

The project ‘Pulse of HKUST’ is to build such a system. The system senses crowds on campus based on Wi-Fi access point usage information and CCTV videos, then visualize such crowd flows in an appealing and engaging form, and feed information to the on-screen display to promote campus data access and facility management.

The project connect to the ITSC system to get campus wifi data, then analyze them by machine learning or other data analytics technologies to figure out patterns behind data and try to predict trend of dynamic data. Finally, combine data visualization and data analytics, that is to say, visualize all these data and patterns, display these data on the screen, so that people on campus benefit from this in two ways. On the one hand, referring to these information when making decisions saves people lot of waiting time. On the other hand, by influencing people’s decision, this system will dispatcher the flow of crowd so as to prevent congestion.

This report shows a function, which creates a heatmap based on the map of library according to the campus wifi data gotten from ITSC. Students

are able to draw a conclusion easily that how crowded each part of the library is and make their decision where to go even before they enter the library. The method of processing data is quite efficient that it can generate a real-time heatmap from the data.

The way to visualize crowd level in library in this report can be also used to visualize crowed level in other places in campus, such as gym and canteen. By this way, it will save people in the campus much time by avoiding to squeeze into crowed.

4 Design

4.1 System structure

In this system, people browse the heatmap of the library on a web page, as figure 1, the client sent a request to server periodically asking for the number of device that is connecting to each access point, then the server sent back these data in a json file as figure 1.

So in order to see this heatmap effect on a web page, you need to build a local server in your computer first. There are many ways to build local server, for example, download phpStudy, add the directory your html file is in and reserve a port for it as figure 3, then enter “localhost:reserved port number” into the browser address bar to open the web page.

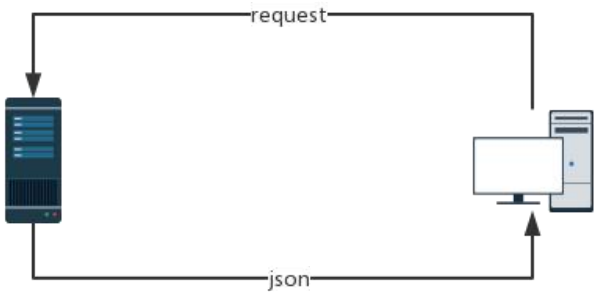


Figure 1

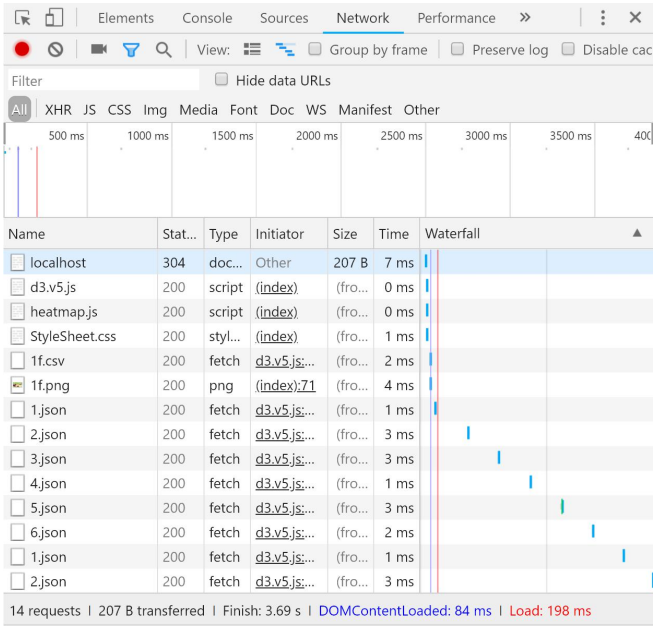


Figure 2

Figure 3

4.2 Data cleaning

To visualize the data provided by ITSC system, accurate coordinate of each wifi access point is needed, but ITSC only offer the photograph of computer screen as figure 4, so data cleaning should be done first.

