10/07/2018

Report of Interdisciplinary Project

WeMap

Course Professor: Professor Fabio Dovis

Tutor: Professor Marco Piras

Students: Stefano Calleris – Matteo Nisi – Navid Yamini

A.A. 2017/2018

SUMMARY

WEMAP REPORT

MOTIVATION

Today, having internet access and Wi-Fi connection is highly important for people.

Using these technologies is becoming more essential and crucial day by day especially for educational purposes. In this project we tried to analyses Wi-Fi power and internet speed in University of Politecnico di Torino's area by creating heat map of the study area.

The leading idea of the project is to use GIS tools to provide a graphical interface of how the internet speed and the internet power work in the Politecnico di Torino (note: we are going to use PoliTo). The result can help the Polito IT crew to understand what weaknesses we have in Wi-Fi coverage and try to find new strategies to solve them.

One of the ideas of this project was to involve students in data gathering part by developing collaborative tool.

The main questions that have been covered in this project are:

- How does Wi-Fi signal strength and internet speed change in different places of the study area?
- How does the signal strength and internet speed change during the class time?
- Is it possible to use GNSS for positioning inside the case study area?
- Is it possible to find alternative solution for indoor positioning?
- How students can help in improving the Wi-Fi coverage?

STATE OF THE ART

On the market, several applications are used to map Wi-Fi in indoor buildings such as Wi Fi Heatmap[1], Ekahau HeatMapper[2], Acrylic Wi-Fi Professional[3], Indoor GPS[4] etc., but unfortunately none of them could do all the tasks that our project required, I.e. data collection and data gathering into a common online database. Moreover, some of them where not for free.
Some of these applications uses GNSS location to track a user while other applications ask for a map of the building.

Another issue we had to deal from the beginning is the different chipset, and especially chipset quality, both for Wi-Fi connection and for GNSS positioning available on different smartphones on the market. About this last point, it's right to make clear that we never dealt directly with the GNSS chip raw data, but we used an Android library which provided the position and an estimate of the measurement accuracy. The use of those data is explained later on in detail.

[1] https://play.google.com/store/apps/details?id=ua.com.wifisolutions.wifiheatmap&hl=en
[2] https://www.ekahau.com/products/heatmapper/overview/
[3] https://www.acrylicwifi.com/en/wlan-wifi-wireless-network-software-tools/wifi-analyzer-acrylic-professional/
[4] https://play.google.com/store/apps/details?id=com.ladiesman217.indoorgps&hl=en

In a huge building like the Politecnico area, it is difficult to give the user the job of locating himself on the map, and plus it would've required a bigger effort to retrieve for each room the map and let the user locate himself with a touch.

We decided then, to develop a hybrid idea which is taking into account the GNSS localization sensors in the smartphones, giving also the user the possibility to locate himself better in a room with a grid system which is going to be explained better below.

Another drawback of the applications present on the market, is that they are done for private purposes: a user willing to measure the Wi-Fi power and speed in his apartment starts recording samples in order to build the heatmap.

So, every user has an offline map with his measurements.

Even if the first idea of the project was of an offline cooperative data gathering, followed by data processing and analysis, we decided to have a real time tool for data gathering in order to exploit the collaborative aim.

STARTING IDEA

As already stated previously, the original idea was to gather offline information through third-party applications and fill shared sheets.
Due to the possible additional error from data manual insertion, we decided to create our own application that would have helped students and professors during the data-gathering phase.

In particular the weaknesses we have seen in the use of an offline method were:

- Difficulty of having a collaborative tool
- Difficulty in having homogeneous data gathering
- Uncertainty given by human errors

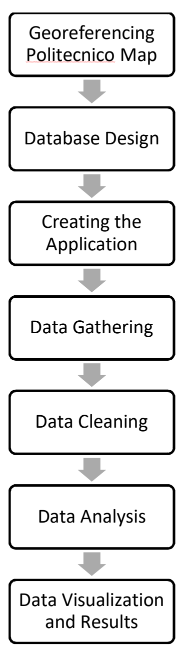

Figure 1: Project's Phases

To overcome these problems, we divided the whole project into seven steps, Figure 1 shows general steps of our solution.

MAPPING POLITO

PoliTo was asked to give the state of the art of the Wi-Fi infrastructure by giving access to:

- Digitalized map of PoliTo
- Georeferenced position of the Access Point (Ap)
- Network information about the AP

Actually, none of these requirements was fulfilled, so we decided to:

- Extract the map from POLIMAP, a service used to see and make some analysis working on an internal GIS
- Try to georeferenced the AP as an outcome of our results and not as a starting point
- Retrieve network information about the AP by measuring them during the data gathering phase.

