AI meets Big Data: User Analytics and Personalized Recommendation Based on Location Data

&

Indoor Localization and Mobile Computing

Students: Mu-Ruei Tseng(Data collection, API design)

Yuxiang Zhu(Data collection, API design)

Professor: CHAN, Gary Shueng Han

UROP 1100x Report

Abstract:

This paper describes the methodology for indoor localization and the management of the data system. The localization method makes use of Arduino's wifi model to collect wifi signals in indoor areas to build the signal database. Once data was collected, we will transfer them to a content management system for better organization and analysis. When a user navigates in the building, his cellphone will receive wifi signals, and by comparing to the database, the actual position of the user can be obtained. Tests have been done in the first and second floor of the Academic building in HKUST. The result appears to be accurate but with a slow refresh rate.

I. Introduction

Indoor localization is an important yet undeveloped area of modern technology. Unlike localization for outdoor areas can be done with the help from the satellite, we need to have

alternatives when it comes to obtaining a precise position for the indoor. The current method we are using is to collect the wifi signal received from user's cellphone and compares it to the signal database in the building, which needs to be done by doing the site survey. A site survey is when we use the Arduino model to collect signals in the entire building, majorly wifi signals. However, we will also observe other information like the gravity and the power of the magnetic field. We create a content management system to help organize the results from each site survey and others can access the database and down building configurations with ease.

II. Current Progress

We are assigned to majorly do two parts, the collecting and generating fingerprints from the site surveys and building a system to help better manage them.

1. Site Survey

a. Collecting data

To do a site survey, we first connect the cellphone to one Arduino for collecting wifi signals. Then, we have to download the target building configuration to the phone and obtain the aisles information that people will travel regularly. We make use of the SigInt site survey app which was developed by the team. It allows us to choose the floor we want to do the survey on and draw polygons or lines to define the region of interest. After all the settings are done, we will simply just travel through the drawing area on the map at a normal pace. The phone will track all the signal accordingly and save the collected data in the specific folder.

b. Generating Fingerprint

To generate the fingerprint, we have to first zip all the data files in the phone and transfer onto a computer with a python environment set up beforehand. We next use the

command line to run the wifimap_legacy.py script that requires input parameters building data directory, project directory, and reference point distance. We can obtain a fingerprint.txt file and a fingerprint map after executing the command.

2. **Build the Content Management System:**

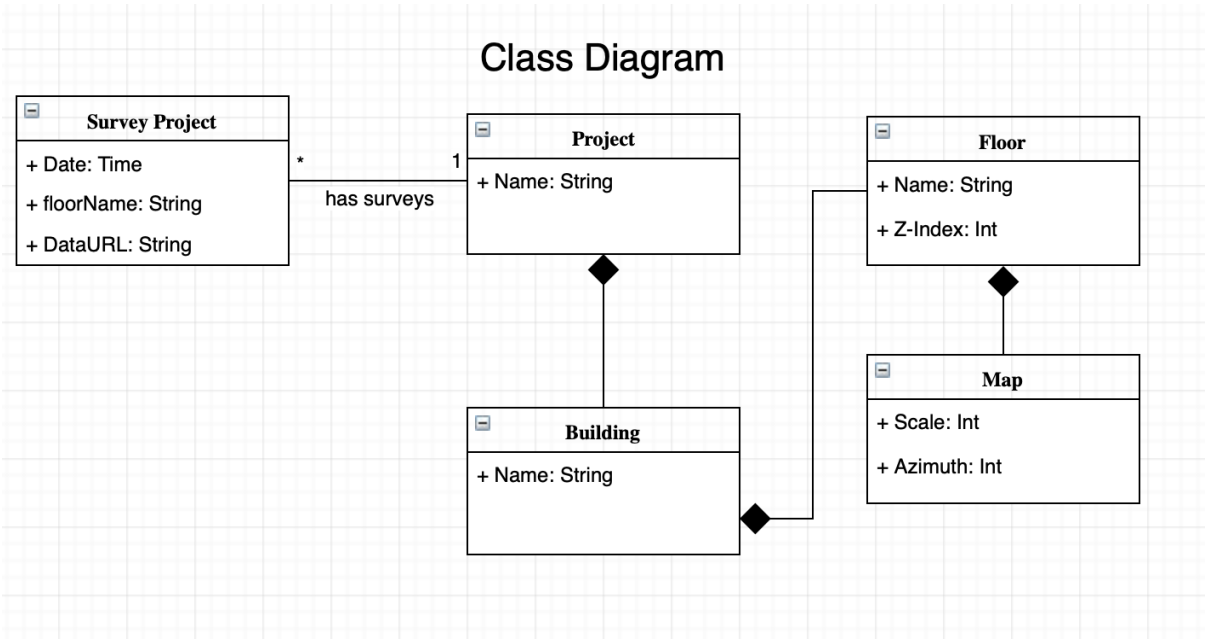
To build the content management system, also called as the CMS, we make use of one of the headless CMS open-source, Genetics Mesh and build the system under the lab's server.

