

Grado en Ingeniería Telemática
2017-2018

Trabajo Fin de Grado

“Captura y análisis de datos para un sistema de localización en interiores basado en Wi-Fi”

Jorge Saucedo Pavón

Tutor/es

Pablo Serrano Yañez-Mingot

Carlos Jesús Bernardos Cano

Lugar y fecha de presentación prevista

Jueves 4 de octubre del 2018. Aula 4.2.E08

ÍNDICE

1. INTRODUCCIÓN	1
1.1. Resumen.....	1
1.2. Contexto.....	2
1.3. Motivación	3
1.4. Marco legal y ética.....	4
1.5. Objetivos.....	5
1.6. Estructura	5
2. METODOLOGÍA DE MEDICIÓN	7
2.1. Introducción	7
2.2. Descripción del entorno	8
2.3. Equipo	9
2.4. Captura de datos.....	10
2.4.1. Información principal.....	10
2.4.2. Información relacionada.....	14
2.5. Almacenamiento de datos	15
3. ANÁLISIS DE RESULTADOS	19
3.1. Pruebas de validación	19
3.1.1. Localizar los AP por la potencia de señal recibida	19
3.1.2. Comparación de señales Wi-Fi recibidas desde un mismo AP.....	21
3.1.3. Verificar ubicación de APs externos y de Rogue APs.....	23
3.2. Aplicación Web para visualización de contenidos	27
3.3. Análisis general.....	30
3.4. Análisis de redes oficiales y no oficiales	35
4. PLANIFICACIÓN Y PRESUPUESTO.....	43
4.1. Planificación	43
4.2. Presupuesto	47
5. CONCLUSIONES Y TRABAJO FUTURO.....	49
5.1. Conclusiones principales	49
5.2. Trabajos futuros	50
6. BIBLIOGRAFÍA.....	51

ÍNDICE DE IMÁGENES

Imagen 1. Uso de los dispositivos móviles [1]	2
Imagen 2. Uso que se le da a los dispositivos móviles [2]	2
Imagen 3. Esquema de la Metodología	7
Imagen 4. Montaje de las plantas del Edificio 4	8
Imagen 5. Características del equipo utilizado	9
Imagen 6. Ejemplo de la división de celdas de la planta 1	11
Imagen 7. Esquema explicativo de la captura de datos	12
Imagen 8. Esquema del funcionamiento del script “juntar_de_tabla_total.sh”	13
Imagen 9. Tabla Prueba	15
Imagen 10. Tabla Lugares	16
Imagen 11. Tabla LugarFechaTotal.....	17
Imagen 12. Tabla EscaneoTotal	18
Imagen 13. Ejemplo localización de un AP	20
Imagen 14. Leyenda de colores según la calidad de la señal	20
Imagen 15. Ejemplos extra de la localización de APs.....	21
Imagen 16. Comparación de los valores de Wifi-UC3M y eduroam	22
Imagen 17. Calidad de la señal según la potencia recibida [6].....	23
Imagen 18. Puntos donde es visible la red WORST	24
Imagen 19. Puntos donde es visible la red TSCWLANG	25
Imagen 20. Puntos donde es visible la red Orange_0B20	26
Imagen 21. Puntos donde es visible la red VodafoneA3EA	26
Imagen 22. Explicación del MVC de este trabajo	27
Imagen 23. Pantalla de inicio de la aplicación web.....	28
Imagen 24. Pantalla de opciones de la aplicación web.....	29
Imagen 25. Pantalla de muestra de una de las opciones de la app.web.....	29
Imagen 26. Distribución de los escaneos por plantas.....	30
Imagen 27. Distribución de las entradas de EscaneoTotal por plantas	30
Imagen 28. Número de APs distintos en cada planta	31
Imagen 29. Distribución de las distintas redes Wi-Fi.....	32
Imagen 30. Clasificación de los distintos APs por tipo de red	33
Imagen 31. Uso de los canales por APs (Banda de 2,4 GHz).....	33
Imagen 32. Uso de los canales por APs (Banda de 5 GHz).....	34
Imagen 33. Redes emitidas por un mismo AP	35
Imagen 34. Canales usados por cada tipo de red	36
Imagen 35. Uso de los canales por APs de la universidad (Banda de 2,4 GHz)	36
Imagen 36. Uso de los canales por APs (Banda de 5 GHz).....	37
Imagen 37. Canal usado en cada por su mejor eduroam	38
Imagen 38. Curva calidades del sótano	39
Imagen 39. Curva calidades de la planta 0	39
Imagen 40. Curva calidades de la planta 1	40
Imagen 41. Curva calidades de la planta 2	40

Imagen 42. Curva calidades de la planta 3	41
Imagen 43. Puntos con calidad deficiente	42
Imagen 44. Desglose de tareas	46
Imagen 45. Diagrama de Gantt.....	46
Imagen 46. Presupuesto.....	47

1. INTRODUCCIÓN

1.1. Resumen

Mediante este estudio y análisis de los puntos de acceso que están localizados en el edificio 4 de la universidad se espera conseguir información suficiente que sirva como primer paso para el desarrollo de un sistema que permita la localización en interiores a través de la red Wi-Fi.

Este estudio se ha dividido en varios bloques acordes a la planificación que se ha hecho a la hora de realizar el trabajo:

- Escaneo de los puntos: Se ha procedido a dividir cada planta del edificio en sectores que debían ser escaneados. Para la realización de esta parte previamente se crearon una serie de scripts que permitirán fácilmente realizar este escaneo de forma automática, así como la extracción de la parte útil de ellos que posteriormente se introduce automáticamente en la base de datos que se ha creado para el almacenamiento de la información para su posterior análisis.
- Herramienta para el análisis: Se ha creado una aplicación web que sirva como apoyo visual e interactivo para el análisis de los datos, consiguiendo una mayor facilidad de acceso a la información almacenada dentro de la base de datos. También se ha usado como apoyo para el proceso de validación.
- Validación de los datos: Para garantizar la consistencia y la fiabilidad de los datos que se han escaneado, se realizan una serie de pruebas para comprobar que el resultado obtenido es el esperado.
- Análisis de la información: Todo lo obtenido durante los escaneos se estudia y se analiza para observar la información más relevante, tanto a nivel general como más concreto para el proyecto que nos ocupa. De forma que consigamos recopilar información suficiente para saber si es posible realizar un sistema final que permita la localización solo con la red Wi-Fi.

Palabras clave: Red Wi-Fi, sistema de localización en interiores, captura de datos, análisis de datos.

1.2. Contexto

Hoy en día, vivimos en una sociedad donde la tecnología está muy presente en nuestra vida. Todos los días vemos noticias o artículos de nuevos avances y el desarrollo de nuevos elementos que tienen que ver con la tecnología. Esto ha propiciado un horizonte completamente novedoso lleno de posibilidades y que era muy difícil de imaginar hace tan solo 20 años, al menos a tan corto plazo.

Es habitual ver a la gente por la calle con sus teléfonos móviles en la mano, interactuando con ellos, ver a personas en cafeterías navegando en tablets o estudiantes utilizando un portátil en la biblioteca para descargarse los últimos apuntes de internet. Esta tecnología se ha hecho parte de nuestra vida, casi tanto que el 78% de la población mundial tiene este tipo de dispositivos móviles [1], y la tendencia es al alza, es decir, cada vez más personas usarán este tipo de tecnología de forma cotidiana.



Imagen 1. Uso de los dispositivos móviles [1]

Otra cosa que ha ido cambiando con el tiempo es el uso que se le da a este tipo de dispositivos. Frente al uso clásico de los móviles o los ordenadores, se han ido ampliando las posibilidades y usos que se le puede dar a estos, de tal forma que ahora es muy normal comprar a través del ordenador, pagar a través del teléfono móvil, hacer transferencias de dinero por internet, jugar a multitud de juegos, consultar las últimas noticias... Esto provoca que las personas estemos utilizando nuestros teléfonos, tablets u ordenadores una gran cantidad de horas al día.

Tráfico de datos móviles por categorías

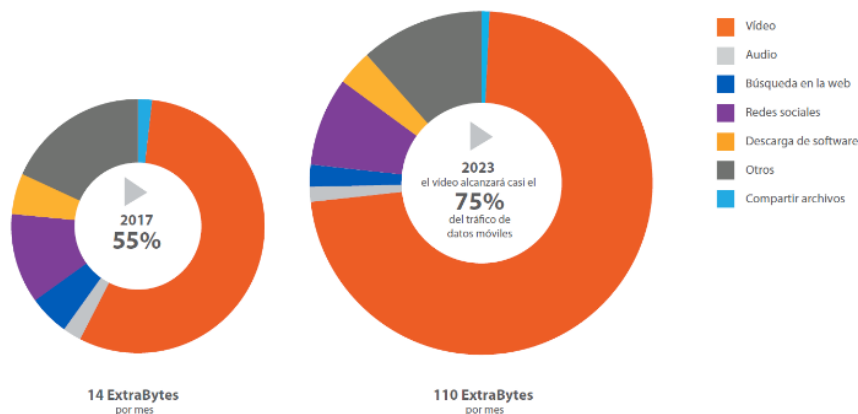


Imagen 2. Uso que se le da a los dispositivos móviles [2]

Todo esto es posible gracias a internet y, sobre todo, a la facilidad que hoy en día tenemos para acceder a él. Ya no dependemos de la conexión que tengamos contratada en nuestras casas, sino que las compañías proporcionan tarifas de múltiples precios y opciones para que podamos conectarnos fuera de casa. Sin embargo, cada vez es más frecuente que nos encontremos redes Wi-Fi gratuitas a las que conectarnos, ya sea en aeropuertos, centros comerciales, bibliotecas, universidades... Este tipo de redes facilitan aún más el acceso a internet ya que personas que no tengan contratada una conexión portátil a la red para alguno de sus dispositivos multimedia, o que no quieran gastar parte de su tarifa en internet, tienen la posibilidad de hacerlo de forma gratuita. Según algunos estudios, de media 7 de cada 10 españoles se conectan al menos una vez al día a una red Wi-Fi [3], lo que significa que esta forma de conectarse a internet es muy prolífica y al día existen millones de conexiones a este tipo de redes.

1.3. Motivación

Dada la cantidad de personas que utilizan los dispositivos móviles para conectarse a internet y de éstos, todas las personas que se conectan a redes Wi-Fi, surge el utilizar estas redes Wi-Fi para obtener un beneficio a cambio de ofrecer el servicio de forma gratuita.

Centros comerciales o museos, por ejemplo, ya han realizado sistemas similares para capturar la información de sus clientes para sus análisis. El museo Louvre por ejemplo, utilizó unos dispositivos que utilizaban el Bluetooth para saber el recorrido que sus clientes hacían por el museo, ver en qué cuadros se detenían más tiempo, si había personas que preferían ver unos u otros [4].

De esta manera surge uno de los posibles motivos para crear un sistema de localización basado en Wi-Fi. Como hemos visto previamente, cada vez son más los centros que proveen de conexión Wi-Fi gratuita, la cual puede ser un foco muy grande de información para el proveedor.

Un centro comercial podría utilizar este sistema para saber qué tiendas son más frecuentadas por las personas que acuden al centro, o una biblioteca podría saber que zonas tienen más uso y así poder destinar más atención a esos lugares para ofrecer mejor calidad en sus servicios.

Otro motivo para realizar este sistema lo hemos sufrido en muchas ocasiones nosotros mismos cuando hemos intentado usar una aplicación que usa la localización dentro de un edificio y este ha sufrido errores de posicionamiento debido a las interferencias que ha podido encontrar el móvil al localizarse con el GPS.

Si pensamos en que usos o beneficios se le podría dar a este sistema en el entorno de un trabajo, podría ser el de, cuando el usuario quisiera, activar la localización de forma real que indique la posición que tiene en cada momento, de forma que personas que puedan necesitar asistencia estén controladas en todo momento dentro del edificio, o personas que deban moverse por el interior del recinto por motivos de trabajo pero que deban ser

localizados rápidamente en caso de una urgencia. Incluso podría servir para las empresas como modernización del sistema de control de asistencia de los trabajadores, como sistema para fichar.

1.4. Marco legal y ética

Como hemos visto una herramienta de estas dimensiones nos ofrece un gran abanico de posibilidades, pero de la misma manera surgen posibles inconvenientes según el uso que se le dé, tanto de manera legal como moral.

Por un lado, cuando nos conectamos a una red Wi-Fi gratuita, lo normal es que en el proceso de conexión tengamos que consentir o confirmar que nos queremos conectar, aceptando una serie de indicaciones o condiciones para su uso. Aquí sería donde debería indicarse que, consintiendo, apruebas que los proveedores de la conexión utilicen ciertos datos que se pueden extraer del uso para realizar el seguimiento de localización. Normalmente aceptamos estas condiciones sin siquiera pararnos a leer qué estamos consintiendo, un grave error si lo que queremos es evitar cualquier tipo de concesión de nuestros datos. Sin embargo, si no se aceptan, no es posible beneficiarse de dicha red.

Es importante tener en cuenta que la normativa legal de cada país [5] puede restringir qué datos pueden extraerse de cada usuario, y, tanto por motivos legales como morales debería ser un uso donde los usuarios sean siempre anónimos para que así ningún dato personal pueda ser usado de manera dañina para esa persona.

A nivel ético, este es uno de los mayores problemas en la realización de una herramienta que permita localizar a través de una red Wi-Fi. Hoy en día vemos como al estar tan familiarizados con la tecnología, los datos se encuentran a merced de las empresas que controlan ciertas aplicaciones o redes sociales de internet y realmente no se conoce el uso real que se le da a ellos. Una aplicación como la que se quiere desarrollar podría usarse para fines contrarios a los planteados, por eso, es importante limitar la capacidad del sistema o de la empresa que lo vaya a controlar para que la privacidad y los datos personales de los usuarios no se vean afectados.

En cuanto a este trabajo se refiere, como parte inicial de esta aplicación se han llevado a cabo una serie de medidas para recoger la información de los puntos de acceso. Para evitar acceder a cualquier tipo de información no deseada, procedente tanto de los puntos de acceso como de las personas que puedan usar esa red, se ha procedido a realizar este escaneo en modo pasivo. Es decir, se ha analizado la información que los distintos puntos de acceso emiten en broadcast para informar de su existencia y de las condiciones de la conexión, de forma que tanto a nivel moral como legal para esta primera parte del sistema no ha existido problema alguno. Además, para una mayor seguridad los datos obtenidos mediante estos escaneos se encuentran almacenados en una base de datos protegidos mediante contraseña.

1.5. Objetivos

Con el fin de realizar el sistema de localización final, este trabajo ha buscado cumplir una serie de requisitos mínimos que garanticen que dicho sistema es realizable:

- Sistematizar en la medida de lo posible el proceso de escaneo en los distintos puntos de la universidad para lograr una mayor agilidad que nos permita evitar tanto posibles errores a la hora de capturar la información y traspasarla a la base de datos como para hacer esta parte más rápida y sencilla.
- Realizar un muestreo en las distintas plantas del Edificio 4 de la universidad para obtener la información visible en cada punto escaneado para posteriormente poder analizarlo.
- Extraer de dicho muestreo los campos que nos interesan con el fin de sacar una información limpia. Dirección MAC, ESSID, calidad, potencia de la señal y canal.
- Creación y manejo de una base de datos donde podamos introducir la información obtenida del filtrado para posteriormente poder analizar y extraer lo que nos sea necesario en un futuro.
- Validación de los datos almacenados para garantizar que estos no son erróneos, y que los resultados que esperábamos obtener de los escaneos son acordes a lo que tenemos.
- Análisis de la información de la base de datos con el objetivo de comprobar que la red Wi-Fi implantada en la universidad nos va a permitir realizar el sistema de localización.
- La creación de una aplicación web interactiva y en tiempo real que sirva de soporte tanto para la validación de los datos escaneados como para su posterior análisis y uso en la continuación del trabajo para la creación de la aplicación.

1.6. Estructura

Esta memoria ha sido dividida en 5 bloques principales para su mejor comprensión y fácil identificación de cada parte.

- Introducción: En ella se puede encontrar información genérica sobre el uso que tienen las nuevas tecnologías y la red Wi-Fi en la actualidad, lo cual motiva la realización de este proyecto. También podemos encontrar los objetivos que se han perseguido durante su realización, así como las posibles trabas legales y éticas que pudiera haber. Por último, se detalla cómo está estructurada la memoria.

- Metodología de medición: Aquí se documenta el equipo utilizado para la realización del trabajo y la manera en que se ha organizado el desarrollo de éste. Primero se ha procedido a realizar las mediciones necesarias y la introducción de estas en la base de datos para obtener la información que posteriormente se iba a analizar. A su vez se han ido haciendo comprobaciones y pruebas de que los datos fueran los esperados.
- Análisis de resultados: En esta parte se analizan los datos que se han ido recabando durante el trabajo, de tal forma que podamos obtener de la misma manera aquella información que pueda resultar útil para el posterior trabajo en el desarrollo del sistema. Además, se detalla el desarrollo de la aplicación web creada para el apoyo y como soporte del trabajo.
- Planificación y presupuesto: En este capítulo se encuentra el desglose de gastos que se necesitan para llevar a cabo el proyecto y como se ha planificado y dividido en el tiempo las distintas tareas para su realización.
- Conclusiones: Finalmente, puede hallarse el resumen de todo el trabajo realizado, así como la información más relevante obtenida gracias a éste.
- Anexos: Aunque no es considerado como un capítulo, hay que mencionar este apartado. Aquí puede encontrarse toda la documentación adicional que se ha usado para la realización de este trabajo, como los distintos scripts que se han usado como apoyo tanto para el escaneo de los datos como para su posterior análisis.

2. METODOLOGÍA DE MEDICIÓN

2.1. Introducción

La manera en que se ha abordado la realización de este trabajo es la siguiente:

- 1- Se ha procedido a la medición de forma manual de los distintos puntos previamente seleccionados para ser objeto de estudio obteniendo la información que nos interesa para su posterior análisis.
- 2- Posteriormente, gracias al sistema de scripts, la medición realizada en cada punto era incluida en la base de datos previamente creada y habilitada como almacén de los datos relevantes que se obtenían en cada lugar.
- 3- Esta información que se encuentra en la base de datos finalmente es analizada para obtener los datos más relevantes para nuestro estudio, y se usa además gracias a una aplicación web para mostrar de forma interactiva en los mapas de las distintas plantas y así poder observar con una mayor facilidad y comprensión los datos que hemos ido obteniendo.
- 4- Existe una 4ª iteración, que, aunque no se trata de una parte del proceso en sí, cabe destacarla. Durante el proceso de captura, se han ido comprobando los datos incluidos en la base de datos con el fin de evitar problemas en el funcionamiento del escaneo, y en caso necesario, para realizar mejoras o cambios en los scripts.

Todo este proceso puede observarse en la imagen inferior, que esquematiza todo ello.