Unfortunately, the POLIMAP database is not available to be downloaded so we decided to took screenshots of these maps and georeferenced them. An example of these maps is shown in Figure 2.

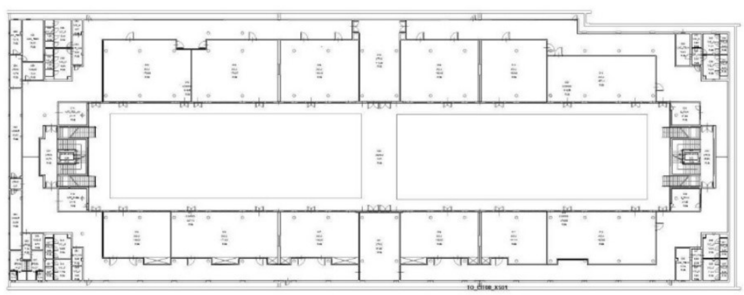

Figure 2- Example of map used at the starting point

In order to fill this lack in knowledge, the first steps were. Being more detailed we:

- Georeferenced the screenshots (POLIMAP)
- created the different entities (buildings, room, corridors…), using the information readable in the image
- Computed area, perimeter and extract coordinates for each entity

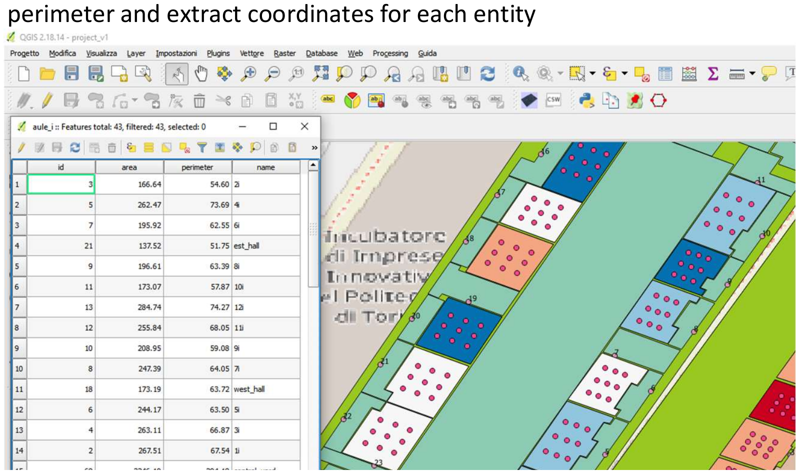

Figure 3- Georeferenced map with relative ID, perimeter and area

All this information has extracted using the QGIS software, and then it has been stored in a MySQL DB in order to be able to retrieve them at any time, from everywhere.

For each room of the study area, we drew a digital polygon, starting from which, we extrapolated the center of the room and the grid points, i.e. the internal room mapping, as explained in the mobile application section (see Figure 3).

The obtained layers have been useful in the design phase already, as they have been the perfect tool to make tests on the data obtained by the mobile application, to validate the system and to perform improvements. mobile application.

DESIGN PHASE

In order to develop the application, we first decided, agreeing with our mentor, which design to implement for this study case.

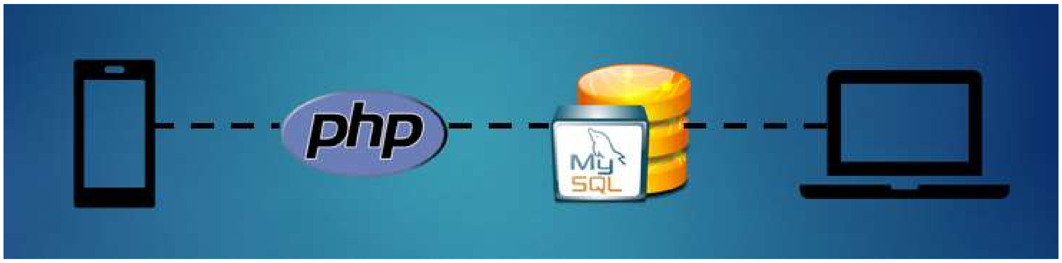

Figure 4– A basic schema of the communication flow

Basically, as reported in Figure 4**Error! Reference source not found.**, the idea was to have 4 main entity:

- <u>A real machine hosting a Server</u>
- <u>A MySQL server indeed</u>
- <u>PhP services to expose the server</u>
- <u>An Android Application for data gathering</u>

SERVER

For sake of simplicity, we decided to create our own server on a personal computer, exposing it on the web to make it accessible.

