



Grado en tecnologías de las telecomunicaciones

Curso Académico 2015/2016

Trabajo Fin de Grado

Análisis de proyectos FOSS

Autor : Joel Peralta Rodríguez Tutor : Dr. Gregorio Robles

Proyecto Fin de Carrera

Análisis de proyectos FOSS

Autor : Joel Peralta Rodríguez **Tutor :** Dr. Gregorio Robles Martínez

La defensa del presente Proyecto Fin de Carrera se realizó el día de
de 20XX, siendo calificada por el siguiente tribunal:

Presidente:

Secretario:

Vocal:

y habiendo obtenido la siguiente calificación:

Calificación:

Fuenlabrada, a de de 20XX

*Dedicado a
mi familia / mi abuelo / mi abuela*

Agradecimientos

Aquí vienen los agradecimientos... Aunque está bien acordarse de la pareja, no hay que olvidarse de dar las gracias a tu madre, que aunque a veces no lo parezca disfrutará tanto de tus logros como tú... Además, la pareja quizás no sea para siempre, pero tu madre sí.

Resumen

Aquí viene un resumen del proyecto. Ha de constar de tres o cuatro párrafos, donde se presente de manera clara y concisa de qué va el proyecto. Han de quedar respondidas las siguientes preguntas:

- ¿De qué va este proyecto? ¿Cuál es su objetivo principal?
- ¿Cómo se ha realizado? ¿Qué tecnologías están involucradas?
- ¿En qué contexto se ha realizado el proyecto? ¿Es un proyecto dentro de un marco general?

Lo mejor es escribir el resumen al final.

Summary

Here comes a translation of the “Resumen” into English. Please, double check it for correct grammar and spelling. As it is the translation of the “Resumen”, which is supposed to be written at the end, this as well should be filled out just before submitting.

Índice general

1. Introducción	1
1.1. Contexto	1
1.2. Motivaciones personales	2
1.3. Estructura de la memoria	2
2. Objetivos	5
2.1. Objetivo general	5
2.2. Objetivos específicos	5
2.3. Planificación temporal	6
3. Estado del arte	7
3.1. Free Open Source Software (FOSS)	7
3.2. Python	8
3.2.1. Django	9
3.2.2. Scancode Toolkit	9
3.2.3. GitPython	10
3.2.4. Pygments	10
3.3. Base de datos	11
3.3.1. MySQL	12
3.4. Front-end	12
3.4.1. JavaScript	13
3.4.2. HTML5 y CSS3	14
3.4.3. Bootstrap	15
3.5. Control de versiones	16

3.6. Interacción entre tecnologías	16
4. Diseño e implementación	19
4.1. Arquitectura general	19
4.2. Back-end	20
4.2.1. Análisis de proyectos	21
4.2.2. Procesado de estadísticas	22
4.3. Front-end	25
4.3.1. Estructura AngularJS	25
4.3.2. Funcionamiento	26
4.4. Adaptación de la información	27
5. Resultados	29
6. Conclusiones	31
6.1. Consecución de objetivos	31
6.2. Aplicación de lo aprendido	31
6.3. Lecciones aprendidas	31
6.4. Trabajos futuros	32
6.5. Valoración personal	32
A. Manual de usuario	33
Bibliografía	35

Índice de figuras

3.1. Funcionamiento del Two Way Data Binding	14
3.2. Interacción entre las distintas tecnologías	17
4.1. Distribución de la base de datos	23

Capítulo 1

Introducción

En este proyecto se ha realizado un análisis del ecosistema de proyectos de software libre que podemos encontrar en la red, centrandose especialmente en la plataforma Github. Mediante este análisis se ha intentado dar a conocer las características de este entorno además de determinar las peculiaridades que aportan los datos obtenidos a la formación y desarrollo del Free Open Source Software.

A continuación, para una mejor comprensión del proyecto, se explicarán el contexto del mismo, su estructura y las motivaciones personales para realizarlo.