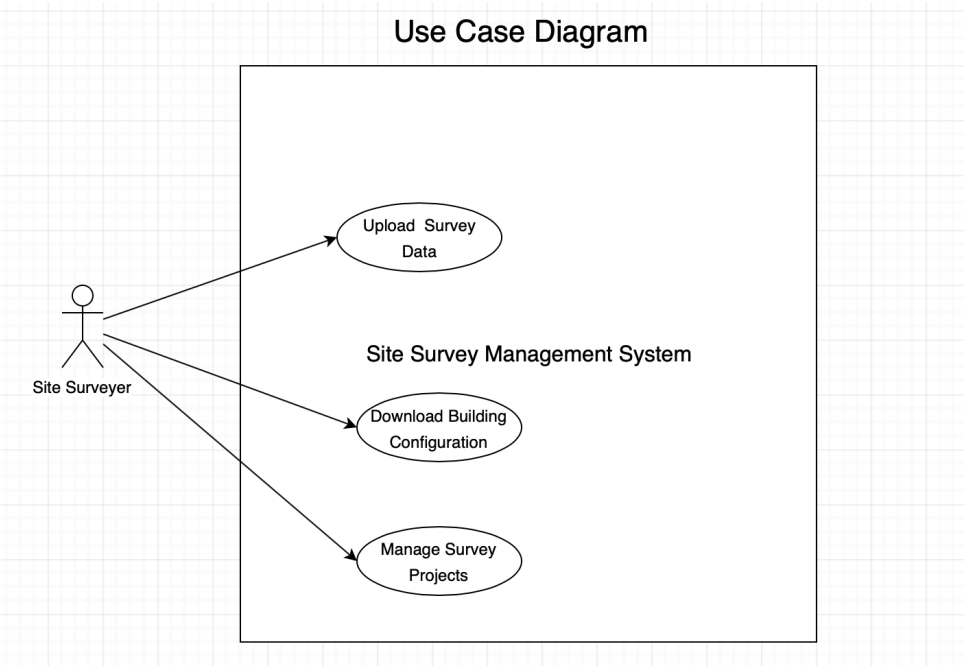
There are several reasons for us to build a CMS. First of all, we can simplify the process of setting up the environment for doing the site survey. With the help of the CMS, we can download the specific building configuration from the website easily instead of manual put the configuration folder to the phone. Secondly, we can upload all the data to the server from our phone after finishing each site survey. This helps better monitor the results. The last reason for having this CMS is to help to generate the fingerprint automatically whenever a site survey data is uploaded. According to the original method of producing the fingerprint and the map, we need to manually run the python script using the command line. This is quite time-consuming and inefficient. If we can incorporate this step in the CMS, we can reduce the time and effort spent significantly.

The headless CMS open source that we are using is Genetic Mesh. Genetic Mesh supports RESTful API that allows users to use post and delete command to alter the website. To better modify the website, instead of using applications like the postman to launch the request, we wrote python scripts to access the contents. To retrieve information from the website, we first need to obtain the Cookie token by performing post command under "/auth/login" with username and password passed in as the data. We next add the Cookie to the

headers. Every time when we give requests, we need to provide with the headers to access the website.

For the structure of our CMS, each project contains the buildings and the survey project. The buildings section will list out different buildings in the area. For example, there is Academic building, CYT building, LSK building under the project named HKUST. Each building contains Floors, which stores its z-index and the map. The map of the floor needs to contain the scale and its Azimuth. This building section will let users straightly download the specific building area configuration when they want to do a site survey. Regarding the survey project section, this part is for the storage of all the data after each site survey is finished. Users are able to upload data from their phones and the website will generate the fingerprint accordingly. Here are the class diagram and use case diagram for better understanding.