The server, reachable through wemapserver.sytes.net, was used to:

- Store physical data about rooms
- Give a unique ID to each entity
- Store real time measurements
- Retrieve measurements in the analysis phase

MYSQL DATABASE

ER design phase

In order to store the data, we decided to use a MySQL database.
Before starting the actual realization of itself, the ER diagram of the database were deigned.

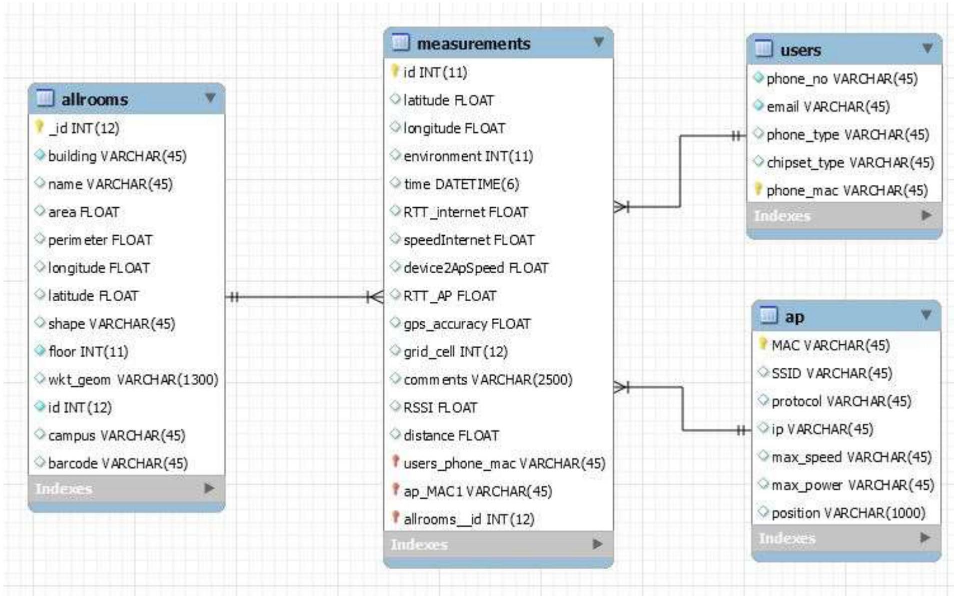

Figure 5 – ER diagram of the DB

The DB is mainly based on four entities:

1. allrooms
   containing all static information about the room and the ID of the room.
   In this table, even corridors are considered, and a unique identifier is assigned to them too.
   In a further step, it has been decided to have a unique identifier for corridors since it was not possible to let the user choose his position in a user-friendly manner. Table 1 shows all the attributes related to this table.
2. measurements
   containing all measurements taken by the users. It is a dynamic table updated in real time whenever the user is inserting the measurement (Table 2).
   Each measurement has a unique identifier and it is connected to the other tables by foreign keys:
   - allrooms_id → room: is the identifier of the room in which the measurement was taken
   - ap_MAC -> ap: is the identifier of the AP to which the user was connected when the measurement was recorded
   - users_phone_mac -> users: is the identifier of the device which recorded the measurement

   All the other features regard the:

   - Wi-Fi information
   - Geographical information
3. users
   Each user is identified by the unique MAC address of his device. If a user records measurement with more device, he can identify them all with himself by registering to the system.
   In this way, by an email address, a merge between the two itemset is done. Table 3 shows all the attributes related to this entity.
4. ap
   It is the table with the information regarding the AP.

Since no information was given by PoliTo on the position and the MAC addresses of this devices, this table is dynamic.

Basically, every time that the user records a data, if the MAC address of the access point that the user is connect to it, is not exist in the table, the application will insert the information of that access point into AP table.

In the AP information, we have SSID attribute, since in PoliTo area two networks are present.

Moreover, having this information, has allowed us to make a first filtering in order to cut off all the measurements that were taken with a different SSID then the known ones. For more information please refer to table 4.

ALL ROOMS		
Attribute	**Type**	**Description**
_id	**INT**	**(PK) the id of the room given by the app**
building	**VARCHAR**	**the name of the building the room belongs to**
name	**VARCHAR**	**the name of the room**
barcode	**VARCHAR**	**the barcode number inside the room**
area	**FLOAT**	**geometrical area of the room**
perimeter	**FLOAT**	**perimeter of the room**
longtitude	**FLOAT**	**longitude of the center point**
latitude	**FLOAT**	**latitude of the center point**
shape	**VARCHAR**	**flat or sloping**
floor	**INT**	**the floor number in which the room is located**
wkt_geom	**VARCHAR**	**latitude and longitude of the vertexes of the room**
id	**INT**	**id of the room from the maps (the same of polito db)**
campus	**VARCHAR**	**name of the campus that the building is located in**