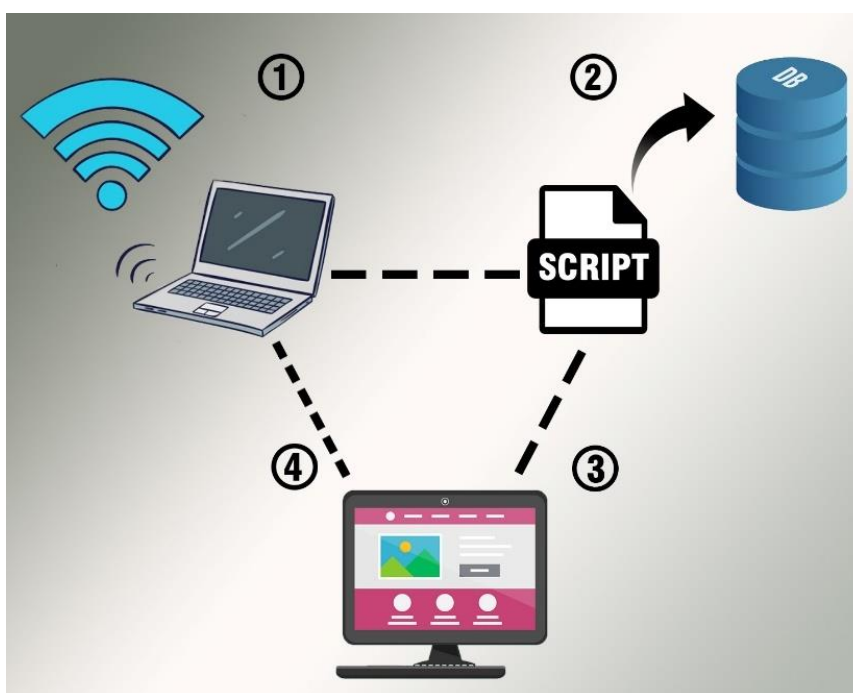


Imagen 3. Esquema de la Metodología

2.2. Descripción del entorno

Para la realización del sistema se ha elegido como entorno de pruebas el edificio 4 del campus de Leganés, debido a que se trata del lugar en el que se encuentra el Departamento de Telemática, en el que se enmarca este proyecto.

El edificio consta de 5 plantas:

- Sótano, en el que podemos encontrar laboratorios docentes y laboratorios de investigación, al igual que una salida de emergencia.
- Planta 0, aquí se encuentra la entrada del edificio, así como salidas de emergencia, aulas docentes, laboratorios docentes y de investigación y también despachos de profesores.
- Planta 1, nuevamente encontramos laboratorios docentes y de investigación, así como aulas y despachos.
- Planta 2, hallamos laboratorios de investigación, aulas y despachos.
- Planta 3, en ella solo hay despachos y laboratorios de investigación.

En la imagen inferior podemos observar tanto la forma, como la distribución de las 5 plantas que se van a estudiar.



Imagen 4. Montaje de las plantas del Edificio 4

Como es evidente, en todas las plantas existen además otros espacios en los que medir, además de los mencionados con anterioridad, como son los pasillos, escaleras y baños.

Es importante añadir que hay dos zonas de cristalerías que ofrecen la visión de, por un lado, pasillos, y por el otro, despachos de otras plantas, lo cual, como se verá posteriormente en el análisis de los datos, hace que sean visibles puntos de acceso de otras plantas debido a estos espacios abiertos.

2.3. Equipo

Para llevar a cabo el estudio de los puntos de acceso de la red Wi-Fi del edificio, el material necesario era principalmente un ordenador portátil que nos permitiera movernos fácilmente por los distintos puntos a escanear y que sirviera como servidor para la base de datos y de la aplicación web, que como hemos visto en la *Imagen 3*, que describe el trabajo realizado, es todo cuanto se necesita para hacer la captura y el análisis que queremos realizar.

Las características de hardware del equipo utilizado son:

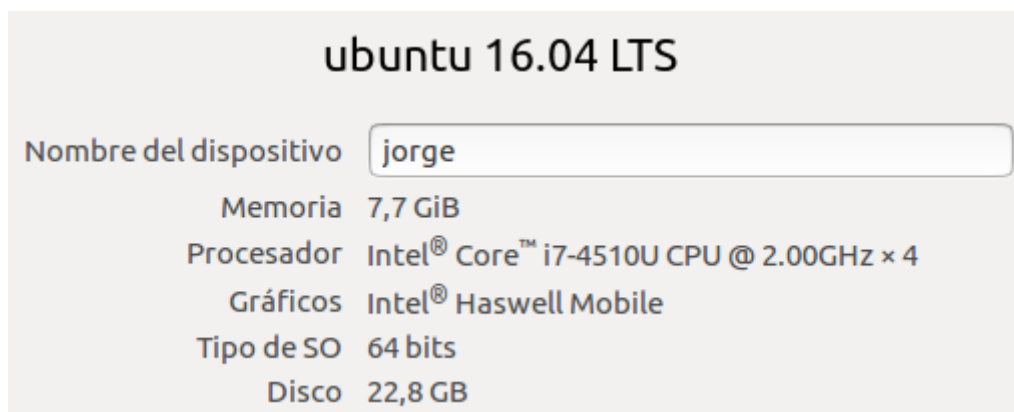


Imagen 5. Características del equipo utilizado

Donde el sistema operativo que se ha utilizado es GNU/Linux con una distribución Debian, concretamente la 16.04.

Se ha elegido esta distribución debido a la facilidad en su manejo y a que se trataba de una distribución con la que se estaba ya previamente familiarizado. Se dispone de 22.8 GB, suficiente para los programas que tenemos que instalar para poder llevar a cabo los distintos objetivos o módulos que nos hemos propuesto para el correcto desarrollo del trabajo.

El software que hemos tenido que instalar y usar es el siguiente:

- MySQL: Es un sistema de gestión de bases de datos relacional considerada una de las bases de datos de código abierto más utilizada del mundo, en especial para el desarrollo web. Este ha sido un motivo para utilizarla junto a que se trataba de un sistema conocido previamente y cuyo uso es bastante sencillo.

- Apache: Para el desarrollo de la anteriormente mencionada aplicación web para la visualización de los datos obtenidos se ha utilizado este servidor web gratuito con el que ya se estaba familiarizado por haberlo utilizado anteriormente durante el grado.
- Geoserver: Se trata de un servidor de código abierto que permite editar datos geoespaciales. Se ha utilizado para que las imágenes que teníamos de las distintas plantas del edificio tuvieran un formato geoespacial y de esa manera que pudiéramos representar y usar los datos almacenados en la base de datos de forma que correspondiera a su localización correcta.
- Python: Es un lenguaje de programación que se ha utilizado para la realización de los scripts que permiten la automatización del proceso de captura de datos y la inserción de estos en la base de datos.

2.4. Captura de datos

2.4.1. Información principal

Para comenzar la parte de captura de datos lo primero de todo es proceder a dividir el espacio que hay que escanear y decidir cada cuántos metros se van a realizar dichos escaneos. Es necesario que este espacio sea del tamaño exacto, de forma que con cada escaneo se reciba una potencia de señal donde sea visible la variación de esta con respecto a otro escaneo. Esto nos servirá, como se verá posteriormente, para localizar la zona en la que se encuentra el punto de acceso que emite dicha señal. Por ello, el espacio entre escaneos debe ser lo más ajustado posible, para no perder información. Finalmente, se han utilizado celdas o divisiones de 4x4 metros, es decir de 16 m^2 .

El edificio, como se puede comprobar en la *Imagen 4*, no es rectangular, de forma que sabiendo como se propagan las señales y la forma de cada lugar, se toman una serie de consideraciones:

- En cada despacho se realiza una medición teniendo en cuenta la posibilidad de acceder a ellos, el objetivo de causar el menor trastorno a los ocupantes de ellos y a que al tratarse de un espacio donde no hay barreras que atenúen la señal, suponemos que la potencia recibida será prácticamente idéntica en cualquier lugar del despacho.
- En los baños, debido a los mismos motivos expuestos para el caso de los despachos, se realiza también un solo escaneo.
- En las aulas se intentan realizar la misma cantidad de escaneos que al hacer la división de cada planta se le han asignado. Sin embargo, hay zonas del edificio, en que el número de medidas se ve condicionado a la forma de este.

- En los laboratorios, tanto docentes como de investigación, al tratarse de zonas de más tamaño que los despachos y con formas variadas, se realizan más de un escaneo por laboratorio. Sin embargo, se tiene en cuenta la forma y el tamaño de cada caso para determinar el número exacto de escaneos y la ubicación en que se realizan.

Gracias a Google Maps y a la posibilidad de obtener el mapa de cada planta, con un gran detalle de cada habitáculo que en él aparece, se elige este como fuente de los mapas tanto para hacer la división en celdas de 16 m^2 como para la aplicación web que se ha mencionado con anterioridad y que veremos posteriormente en el apartado del análisis de resultados.

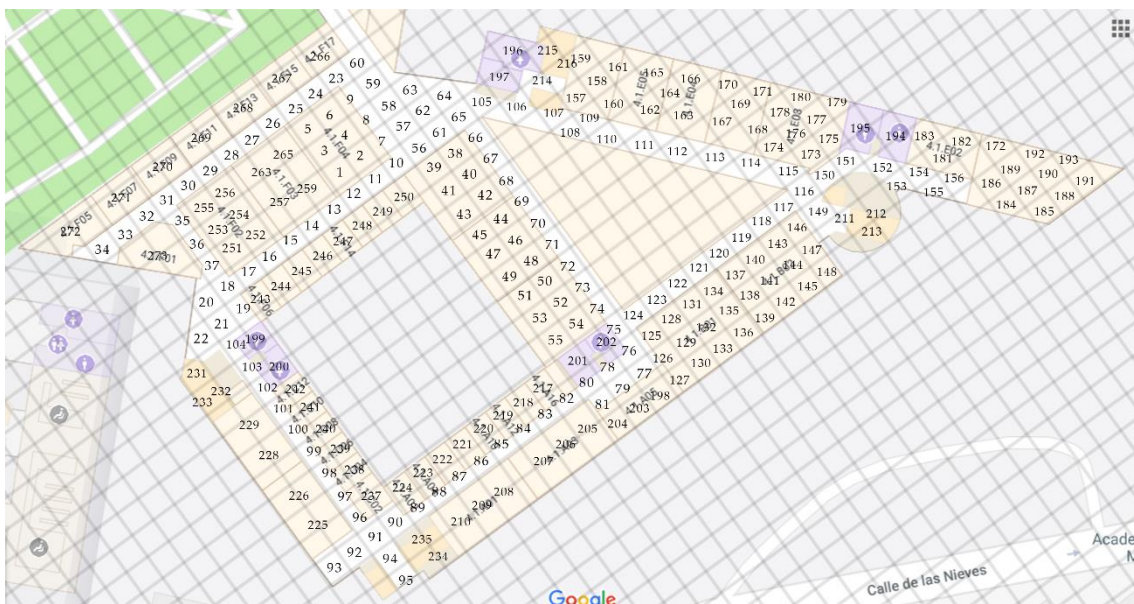


Imagen 6. Ejemplo de la división de celdas de la planta 1

Como podemos observar en el mapa, las celdas han sido numeradas, ya que necesitamos poder identificar cada escaneo al lugar en el que se ha realizado. En la sección 2.5., en la que se detalla el almacenamiento de la información obtenida en cada uno de los procesos de captura de datos, se explica dónde y para qué se usa esta numeración concretamente.

Una vez se ha hecho la partición del espacio total a escanear, y tenemos ya la ubicación exacta que corresponde a cada escaneo se procede a realizarlos.

Para hacer los escaneos se decide crear una serie de scripts para automatizar y acelerar el proceso de captura de información que se explican y detallan a continuación:

- El primero de ellos, *filtrado.sh*, se encarga de realizar el escaneo de los puntos de acceso que se ven desde el lugar en que nos encontramos. Para ello utiliza el comando `sudo iwlist wlp2s0 scanning` y posteriormente obtiene los campos que

nos interesan para nuestro proyecto. De esta forma se hace un filtrado de toda la información relevante y la obtenemos en un formato fácilmente manejable.

- Posteriormente, aunque no se trata de un script en sí mismo, se trabaja sobre el archivo obtenido del script anterior, eliminando la primera línea que corresponde al nombre de los campos de cada columna y almacenando el resultado en un nuevo archivo de texto.
- El siguiente paso lo realiza el script, *ordenado.sh*, que se encarga de leer el archivo de texto que hemos obtenido del anterior paso. Con cada punto de acceso en una línea, descompone cada línea en los distintos campos que introducimos en la base de datos, concretamente en una tabla auxiliar correspondiente a este escaneo.

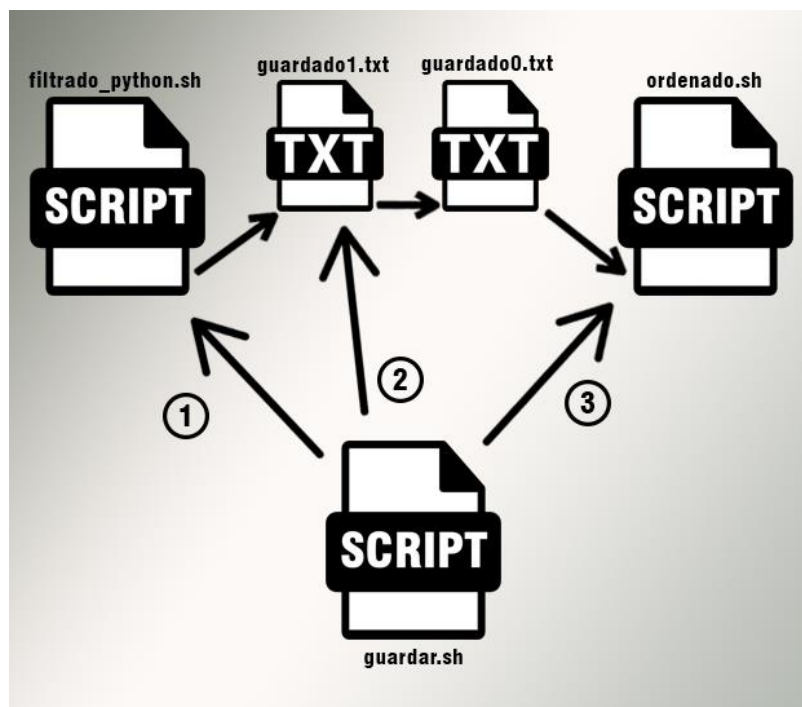


Imagen 7. Esquema explicativo de la captura de datos

Como se indicaba en la explicación del último script, este escaneo realizado se introduce en una tabla auxiliar, llamada *Prueba1*, ya que el proceso de captura de datos en cada punto consta de 3 escaneos, y los datos obtenidos aún no son los definitivos que se almacenarán para el lugar que acaba de ser escaneado.

Esto se lleva a cabo ya que tras realizar las primeras pruebas se evidencia que en ocasiones surgen problemas a la hora de capturar los puntos de acceso. En ocasiones la tarjeta de red se encuentra saturada y no obtenemos ningún punto, o que no siempre capturamos la misma cantidad de puntos en el mismo sitio. Para solventar esto, se procede a realizar, como se indica anteriormente, 3 escaneos en cada punto, de forma se minimicen los posibles problemas o fallos a la hora de capturar la información.

A continuación, se detalla como sigue el proceso de captura, una vez que se ha explicado la razón de realizar tal cantidad de medidas:

Los scripts anteriormente mencionados, funcionan dentro de otro que es el que realmente une todo el proceso y sistematiza toda la captura de datos. Este script, *guardar.sh*, se encarga de repetir estos scripts 3 veces, separando cada operación en 30 segundos, para intentar que la tarjeta de red le dé tiempo a recibir posibles nuevos mensajes de broadcast de los distintos puntos de acceso y así capturar aquellos que pudieran haberse perdido en los otros escaneos. Los datos de cada captura se introducen en 3 tablas auxiliares distintas, debido a que así resultaba más sencillo para manejarlos durante los siguientes pasos que se detallan en el siguiente párrafo.

Una vez que los datos se encuentran en las 3 tablas, el siguiente paso es comprobar estos datos, compararlos entre sí e incluirlos en la tabla definitiva de la base de datos. Este trabajo lo realiza el script *juntar_de_tabla_total.sh* cuyo funcionamiento vamos a ver a continuación.

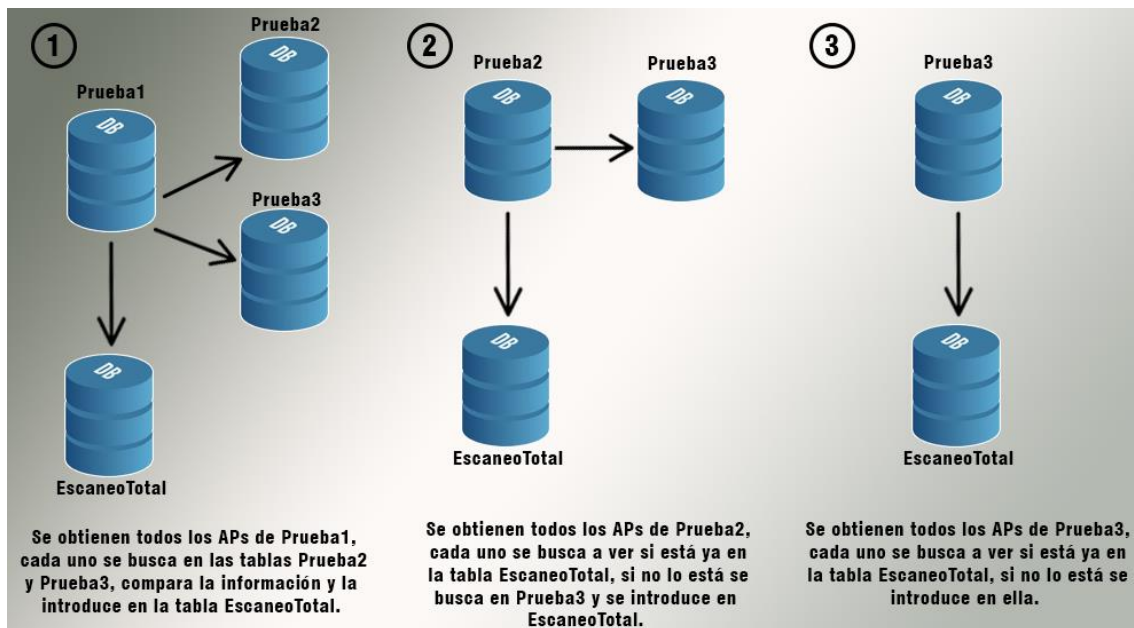


Imagen 8. Esquema del funcionamiento del script "juntar_de_tabla_total.sh"

- Paso 1: lo primero que hace es ir extrayendo de la tabla *Prueba1*, cada punto de acceso uno a uno para comprobar si es un punto de acceso que tenemos también en las otras tablas, si es así, se procede a comparar los valores que pueden cambiar entre un escaneo y otro, esto es, la calidad y el valor de la señal recibida. Se hace la media de éstos para, como anteriormente se mencionaba, solventar los posibles errores que pueda haber de variación entre cada captura de datos. Una vez comparados los valores de cada punto de acceso de la primera tabla con los de las otras dos, se procede a añadir esta información a la tabla *EscaneoTotal*, que tiene

los datos definitivos de los escaneos en cada punto y que se verá en el próximo capítulo.

- Paso 2: Después se realiza lo mismo con la tabla 2, aunque en esta ocasión se comprueba si cada punto de acceso se encuentra ya en la tabla *EscaneoTotal*, y se obvian los que ya se encuentran introducidos. Aquellos que no han sido introducidos ya, son los que el primer escaneo no ha detectado, y solo se comparan con los almacenados en la tabla 3, pues es la única tabla que puede contenerlos, ya que los de la uno se estudiaron en el paso anterior. Tras esta comparación son introducidos en la tabla *EscaneoTotal*.
- Paso 3: Para finalizar se comprueba que los puntos de la tercera tabla estén ya en la tabla definitiva, si no lo están se introducen directamente en ella, pues sabemos que ni el primer ni el segundo escaneo los detectó y no hay con que compararlos.

De esta manera se consigue que los 3 escaneos realizados queden resumidos a uno, donde se minimiza la probabilidad de que existan puntos de acceso que se hayan pasado por alto.

- Finalmente queda un último script, *borrar_tablas.sh*, que es llamado por el script principal del proceso de captura. Se encarga de borrar el contenido de las 3 tablas auxiliares, y crearlas de nuevo para que estén listas para usarse en el siguiente escaneo.

2.4.2. Información relacionada

Una vez explicado el proceso de captura de datos, así como la función de cada uno de los scripts desarrollados hay que aclarar una serie de cosas acerca de la captura de datos.