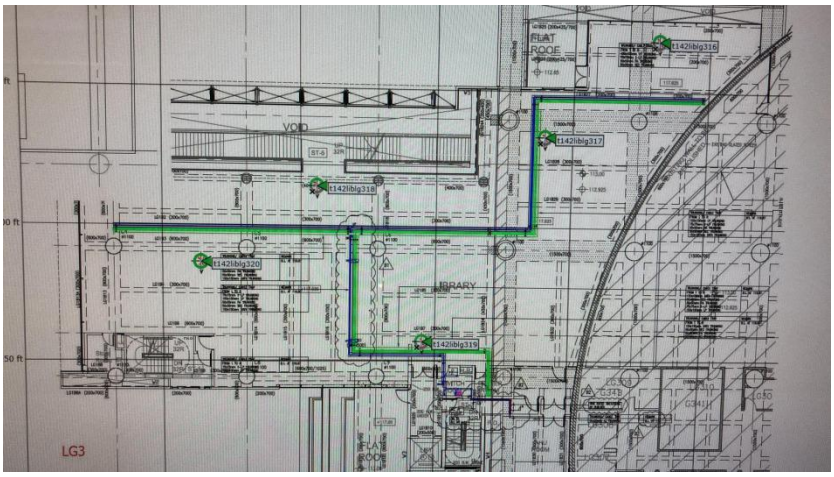


Figure 4

To extract coordinate of each wifi access point in the picture, use houghcircle function in opencv library to localize the center of green circle in the picture and output their coordinates, then align this picture with figure 5, convert coordinate in figure 4 to coordinate in figure 5. The converted coordinates are shown as figure 6.

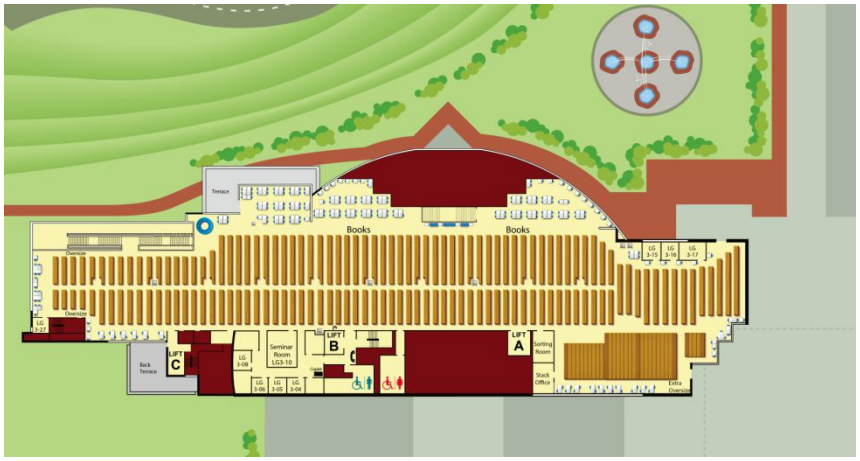
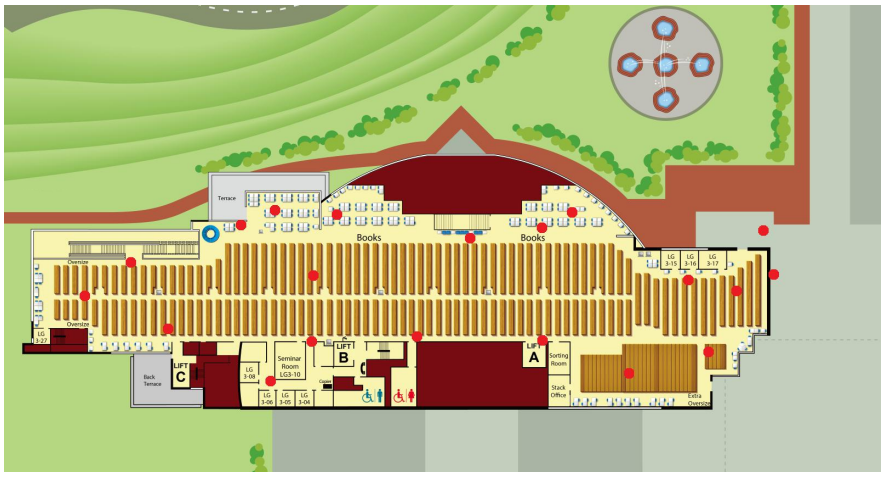