1.1 Contexto

En un sector tan amplio como el de la programación, en el cual se pueden diseñar y desarrollar soluciones a problemas muy diversos, surge el planteamiento de este proyecto fin de grado para tratar de buscar las características comunes de diferentes proyectos de software libre y de sus desarrolladores.

Con la obtención de la información presente en estos proyectos, se pretende realizar una comparativa de los diferentes tipos de software libre que se están desarrollando en la actualidad, además de sus características en cuanto a copyrights y licencias. De esta manera se podrán entender mejor las cualidades que ha de tener un determinado software para ser libre y abierto, y para poder progresar y obtener mayor relevancia en el mundo del Free Open Source Software.

Esta información legal, aunque secundaria, puede suponer un punto clave para un determinado proyecto, puesto que en la comunidad Free Open Source, un aspecto muy importante del

software debe ser su libertad de uso, acceso al código y distribución y modificación del mismo.

Este proceso de análisis se realiza en el contexto de una aplicación web, por tanto, el resultado del mismo podrá ser accesible a múltiples usuarios que busquen el tipo de proyecto que se está realizando en la actualidad, o a colaboradores activos de proyectos de software libre, o características de distintos proyectos que les permitan tomar conciencia de la importancia de que el nuevo software que están desarrollando sea libre y accesible para todo el mundo.

1.2 Motivaciones personales

La principal motivación por la que se decidió realizar este proyecto es el proceso de análisis, filtrado y representación de grandes cantidades de datos. Además al tratarse de una aplicación web, engloba diferentes aspectos importantes del desarrollo de software, como son el desarrollo web, técnicas de Big Data y también el desarrollo de soluciones a problemas de índole académica.

Por otra parte, los diferentes tipos de tecnologías utilizadas para desarrollar el proyecto, hacían que este fuera más interesante desde un punto de vista educativo y de desarrollo de las habilidades programáticas.

1.3 Estructura de la memoria

Para una mejor lectura, se desglosa a continuación la estructura de este documento:

- **1. Introducción.** Con el objetivo de enmarcar el proyecto en su contexto.
- **2. Objetivos.** Sección en la que se muestran los objetivos finales y parciales del proyecto.
- **3. Estado del arte.** Donde se muestran las tecnologías y metodologías utilizadas para elaborar las distintas partes del proyecto.
- **4. Diseño e implementación.** En esta sección se elabora una explicación del proceso realizado y el diseño previo de este proceso, para una correcta realización del proyecto.
- **5. Resultados.** Para mostrar los resultados obtenidos tras la realización del proyecto.

- **6. Conclusiones.** Donde se enumeran las conclusiones obtenidas de los resultados de la sección anterior.
- **7. Apéndices.** Sección en la que se muestra cualquier información adicional para la adecuada comprensión de la memoria.
- **8. Bibliografía.** Lista de fuentes de informacion en las cuales se basa la este documento.

Capítulo 2

Objetivos

2.1 Objetivo general

El objetivo principal del proyecto es dar una visión general de distintos proyectos de software libre que se desarrollan en la actualidad, para comprender mejor el ecosistema Open Source y ofrecer a los posibles usuarios una manera de encontrar proyectos populares en los que colaborar, usuarios activos que puedan unirse a nuevos proyectos y de observar las licencias, herramientas y lenguajes de programación más populares entre los usuarios.

2.2 Objetivos específicos

Para llevar a cabo el objetivo principal, se han cumplido una serie de objetivos secundarios que se describirán a continuación:

- Familiarizarse con la herramienta Scancode Toolkit que servirá para analizar proyectos desde la terminal de Unix y posteriormente analizar la información devuelta por dicha herramienta.
- Inicializar el proyecto Django/Python que albergará la aplicación web en la que se engloba el análisis de proyectos mencionado anteriormente.
- Adaptar la herramienta para ser usada como un módulo de Python y poder tener mayor versatilidad al analizar proyectos en distintos formatos.