Table 1: Attributes related to All Rooms Entity

MEASUREMENTS		
Attribute	**Type**	**Description**
_id	**INT**	**(PK) the id of every measurement given by the app**
latitude	**FLOAT**	**latitude of the measured point**
longitude	**FLOAT**	**longitude of the measured point**
environment	**INT**	**where is the location of the point, 0 for corridors, 1 for inside rooms**
time	**DATETIME**	**time of the measurement taken**
RTT_internet	**FLOAT**	**round trip time related to the pinging of 8.8.8.8**
internetSpeed	**FLOAT**	**the speed of internet that we can reach in the point**
device2ApSpeed	**FLOAT**	**speed related to the access point that we can get**
RTT_AP	**FLOAT**	**round trip time related to the pinging of access point ip**
gps_accuracy	**FLOAT**	**accuracy that we can get from the GPS device in the point**
grid_cell	**INT**	**the position of the user in the room (general grid with position from 0 to 9)**
comments	**VARCHAR**	**user comments**
allrooms_id	**INT**	**(FK) coming from the allrooms table**
phone_mac	**VARCHAR**	**(FK) coming from the users table**

ap_mac	VARCHAR	(FK) coming from the ap table
distance	FLOAT	user distance from the point to the access point
satellite_in_view	INT	Number of satellite in view
satellite_in_use	INT	Number of satellites that the device uses for positioning

Table 2: Attributes related to Measurements Entity

USER		
Attribute	Type	Description
phone_no	VARCHAR	user phone number
phone_type	VARCHAR	the type pf the phone that user has.
email	VARCHAR	user email address
chipset_type	VARCHAR	the GNSS receiver chipset type inside the phone
phone_mac	VARCHAR	(PK) user phone mac address

Table 3: Attributes related to the User Entity

ACCESS POINT		
Attribute	Type	Description
MAC	VARCHAR	(PK)the MAC address of the access point
SSID	VARCHAR	the Service Set ID of the AP (name of the network)
protocol	VARCHAR	the IEEE protocol that the access point uses
ip	VARCHAR	the IP address of the access point
max_speed	VARCHAR	maximum speed measured of the access point
max_power	VARCHAR	maximum power measured of the access point
position	VARCHAR	latitude and longitude of the access point

Table 4: Attributes related to Access Point Entity

ANDROID PROGRAMMING AND IMPLEMENTATION

In order to realize a real time data gathering, we decided to create our own application that could be able to retrieve information about:

- Wi-Fi speed
- Wi-Fi power
- GNSS position
- GNSS accuracy
- Distance from the AP

Wi-Fi speed is computed by starting a download of a file of given size and measuring then the time needed to download it.

Being every device associated to a unique MAC address, we were able to create also an history of the measurements taken by the same device.

If the user registered to the system, he could associate more devices to the same device and so be able to read from the history the measurements related to all of them.

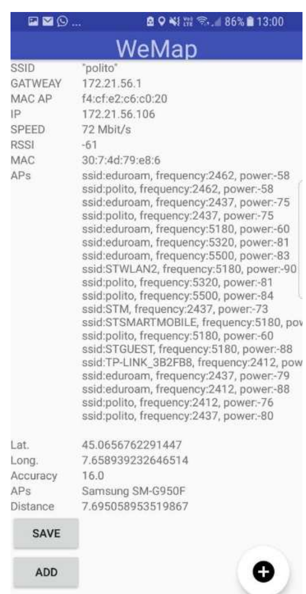

Figure 6: Main page of the Application

Wi-Fi Power is directly computed through an Android Library which returns the power in decibel.

GNSS Position and GNSS Accuracy are both retrieved using an Android library.

Distance from the AP Since PoliTo did not give us any information about where the access points were located, we tried to do the opposite by locating them based on the signal strength.
We used the *Free Space Path Loss* equation in order to find a qualitative distance from the access point.

$$FSPL(dB) = 20log_{10}(d) + 20log_{10}(f) - 87.55$$

We could know the power because it is given from the Android library and we know the frequency at which the Wi-Fi is working so through an inverse equation we are to find d[m].

Testing Phase

Before promoting the application to the student of Politecnico, we started a test phase in order to understand the weaknesses of our system.

DB STRESS TEST

We first tested the DB by a massive querying done through Python REST APIs, that were basically sending batches of request to our machine to understand if it was able to handle a big amount of traffic.
We chose to send batches on increasing size until 5000 packets simultaneously. Due to the research purposes, we thought that this number could be enough and that for more requests, a better infrastructure should have been deployed.