A la hora de realizar cada escaneo, al añadir la información obtenida en la tabla principal con el resto de los datos que ya se encuentran guardados, se indica el lugar en el que se han realizado. Esto se hace gracias a un parámetro que se encuentra en el script *juntar_de_tabla_total.sh*, que va modificándose según se va avanzando por los distintos puntos de la universidad.

De la misma manera además de introducir los nuevos datos obtenidos en el escaneo junto al punto que identifica dicho escaneo con el lugar físico al que corresponde, se añade la información relativa a ese escaneo, es decir, la fecha y la hora del escaneo, que como veremos posteriormente nos puede ser de gran utilidad tanto para analizar los resultados obtenidos, como para el futuro del sistema de localización.

En relación con esto último, hay que decir que al principio del proyecto no se tuvo en cuenta la posibilidad de escanear en un mismo punto más de una vez, pero fue una mejora que se implementó durante el trabajo. Posteriormente se entendió que se trataba de una mejora que debía ser realizada, tanto para la validación de los datos almacenados, que en

capítulos posteriores abordaremos, como para asegurar la escalabilidad y permitir que la base de datos obtenida pueda actualizarse a lo largo del tiempo.

2.5. Almacenamiento de datos

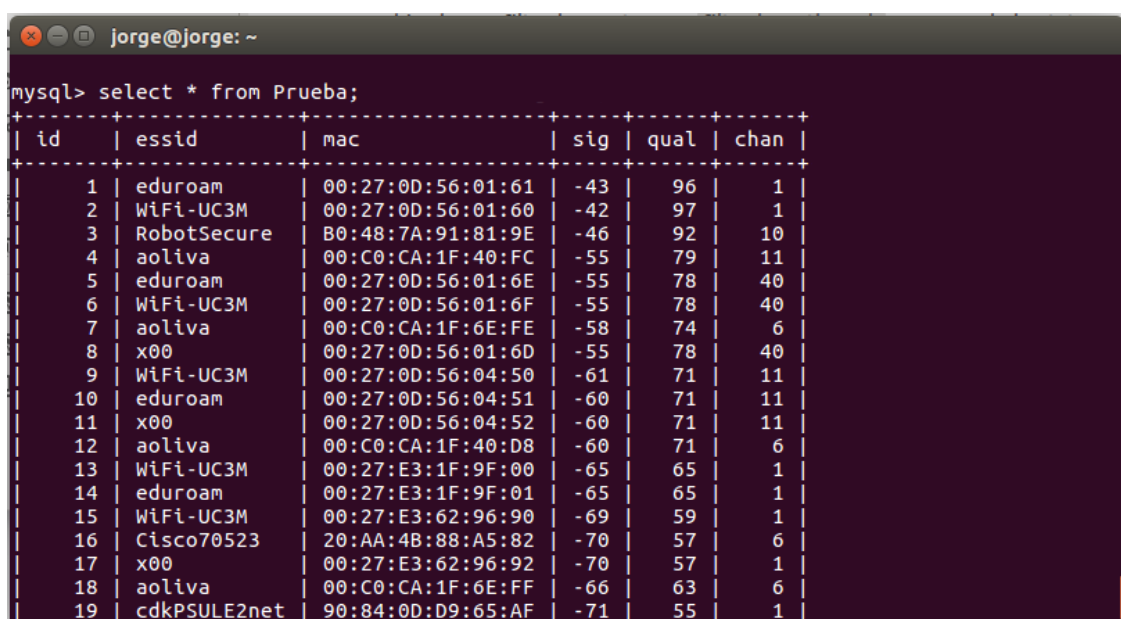
Los datos que nos proporcionan los puntos de acceso en los distintos lugares de la universidad por medio del sistema descrito en el capítulo anterior deben ser almacenados en algún lugar para posteriormente consultarlos fácilmente. Por esta razón se decide crear una base de datos que contenga toda la información relevante a ellos.

Esta base de datos estará formada por distintas tablas, que estarán relacionan entre sí. A través de las similitudes de algunos campos de cada tabla, se creará un sistema de almacenamiento idóneo para el trabajo que se va a realizar. De esta forma podrá seguir utilizándose en los futuros trabajos que continuarán el presente proyecto.

A la hora de construir la base de datos se elige hacerlo a través de MySQL, ya que es un sistema de gestión de bases de datos muy extendido, fácilmente manejable, del que ya se tenía conocimiento previo, como se ha mencionado con anterioridad, y que se puede relacionar fácilmente tanto con el sistema de captura de datos creado como con la aplicación web que se ha desarrollado y de la que se hablará en próximos capítulos.

Para almacenar la información relevante obtenida en los escaneos, junto con la información necesaria para identificarlos, son necesarias una serie de tablas auxiliares, que como ya se mencionó en el capítulo anterior, son un total de 3. Estas tablas, las tablas *Prueba*, de idéntico formato, se van creando y borrando con cada proceso de captura, y sirven para almacenar temporalmente los datos obtenidos en cada uno de los escaneos que se realizan en cada lugar durante el ya mencionado proceso de captura.

Un ejemplo de estas tablas se detalla a continuación:



```
jorge@jorge: ~
mysql> select * from Prueba;
```

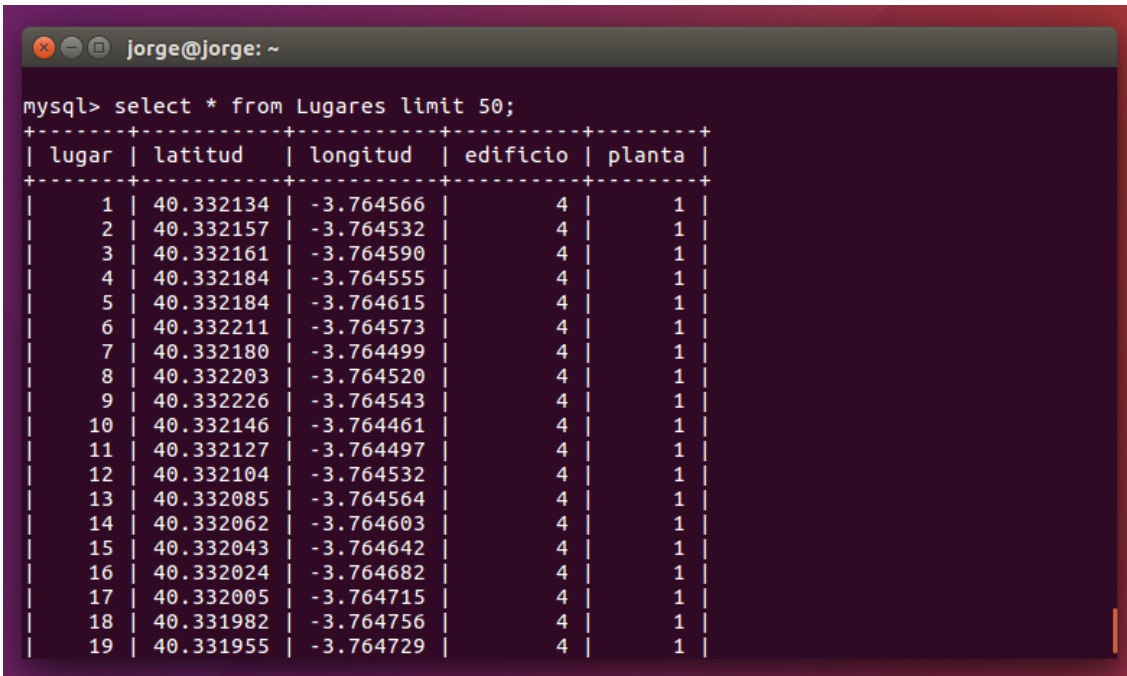
id	ssid	mac	sig	qual	chan
1	eduroam	00:27:0D:56:01:61	-43	96	1
2	WiFi-UC3M	00:27:0D:56:01:60	-42	97	1
3	RobotSecure	B0:48:7A:91:81:9E	-46	92	10
4	aoliva	00:C0:CA:1F:40:FC	-55	79	11
5	eduroam	00:27:0D:56:01:6E	-55	78	40
6	WiFi-UC3M	00:27:0D:56:01:6F	-55	78	40
7	aoliva	00:C0:CA:1F:6E:FE	-58	74	6
8	x00	00:27:0D:56:01:6D	-55	78	40
9	WiFi-UC3M	00:27:0D:56:04:50	-61	71	11
10	eduroam	00:27:0D:56:04:51	-60	71	11
11	x00	00:27:0D:56:04:52	-60	71	11
12	aoliva	00:C0:CA:1F:40:D8	-60	71	6
13	WiFi-UC3M	00:27:E3:1F:9F:00	-65	65	1
14	eduroam	00:27:E3:1F:9F:01	-65	65	1
15	WiFi-UC3M	00:27:E3:62:96:90	-69	59	1
16	Cisco70523	20:AA:4B:88:A5:82	-70	57	6
17	x00	00:27:E3:62:96:92	-70	57	1
18	aoliva	00:C0:CA:1F:6E:FF	-66	63	6
19	cdkPSULE2net	90:84:0D:D9:65:AF	-71	55	1

Imagen 9. Tabla Prueba

Como podemos ver en la *Imagen 7*, estas tablas están formadas por 6 campos. El campo *id* no tiene uso futuro, y simplemente identifica la entrada en la tabla. El campo *ssid* nos indica el nombre público del punto de acceso, así como *mac* hace referencia a la dirección MAC de este. Los campos *sig* y *qual* hacen referencia a la potencia de señal recibida y a la calidad de la señal respectivamente. Por último, *chan* nos identifica el canal en que emite dicho AP.

La tabla *Lugares*, como se puede sobreentender del nombre es la que contiene la información de los puntos que, previamente cuando dividimos el mapa en celdas, asignamos a cada lugar. Estos puntos, tienen asociados su latitud y longitud, lo que nos permitirá en un futuro, poder relacionar la información a un sitio físico, ya sea en un mapa, una aplicación u otro método.

Estos datos de geolocalización han sido añadidos y relacionados con su lugar correspondiente de forma manual. Lo ideal hubiera sido añadirlos a la vez que se realiza el escaneo en cada ubicación, sin embargo, en interiores la señal recibida del GPS no es fiable, y la localización varía. Por ello se toma la decisión de, a pesar de ser una tarea más laboriosa, que se extraigan de Google Maps y se añadan a mano.



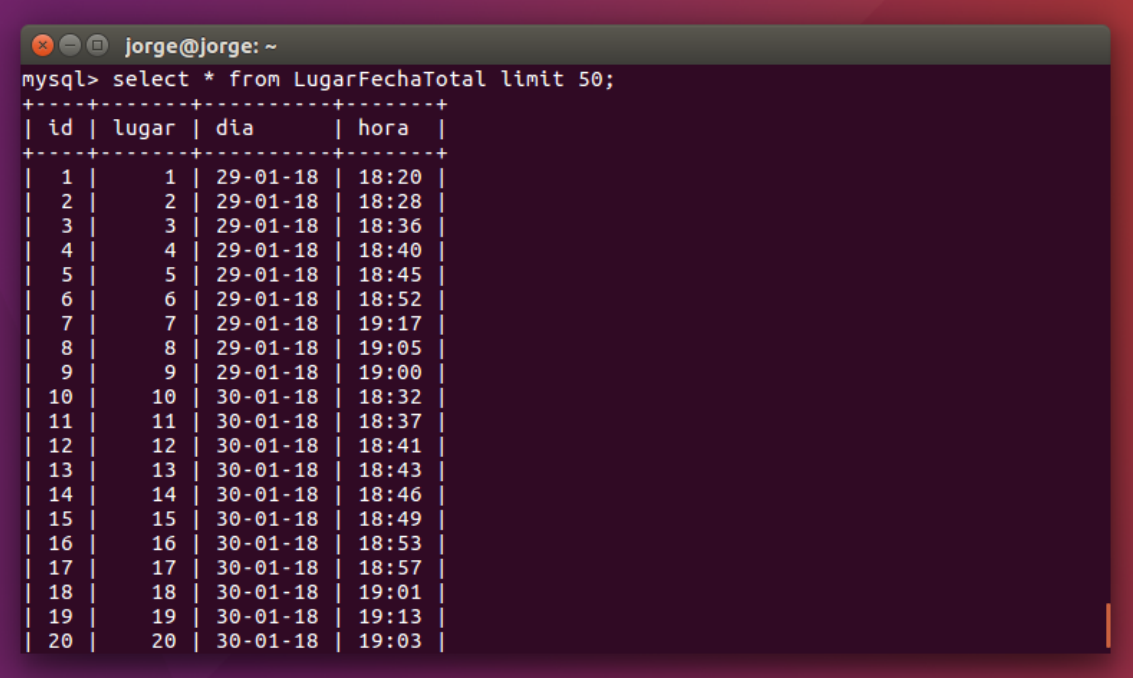
```
jorge@jorge: ~
mysql> select * from Lugares limit 50;
+-----+-----+-----+-----+-----+
| lugar | latitud | longitud | edificio | planta |
+-----+-----+-----+-----+-----+
| 1 | 40.332134 | -3.764566 | 4 | 1 |
| 2 | 40.332157 | -3.764532 | 4 | 1 |
| 3 | 40.332161 | -3.764590 | 4 | 1 |
| 4 | 40.332184 | -3.764555 | 4 | 1 |
| 5 | 40.332184 | -3.764615 | 4 | 1 |
| 6 | 40.332211 | -3.764573 | 4 | 1 |
| 7 | 40.332180 | -3.764499 | 4 | 1 |
| 8 | 40.332203 | -3.764520 | 4 | 1 |
| 9 | 40.332226 | -3.764543 | 4 | 1 |
| 10 | 40.332146 | -3.764461 | 4 | 1 |
| 11 | 40.332127 | -3.764497 | 4 | 1 |
| 12 | 40.332104 | -3.764532 | 4 | 1 |
| 13 | 40.332085 | -3.764564 | 4 | 1 |
| 14 | 40.332062 | -3.764603 | 4 | 1 |
| 15 | 40.332043 | -3.764642 | 4 | 1 |
| 16 | 40.332024 | -3.764682 | 4 | 1 |
| 17 | 40.332005 | -3.764715 | 4 | 1 |
| 18 | 40.331982 | -3.764756 | 4 | 1 |
| 19 | 40.331955 | -3.764729 | 4 | 1 |
```

Imagen 10. Tabla Lugares

La tabla consta de 5 campos, donde *lugar* es el número que define ese sitio y nos sirve para relacionarlo con las distintas tablas de la base de datos y con los mapas marcados y divididos previamente en celdas. *Latitud* y *longitud* como su propio nombre indican, son los campos destinados a los valores de geolocalización de dicho lugar. *Edificio* es un campo creado con la idea de realizar escaneos en otros edificios del campus, y así poder

diferenciar unos de otros fácilmente. Dado que este trabajo solo ocupa el edificio Torres Quevedo, el número 4, todos los lugares poseen esta numeración. Por último y dado que solo tenemos los datos de latitud y longitud, usamos el campo *planta* para poder tener también el valor de la altura a la que está realizado dicho escaneo. Este además servirá para poder filtrar fácilmente los datos a la hora de representarlos por plantas en los mapas interactivos de nuestra aplicación web.

LugarFechaTotal es la tabla que permite identificar cada escaneo con la fecha y la hora en que se realizó. Se puede observar en la parte inferior un ejemplo de lo que hay almacenado en la tabla. El primer campo, *id*, permite identificar de forma única un escaneo con una fecha y hora dada, de tal forma que es posible tener varios escaneos en un mismo punto, con distintas horas. Como se dijo con anterioridad, esta mejora fue implementada durante el desarrollo de la fase de captura de datos. De esta forma, gracias a este campo es posible identificar los datos capturados en la última tabla, que veremos a continuación. El campo *lugar*, nos indica el punto donde se realiza el escaneo. Como es evidente, finalmente están los campos *día* y *hora*, que localizan en el tiempo dicho escaneo.



```
jorge@jorge: ~
mysql> select * from LugarFechaTotal limit 50;
+----+-----+-----+-----+
| id | lugar | dia      | hora  |
+----+-----+-----+-----+
| 1  | 1     | 29-01-18 | 18:20 |
| 2  | 2     | 29-01-18 | 18:28 |
| 3  | 3     | 29-01-18 | 18:36 |
| 4  | 4     | 29-01-18 | 18:40 |
| 5  | 5     | 29-01-18 | 18:45 |
| 6  | 6     | 29-01-18 | 18:52 |
| 7  | 7     | 29-01-18 | 19:17 |
| 8  | 8     | 29-01-18 | 19:05 |
| 9  | 9     | 29-01-18 | 19:00 |
| 10 | 10    | 30-01-18 | 18:32 |
| 11 | 11    | 30-01-18 | 18:37 |
| 12 | 12    | 30-01-18 | 18:41 |
| 13 | 13    | 30-01-18 | 18:43 |
| 14 | 14    | 30-01-18 | 18:46 |
| 15 | 15    | 30-01-18 | 18:49 |
| 16 | 16    | 30-01-18 | 18:53 |
| 17 | 17    | 30-01-18 | 18:57 |
| 18 | 18    | 30-01-18 | 19:01 |
| 19 | 19    | 30-01-18 | 19:13 |
| 20 | 20    | 30-01-18 | 19:03 |
```

Imagen 11. Tabla *LugarFechaTotal*

Por último, la tabla más importante, *EscaneoTotal*, ya que es la que contiene la información que se ha ido obteniendo mediante las distintas capturas realizadas. Cada línea es un punto de acceso distinto. Como puede observarse en el ejemplo inferior, tiene más campos que las otras tablas.

```

jorge@jorge: ~
mysql> select * from EscaneoTotal limit 50;
+-----+-----+-----+-----+-----+-----+-----+-----+
| id | essid      | mac          | sig | qual | chan | lugar | idlugarfecha |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 | WiFi-UC3M | 00:27:0D:56:01:60 | -38 | 100 | 1 | 2 | 2 |
| 2 | eduroam   | 00:27:0D:56:01:61 | -39 | 100 | 1 | 2 | 2 |
| 3 | aoliva    | 00:C0:CA:1F:6E:FF | -40 | 100 | 6 | 2 | 2 |
| 4 | aoliva    | 00:C0:CA:1F:40:FC | -32 | 100 | 11 | 2 | 2 |
| 5 | aoliva    | 00:C0:CA:1F:6E:FE | -42 | 98  | 6 | 2 | 2 |
| 6 | aoliva    | 00:C0:CA:1F:40:D8 | -45 | 93  | 6 | 2 | 2 |
| 7 | eduroam   | 00:27:0D:56:01:6E | -48 | 89  | 40 | 2 | 2 |
| 8 | WiFi-UC3M | 00:27:0D:56:01:6F | -48 | 89  | 40 | 2 | 2 |
| 9 | x00       | 00:27:0D:56:01:6D | -48 | 89  | 40 | 2 | 2 |
| 10 | RobotSecure | B0:48:7A:91:81:9E | -57 | 76  | 10 | 2 | 2 |
| 11 | eduroam   | 00:27:E3:1F:9F:01 | -58 | 74  | 1 | 2 | 2 |
| 12 | WiFi-UC3M | 00:27:E3:1F:9F:00 | -59 | 73  | 1 | 2 | 2 |
| 13 | x00       | 00:27:E3:1F:9F:02 | -59 | 73  | 1 | 2 | 2 |
| 14 | WiFi-UC3M | 00:27:0D:56:04:50 | -68 | 61  | 11 | 2 | 2 |
| 15 | eduroam   | 00:27:0D:56:04:51 | -67 | 62  | 11 | 2 | 2 |
| 16 | eduroam   | A4:6C:2A:86:70:71 | -68 | 60  | 11 | 2 | 2 |
| 17 | WiFi-UC3M | A4:6C:2A:86:70:70 | -69 | 59  | 11 | 2 | 2 |
| 18 | x00       | A4:6C:2A:86:70:72 | -71 | 57  | 11 | 2 | 2 |
| 19 | x00       | 00:27:0D:56:04:52 | -70 | 58  | 11 | 2 | 2 |
| 20 | x00       | 00:27:E3:62:96:92 | -70 | 57  | 1 | 2 | 2 |

```

Imagen 12. Tabla EscaneoTotal

El primer campo, *id*, es un campo que identifica de forma única cada punto de acceso en la tabla. *Essid* y *mac* son, por un lado, el nombre público y por otro la dirección MAC que tiene cada punto de acceso. El campo *sig* hace referencia a la potencia de señal recibida desde cada punto de acceso, en una escala en decibelios, mientras que *qual*, equivale a la calidad de señal recibida, en una escala del 0 al 100, donde afectan tanto el valor del campo anterior como el ruido. Como ya se ha mencionado en las tablas anteriores, el campo *lugar* que hace referencia a la ubicación donde se realizó el escaneo correspondiente e *idlugarfecha* que corresponde al *id* de la tabla *LugarFechaTotal* que sirve para identificar cada escaneo con el día y la hora en que se realizó, en caso de haber más de uno en dicha ubicación.