- Desarrollar una API REST cuyos recursos servirán para analizar proyectos y devolver los resultados en un formato fácilmente representable en el navegador.
- Crear los modelos de la base de datos y almacenar los primeros proyectos FOSS que serán posteriormente representados.
- Montar la estructura del framework web AngularJS para realizar de manera correcta y ordenada la representación de los datos obtenidos de la API.
- Representar los datos obtenidos de la API REST mediante la librería nvd3 basada en d3.js en diferentes gráficos.
- Corregir la estética de la página, además de distintos problemas de adaptación entre el back-end (Python) y el front-end (JavaScript).
- Analizar mayor número de proyectos para obtener mayor número de datos que ayuden a sacar conclusiones del entorno Open Source.

2.3 Planificación temporal

La planificación temporal del proyecto queda representada por el gráfico del apéndice X para una mejor comprensión de los plazos tomados para la consecución de los distintos objetivos.

Capítulo 3

Estado del arte

En esta sección se presentarán las tecnologías utilizadas en el desarrollo del proyecto, tanto en el análisis de datos como en la representación de los mismos.

Para ello se realizará una descripción general de cada herramienta y las cualidades específicas por las cuales han sido utilizadas.

3.1 Free Open Source Software (FOSS)

Para entender el concepto de FOSS, se explicarán a continuación los requisitos del software para ser libre y abierto, estos dos términos siguen las directrices de dos organizaciones, Free Software Foundation [4] y Open Source Initiative [9] respectivamente.



A continuación se especificarán las características de cada uno de ellos.

Para que el software se considere libre, ha de cumplir estas cuatro libertades:

- Libertad de uso del programa por cualquier razón.
- Libertad de acceso, estudio y modificación del código.
- Libertad de copia y distribución.
- Libertad de mejora y publicación de las mismas.

Para las libertades 1 y 3, es necesario que el código fuente del software sea de libre acceso al público.



Por otro lado, la iniciativa Open Source se centra menos en las características éticas del software libre, y más en las técnicas y económicas. Sin embargo, son muy pocos los proyectos de software que siendo reconocidos por la Ola Open Source Initiative, no sea considerado Free Software. Por esta razón se suele utilizar el término Free Open Source Software.

3.2 Python

Python es un lenguaje de programación multiparadigma, al permitir la programación orientada a objetos, declarativa y funcional. Es interpretado, por no requerir compilación, con tipado dinámico y multiplataforma. [11]

Este lenguaje fue publicado por Guido van Rossum en 1991 en los Países Bajos como un proyecto de software libre. Actualmente se encuentra bajo una licencia Python Software Foundation License, compatible con la licencia general de GNU a partir de su versión 2.1.1.

Debido a su facilidad de aprendizaje y su uso sencillo, además de por su versatilidad al permitir la combinación de programación declarativa con programación orientada a objetos y el gran número de módulos utilizables en este lenguaje, hacían de Python la elección adecuada como lenguaje de programación para este proyecto.

Actualmente existen dos versiones de Python, las 2.X y las 3.X. Para este proyecto, se utilizó la versión 2.7.6 de Python. Esto es debido a que, aunque las versiones 3.X se están desarrollando cada vez más, existen incompatibilidades entre estas y las versiones 1.9 del framework Django que se describirá a continuación.



3.2.1 Django

Django [3] es un framework web de alto nivel para el lenguaje de programación Python. Este framework permite el desarrollo rápido, limpio y estructurado de cualquier aplicación web sin importar su grado de complejidad. Django es al igual que Python es de código abierto y cuenta con una licencia BSD.



Este framework estructura sus proyectos siguiendo el sistema Modelo-Vista-Controlador. Esto, sumado a su filosofía DRY (Don't Repeat Yourself), hacen que Django sea la mejor opción para entrar en contacto con el desarrollo de aplicaciones web de una manera estructurada, concisa y abordable para programadores de cualquier nivel.