After the system sizing test phase, we tested the application in the real situation to understand what is the right process that we can provide for the user.

WIFI INFORMATION TEST

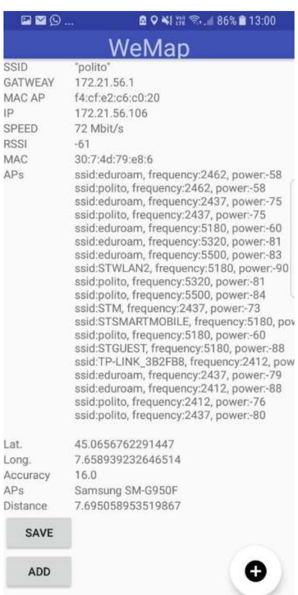

Figure 7

In this test phase, we have seen that the speed given by the Wi-Fi library of Android is retrieving the speed of the channel and not the actual speed of download.
Since our purpose was to give the students and the infrastructure owner an idea of which were the weak link of the network, this was not enough.

We decided then to emulate well known systems like SpeedTest and apply the definition of speed for digital purposes.
In the final version of the app, Figure 7, when a user presses on the plus button, a file of known size is downloaded.

The initial size was set at 1 Mb in order to have a fast download.
By the way, increasing the file size, you get a better picture of the speed.
If you imagine a “normal” Wi-Fi network speed is about 20 Mbps, a 1Mb file is downloaded in 0.05 seconds. Given the variability of a wireless network, you can download a file when you get a speed of 100 Mbps but the second after is 20Mbps and so on.
We decided then to increase the size of the download to 20 Mb in order to have a better picture and not to add a big overhead.

GNSS INFORMATION TEST

Due to the nature of the environment, the collection of the position through GNSS sensors in smartphones led to some issues.
Indeed, in an indoor environment like the rooms, the GNSS accuracy is not high and in the corridors without any window, the coverage is even worse.

Figure 8 – GPS Logging in the study area

A general idea of the coverage in the study case area is given in Figure 8. In the image is shown the accuracy of the GNSS measurements from a best case (green) to the worst case (red).

It has to be noticed that during this test, we went inside the building even if this was not recorded at all by the GNSS logger.
Since this project was thought for a test of the indoor environment, this behavior put at risk the entire project.

Solutions

To overcome the GNSS issues we came out with two different solutions for room and corridors.
Since We Map is a collaborative tool, we'd like to think to base our application on the trust.

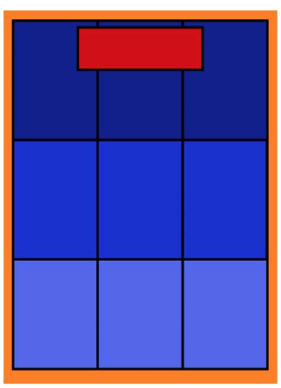
Figure 9 Room Grid

Room

For the rooms we decided to make use of a grid that divides the area in 9 smaller zones. When the user takes the measurement, it also chooses a square where to locate himself in the room.

In Figure 9 it is shown a basic idea of the implementation. The red rectangular shape indicates the blackboard of the room, so the user is helped in locate himself.

This information is then stored in the database with a number from 1 (top left corner) to 9 (bottom right corner). Moreover, in the allrooms table the coordinates of squares' centers are stored in order to locate the user measurements.

Corridors

{'id':QRcode1,
'lat': 45.062492,
'lon': 7.662203}

Since as we said before, having a simple way to locate the user in the corridors is more difficult than for the rooms, we thought of another way to solve this issue.

The final decision was to put QR codes in some specific point. The user that is willing to take measurements for corridors, can locate itself through a scanning of this QR codes. As it can be seen in an example of QR code, the information contained in it, identifies the QR code itself and gives the location of it.

When the measurement is recorded, only the QR id is stored in the measurement row. Then in the analysis phase, through a merge of the tables, the position of the user can be retrieved.

DATA GATHERING

The data gathering phase can be depicted in three phases:

1. Testing phase of the system by gathering speed by our own
2. Collaborative Gathering by:
    a. Asking our colleagues to help us
    b. Presenting the project in other classes asking for further help

    These approaches led to a total number of 16 different devices measuring the Wi-Fi speed and power during the gathering phase.