3. ANÁLISIS DE RESULTADOS

Una vez se ha explicado como son los lugares en los que debemos escanear, el material con el que vamos a realizar el proyecto, la forma en la que se ha procedido a realizar las mediciones, cómo está diseñada la base de datos, y cómo se introduce la información en ella para su posterior estudio, podemos comenzar a realizar las distintas operaciones para el estudio de los datos, empezando por su validación para comprobar que son fiables y a continuación su posterior análisis más minucioso.

Para ello, como se ha comentado en capítulos anteriores, además se ha desarrollado una aplicación web como herramienta de apoyo a este análisis, haciendo de forma más interactiva y visual la obtención de resultados.

3.1. Pruebas de validación

Como hemos explicado anteriormente, a pesar de haber llevado a cabo una serie de medidas para intentar minimizar los posibles errores o fallos que pueda haber a la hora de realizar el escaneo, se llevan a cabo una serie de pruebas durante el desarrollo del trabajo con el fin de lograr un trabajo lo más sólido y veraz posible y tratar de asegurar de esa manera que los datos almacenados son consistentes y se pueden utilizar para el desarrollo del sistema de localización. Dado que el objetivo es que se trate de un sistema que utilice el sistema Wi-Fi de la universidad se usará como referencia tanto a la hora de analizar los datos como para las pruebas de validación, los datos obtenidos referentes a *eduroam*. Para conectarse a dicha señal es necesario registrarse previamente, de forma que se supone que será una señal Wi-Fi con menos uso de lo que puede ser *Wifi-UC3M*, ya que a ésta puede conectarse cualquier persona, sin necesidad de pertenecer a la comunidad universitaria.

3.1.1. Localizar los AP por la potencia de señal recibida

Se trata de una de las pruebas más evidentes, necesarias e importantes a realizar. Gracias a los datos que se han ido recopilando mediante los escaneos, analizando estos, dada una dirección MAC concreta que identifica a un punto de acceso, vemos como dependiendo del lugar en el que se ha encontrado ese AP tiene distintos valores de *quality* de la señal. De esa forma se usará este valor para estimar si el punto de origen de la señal se encuentra más cerca o más lejos del lugar del escaneo. Cuanto más alto sea, más cerca estará.

Por ello, se desarrolla en la aplicación web que se ha creado como apoyo al trabajo, la opción de mostrar el área en el que se observa un AP en cada planta dada una MAC concreta.



Imagen 13. Ejemplo localización de un AP

En la *Imagen 11*, se observa el mapa de la planta 1 y una serie de puntos que representan los lugares en el mapa donde al escanear hemos visto dicha dirección MAC. Los puntos tienen distinto color según el valor que tiene el campo *quality* de la señal, de forma que este indica cómo de próximo se está al AP, ya que cuanto más cerca, la calidad será mayor. En la imagen inferior, se puede observar la leyenda de colores según los valores de calidad.



Imagen 14. Leyenda de colores según la calidad de la señal

Así se puede deducir de forma aproximada el lugar en el que debe estar el punto de acceso cuya MAC es la que se ha estado observando, ya que será la zona en la que se encuentran los puntos de color azul oscuro.

Para comprobar si la ubicación aproximada que se obtiene del AP es la correcta, gracias a Servicios Informáticos de la universidad, que nos ha proporcionado la ubicación real de cada punto de acceso, es posible comprobar que efectivamente los datos obtenidos muestran de una forma casi exacta, el lugar en el que se encuentra el punto de acceso.

A continuación, se exponen varios ejemplos más correspondientes a distintos APs, cada uno de ellos perteneciente a una planta distinta. Todos ellos han sido comprobados nuevamente con la información que Servicios Informáticos nos proporcionó y ha sido posible cerciorarse de que son correctos.

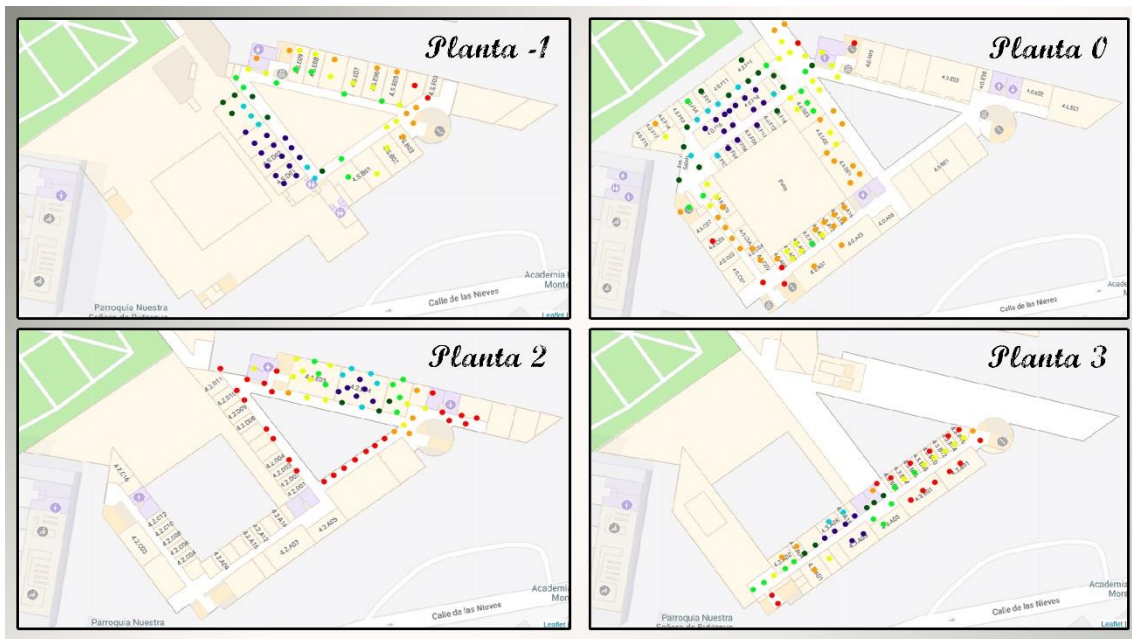


Imagen 15. Ejemplos extra de la localización de APs

3.1.2. Comparación de señales Wi-Fi recibidas desde un mismo AP

Como se verá posteriormente en el capítulo de análisis de resultados, los puntos de acceso oficiales de la universidad dan soporte a más de una red Wi-Fi a la vez. Este es el caso de *Wifi-UC3M* y *eduroam*. Como las dos redes Wi-Fi parten del mismo punto de acceso, implica que ambas deben tener los mismos valores de señal y calidad de ésta en un mismo lugar.

Es por ello por lo que se han analizado todos los datos que tenemos de estas dos redes Wi-Fi en todos los puntos donde se ha escaneado. Así, es posible comprobar que ambos valores son aproximadamente iguales, lo que servirá para comprobar nuevamente que lo teóricamente esperado antes de comenzar este estudio se cumple finalmente al realizarlo.

En nuestra base de datos, tenemos un total de 7827 puntos de *eduroam* mientras que de *Wifi-UC3M* tenemos 7921. Esta diferencia atiende al problema comentado en el apartado anterior, el referido a la captura de datos. Se entiende que al escanear habrá una serie de puntos que se perderán, ya sea porque la tarjeta de red esté saturada, porque al ser dos redes Wi-Fi distintas el mensaje de anuncio de la red llegue en distinto momento o que la señal en ese punto sea baja y el hecho de encontrar una de las dos redes sea ya algo inusual.

Una vez visto esto, se procede a emparejar cada punto de *Wifi-UC3M* con su correspondiente en *eduroam*, utilizando los campos *mac* (que no son iguales, pero se diferencian en solo el último número o letra de la dirección MAC), y los campos *lugar* e *idlugarfecha* para obtener la pareja de cada dirección MAC con los datos obtenidos ese mismo día, y no el obtenido en otro hipotético escaneo donde puede haber cambiado el valor. Así se comprueba otra de las cosas que se podían intuir ya al saber la diferencia entre la cantidad de puntos de cada red Wi-Fi, que hay puntos correspondientes a un AP

en un lugar dado del que solo tenemos una de las dos redes. Tenemos un total de 5817 parejas *Wifi-UC3M* y *eduroam*. Esto quiere decir que tenemos almacenados 2010 puntos donde sí tenemos datos de *eduroam* pero no de *Wifi-UC3M* y un total de 2104 donde tenemos *Wifi-UC3M* y no *eduroam*.

Primero se van a analizar los puntos en los que sí hay ambas redes, para comprobar que tienen valores similares para un mismo AP en el mismo lugar. Para ello se utiliza esta gráfica donde cada punto corresponde a un par del anteriormente mencionado. En el eje X tenemos representado el correspondiente a *eduroam* y en el eje Y a *Wifi-UC3M*. Los 5817 pares de puntos de ambos se distribuyen en forma de una recta, que era lo esperado, ya que, en cada lugar, según la cercanía a la que nos encontremos del punto de acceso tendrá una calidad u otra, teniendo un abanico de datos desde el 100% de la calidad hasta el 35% aproximadamente. La recta que forman los puntos esta sobreimpuesta a estos en rojo para que resulte más sencillo observarla.

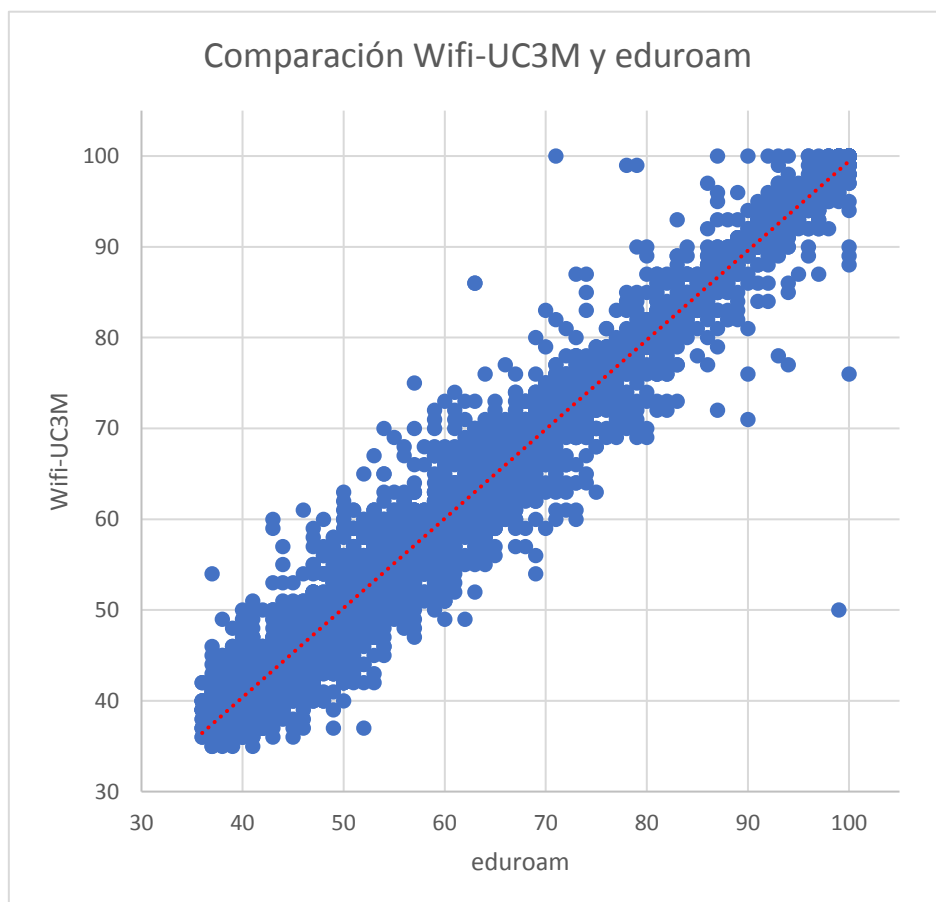


Imagen 16. Comparación de los valores de *Wifi-UC3M* y *eduroam*

Una vez visto esto, se procede a analizar los puntos en los cuales solo hay una de las dos redes. Como se ha dicho anteriormente existen varias razones para que haya escaneos donde una de estas dos redes no se haya recibido. Una posible saturación de la tarjeta de red, que el mensaje de anuncio que envían los APs llegue en distinto momento, de forma

que se pierda uno de ellos... Otro posible motivo es que la distancia al punto de origen sea grande, y la potencia de señal que se reciba sea baja, de forma que el hecho de recibir una de las dos redes Wi-Fi sea ya algo extraño. Es por ello por lo que se ha decidido contar cuantos de estos puntos tienen una potencia de señal inferior a lo que consideramos mínimo. Como se puede ver en la imagen inferior, se considera mala calidad para valores de señal inferiores a -70 dBm, que es aproximadamente el 57% de calidad. Del total de los puntos desemparejados, el 56,11% de *eduroam* y el 55.18% de *Wifi-UC3M* son inferiores a este valor, por lo que es lógico pensar que estos datos se escapen al escaneo debido al motivo que se ha explicado con anterioridad. En cuanto al resto de puntos, es de suponer que el motivo se encontrará entre las otras razones mencionadas.

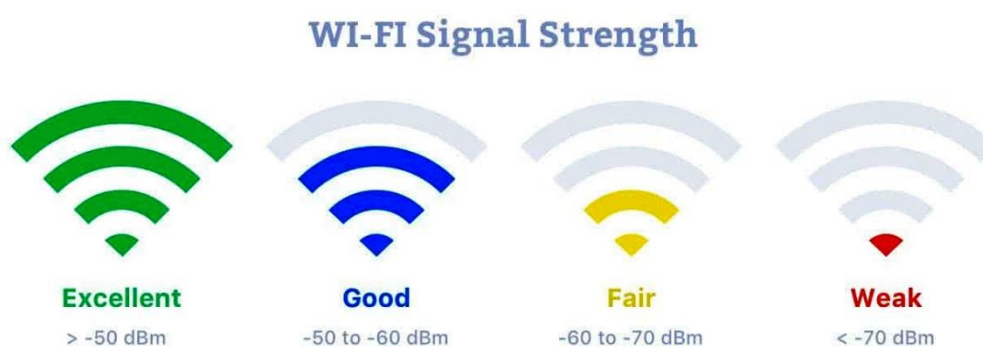


Imagen 17. Calidad de la señal según la potencia recibida [6]

3.1.3. Verificar ubicación de APs externos y de Rogue APs

La tercera prueba que se ha llevado a cabo para garantizar que los datos almacenados son fiables es, apoyándonos de nuevo en la aplicación web desarrollada, mostrar los lugares en los que una *ssid* concreta es visible. Sin embargo, estas *ssid* son externas a la universidad o no oficiales (Rogue APs), de forma que sabiendo o bien el departamento al que pertenecen o, si no, que al tratarse de redes externas deben encontrarse en una zona del edificio concretas, se puede asegurar que los puntos donde son visibles son los esperados.

Se han realizado 4 pruebas de este tipo, dos con redes no oficiales y otras dos con redes externas. Primero se van a explicar las realizadas en redes no oficiales para más adelante atender a las redes externas.

Como primer ejemplo tenemos la red Wi-Fi *WORST*. Se trata de una red Wi-Fi ubicada en la planta 0, cerca del despacho que pertenece a la Asociación BEST. Como se puede observar en las imágenes, es en el plano de la planta 0 donde se encuentran la mayoría de los puntos. Debido a que el AP se encuentra en un despacho que da al patio cuadrado del edificio y al tratarse éste de un espacio libre, observamos en el resto de las imágenes como

en los puntos que bordean este patio es visible, pues no tiene elementos que destruyan la señal, siendo principalmente y como es esperado la planta 1 la siguiente con mayor cantidad de lugares donde al escanear hemos podido observar dicha red Wi-Fi.

WORST



Imagen 18. Puntos donde es visible la red WORST

La red Wi-Fi *TSCWLANG* pertenece al Departamento de Teoría de la Señal y es la red wifi utilizada por él, por tanto, tiene puntos de acceso en distintas plantas del edificio, al contrario que el caso anterior. Como el Departamento de Teoría de la Señal tiene despachos y laboratorios tanto en la planta 0, como 2 y 3 hay puntos de acceso en estas plantas, lo que provoca que sea visible prácticamente en la totalidad del edificio. En el sótano al no tener un punto de acceso, pero sí tener el patio triangular en medio, observamos esta red en casi todos los puntos. En la planta 0 y 1 es visible prácticamente en la totalidad de ellas, menos en la parte que está más cerca de los jardines, el ala *F*, debido a que esa zona no es del departamento y hay aulas y despachos que atenúan la señal procedente o bien de otras plantas a través de los patios interiores, o de la misma planta, pero del ala contraria, el *A*.

TSCWLANG



Imagen 19. Puntos donde es visible la red TSCWLANG

Los dos siguientes ejemplos son de redes Wi-Fi externas a la universidad. En los puntos cercanos a los edificios que rodean la universidad se detectan una gran cantidad de redes Wi-Fi de particulares, siendo éstos de las principales compañías de telecomunicaciones, sobre todo. Ambos ejemplos corresponden a dos de estas compañías, *Orange* y *Vodafone*, lo que casi con total seguridad garantiza que son de alguno de los pisos vecinos, pues es muy poco probable que sean de la universidad. La zona donde se encuentran los edificios, el ala *A*, *B* y el *E* son precisamente las zonas donde son visibles los puntos de estas dos redes Wi-Fi. En la imagen superior, *Imagen 20*, la cual corresponde a la red *Orange*, vemos como los puntos donde captamos la señal están en la parte *A*, *B* y final de la *E* del edificio, lo que supone que el punto de acceso estará en esa zona, la parte que da a los edificios que se encuentran enfrente del ala *A* y *B* principalmente. Se ha omitido el mapa del sótano pues ahí no tenemos ningún punto registrado. Por el contrario, en la imagen inferior, la que corresponde a la red *Vodafone*, *Imagen 21*, vemos cómo ésta es visible en casi la totalidad del ala *E*, lo que nos hace pensar que el punto de acceso está cercano al polideportivo, posiblemente en los edificios que se encuentran detrás de éste. Otro motivo es que en la planta 3, que no tiene ala *E* no tenemos ningún lugar desde el que se ve esta

red, lo cual nos reafirma a pensar que está en la zona anteriormente comentada. Por esta razón se ha decidido omitir el mapa de la planta 3 en la imagen.