Estas cualidades además de la sencillez de interacción con bases de datos relacionales supusieron un punto clave para la elección de Django como framework sobre el que desarrollar la aplicación web en la que se basa este proyecto.

Para la realización del mismo se ha utilizado la versión 1.9 de Django, por una mayor sencillez y una mejor interoperabilidad con las bases de datos relacionales como puede ser MySQL. Esto supone como ya se ha comentado anteriormente, que existan restricciones de compatibilidad de esta versión con las versiones 3.X de Python.

3.2.2 Scancode Toolkit

La herramienta Scancode Toolkit, es un proyecto de software libre desarrollado en Python y que permite 'escanear' carpetas y ficheros para obtener entre otros, las licencias y copyrights contenidos dentro de cada fichero.

Esta herramienta permite su uso desde la terminal de Linux, o en el caso de este proyecto, la utilización del código y una interfaz API contenida en el mismo para obtener la información requerida en cada caso.

Para poder utilizar Scancode Toolkit en el proyecto, se tuvo que adaptar el código para la obtención de la información legal del archivo directamente desde su contenido, en lugar de facilitar la localización del mismo tal y como se realizaba por defecto en el código que se puede ver en la plataforma Github. [8]

Esta herramienta además cuenta con otras funcionalidades que no han sido utilizadas para realizar el análisis de los archivos en este proyecto. Estas funcionalidades son la obtención del lenguaje en el que están escritos, el número de líneas, etc.

3.2.3 GitPython

Esta librería Python sirve como interacción de alto nivel con repositorios git.

La interacción de esta librería con los repositorios abarca desde la funcionalidad de clonar un repositorio a obtener el 'blame' de un fichero dentro del mismo. Y precisamente estas dos funcionalidades son las que se han utilizado en este proyecto.

La librería GitPython es de código abierto y esta disponible para su uso y modificación en su repositorio de Github [5].

3.2.4 Pygments

Pygments es un paquete escrito en Python que permite resaltar las características sintácticas principales de un código fuente dado.



Esta librería nos sirve por tanto para poder encontrar patrones característicos de un determinado lenguaje dentro del código, y de esta forma, determinar en que lenguaje está programado un determinado archivo.

Esta diferenciación será usada en el análisis planteado en este proyecto, para determinar cuales son los lenguajes de programación más utilizados en los proyectos FOSS a analizar.

3.3 Base de datos

Para la realización de este proyecto, se realizó una comparativa previa del tipo de base de datos que se utilizaría [12].

A la hora de escoger una base de datos adecuada para un proyecto, se pueden escoger dos tipos de base de datos, relacional y no relacional.

Una base de datos relacional, normalmente con SQL como lenguaje base, es aquella que cumple con el modelo previamente fijado sobre ella. Este tipo de bases de datos permiten establecer relaciones entre los datos almacenados en distintas tablas.

Estas características suponen una serie de ventajas de uso:

- Mayor simplicidad de la base de datos
- Robustez.
- Flexibilidad y mejor performance.
- Escalabilidad.
- Mayor eficiencia al realizar peticiones sobre ella.

Por otro lado, las bases de datos no relacionales se caracterizan por no guardar relaciones entre sus distintas colecciones. Su modo laxo de almacenamiento permite guardar en sus colecciones, cualquier tipo de dato o estructura sin seguir ningún modelo fijado previamente.

Dichas particularidades otorgan a las bases de datos no relacionales, las siguientes ventajas:

- Sencillez en su diseño.
- Mayor escalabilidad horizontal.
- Mejores cualidades para usos de Big Data.
- Mayor rapidez en la obtención de los datos.

Observando las cualidades de ambas, se decidió usar una base de datos relacional, ya que a pesar de su mejor funcionamiento en entornos de Big Data y su rapidez al devolver los datos almacenados, las bases de datos no relacionales tienen numerosos problemas a la hora de hacer

cálculos de datos agregados. Al tratarse nuestra aplicación web un dashboard interactivo en el que se mostrarán datos sumados de los distintos proyectos almacenados, las bases de datos relacionales se adaptan mucho mejor al uso que se dará a los datos recogidos.