3. Monitoring the power and speed for all the duration of a lecture to understand common pattern and behavior.

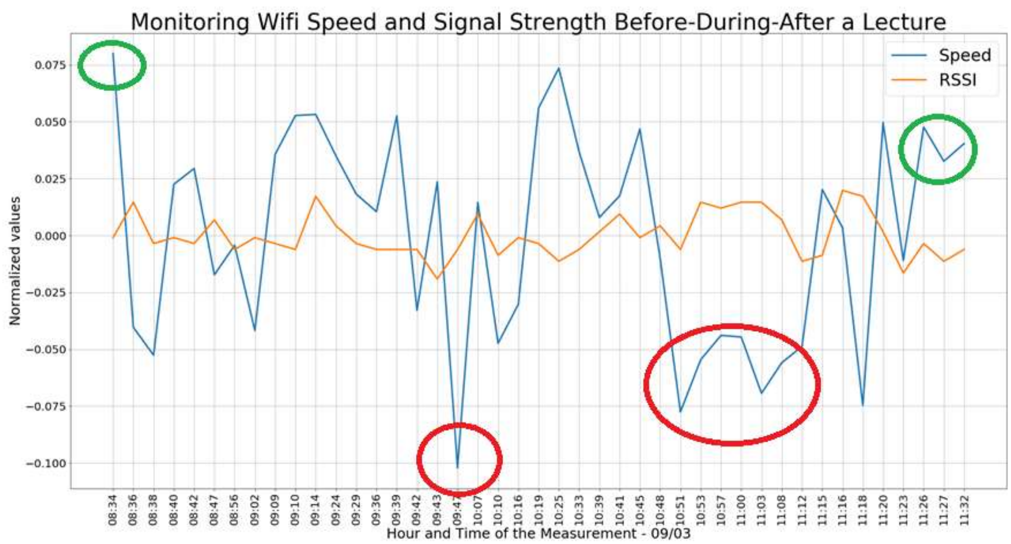

Figure 10 - Wi-Fi signal strength and Internet Speed behavior during the lecture

An example of analysis of point 3 is shown in Figure 10. This plot is describing the change in Wi-Fi speed and strength during the whole duration of a lecture.

It is important to take into account that the results are normalized by the maximum value for a better visualization but also in this scale it can be appreciated a variable speed and power during the three hours.

Having a priori knowledge of this situation (being of our teammate at that lecture) we know that from 10.51 to 11.12 the professor was accessing a cluster and so taking a lot of bandwidth.
This shows the importance of having an a priori knowledge of the environment and of the field in the analysis of big amount of data.

DATA CLEANING

Due to the number of people involved and due to the test phase, we applied some data cleaning to improve the dataset before applying any analysis algorithm.

First, we cleaned base on the Service Set ID (SSID). All the measurements taken with an SSID different than “polito” or “eduroam” were filtered out.

Then, we filtered out all the measurements that, given the GNSS position, were taken too far from Politecnico. It is indeed possible to name a private network “polito” or “eduroam” and this will have affected our analysis. Figure 11 shows the steps of data cleaning process.

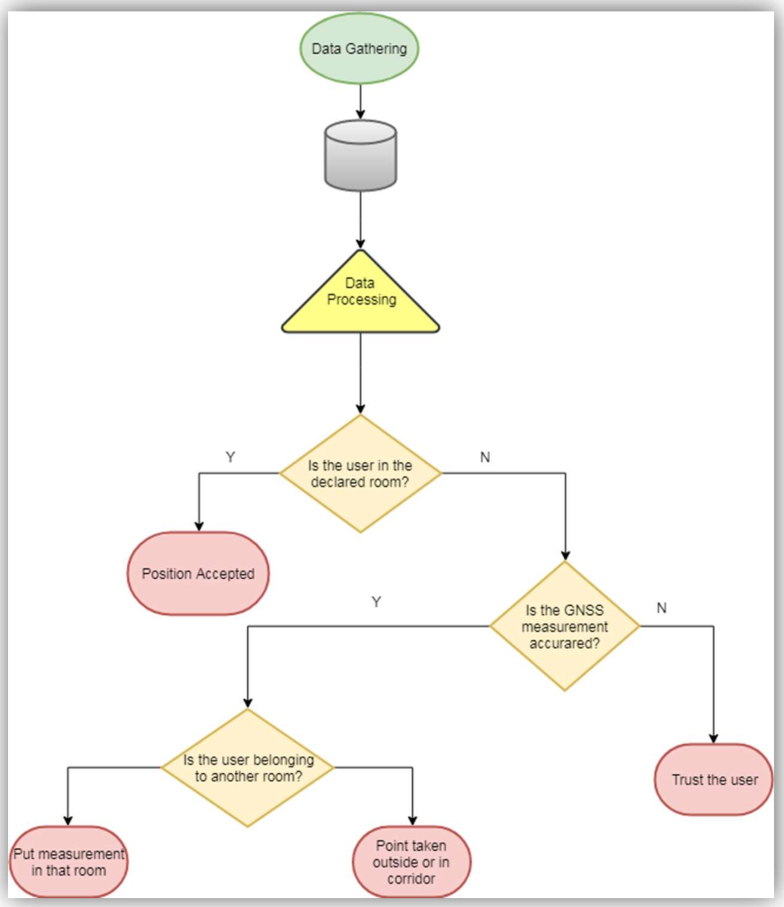

Figure 11- Data Cleaning Process

dfghIt has to be noticed that, if PoliTo is going to give data about their access points concerning their MAC addresses, a better filtering can be done.