Figure 5

t142liblg	362	471
t142liblg	327	516
t142liblg	385	325
t142liblg	385	325
t142liblg	556	411
t142liblg	381	607
t142liblg	492	564
t142liblg	612	561
t142liblg	684	486
t142liblg	424	850
t142liblg	603	753
t142liblg	377	1036
t142liblg	405	981
t142liblg	610	982
t142liblg	670	1140
t142liblg	631	1285
t142liblg	500	1248
t142liblg	520	1336
t142liblg	490	1404
t601all10	410	1385

(a)



(b)

Figure 6

4.3 Related library

The implementation of the project use two javascript library, one is D3.js, another is heatmap.js.

4.3.1 D3.js

The full name of D3 is data-driven documents, which can be understood as the programming model whose drawing process is determined by data. D3 is a JavaScript library that is common used for data visualization.

Turning data into graphics is cumbersome and difficult to implement with native HTML, SVG, and Canvas. D3 encapsulates these for us, allowing developers to focus more on the layout and logic of the chart.

D3 support both SVG and Canvas. SVG and Canvas are HTML5 elements for drawing. SVG is used to draw vector graphics, while Canvas is used to draw scalar graphs. D3 3.x was previously based on SVG and mainly provides SVG drawing functions. D3 started support Canvas from 4.x.

D3 is commonly used to draw choropleth map to display population distribution, this might be suitable to draw the crowd distribution in every part of library, but choropleth map required a SVG map to base on, and most function in D3 is operate on SVG image. Thus this project use the functions in D3 that operate on Canvas only, combined with the functions in another library called heatmap.js to draw a heatmap on the map of library.

4.3.2 heatmap.js

Heatmap is a lightweight javascript library, which enable users visualize their data into dynamic heatmap for the web. It is quite efficient, all users need to do is just input several parameter, then they can draw an interactive web heatmaps within few lines of code.

4.4 Workflow

The workflow of this project is shown as figure 7.

First step is to clear interval, clearInterval function will interrupt the loop set by setInterval function. Then load image to Canvas. If the coordinate data corresponding to current floor hasn't been loaded from csv file, then load them. If the coordinate data has been loaded, then request for wifi data, ITSC will send present wifi data back, which contains the list of wifi access points, the number of devices connecting each wifi access points, and a time stamp, as shown in figure 8. This wifi data will be store in a json file, according to this json file, generate heatmap data and set them for a heatmap instance. Finally, set the interval, repeat the procedure that requesting wifi data, converting wifi data to heatmap data and setting heatmap data for heatmap instance if there's no click action, if there's a click event, then the current floor changed, image should be reloaded to Canvas and coordinate data should be renewed.