Orange_0B20

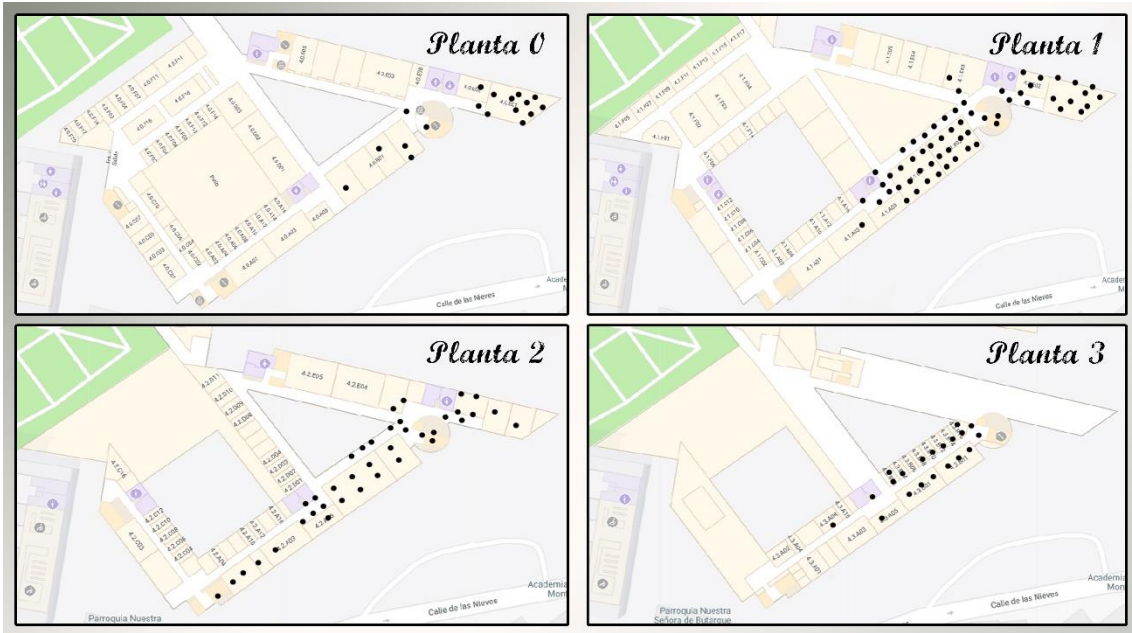


Imagen 20. Puntos donde es visible la red Orange_0B20

VodafoneA3EA

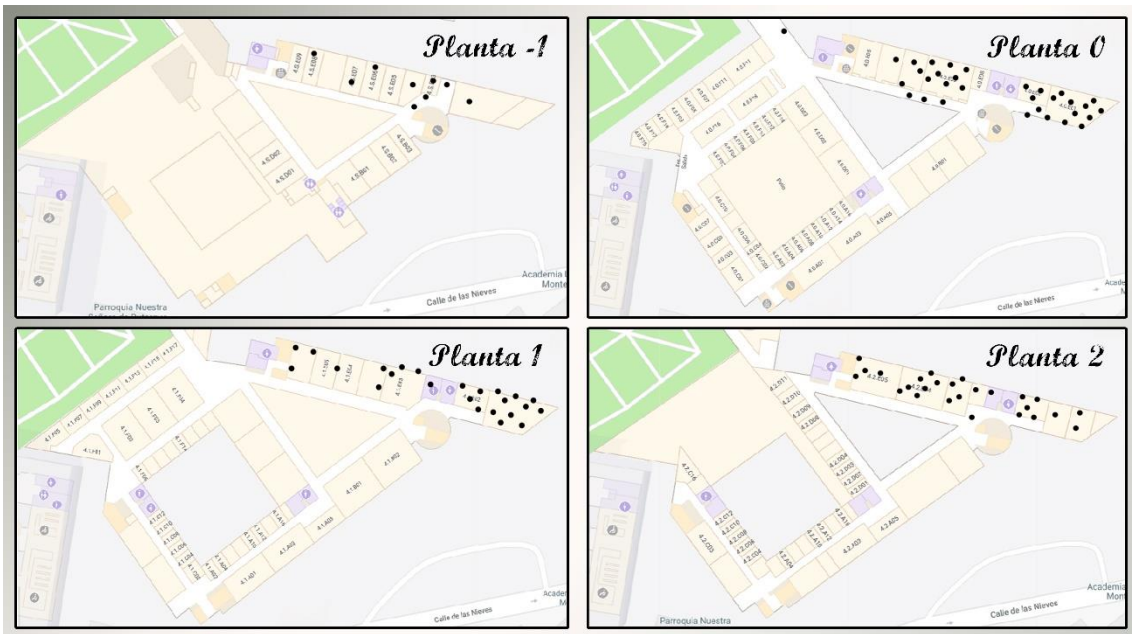


Imagen 21. Puntos donde es visible la red VodafoneA3EA

3.2. Aplicación Web para visualización de contenidos

Para el desarrollo de esta herramienta se han utilizado un conjunto de Servlets de Java así como JSPs relacionados entre sí para formar una misma aplicación.

El funcionamiento de nuestra aplicación sigue el patrón del Modelo Vista Controlador (MVC) donde:

- Modelo: Es la parte encargada del almacenamiento de la información, en nuestro caso se trata de la base de datos que hemos creado y que hemos ido llenando de información según íbamos realizando los escaneos.
- Vista: Es la interfaz de usuario, es decir la parte visible. Se realiza en forma de páginas HTML (que es básicamente un JSP aunque nos proporciona más posibilidades).
- Controlador: Nos permite relacionar la parte visual con la parte que almacena la información. Éste recibe las peticiones que hace el usuario en la interfaz y se encarga de manejar la base de datos para obtener la información que nos interese para después mostrarla al usuario en una nueva vista.

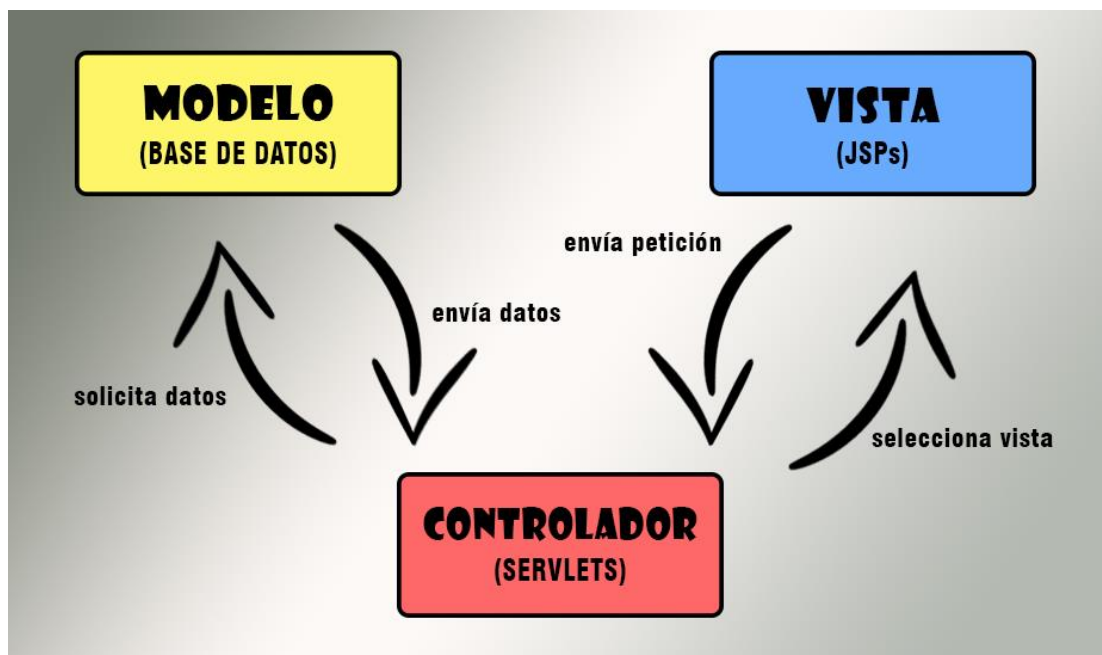


Imagen 22. Explicación del MVC de este trabajo

En el caso de nuestra aplicación el principal uso que se le da es el de mostrar los datos que tenemos almacenados en la base de datos de una forma visual, a través de los mapas de las distintas plantas del edificio. Como tenemos una gran variedad de datos, los datos a mostrar son diversos, desde, como se mencionaba en el capítulo anterior, la calidad con la que se ve en cada lugar un punto de acceso, al canal en el que emite el *eduroam* con mejor calidad en cada punto, los lugares en los que la calidad es mala...

Sin embargo, para poder utilizar una imagen de una planta como mapa, con unos parámetros de geolocalización que nos permitieran representar de forma adecuada la información que teníamos guardada en la base de datos, la cual previamente ya habíamos marcado con unas posiciones concretas, se ha tenido que utilizar otra herramienta, *Geoserver*. Este servidor nos permite a través de una imagen guardada en formato *tiff* a la que previamente hemos insertado unos datos referentes a su dimensión, almacenarla de forma que podamos utilizarla en nuestra aplicación con unos parámetros geoespaciales que nos facilitan representar los datos recopilados simplemente teniendo que hacer una conversión sencilla que relaciona el tamaño del mapa con los puntos de latitud y longitud que lo delimitan.

A la hora de crear nuestra aplicación web hemos elegido un diseño simple e intuitivo donde lo importante es el contenido. Como podemos ver en la imagen inferior, al arrancar la web tenemos una portada que nos presenta que se trata de una página web de la universidad, con el logo y el nombre de ésta. Tenemos un botón que nos sirve para avanzar a la siguiente vista.



Imagen 23. Pantalla de inicio de la aplicación web

En la siguiente vista disponemos de la lista de opciones. Estas se encuentran ordenadas según la planta que deseamos observar. Para cada planta tenemos las mismas opciones, ya que los estudios que se han realizado son los mismo para cada una de ellas. Tenemos, tanto pruebas de validación, concretamente la mencionada en el apartado 3.1.1. que corresponde a la opción *Rango Cobertura de AP* y la mencionada en el apartado 3.1.3. que corresponde a la opción *Verificación Zona Essid*, como otros análisis que se han realizado. En la imagen inferior podemos observar esta vista.



Imagen 24. Pantalla de opciones de la aplicación web

Por último, está un ejemplo de una de las vistas que hay si seleccionamos alguna de las opciones que tenemos en la vista mostrada por la *Imagen 24*. En ella podemos ver, tanto un mapa que muestra la planta seleccionada, como los puntos que obtenemos con esa opción. En este caso, se ha seleccionado ver el mapa de canales, y dado que muestra un color según el canal que corresponda, en la parte inferior aparece la leyenda de colores, que relaciona cada color con su canal correspondiente. Debajo de la leyenda, aunque no es visible, hay un botón que sirve para regresar a la página anterior.



Imagen 25. Pantalla de muestra de una de las opciones de la app.web

3.3. Análisis general

Una vez ha llegado el momento de analizar la información que tenemos en nuestra base de datos, hay que hacerlo de una manera ordenada para así tener una mayor comprensión de éstos, partiendo de los datos más generales para ir concretando poco a poco en los más específicos.

En total se han realizado un total de 785 escaneos repartidos en las distintas plantas del edificio. De todos ellos, tan solo 6 han sido en un punto de que ya habíamos escaneado previamente. Principalmente, estos escaneos se realizaron para comprobar que era posible realizar más de un escaneo en cada lugar, accediendo de forma correcta a la información, y distinguiendo sin problema los datos almacenados en cada escaneo en esa misma ubicación en distintas fechas.

La distribución de estos escaneos es acorde al tamaño de cada planta ya que como se ha explicado en la metodología de medición, éstas se iban haciendo de forma ordenada y precisa según el tamaño del lugar a escanear.

Estos 779 escaneos en puntos distintos se distribuyen de la siguiente manera:

Planta -1	Planta 0	Planta 1	Planta 2	Planta 3
86	230	256	153	54

Imagen 26. Distribución de los escaneos por plantas

Al realizar estos escaneos, más aquellos realizados doblemente en ciertos lugares, se obtuvieron un total de 36078 entradas en la tabla *EscaneoTotal*, pertenecientes a los distintos puntos de acceso que se ven en los 779 lugares donde se escaneo. Estos puntos nuevamente se distribuyen de una forma similar que los escaneos, pues como es evidente, el número de puntos añadidos es proporcional al número de escaneos realizado.

Planta -1	Planta 0	Planta 1	Planta 2	Planta 3
4284	11318	12317	6014	1910

Imagen 27. Distribución de las entradas de EscaneoTotal por plantas

Es evidente que no hay tanta cantidad de puntos de acceso, sino que los mismos puntos de acceso son registrados en los distintos puntos en los que son visibles, variando los valores de potencia de señal recibida y calidad, que dependen de la distancia a la que se encuentren del punto de acceso.

Para saber exactamente el número de puntos de acceso distintos que hemos capturado durante el proceso de escaneado hemos utilizado dos métodos distintos, los cuales, aunque en un primer momento sorprendentemente daban resultados distintos, tras analizarlos detalladamente se comprobó cual era el motivo.

Por un lado, gracias a la posibilidad que nos da la base de datos utilizada, *mysql*, realizar consultas es muy sencillo y de esa manera pudimos, como hemos hecho en los datos que hemos mencionado justo antes, obtener del número de datos obtenido (36078) cual es la cantidad de puntos de acceso distintos, gracias al campo *mac*, así hemos filtrado la cantidad de resultados distintos para este campo, obteniendo un total de 794. Esto significa que hay en total 794 puntos de acceso distintos en nuestra base de datos.

En esta ocasión para ver la distribución de estos AP por planta, se ha seguido una lógica distinta. Como ya se ha explicado anteriormente, se capturan datos acerca del mismo punto de acceso en distintos lugares, lo que significa que este mismo se verá en distintas plantas, por tanto, en vez de intentar saber cuál es la cantidad de puntos de acceso ubicados en cada planta, obtendremos la información de, sobre los puntos totales de esa planta, cuantos puntos de acceso distintos se ven en cada una. Esto como se ve en la tabla de abajo, hace que la suma no sea 794, sino un número mucho más alto.

Planta -1	Planta 0	Planta 1	Planta 2	Planta 3
235	489	585	489	256

Imagen 28. Número de APs distintos en cada planta

Esto corresponde al primer método, sin embargo, no es el único utilizado. Otra forma de analizar esta misma información es a través de las distintas redes wifi que tenemos.

Si se utiliza el filtrado de datos para obtener las distintas redes Wi-Fi que hay, es decir, a través del campo *ssid*, los distintos nombres almacenados, y vemos a través de cada uno de estos, el número de direcciones MAC distintas que tienen este nombre, podemos obtener de igual manera el número de puntos de acceso. Sin embargo, al utilizar este método, pudimos comprobar como no cuadraba con los resultados mencionados anteriormente. De esta forma se obtiene un total de 834 puntos de acceso distintos. Dado que el primer método era simple e inequívoco, algo anómalo debía ocurrir al hacerlo del segundo método. De esta manera se comprobó como existían ciertas direcciones MAC que daban cobertura a dos ESSID distintas. Tras analizar este dato, se llegó a la conclusión de que la razón más plausible era que debido a que el proceso de captura de análisis duró varios meses, es posible que se produjera alguna sustitución de algún punto de acceso, o que este fuera reconfigurado durante este periodo, provocando esta anomalía. Como se verá en el capítulo 3.4. esta posibilidad es algo bastante probable.

Este segundo método que habíamos utilizado, hacia uso de las distintas redes Wi-Fi que tenemos, que son un total de 289 distintas. Estas redes Wi-Fi las podemos clasificar en oficiales de la universidad, no oficiales (Rogue AP), y externas según si dan cobertura a las redes públicas de la universidad, si son privadas de los departamentos, o si por el contrario pertenecen a particulares de los pisos que rodean la universidad.

Oficiales	No oficiales	Externos	No identificados
13	7	186	83

Imagen 29. Distribución de las distintas redes Wi-Fi

Las principales redes Wi-Fi registradas, las que hemos considerado oficiales y no oficiales son:

Oficiales:

- *Wifi-UC3M*
- *Eduroam*
- *WLIT-WPA*
- *x00*, así como sus variantes.

Así como las que contienen la palabra UC3M como son *UC3M-Events* y *UC3M-Polideportivo* o *eduroam-test*.

No oficiales:

- *TSCWLANG* y su variante *TSCWLANG-5G*
- *WORST*
- *aoliva*
- *riasbaixas*
- *cortijo*

El motivo de registrar éstas como oficiales y no oficiales, además de su dirección MAC, que comentaremos próximamente, es el fabricante de estos puntos de acceso, que a través de su dirección MAC hemos podido comprobar cómo para el caso de las oficiales siempre es *Cisco*, y para el caso de las no oficiales es *Apple* [7].

Por último dentro de las externas hemos considerado todas aquellas que sean de las compañías telefónicas, *Orange*, *Movistar*, *Vodafone*, *Ono*... así como aquellas cuyo nombre este incluido la palabra *MiFibra*, *MIWIFI* o *WLAN*, ya que suelen ser también nombres comunes dentro de las redes caseras procedentes de las compañías normales de comunicación.

En cuanto al fabricante de este último tipo de redes, depende completamente de la marca, variando también dentro de éstas, por lo que no es raro ver dentro de una misma compañía entre 5 y 10 fabricantes distintos, nunca siendo *Apple* o *Cisco*.

Por otro lado, hemos clasificado las 289 redes Wi-Fi distintas, para ver qué cantidad de cada una de estas redes procede de cada punto de acceso de los 794 antes mencionados. Considerando de nuevo las redes que forman cada grupo, tenemos:

Oficiales	No oficiales	Externos	No identificados
465	24	212	93

Imagen 30. Clasificación de los distintos APs por tipo de red

Como podemos ver, principalmente se acumulan en dos de los grupos, en el de los oficiales y en el de los externos. Por un lado, es lógico que haya una gran cantidad de puntos de acceso que den soporte a redes oficiales, ya que al haber estado escaneando en la universidad, lo lógico es que la mayoría sean de este tipo, ya que, además, como cada red tiene una dirección MAC distinta hace que el número de puntos de acceso se multiplique. Por otro lado, al dar soporte a toda la comunidad universitaria, para garantizar que todo el edificio este cubierto, hace falta una gran cantidad de puntos de acceso. Sin embargo, para los no oficiales no ocurre igual, ya que como hemos visto son utilizados para fines concretos y su área de cobertura es solamente la parte que interesa a cada departamento o investigador, no todo el edificio. En el caso de las externas, y podríamos decir de las no identificadas, al ser redes de particulares, lo lógico es pensar que cada una de las redes proviene de un punto de acceso distinto, pues uno de estos da cobertura suficiente para una vivienda, al contrario que lo que nos ocurre con la universidad.