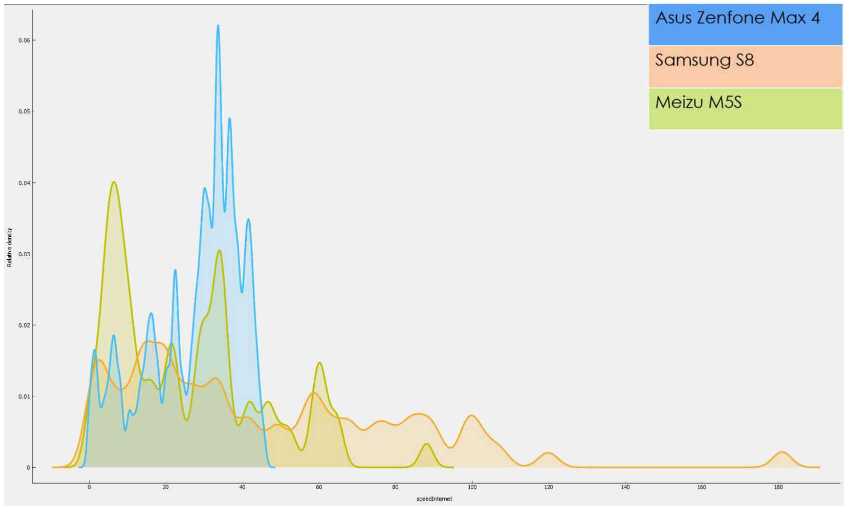

Figure 12 - PDF of Internet Speed

DATA ANALYSIS

From the analysis point of view, after the cleaning part, we had to decide which parameters to use for the visualization.

As it can be seen in **Error! Reference source not found.** the maximum speed reached by three different devices of three different market ranges it's way different.

The orange shape shows a maximum speed of 180 Mbps while, the blue shape even if it is clearly the most active one, reaches maximum speeds of 45 Mbps.
Taking the maximum speed of the one recorded for each grid, would have misled our results. Also, the mean is not good because it is influenced by the number of measurement taken with the same device.
We decided then, in order to cut off the outliers, to take the median value for each entity we wanted to analyze.

The same approach was followed in order to evaluate the accuracy of GNSS measurements in the study area. This is a consequence of the GNSS behavior already shown in Figure 8 – GPS Logging in the study area, in order to understand the GNSS accuracy in the area.

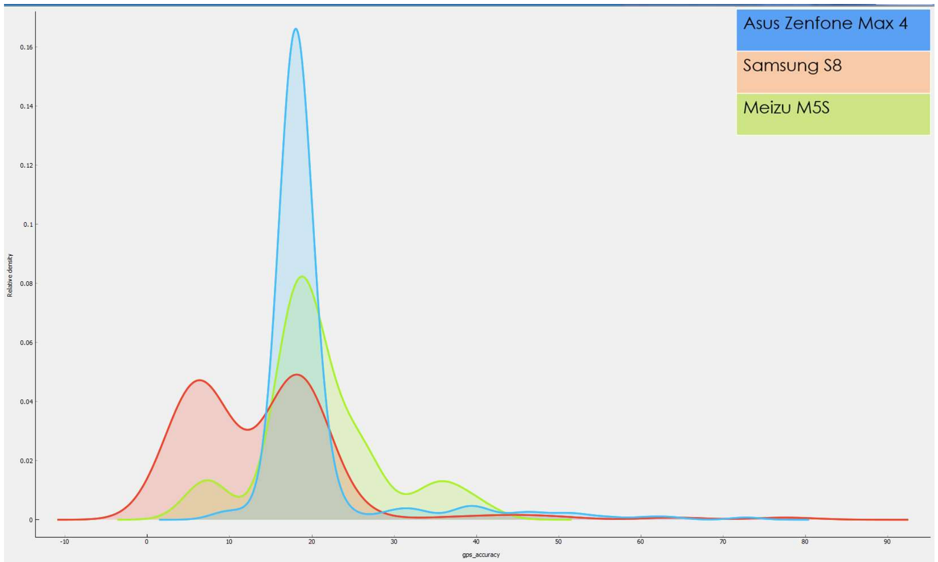

Figure 13 – GNSS accuracy evaluation

As for the Wi-Fi speed, it can be seen how newest and more powerful devices like Samsung S8 are able to reach level of accuracy that it's impossible to reach for the other devices.
Again, the width of the Asus Zenfone Max4 shape it's the narrowest one.
This could be affected by the fact that with this device all the measurements of the Wi-Fi speed during classes were taken. So, the device has been in the same place taking all the measurements (about 60 measurements).