3.3.1 MySQL

MySQL [10] es un sistema de gestión de bases de datos relacionales multiplataforma, desarrollado en C y C++ bajo una licencia dual GPL/Licencia Comercial de Oracle Corporation y es considerada la base de datos relacional open source más popular del mundo.



La simplicidad de MySQL como gestor de bases de datos no relacionales, supone una cualidad esencial para la gestión de páginas web con representación de datos dinámicos.

Además, MySQL soporta una gran cantidad de tipos diferentes para almacenar datos y al estar tan extendido, tiene multiples APIs para interactuar con la base de datos desde diferentes lenguajes de programación.

Estas cualidades además de se escalabilidad, y su facilidad de uso utilizando Django como framework web, son características clave para el uso de este gestor de bases de datos en este proyecto.

3.4 Front-end

En el código del cliente o front-end, se han utilizado múltiples tecnologías para la visualización de los datos y la realización de una adecuada interfaz de usuario. Esas tecnologías se pasan a describir en las siguientes secciones.

3.4.1 JavaScript

JavaScript [7] es un lenguaje de programación interpretado, dialecto del estandar ECMAScript y fue desarrollado por Brendan Eich, trabajador de Netscape en 1995. Este lenguaje es orientado a objetos, imperativo y con tipado débil y dinámico.



Este lenguaje es usado normalmente en el lado del cliente, como parte del código utilizado por el navegador, permitiendo la realización de páginas web dinámicas e interfaces de usuario.

JavaScript es el único lenguaje apoyado comunmente por todos los navegadores modernos. Por esta razón, es el más utilizado en el desarrollo de aplicaciones web en el lado del cliente, y el que se utiliza en este proyecto.

AngularJS

AngularJS [1] es un framework web abierto de JavaScript, mantenido por Google y que permite la creación de SPA o Single Page Applications.



El nivel de organización que permite Angular en el lado del cliente, además del dinamismo que otorga su principal característica, el two way data binding, cuyo funcionamiento permite que cualquier modificación realizada sobre la vista repercute en el modelo y viceversa, tal y como se muestra en la figura 3.1, lo han convertido en uno de los frameworks web más populares en la actualidad, al librar al back-end de funciones de renderizado de templates.

Para la representación de los datos obtenidos del análisis, se utilizarán gráficos que se generarán mediante la modificación del DOM de la página web. Esto supondría una gran cantidad de código desestructurado si se utilizara JavaScript con jQuery como librería principal, pero

con la utilización de AngularJS, además de ordenar el código en una estructura MVC, se reduce drásticamente la cantidad del mismo.

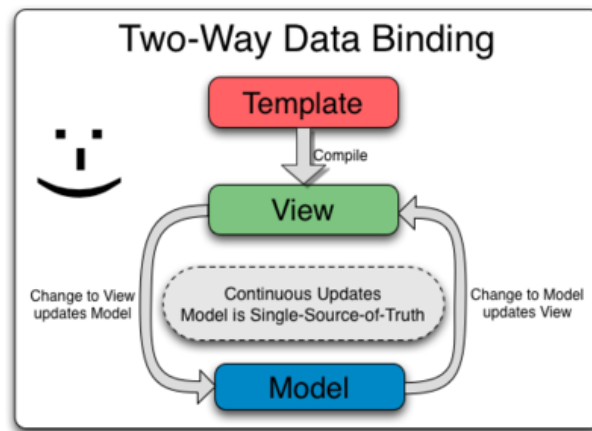


Figura 3.1: Funcionamiento del Two Way Data Binding

D3.js

La librería de JavaScript D3 [2], llamada así por las siglas Data-Driven Documents, permite producir a partir de datos, gráficos dinámicos e interactivos mediante SVG.