En cuanto a los canales en los que emiten los puntos de acceso de las distintas redes wifi que hemos capturado, tenemos un total de 29 canales distintos en los que nuestros AP escaneados emiten. En las gráficas de abajo tenemos como se encuentran configurados los puntos de acceso en lo que al canal se refiere:

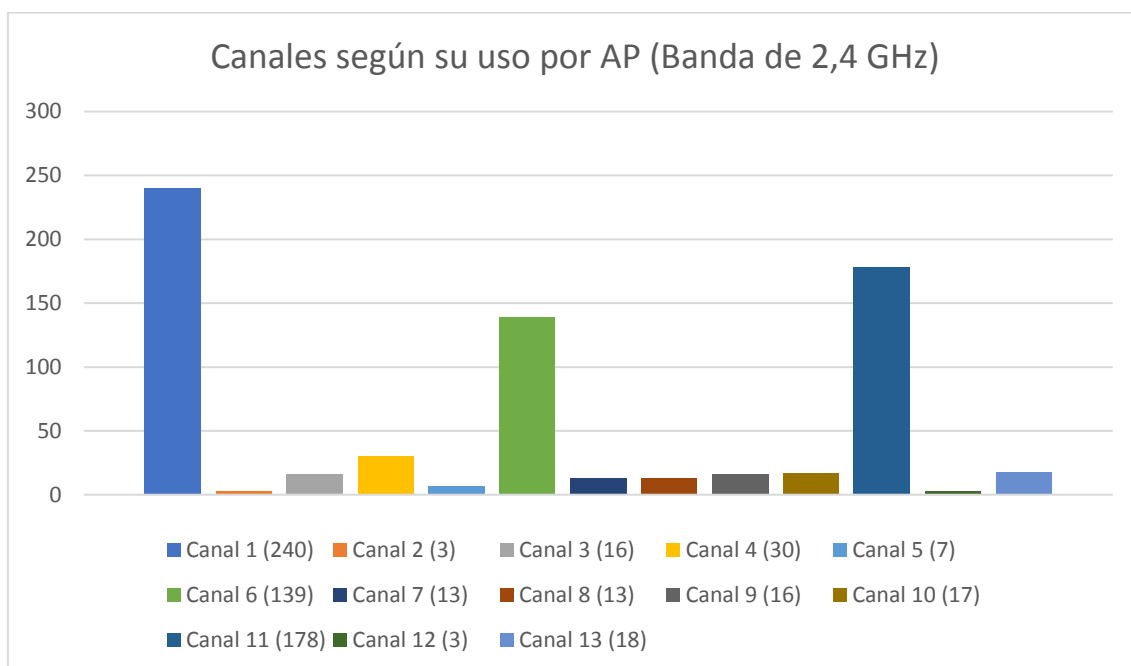


Imagen 31. Uso de los canales por APs (Banda de 2,4 GHzs)

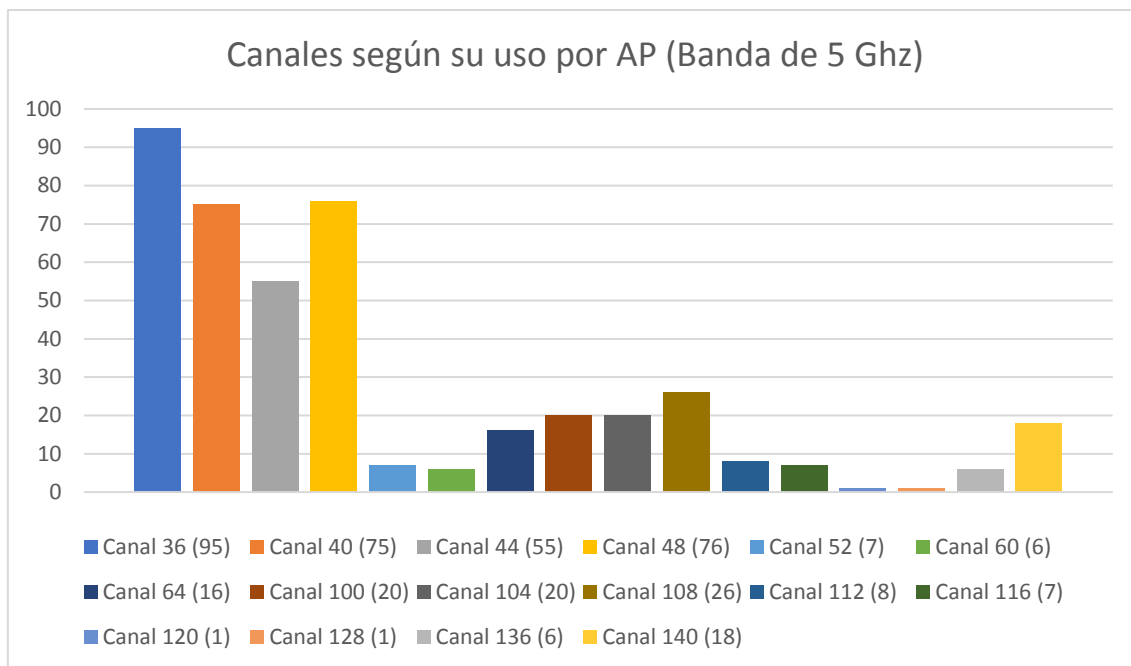


Imagen 32 Uso de los canales por APs (Banda de 5 GHzs)

En las gráficas anteriores podemos observar la distribución de los canales en cada uno de los 794 APs diferentes, tanto para la banda de 2.4 GHzs como para la de 5 GHzs. La banda de 2.4 GHzs está utilizada en toda su extensión. Sin embargo, al observar la *Imagen 31* puede verse como los puntos de acceso están configurados en 3 canales principalmente. Estos canales son utilizados por el 80.4% de los APs, donde puede observarse que el primero es el canal 1, luego el 11 y por último el 6 en cuanto a utilización por parte de los APs.

En cuanto a la gráfica de la banda de 5 GHzs, de ella se pueden sacar varias conclusiones. Por una parte, al contrario de lo que ocurre en la banda de 2.4 GHzs, se observa como la ocupación de canales no está tan distribuida como lo estaba en la otra banda, sino que los canales más ocupados son los 4 primeros. Esto puede deberse a que al tratarse de la banda de 5 GHzs, al tener menor propagación la señal, no sea necesario configurar los canales con tanta separación como hace falta para la banda de 2.4 GHzs, ya que es menos probable que exista interferencia de señales y por ello se configuren de forma que se seleccione el primer canal libre que encuentra.

Es interesante darse cuenta como si se suman la cantidad de ocupación de cada canal la suma es muy superior a los 794 APs que era de esperar. Esto se debe a que algunos APs, concretamente 195, su configuración durante el proceso de captura ha ido variando y tenemos almacenados datos suyos para distintos canales. Esto puede deberse tanto a que la configuración fuera automática para que siempre estuviera en el canal menos saturado, una posibilidad muy utilizada por los APs comerciales, o a que fueran reconfigurados durante dicho proceso.

3.4. Análisis de redes oficiales y no oficiales

Una vez que hemos estudiado los datos almacenados a nivel general, para poder distinguir entre los distintos tipos de redes, los puntos de acceso que tiene cada una, o la cantidad de escaneos que se han realizado y como se han distribuido por plantas, podemos centrarnos en los datos que nos interesan más para nuestro objetivo final, que es la creación del sistema de localización dentro de la universidad. Para ello nos centraremos a partir de ahora solo en las redes Wi-Fi oficiales.

Anteriormente ya mencionamos cuales eran las que formaban este grupo, donde las más importantes por su uso son *Wifi-UC3M* y *eduroam*, ya que son las dos redes a las que cualquiera que pertenezca a la comunidad universitaria puede conectarse. Sin embargo, aunque nunca tienen una coincidencia de dirección MAC entre ellas, se observa como las MACs son similares, solo cambiando el último número o letra, es decir son contiguas. Esto nos hace suponer que provienen del mismo punto de acceso. Esta lógica fue la utilizada para relacionar cada MAC de *eduroam* con su pareja de *Wifi-UC3M* procedentes del mismo punto de acceso para la prueba de verificación de que ambas debían tener unos valores de potencia de señal y calidad similares.

Sin embargo, un punto de acceso no da soporte solo a estas dos redes, sino a alguna más. Prácticamente la totalidad de los puntos de acceso oficiales de la universidad dan soporte al menos a otra red wifi más, en algunos casos a *WLIT-WPA* y en otros a *x00* o su familia de ESSIDs, que son este mismo nombre, pero repetido varias veces, (ej: *x00x00x00x00*).

Además, hay algunos casos en los que el punto de acceso da soporte también a ciertas redes Wi-Fi temporales de la universidad como es *UC3M-Events*. En la imagen inferior podemos observar un ejemplo de un punto de acceso con las distintas redes Wi-Fi que soporta y las diferencias entre las direcciones MAC de cada una, donde además podemos observar como es lógico que todas las redes emiten en el mismo canal.

id	ssid	mac	sig	qual	chan	lugar	idlugarfecha
10825	UC3M-Events	84:3D:C6:46:5B:A4	-46	91	1	379	354
10826	WLIT-WPA	84:3D:C6:46:5B:A2	-46	91	1	379	354
10827	eduroam	84:3D:C6:46:5B:A0	-46	91	1	379	354
10828	WiFi-UC3M	84:3D:C6:46:5B:A1	-46	92	1	379	354

Imagen 33. Redes emitidas por un mismo AP

Como se explicó en el apartado de la metodología de medición, durante el proceso de captura, primero se escanearon los pasillos y aulas del edificio, para posteriormente escanear los despachos, que fue cuando se entendió que el formato de captura necesitaría una mejora para permitir escanear varias veces en un mismo punto y que se pudiera diferenciar cada escaneo. Al realizar las pruebas de esta mejora, se descubrió que había puntos donde las direcciones MAC de las redes oficiales de la universidad habían cambiado y no había ya presencia de las que en ese mismo lugar se habían capturado al principio del escaneo. Este problema aparecía en la planta 1, en el ala *F* del edificio y afectaba a 4 conjuntos de direcciones MAC. La explicación que se encontró a este hecho

fue que se habían sustituido los puntos de acceso, ya que además de cambiar la dirección MAC de todas las redes que soportaba dicho AP, antes la tercera red que había era *x00* y fue sustituida en todos los casos por *WLIT-WPA*.

Por este motivo se demostró la necesidad que había y el buen criterio a la hora de habilitar la posibilidad de escanear más de una vez en cada punto, ya que los puntos de acceso pueden verse modificados como se ha visto, y será útil para poder seleccionar el ultimo escaneo hecho en un punto dado, para así, tener la última información registrada y tratar de garantizar así que estén actualizados al estado actual de la red Wi-Fi.

Dentro de los canales, ahora nos interesan los están configurados en los APs oficiales y no oficiales de la universidad, ya que los puntos de acceso de estas redes se encuentran dentro del edificio y podría darse el caso de que hubiera interferencia entre ellos.

Si restringimos el número de canales a estos grupos de redes Wi-Fi, se observa que es posible saber el número de canales distintos en que emiten los APs de estas redes.

Oficiales	No oficiales
15	12

Imagen 34. Canales usados por cada tipo de red

De estos canales, hay 9 que coinciden entre sí, lo cual podría suponer que los Rogue AP podrían interferir en las redes Wi-Fi oficiales de la universidad.

Anteriormente, vimos la distribución de los canales según los APs para todos los datos registrados. Ahora vamos a hacerlo solo para el caso de los APs oficiales y no oficiales.

Tenemos un total de 18 canales distintos entre los oficiales y no oficiales de la universidad cuyo reparto en puntos de acceso es el siguiente:

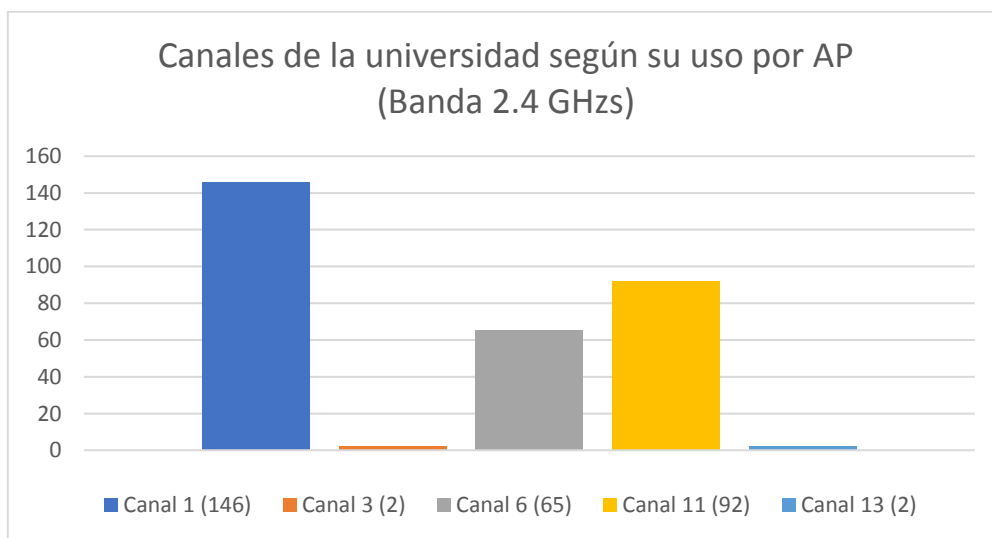


Imagen 35. Uso de los canales por APs de la universidad (Banda de 2,4 GHzs)

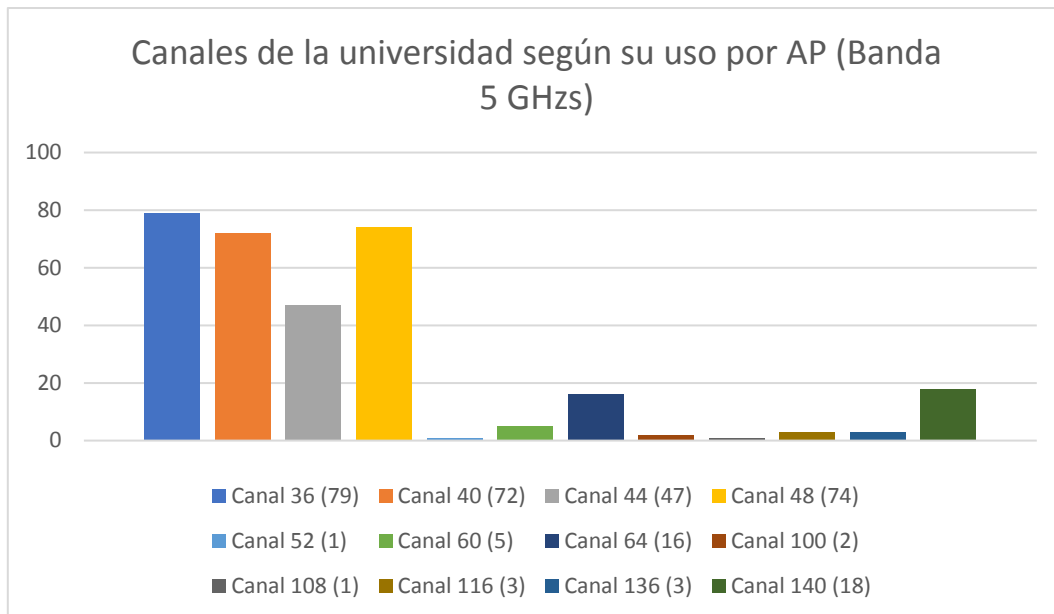


Imagen 36. Uso de los canales por APs (Banda de 5 GHz)

Gracias a las dos gráficas anteriores es posible observar el uso de los canales en los 489 puntos de acceso que dan soporte a las redes oficiales y a las no oficiales de la universidad (Rogue APs). De nuevo es posible observar como si se suman el número de APs que ocupan cada canal el número es mayor al de APs, como se dijo al analizar los canales en todas las redes, el motivo es que o bien hay APs configurados en modo automático o que se han reconfigurado durante el escaneo.

Si se compara estas imágenes con las imágenes *Imagen 31* e *Imagen 32*, que analizaban el uso de los canales, pero en la totalidad de los APs capturados, podemos observar cómo, por un lado, en la gráfica referente a la banda de 2,4 GHzs, los canales más ocupados siguen siendo los mismos. Por otro lado, en la gráfica referente a la banda de 5 GHzs, se observa que es muy parecida a la *Imagen 32*, lo que confirma que los Access Points que ocupan estos canales son, en gran medida, los que dan soporte a la universidad, ya que el número de APs configurados en estos canales es muy similar al de la *Imagen 32*.

Una vez observado el caso para las redes oficiales y no oficiales, ahora vamos a ver un ejemplo aún más concreto. Utilizando nuevamente la aplicación web para mostrar en los mapas de cada planta la información que más nos interesa, hemos representado en esta ocasión el canal en el que está emitiendo el punto de acceso del *eduroam* con la potencia de señal más alta que percibimos en cada lugar, esto es, el punto de acceso al que teóricamente estaremos conectados en cada caso si estamos utilizando la red *eduroam*.



Imagen 37. Canal usado en cada por su mejor eduroam

La imagen anterior, *Imagen 37*, nos muestra el mapa de la planta 1 donde cada punto representa un lugar donde se ha escaneado, y cada color hace referencia a un canal distinto. Como podemos observar hay una gran variedad de colores, y aunque vemos que hay zonas donde hay una mezcla considerable de colores, en gran parte del mapa podemos encontrar zonas donde hay un color predominante, lo que nos demuestra que, en efecto, hay zonas en las que hay un punto de acceso cuya potencia de señal es más diferenciada del resto. Las zonas en las que hay una mezcla de colores puede ocurrir precisamente a que en esta zona haya una convergencia del área de varios puntos de acceso, de forma que la calidad que se recibe es parecida y elegir uno u otro depende del momento de saturación en que se realizó el escaneo.

Para el ejemplo se ha utilizado solo uno de los mapas, aunque el resto de los mapas de color de los canales se encuentran disponibles en la aplicación web para usarse en el momento necesario en el desarrollo del futuro trabajo.

Finalmente nos queda detallar uno de los factores más importantes que hemos analizado, **la calidad**.

Para ello se ha obtenido en cada punto escaneado el valor más alto de la calidad de la señal, y de esa forma poder analizar la potencia de señal recibida por todo el edificio, y poder corroborar el buen despliegue de la red, o si, por el contrario, hay zonas donde la cobertura de la señal Wi-Fi es deficiente. Nuevamente se ha utilizado la señal procedente de la red Wi-Fi *eduroam* por los motivos que ya se explicaron con anterioridad.

Para analizar de forma más concreta la calidad de la señal, vemos la curva de calidad en cada una de las plantas gracias a las siguientes gráficas. En cada una de ellas se representa en el eje de las X los distintos lugares escaneados, y en el eje de las Y la calidad de la señal. El número de lugares está normalizado, de forma que podemos saber el porcentaje de puntos que se encuentran por encima o por debajo de un porcentaje de calidad.

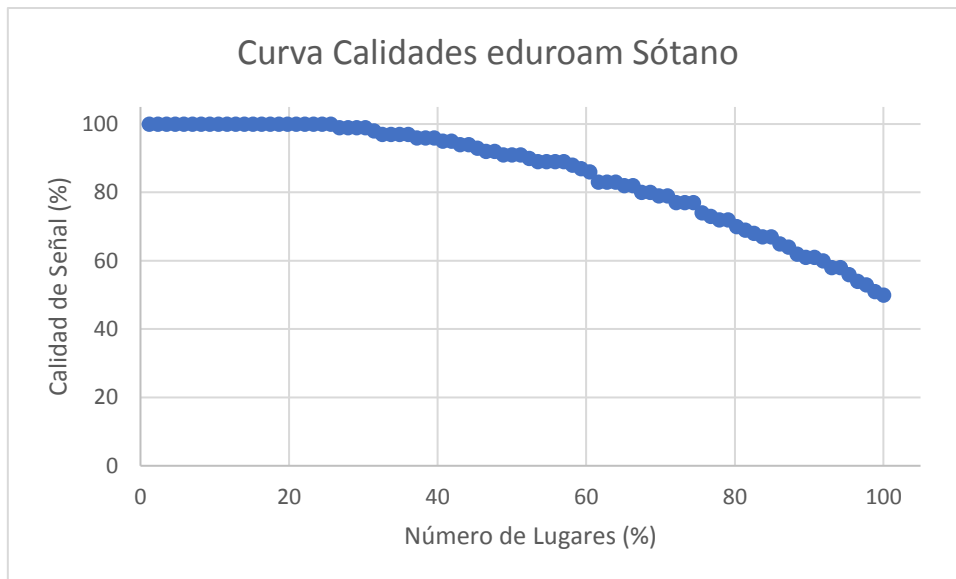


Imagen 38. Curva calidades del sótano

En el sótano tenemos 86 puntos, de éstos el 7% están por debajo de los -70dbm, es decir del 57%-58% de calidad, que es el valor mínimo para poder usar la red Wi-Fi. La gran mayoría están en valores donde la calidad es muy buena, concretamente el 70% de los puntos escaneados, que es de suponer que corresponden a los realizados en pasillos y aulas, mientras que los de calidad regular corresponderán a los realizados dentro de los laboratorios y demás sitios con barreras arquitectónicas.