Figure 14 – 2D Accuracy Evaluation

Still for these results the plot in Figure 14 shows some uncertainty in the accuracy evaluation. For example, as it can be seen in the rounded points, the accuracy in the corridors of rooms I building (a corridor with no windows placed underground) it is recorded as in between 0 and 10 meters.

This is unlikely true, except if you take into account the logging shown in Figure 8. What we think is happening indeed, is that the GNSS sensor is memorizing the last valid accuracy recorded and put it as the real time one.

RESULT VISUALIZATION

To return the result of our work in an intuitive and user-friendly way, we went along two different paths: we used concentration maps by Google for the 2D data representation and a QGIS plugin for the 3D data plotting.

Figure 15 – Speed per Room

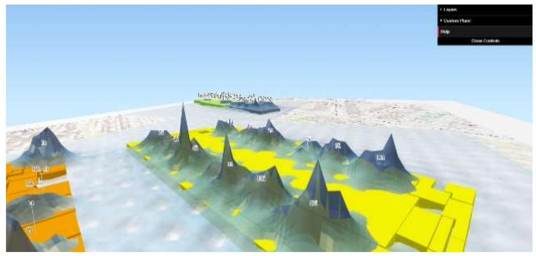

Figure 16 - Speed per grid element

All the results can be seen on a http://wemapserver.sytes.net/, a website we created to let the user see the result of their cooperation.

NEXT STEPS

Indeed, this was just a preliminary work: the mobile application has to be improved to reach Android standards, and some help from the Politecnico is needed to improve data gathering and website security.

Eventually, the project has to be scaled to be able to cover a greater variety of structures (multi-floor buildings, complex rooms).

Starting from the idea of collaborative tool, at the following links:

- https://github.com/nisimatteo834/WeMap_Android the source code of the Android Application
- https://github.com/nisimatteo834/WeMap_Python_GIS the code used for analysis and preprocessing for GIS software
- http://wemapserver.sytes.net/ website to access the results

CONCLUSIONS

The result of this work is definitely different from the initial requirements: while proceeding with the tasks and gaining confidence with the project, we often decided to reach further than required to apply modifications to the starting specifications. We aimed to improve the system, implementing our own set of tools from low to high level, reducing the manual user interaction and increasing the cooperative side of it.

We had to cover a wide range of topics and problematics, from the lack of initial information to GNSS issues, learning meanwhile how to program a mobile application. In this context, we could experience and exploit the multidisciplinary soul of the project, which matched perfectly with the “ICT for Smart Societies” guidelines. We

exploited the knowledge acquired our Master's Degree in full, touching Geomatics, high-level programming, web service creation, database management and analysis of complex data.
Below, examples of the interdisciplinarity of the project are reported:

- ICT for Geomatics: GNSS knowledge, GIS
- Programming for IoT: REST request, webservices
- Statistical Signal Processing: Signal Analysis
- ICT For Transport System: Handle big amount of data and results classification
- Mobile Application Development: Android Application
- Management and content delivery for Smart Networks, Mobile and Sensor Networks: Wi-Fi knowledge

Proceeding with the implementation, we tried to shift the user role in two ways, looking for the right balance on user interaction. On one hand, we pushed the cooperative side, letting free-users to feel involved and part of an ongoing project which is useful and provides updated results; on the other hand, we tried to reduce the error sources related to the user itself, such as wrong positioning, data gathering and data insertion.

The cooperative side of the project is doubtlessly the most powerful one, in order to have updates on real conditions of the network, virtually for free. However, it was not an easy task to motivate users to join the data collection, which could probably become more interesting with some reward mechanism.

What's more, we found of great help the recurrent meeting with our reference professor and the updates in front of our colleagues we had during the project time: They have been useful to have a periodical picture of the state of the work, see objectives, problems and challenges, and to push us to keep on the right schedule and exploit time at best.

The final product is a complete system, based on cooperative paradigm and easily scalable, that can help the Politecnico di Torino and other institutes to find Wi-Fi network weaknesses and improve the system at a very low cost and in the short-term horizon.