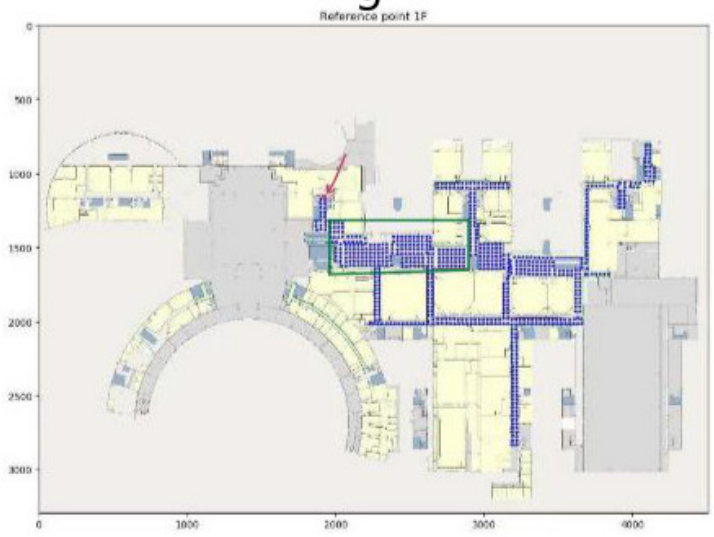
III. Limitation and Problem

- **Fingerprints accuracy:**

When we were doing the site survey and generating fingerprint through wifi data, two factors will affect the result of the fingerprint.

- The first problem is the number of wifi ports varies from place to place. We can generate accurate fingerprint points that are 2 meters away from each other in the parts like the Engineering center and Science center (Fig. 1 green part). And we can also get very accurate data of corridors located near offices (Fig. 2 green part). However, the wifi signal near the Nano Technology center is too weak to detect by our devices (Fig. 2 purple part).

Fig 1



- The second problem is that there are apparatus and cargos on the passage which diminishes the number of paths our devices will cover and make it as around ⅓ wide as the original passage (Fig. 2 red part). This situation appears near the laboratory on the 2F of the academic building and fails collecting wifi data.

Fig 2

- **Localizing SDK accuracy**

There are two problems in the current version SDK.

- The first one is that the localization pin may trespass rooms while we are walking around lift 18-19, 24-25. This should be a problem of boundary property when we're building the SDK model. One possible solution is to set walls passing forbidden.
- The second problem is that the location pin may stay unchanged while the device is moving and floats quickly to your current location. This problem is obvious when we travel to a large area such as the Engineering Common. One possible reason is that the number of wifi ports there is limited.

- **CMS API limitation**
  - Due to the instability of the wifi collecting device, one's project may introduce in inaccurate data. In our current version CMS, there is no function to tracking the user who pushes the data and the data of it.
  - Another problem results from continuous renewal of our campus. People who collect data need to use old version map while collecting the data in a new environment, which will cause inaccurate in the final fingerprint.
  - Current version "SigInt", the application used for collecting wifi data, will not help a user to check if a data session(the minimum unit for data collection) is empty. And users will get an error message when generating fingerprint due to an empty folder.

IV. Future Plan

We now finish site survey for the first and second floor in the academic building and are currently developing APIs for Mesh CMS system. Here are several plans for future work.

As for the problems of fingerprint inaccuracy, we devise the following possible solution.

- For the area where the signal is very weak, to deal with the problem of data missing problem, we can try to collect data there for several more times. After that, we can use machine learning algorithms to generate wifi data according to the nearby signal condition.
- For the problem that the localization pin will trespass the room area and wall, we plan to find the model and android source code and add the constraint that the area outside data collecting boundaries is forbidden to get into.

For the part of CMS API, we currently are working login, adding and getting project, adding modifying nodes and adding and modifying schema API. Here are some plans about functions that our content management system will have in the future:

- **Version Control:**

  To delete the inaccurate data session and keep the integrity of data for the specific version map, we devise a plan to introduce version control feature to our CMS.

  For every project, it will maintain a list to keep track of submission history of all participating users. Each user submission will contain data, timestamp, user information and floor information. We will derive submission serial number from the timestamp, user information. If there occurs a problem about data, the administrator will be able to find all the data sessions shared the same submission serial number with that erroneous data folder. It will help users to find all possible inaccurate data and also enable them to revert the version.

To check a data session with an empty data folder or duplicated data folder, we also devise that the CMS system will check it and detect them and return warning information about that.

V. Conclusion

During this summer, we finished collecting data, development of part of Content Management System for data storage and versioning and we also tried to learn the knowledge used in the indoor localization entrepreneurship area. In the future study, we plan to learn more about the training algorithm and data processing method used in the indoor localization project and we will try to build our model to realize localization.