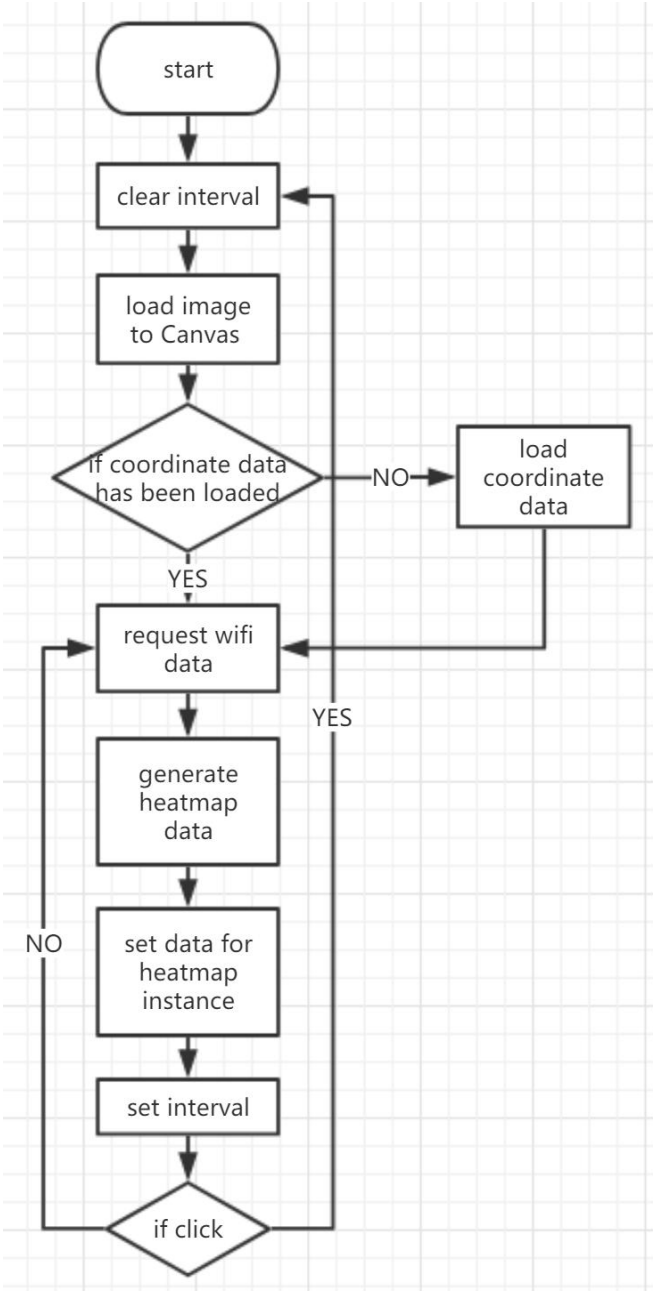


Figure 7


```
"data": {
  "t142liblg417": 0,
  "t142liblg419": 0,
  "t700lib108": 16,
  "t700libg09": 20,
  "t802liblg402": 1,
  "t700libg04": 61
},
"time": "201809121651"
}
```