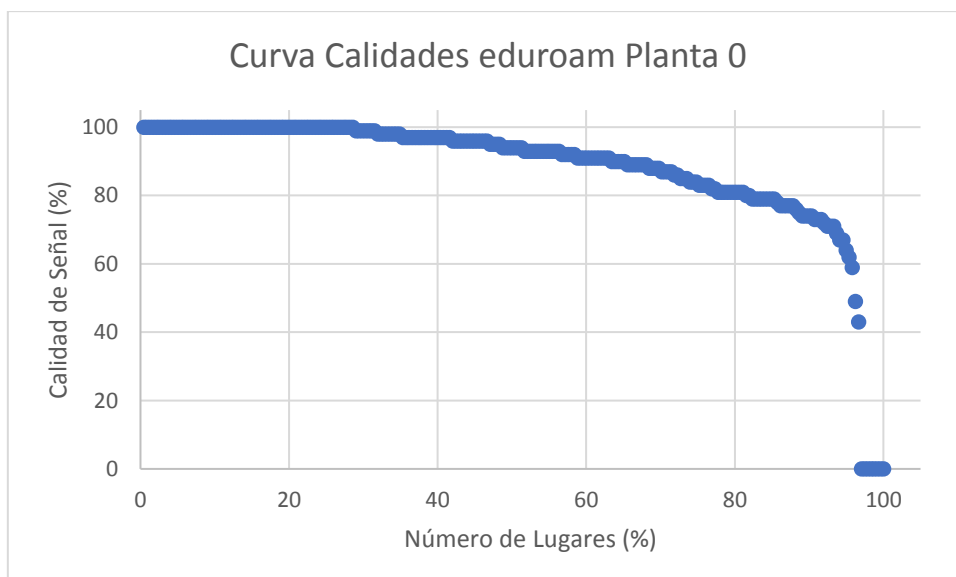


Imagen 39. Curva calidades de la planta 0

En cuanto a la planta 0, podemos observar un total de 238 puntos, de los cuales 8 son nulos debido a que no hay datos. Del resto de puntos vemos como un 82% de los puntos

superan la barrera del 80% de calidad, garantizando una conexión muy buena, tan solo una pequeña cantidad de lugares tiene una calidad regular, y solo dos tienen una calidad mala, que serán analizados más adelante.

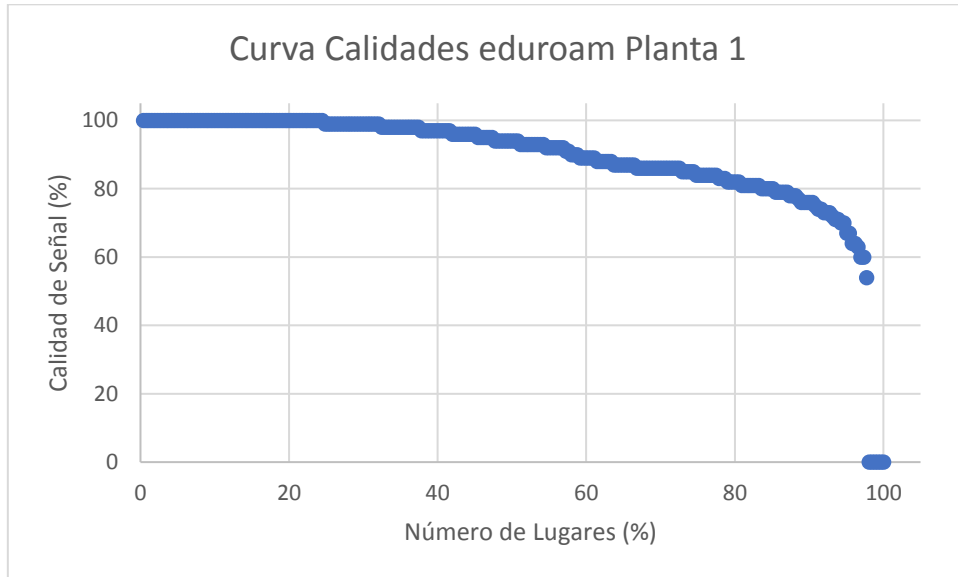


Imagen 40. Curva calidades de la planta 1

Como vemos en la *Imagen 40*, correspondiente a la planta 1, tenemos un caso de nuevo muy similar al de la planta 0, donde de los 262 puntos escaneados, el 85% superan el 80% de calidad. Tan solo un punto tiene una calidad inferior a 57%, y tenemos un total de 6 puntos sin datos.

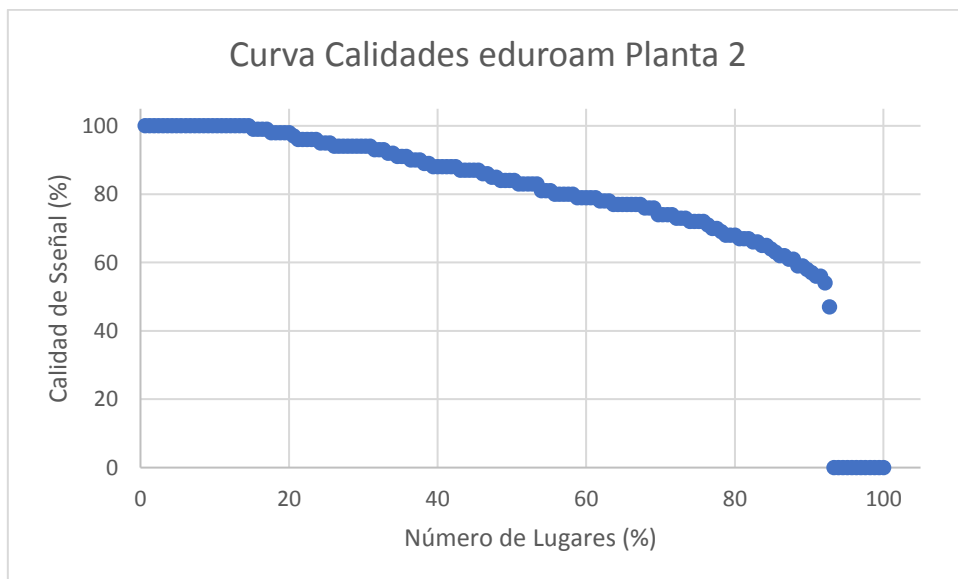


Imagen 41. Curva calidades de la planta 2

En la planta 2 hay 165 escaneos, como vemos, en esta ocasión la gráfica desciende de forma más pronunciada, y solo el 60% de los puntos supera el 80% de calidad. En esta planta es donde más puntos de calidad regular hay. También es, junto con la planta 3, donde hay un mayor porcentaje de puntos sin datos, en este caso, el 7%. Finalmente hay un total de 6 puntos donde la calidad es mala.

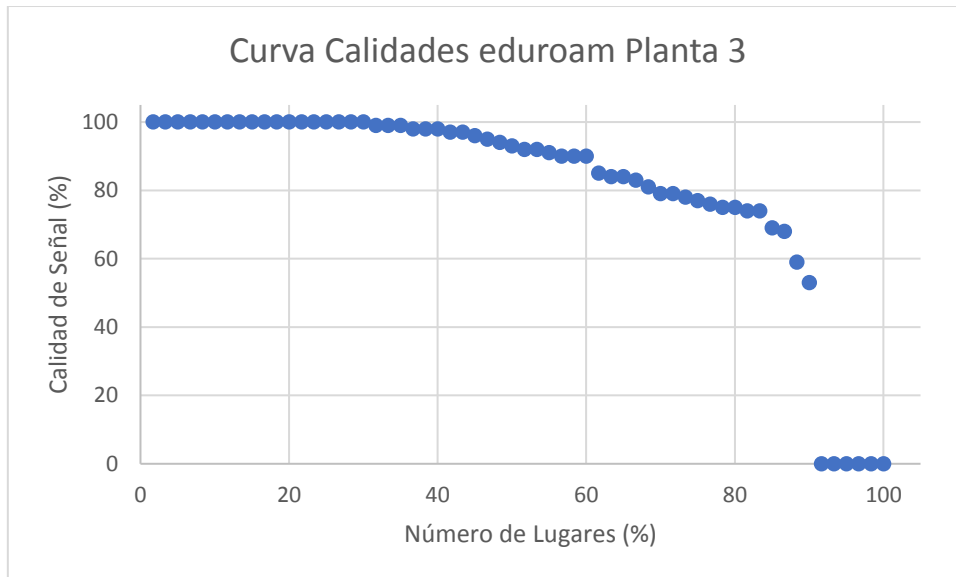


Imagen 42. Curva calidades de la planta 3

Por último, en la planta 3, donde tenemos un total de 60 puntos, 6 de ellos sin datos, ya que no se ha realizado escaneo, tan solo uno está por debajo del valor mínimo para poder usar la red Wi-Fi. El 70% de los puntos superan el 80% de calidad.

Tras analizar la curva de calidad en cada una de las plantas vemos como en la mayoría de los lugares de la universidad la calidad de la señal de *eduroam* es muy buena, existiendo también puntos en los que la calidad es algo peor pero lo suficiente como para poder utilizar la red. Estos puntos suponemos que se encuentran en las zonas más alejadas a los puntos de acceso donde puedan existir barreras arquitectónicas. Finalmente, la existencia de puntos donde la calidad es mala, impidiendo usar la red Wi-Fi es preocupante. Por ello, utilizando la herramienta para representar los puntos en el mapa, vamos a estudiar estos lugares. Así podremos localizarlos y determinar si hay algún motivo lógico para este problema o si por el contrario puede deberse a un error a la hora de desplegar la red Wi-Fi por el edificio.

Como vemos en la imagen inferior, los puntos con mala calidad se encuentran principalmente en lugares donde la red Wi-Fi no llega de forma directa, es decir, dentro de ciertos despachos, escaleras, baños o laboratorios. Esto es algo bastante lógico, ya que dependiendo donde se encuentren los puntos de acceso cercanos, al tratarse de zonas donde existen barreras arquitectónicas en los alrededores, provoca que la señal recibida sea débil, de forma que la calidad es baja. Es de suponer que, en otros lugares similares,

aunque el nivel de señal sea más elevado, tendremos problemas para obtener una señal de una gran calidad, aunque sea mejor que la que encontramos en los lugares marcados por esta imagen. Estos lugares se corresponderán con los puntos donde la calidad era regular y que habíamos visto en las gráficas correspondientes a las curvas de calidades anteriormente analizadas.



Imagen 43. Puntos con calidad deficiente

4. PLANIFICACIÓN Y PRESUPUESTO

4.1. Planificación

La realización del trabajo se ha dividido en distintas tareas acordes a los objetivos que había que ir cumpliendo para el correcto desarrollo del proyecto y para poder continuar con las siguientes tareas.

Podemos dividirlo de la siguiente manera:

1. Búsqueda y lectura de información referente al proyecto

1.1 Búsqueda de artículos, papers y noticias sobre las técnicas de escaneo

Para arrancar el proyecto lo primero que debemos hacer es conocer el tema, para saber las diferentes técnicas de escaneo que hay, junto a las diferencias entre una y otra, para poder elegir la que más nos interesa para nuestro proyecto. De la misma manera nos sirve para entender mejor para qué se suelen realizar, así como proyectos similares al nuestro que se hayan podido realizar en el pasado.

1.2 Lectura y análisis de la información recopilada

El material que hemos encontrado, y que se nos ha suministrado es analizado y organizado según el uso que le vamos a dar, o cuándo se lo vamos a dar, ya sea como información general sobre el tema, información para el proceso de captura o información para analizar posteriormente los datos escaneados.

2. Creación de los Scripts para la automatización del escaneo

2.1 Búsqueda de comandos para el escaneo

Una vez que conocemos qué es lo que necesitamos a la hora de realizar los escaneos, debemos buscar los comandos y la forma de obtener esta información de la manera más fácil y completa posible. Para ello realizamos distintas búsquedas tanto en páginas web como en foros relacionados con nuestro interés.

2.2 Escritura de los Scripts

En esta parte, teniendo ya los útiles necesarios para escanear las redes que hay a nuestro alrededor, comenzar a crear el conjunto de scripts que nos permitirán capturar esta información, filtrarla e introducirla en la base de datos, previa comparación con otros dos escaneos más.

2.3 Prueba y mejora de los Scripts

Como es lógico, el resultado esperado no se conseguirá en el primer intento, y una vez que tengamos ya una serie de scripts que nos permitan comenzar a probar la funcionalidad de éstos, veremos si realizan las funciones esperadas, y si no, iremos viendo, modificando y mejorándolos hasta que realicen exactamente lo que queremos.

3. Creación de la Base de Datos

3.1 Diseño de la Base de Datos

Como método de almacenamiento debemos pensar en qué programa vamos a usar y a continuación como va a ser el diseño de tablas que más nos interesa según la información que va a haber en ella, estática, más dinámica...

3.2 Creación del sistema de tablas

Una vez que hemos decidido las tablas que va a tener nuestra base de datos, así como los campos que va a haber en cada una de ellas, procedemos a configurar nuestra base de datos y nuestras tablas, de forma que estén ya operativas para el proceso de escaneo.

4. Escaneo de los Puntos de Acceso

4.1 Escaneo en lugares de fácil acceso

Para comenzar el proceso de captura de datos, empezamos por los lugares en los que escanear va a ser un proceso fácil y vamos a poder escanear muchos puntos de seguido, estos lugares son los pasillos y aulas.

4.2 Depuración y mejoras del proceso de captura de datos

Una vez que hemos escaneado en diversos puntos de la universidad, podemos hacer un primer estudio sobre si la manera en que estamos realizando los escaneos, o si los scripts están funcionando de la forma que queremos. Además, dado que aún nos encontramos en una fase temprana de este proceso, podemos realizar algún cambio o mejora sin que resulte peligroso de cara a perder muchos datos.

4.3 Escaneo en lugares de difícil acceso

Una vez finalizado el escaneo en las zonas más accesibles, se procede a escanear en los lugares donde no es tan sencillo. Este proceso se prolonga más tiempo pues es necesario que haya gente en esos lugares a escanear, y no es tan fluido y continuo como los escaneos del principio del proceso.

5. Creación de la aplicación web

5.1 Estudio de sistemas para visualizar los datos almacenados

Queremos diseñar una aplicación web que nos permita visualizar la información referente a los puntos de acceso almacenados en cada lugar escaneado. Para ello tenemos que pensar qué tipo de aplicación necesitamos, así como la tecnología que vamos a usar en la página para conseguirlo.

5.2 Desarrollo del sistema de visualización

Una vez que hemos decidido la tecnología y las librerías y programas que vamos a utilizar procedemos a instalar el software necesario, así como la creación de las carpetas, y código necesario para su funcionamiento.

5.3 Creación de las imágenes con datos geoespaciales

Durante la tarea 5.1 se estudió también no solo el formato de la aplicación web, sino también como conseguir usar un mapa a partir de una imagen. Una vez encontrada la manera, se instala el software necesario, y se crean y configuran las imágenes para dotarlas de características geoespaciales. Una vez conseguido solo hay que importarlas en nuestra página web desde el servidor.

5.4 Pruebas de la Aplicación Web

Realizamos las pruebas necesarias para garantizar que todo funciona de manera adecuada y modificamos lo necesario para conseguir que la aplicación web se visualice de la manera que queremos.

6. Análisis de datos

6.1 Creación de Scripts para el análisis

Para analizar los datos más relevantes y poder extraerlos, creamos una serie de scripts que acceden a la base de datos filtran los resultados para obtener cada uno de los resultados que nos interesa saber para ver si el proyecto ha sido realizado correctamente.

6.2 Uso de la aplicación web para el análisis

De la misma manera que ciertos datos y estadísticas las obtenemos a través de scripts, otros los representamos en los mapas que tenemos en nuestra aplicación web de forma que podemos observar de una manera fácil y rápida los datos que hemos recabado en los escaneos.

7. Elaboración de la memoria

Gracias a los datos, estadísticas y a las representaciones en los mapas de la aplicación web ya tenemos lo suficiente para poder redactar la memoria, último paso para finalizar el proyecto.

Nombre de la tarea	Inicio	Fecha Final	Duración (días)
Búsqueda y lectura de información	08/01/2018	18/01/2018	10
Búsqueda de papers, artículos y noticias sobre las técnicas	08/01/2018	13/01/2018	5
Lectura y análisis de la la información recopilada	15/01/2018	18/01/2018	3
Creación de Scripts para la automatización del escaneo	18/01/2018	28/01/2018	10
Búsqueda de comandos para el escaneo	18/01/2018	20/01/2018	2
Escritura de los Scripts	20/01/2018	26/01/2018	6
Prueba y mejora de los Scripts	23/01/2018	28/01/2018	5
Creacion de la Base de Datos	24/01/2018	28/01/2018	4
Diseño de la Base de Datos	24/01/2018	26/01/2018	2
Creacion del sistema de tablas	26/01/2018	28/01/2018	2
Escaneo de los Puntos de Acceso	29/01/2018	01/06/2018	123
Escaneo en lugares de fácil acceso	29/01/2018	01/04/2018	62
Depuración y mejoras del sistema de captura de datos	24/02/2018	12/03/2018	16
Escaneo en lugares de difícil acceso	01/04/2018	01/06/2018	61
Creacion de la Aplicación Web	01/05/2018	15/06/2018	45
Estudio de sistemas para visualizar los datos almacenados	01/05/2018	01/06/2018	31
Desarrollo del sistema de visualizacion	08/05/2018	08/06/2018	31
Creacion de las imágenes con datos geoespaciales	24/05/2018	05/06/2018	12
Pruebas de la Aplicación Web	29/05/2018	15/06/2018	17
Análisis de los Datos	16/06/2018	05/09/2018	81
Creación de Scripts para el análisis	20/07/2018	05/09/2018	47
Uso de la aplicación web para el análisis	16/06/2018	12/08/2018	57
Elaboración de la memoria	24/08/2018	14/09/2018	21

Imagen 44. Desglose de tareas

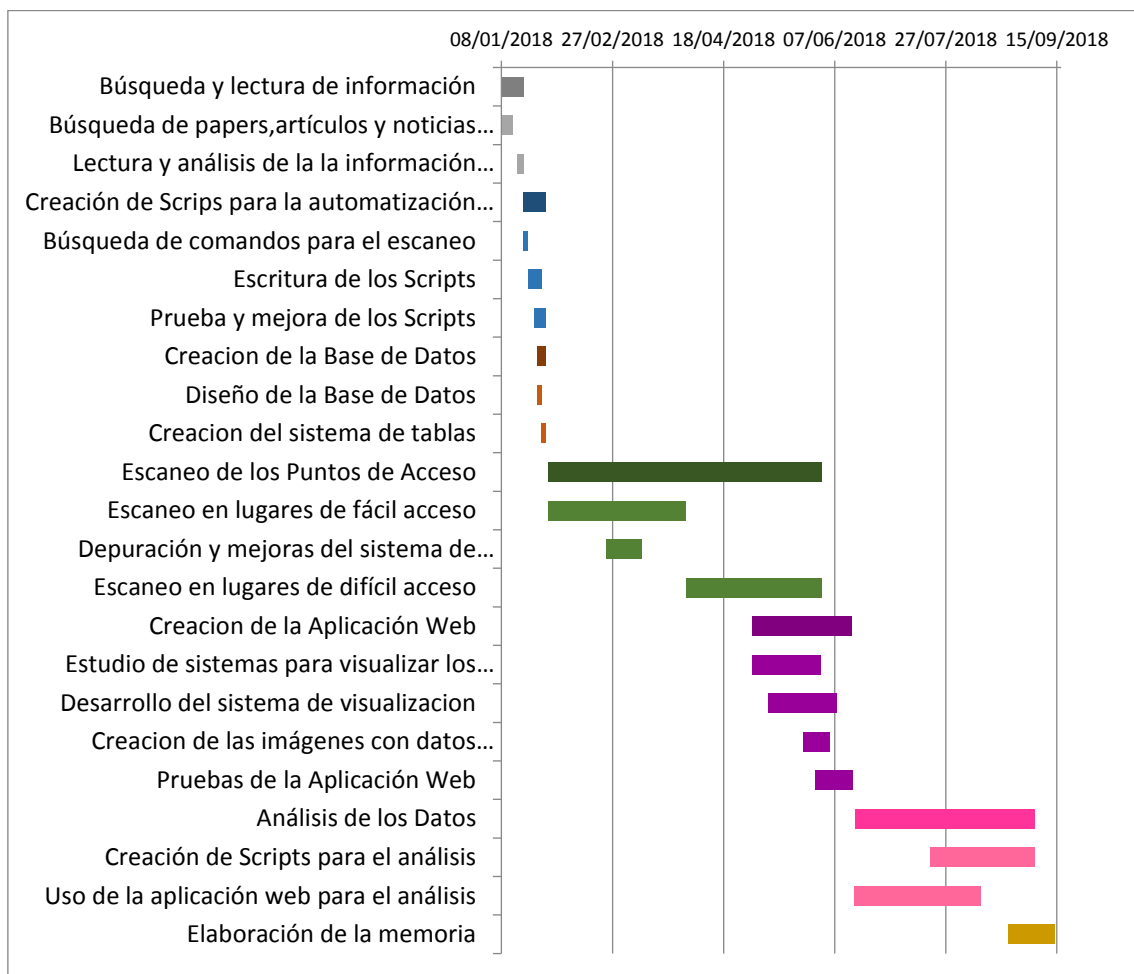


Imagen 45. Diagrama de Gantt

En la *Imagen 44* podemos observar el desglose de tareas con sus fechas de inicio y fin, mientras que en la *Imagen 45* tenemos el diagrama de Gantt de la planificación del proyecto.