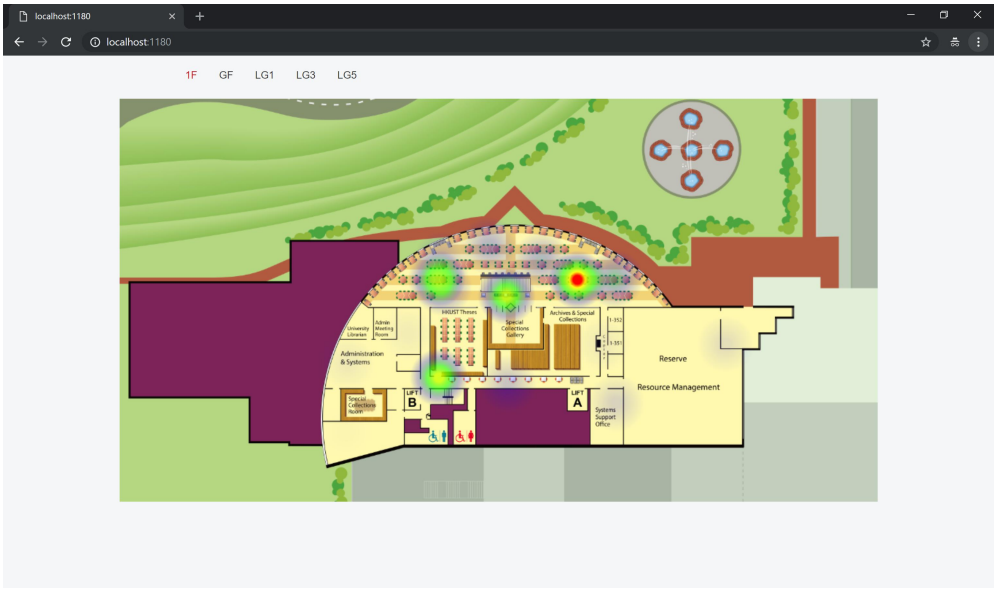

Figure 8

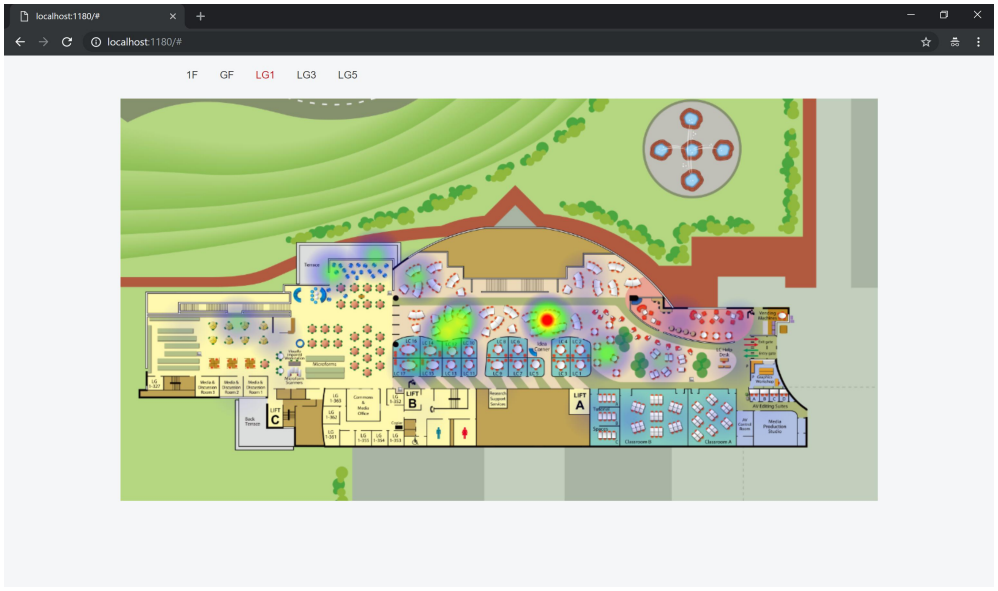
4.5 Experimental results

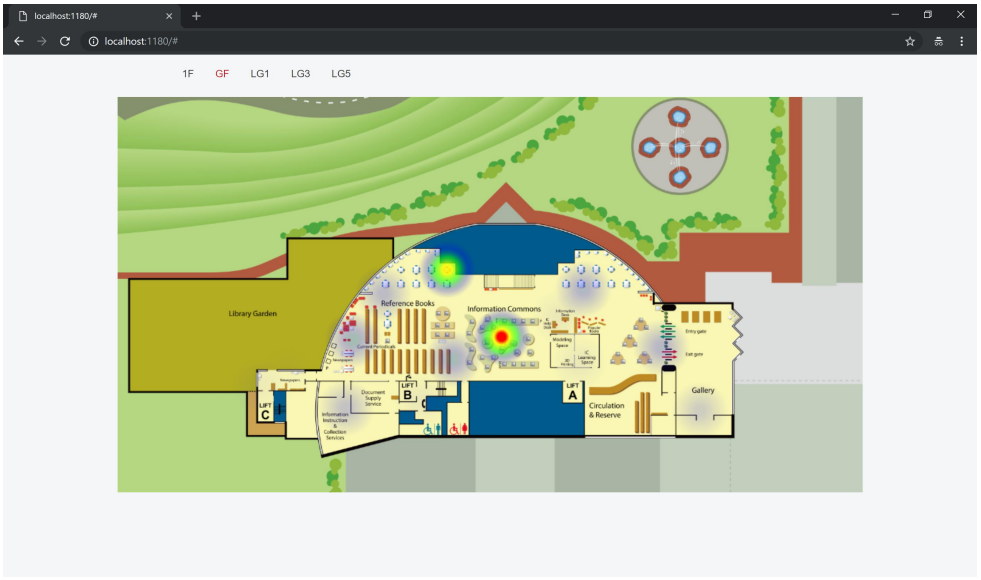
The library has five floors in total, heatmaps for each floor are shown as follows.