4.2. Presupuesto

Nuestro presupuesto lo vamos a dividir en 3 distintas categorías:

Recursos Materiales:

- Lo único que necesitamos es un ordenador portátil para poder realizar los distintos escaneos por el edificio.

Recursos Personales:

- Ingeniero Telemático Junior, en este caso, el autor del TFG, que será el encargado de llevar la principal carga de trabajo del proyecto.
- Jefe de Proyecto, será el encargado de supervisar el correcto desarrollo del trabajo, así como proporcionar apoyo y solventar las dudas que se produzcan.

Otros Recursos:

- Conexión a internet, necesaria para poder realizar las búsquedas necesarias en las distintas tareas del proyecto, así como obtener localizaciones, y otros datos necesarios.
- Electricidad, tendremos un cierto gasto en luz para cargar el ordenador que debemos tener en cuenta.

PRESUPUESTO			
MANO DE OBRA			
Personal	Coste/mes	Meses	Total
Ing. Telemático Junior	1.750 €	9	15.750 €
Jefe de Proyecto	2.916,70 €	9	26.250 €
			42.000 €
MATERIALES			
Equipo	Coste	Cantidad	Total
Ordenador Portatil	700 €	1	700 €
			700 €
OTROS GASTOS			
Gasto	Coste/mes	Meses	Total
Luz	2,79	9	25,11
Conexión a Internet	35,84	9	322,56
			347,67
			43.048 €

Imagen 46. Presupuesto

*Como fuente se ha buscado en distintas páginas de internet para obtener los salarios [8] y el coste de la luz [9].

5. CONCLUSIONES Y TRABAJO FUTURO

5.1. Conclusiones principales

El objetivo principal de este trabajo era el de servir de primer paso y de base para la realización en un futuro de un sistema de localización en interiores basado en Wi-Fi. Este primer paso era básicamente lograr una gran base de datos con información sobre los distintos puntos de acceso de la red Wi-Fi de la universidad en los distintos puntos del edificio 4 de la universidad.

Para ello, este objetivo se dividió en dos tareas principales. Por un lado, la obtención de la información de cada lugar del edificio a través de escaneos pasivos de las redes Wi-Fi a nuestro alcance en cada uno de ellos. Por otro, el análisis de esta información para garantizar que ésta fuera válida para trabajos futuros.

Para el escaneo de la información se crearon una serie de scripts que nos permitieran sistematizar todo el proceso, para así obtener de forma rápida toda la información que nos interesa, realizando además varios escaneos para obtener la mayor consistencia posible de los datos.

Una vez obtenida la información que nos interesaba y realizados los escaneos extra, la información se introduce dentro de la base de datos, donde tenemos almacenada la información relativa al punto en el que estamos escaneando, y la fecha en que se realizó.

Una vez avanzado todo el proceso de captura de datos, se procedió a analizarlos. Para ello antes se creó una aplicación web, en ella se alojan mapas de las distintas plantas y es posible representar la información que más nos interese utilizando la información relativa al lugar en el que se escaneó.

Esta herramienta nos permite mediante distintas consultas y representando la información que nos interesa en los mapas, comprobar que la información que hemos obtenido en la parte de escaneo es consistente y fiable.

Una vez comprobado mediante estas pruebas de validación que los datos cumplen esta condición, pasamos a analizar la información ya considerando que ésta es válida. Con la aplicación web podemos observar cosas tales como el canal en el que emite la *eduroam* con más potencia, los puntos donde la calidad es mala...

Complementando este análisis visual con el realizado a través de gráficas y demás observamos que la información que tenemos nos hace pensar que sí se puede utilizar para el desarrollo del sistema que queremos. Esto es así ya que vemos que realmente se puede identificar mediante los datos almacenados la ubicación aproximada de los puntos de acceso de la universidad, vemos que los canales utilizados están lo suficientemente bien configurados para evitar posibles interferencias entre la señal de unos puntos de acceso u otros. Además, nos permite ver que el edificio se encuentra cubierto en una gran totalidad de su superficie por la red wifi con una gran calidad...

Por todo ello vemos que hemos cumplido los objetivos que nos marcamos al principio del trabajo, lo cual nos permite afirmar que la primera parte del sistema, desarrollada por este TFG se ha realizado de forma satisfactoria.

5.2. Trabajos futuros

El desarrollo de este TFG, con un resultado positivo nos permite partir de él en el futuro para continuar con el sistema de localización que queremos realizar.

Un buen punto de partida es tratar de crear un algoritmo que nos permita predecir la posición en que nos encontramos según lo que estamos viendo en ese momento, desarrollando un programa que escanee los puntos que vemos a nuestro alrededor, y comparándolo estos y sus valores con la base de datos que hemos creado en este trabajo, que nos indique el lugar aproximado en el que se ha realizado este escaneo.

Si funcionara el algoritmo, el siguiente paso podría ser el de predecir la ruta que ha podido seguir el usuario a través de la información relativa a los puntos de acceso a los que estaba conectado en cada momento.

De la misma manera sería realmente interesante desarrollar un nuevo sistema para escanear datos de forma automática e introducirlos en la base de datos, para tener estos siempre lo más actualizados posible, ya que como hemos visto durante el análisis de los datos, estos pueden quedar obsoletos por la sustitución por averías o cambios de configuración de los APs.

6. BIBLIOGRAFÍA

- [1] Deloitte, «Tendencias globales de consumidores de dispositivos móviles,» [En línea]. Available: <https://www2.deloitte.com/do/es/pages/technology-media-and-telecommunications/articles/pr-global-mobile-consumer-trends.html>.
- [2] Ditrendia, «Informe Ditrendia: Mobile España y en el Mundo 2018,» 2018.
- [3] Data Center Market, «España, el segundo país europeo con mayor conectividad a redes Wifi,» [En línea]. Available: <http://www.redestelecom.es/comunicaciones/noticias/1096012000303/espana-segundo-pais-europeo-con-mayor-conectividad-a-redes-wi-fi.1.html>.
- [4] Austrian Institute of Technology: Yuji Yoshimura, Stanislav Sobolevsky, «Art traffic at the Louvre,» [En línea]. Available: <http://senseable.mit.edu/louvre/>.
- [5] España, *Real Decreto-Ley 5/2018 por la Protección de datos de carácter personal*.
- [6] NetSpot, «Qué es RSSI y su relación con una red Wi-Fi,» [En línea]. Available: <https://www.netspotapp.com/es/what-is-rssi-level.html>.
- [7] HWaddress.com, «HWaddress: MAC address and OUI lookup,» [En línea]. Available: <https://hwaddress.com/>.
- [8] JobTonic.es, «Buscar trabajo en todas las páginas web de empleo,» [En línea]. Available: <http://www.jobtonic.es/>.
- [9] E. Energética, «Cuánto consume un ordenador,» [En línea]. Available: <http://www.etiquetaenergetica.com/cuanto-consume-ordenador/>.

ANEXO A. RESUMEN EN INGLÉS

This essay, *Captura y análisis de datos para un sistema de localización en interiores basado en Wi-Fi (Capture and data analysis for a Wi-Fi-based localization indoor system)*, is the first part of a bigger project. The title explains itself, the ultimate goal is to develop a localization system based on Wi-Fi. Nevertheless, for that purpose it is necessary to start doing a research about the Wi-Fi network of the university: how it is distributed, its floors, access points...

Everyday more people use different kinds of media devices, which are very integrated in our lives, and more everyday. Reading books, listening to music, watching the news, doing banking operations, keeping in touch with friends or family members are some of the possibilities that these technologies offer to us and that make them be in a continuous growth. However, all of them have something in common: the use of internet connection is key for them. Thus, Wi-Fi networks —private or public— are increasing and they are used every day by more and more people.

In this context, our idea of the project emerges. These Wi-Fi networks have a wide use, and this fact presents the possibility of obtaining a certain use of them, apart from the evident, which is providing wireless support to internet connections.

The main goal set in our project is to create a base which will be used in the future to continue with the develop of the localization system. In order to do that, in one hand, we create the database full of information related to the access points which we will find in different locations of the university and, in the other hand, we analyze and validate all this information to guarantee that it will be useful to continue with this project in the future.

Our research has been made in the Building 4 of the university, since the Department which frames the project sets there. This building has 5 floors, with different shapes and sizes each one. For the process of data collection, the location was divided in cells of 16 square meters to scan them, so the scan would be uniform and with the same gap between each of them. Then, all of the cells in which we were collecting information were numbered, so this way, each scan and the data found could be identified with its corresponding physical location.

On the other side, a set of scripts which automatizes the whole process of data collection is created, due to the number of scans we were performing. Firstly, we scan the access points that we can see in the place where we are, and then, we extract from each of them the data which are relevant to us, that is, the essid name, MAC direction, channel in which it can be found, signal strength and its quality. We do this 3 times in order to make the snapshot more consistent and to avoid eventual network card overloads or the loss of information due to a bad synchronization when the access points are sending its

announcement message. Hereafter, these 3 scans are compared and the information of the different access points in the database are inserted in them. This makes each of the scans of each point to be made in about two minutes, saving a great amount of time and ensuring that the captured information is the one inserted in the database, free of the errors that may occur if it was manually done.

As said, the data obtained during the capture are inserted in the database we have created as storing method. This database is formulated by a complex of interconnected tables that allow us linking each scan with the location where it was made, the information obtained and the date when it was made.

In total we have 6 tables, 3 of them used in the capture process for the 3 scans made in each data capture. These are the *Prueba (Test)* tables, auxiliary tables which are deleted once all the data of that point has been inserted into the main table, which contains information about all the scans. This table is called *EscaneoTotal (TotalScan)* and, as said, it contains the information of all the access points which we have been obtaining on each scan, linking this information with the location, date and time in which it was obtained. The information of the scanning point can be found on the table *Lugares (Locations)*, which identifies a numeric point with its real geospatial localization. This information has been collected manually via Google Maps, which allows us to access to the university map of every floor, and there we can mark the location obtaining its latitude and longitude. Finally, we find the table *LugarFechaTotal (LocationDateTotal)*, which identifies the location with the date and hour in which it was scanned. Furthermore, it allows us to do different scans in the same place and identify them in a way that we can select the most updated information of each point in the future.

The chosen storing system is MySQL, since it is a well-known and wide-used server. Its easy use it through the command screen as well as accessing to it through scripts or web apps. It makes this system perfect to our work.

Once all the scripts for the scan have been written, maps are divided in cells, the locations to scan have been numbered, the tables have been created and everything has been checked to work the way we desire, we can start the capture process.

It will be divided in two kind of places to scan. In the one hand, we have the halls and classrooms of easy access, where the data capture is practically continuous, since as we finish with one point, we have to move few meters to start with the next one. Moreover, the access to them is quick and easy, since they are public spaces. On the other hand, we have the offices and research laboratories, which are private, we need permission to access them and they are isolated one from the other. We also find the problem that, sometimes, the offices are closed or occupied, and here the data capture is not that mechanical and quick.

After the data capture process, we have the second main task to the development of this final project, the analysis of the information. This analysis is divided in two parts. The

first is based in the validation of the data we have been inserting in our database from the scans and the other consists in analyzing these data in order to obtain relevant information about the quality of the Wi-Fi network, its channels and possible interferences...

However, before starting the data analysis, we have created a tool which will be useful as a support of that analysis. Thanks to a web application based on servlets and JSPs and in the model view controller, we can display maps of the different floors of the building, where the information which is interesting to us depending on the scanning location can be represented. For that, we only need to import the images of the floors which we have previously geospatially located in our app, in a way that we can convert the latitude and longitude of each point in the place referred in the map. This is made in our server from *geoserver*, which we have previously configured and transformed.

The validation tests, as we have said, try to confirm the data obtained are as expected. We have made 3 tests.

Since it indicates that the different fields of the scans can make us deduce the proximity or remoteness with the access point, the most evident test is the approximate reaching of the location of the access point through the information obtained about that AP in the different places where we have detected it. Through the signal power and quality, we can sense that depending on how high these values are, the closer we are to the origin point of the signal, although it is evident that the architectural barriers affect this signal. Thanks to the developed web app we can represent the value of quality of a specific access point in the maps of each floor, in a way that, depending on this value, the point will have a color or the other. This visualization allows us in a quick way to check the place in which the AP is found and to verify that it corresponds to the indicated location in the access point list we have received to confirm our work.

Another test we have made to check that the theoretically expected is the same as the practically obtained is the following: knowing that access points support more than one Wi-Fi network, we understand that these Wi-Fi networks should have the same value of signal power, which means that, identifying these networks which come of the same access point should be practically the same in each place where they have been scanned. Obviously, the practice confirms theory and, through an analysis using a graphic in which each axis represents one of the networks, we can observe that the whole of the point couples create a straight line, which means the value of booth axis is practically the same. The networks we have represented in the graphic are *eduroam* and *Wifi-UC3M*.

Finally, the last check we do is, using again our tool of map view, checking that certain non-official networks of the university or the neighbors APs can be found in the expected area. As we will see in the analysis of the results of the database, there are different kinds of Wi-Fi networks. We have the official ones, which give support to the university networks and on the opposite, we have other networks, which belong to the departments and to which not everybody has access, they are called Rogue APs. These networks even so can be found inside the building and we expect them in specific areas or floors. Using

this web app we confirm that they appear in the expected areas. We also have the private networks, which can be found in the university but carry the names of typical telecommunication companies and thus, come from the adjacent buildings. Henceforth, we can expect that we will see them from the building limits and nearby areas. We confirm this again thanks to our app.

Now that we know that the data inserted during the capture process which is stored in our database are reliable and match with the expected, we start analyzing them taking the more general data as basis to focus gradually in those points which are more interesting to our project.

The first data we can observe is the number of scans and how they are distributed in the different floors of the building. In addition, it is interesting to see how much they are in different points. In other words, how many repeated scans we have, which are only 6 since the scan process was specifically thought to cover the whole building in order to have data from every point and to serve as initial point for further projects. The update of the data will remain open for future projects.

Another data easy to observe and comment is the number of access points that we find on these scans. On the one side we have the totality of points and how they are distributed in the different floors and, on the other side, the different APs we find in every scan. From these access points we can identify the different networks to which they belong and how many of them are visible from each floor.

As we said, we have found out the number of Wi-Fi networks in our database, which is useful to do the first interesting classification. We find 4 different kinds of Wi-Fi networks. Firstly, the official ones, which give support to the university from a same access point, as *eduroam* and *Wifi-UC3M*, and we observe that from that access point there is usually another extra Wi-Fi network, which is an official network of the university although its use is unknown. These two extra networks are *WLIT-WPA* and *x00*.

We also have the networks of the university departments, but they are not available to every member of the university. The main ones are *WORST* y *TSCWLANG*, which belong to the Telematic and the Signal Theory Department, respectively.

Next, we have the neighbor's networks, those with names related to telecommunication companies as *Movistar*, *Vodafone*, *Jazztel*... in addition to those we have others with names that do not refer to them directly but carry names like *WLAN*, *MIFIBRA*, *Miwifi*...

Finally, we have those networks which we could not identify and belong to private users or are non-official university networks. They can even belong to students or teachers who, using their own device and network connection, have enabled their device as a Wi-Fi router in the moment of the analysis capture.

Once we know which one belongs to each group is easier to do new searches and filters in our database to obtain, as we have said, how many networks of each kind do we have or how many access points provide each kind of Wi-Fi network.

It is important to note that at this point of the analysis, new scripts have been created, which filter results and analyze them accessing to the database in order to obtain quick statistics.

Another interesting point which shows which Wi-Fi networks and thus which access points belong to each group is studying the company that have made each AP through the MAC address. This way we have seen how the whole of the official networks are Cisco; those which are not official are Apple and lastly, the endpoints belong to a huge variety of companies.

In the end, another interesting data to study is the total amount of discovered channels, as well as their distribution depending on the access points. This way we can see that the most used channels, as well as the gap between channels, which allows us to observe the probability of an interference between channels.

About the most specific analysis, now we pay more attention to the information which come from the university building, that is, those who come from the official and non-official networks of the university.

As we have said before, the official access points of the university support more than one Wi-Fi network. When the table and data capture system was improved increasing the scan capability in a same access point several times and distinguishing each scan, it was discovered that some access point have changed. Specifically 4 different ones and, with them, the MAC addresses of those 4. This is useful to confirm the need in future projects to create a system which allows us to update the information on each location in a quick way and automatically, to avoid the database information to become obsolete.

Talking about the channels, we can see the amount of different channels depending on whether they are official or non-official networks of the university, and how they are distributed in the different access points which support them. Again, it allows us to see its settings and the number of APs configured in each channel. We see how similar are these informations with the one we have seen for the total information of the networks.

On the other side, our main interest is focused in the *eduroam* network, since this is more restricted than *Wifi-UC3M*, to which everybody can get connected without belonging to the university community. Thanks to the web app we can represent in the maps of each floor which AP uses *eduroam* with the best signal on each point, that most of the times, it will be the one we are connected in that points. This helps us to see the distribution of channels on each area, as well as the areas where there are more quantity of channels because they are farther from an access point.

The last analysis we have made is about the quality of the signal we find in the *eduroam* network in this building. Thanks again to the different scripts created to make our work easier, we can filter the information and obtain the quality of the signal in the *eduroam* access point with best signal strength that we find in each place. This way we can see whether all the building is covered by a good quality signal, whether some areas should be improved and the reason of this signal. Indeed, we can see that, thanks to the quality

graphics of each floor, that the biggest part of points are found with a signal above the 80%. Nevertheless, approximately 150 points are under that number, some of them being even under a minimum level required to use internet.

We can see that these points, thanks to our tool to display, are found in areas where maybe the signal is weaker because of architectural problems like stairs or inside of laboratories.

Finally, as we have seen in the development of this project, the two main tasks we wanted to fulfill, have been done successfully. We could scan in practically the whole places we set in our goal, getting to reach information about them and reaching a complete and full database. Once this task is fulfilled and having a consistent database, we proceed with the analysis of its information, verifying that it is reliable and according to what we expected, obtaining different statistics and data interesting to continue with this project.

This way, fulfilling the two tasks, we have completed the first step to start designing this localization system. Referring to this project and essay, it has served as an initial point to the development of a system, and some colleagues will continue with it in the future. Thanks to this essay we know which path this project should take.

It is evident that, the procedure of capture must be refined, in a way that regular scans are always made automatically and in a more efficient way to allow the database to be updated, avoiding the information to become obsolete and useless. In addition, knowing that the data allow us to establish the area in which we should be according to the different signal strength, with which we see the different access points. If we create an algorithm which compares what we see from a point with what we have stored in our database, we could guess the point from where the scan was made and guess the area.