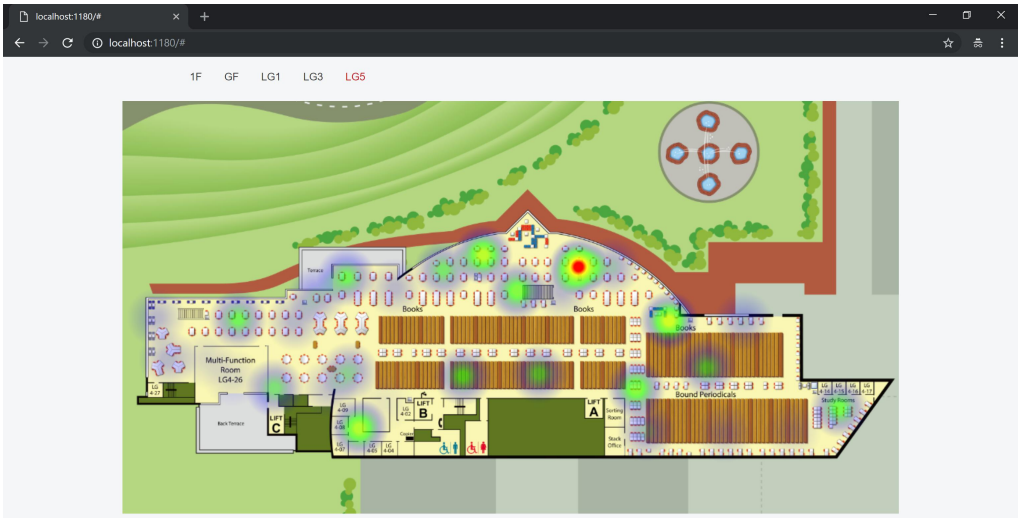
The color and the radius of the circle can be adjusted by changing the parameters.

The implementation of this project only use several example json file to create a simple dynamic heatmap.






localhost:1180/#
localhost:1180/#
1F GF LG1 LG3 LG5
Library Garden
Reference Books
Information Commons
Gallery
Circulation & Reserve
LIFT A
LIFT B
LIFT C


localhost:1180/#
localhost:1180/#
1F GF LG1 LG3 LG5
Books
Multi-Function Room LG4-26
Bound Periodicals
LIFT A
LIFT B
LIFT C


localhost:1180/#
localhost:1180/#
1F GF LG1 LG3 LG5
Books
LIFT A
LIFT B
LIFT C

5 Conclusion and future work

This project successfully visualize some sample wifi data into a heatmap on the map of library to enable people to figure out the crowd level of each part of library intuitively.

However, there's still much left to be improved.

First, the coordinates of wifi access points is not accurate, the picture is a photo of compute screen, because of the view angle of camera, the map might be distorted, so the coordinate might be not so accurate. Furthermore, the map given by ITSC and the map used in this project are not same, the difference between their scale would increase this distortion.

Second, the dynamic heatmap is only based on several sample data, an API is needed to get real wifi data from ITSC, so as to test the feasibility of this project. Maybe after modification, this function can be appended to school's library web page for students to refer.

Third, this heatmap is based on fixed wifi access points, which means the crowd level of each part of library is not so accurate for there are distances between people and wifi access point. People moves while wifi access points never move, thus this heatmap can not represent real crowd distribution on the map. The way to improve might be requesting explicit location of each devices to generate the heatmap.

6 Reference

1. Patrick-wied.at. (2018). *heatmap.js Documentation*. [online] Available at: https://www.patrick-wied.at/static/heatmapjs/docs.html [Accessed 19 Dec. 2018].

2. Kanit Wongsuphasawat, D. (2018). *D3 Tutorial by the Interactive Data Lab, University of Washington*. [online] Uwdata.github.io. Available at: http://uwdata.github.io/d3-tutorials/ [Accessed 19 Dec. 2018].

3. GitHub. (2018). *d3/d3*. [online] Available at: https://github.com/d3/d3/wiki/Tutorials [Accessed 19 Dec. 2018].

4. Bostock, M. and Bostock, M. (2018). *Let's Make a Map*. [online] Bost.ocks.org. Available at: https://bost.ocks.org/mike/map/ [Accessed 19 Dec. 2018].

5. D3.decembercafe.org. (2018). *Learning D3.JS*. [online] Available at: http://d3.decembercafe.org/ [Accessed 19 Dec. 2018].