Para este proyecto, utilizaremos D3 con otra librería basada en d3.js llamada nvd3.js, que permite crear de manera sencilla y uniforme, gráficos de distintos tipos y para cualquier tipo de datos, que además se actualizan conforme estos datos vayan cambiando.

3.4.2 HTML5 y CSS3

Para la realización de este proyecto, juegan un papel importante tecnologías como HTML5 o CSS3, que permiten la creación de una interfaz de usuario llamativa y fácil para el correcto funcionamiento de la aplicación web.

HTML5

HTML5 [6] es un lenguaje de marcado que nos permite realizar plantillas en las que representar nuestra interfaz de usuario.



Además el DOM de la página, que es una representación en árbol de los distintos elementos de la plantilla HTML, nos permite modificar mediante distintas herramientas como puede ser Angular, el aspecto y organización de nuestra interfaz.

CSS3

Como complemento a HTML, tenemos CSS, que es un lenguaje creado para definir la presentación y estilo de documentos HTML o XML.



Mediante este lenguaje definiremos el estilo de nuestra interfaz de usuario, dotandola de las características deseadas para hacerla más atractiva al usuario.

3.4.3 Bootstrap

Bootstrap [13] es un conjunto de herramientas de código abierto, desarrollado y mantenido por Twitter, que permite hacer nuestra página HTML/CSS, responsiva. Es decir, permitir que esta se adapte al tamaño del navegador o el dispositivo en que esta siendo mostrada.



Esta funcionalidad la otorga el sistema de celdas de Bootstrap que permite una mejor maquetación de nuestra página web.

En este proyecto, se utiliza Bootstrap por cuestiones estéticas, pero además por seguir con la filosofía de organización y estructuración del código que se asume al utilizar AngularJS.

3.5 Control de versiones

Para la un mejor seguimiento y control de los avances de este proyecto, se ha utilizado Git como herramienta de control de versiones.



Este software fue diseñado por Linus Torvalds en 2005, para permitirle a él y a su equipo un desarrollo más ágil y productivo.

Por este motivo y por la función de backup que estas versiones permiten, se ha utilizado este software para registrar cada hito conseguido en el proyecto.

3.6 Interacción entre tecnologías

Por último y para sintetizar, en esta sección explicaremos las interacciones entre las tecnologías anteriormente descritas.

Tal y como se muestra en la figura 3.2, al acceder un usuario a la página web y realizar el análisis o la búsqueda de ciertos datos, se enviará una petición desde el cliente Angular hasta el

servidor Django, que tras obtener los datos de la base de datos, los devolverá en la respuesta al cliente, y este actualizará mediante el two way data binding los datos contenidos en los gráficos de la librería D3.

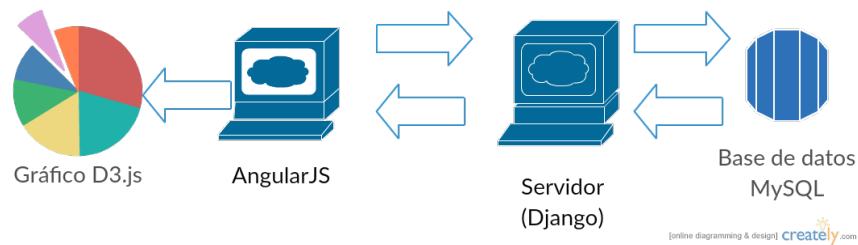


Figura 3.2: Interacción entre las distintas tecnologías

Capítulo 4

Diseño e implementación

En esta sección se desarrollará el proceso seguido para realizar el diseño preliminar del proyecto y su implementación.

Primero se dará una visión general de la arquitectura de la aplicación para posteriormente centrarse en las dos partes principales de la aplicación, el front-end y el back-end, y debido a las diferentes tecnologías utilizadas en ambos, se explicará por último el proceso de adaptación entre una y otra parte, para el correcto funcionamiento de la aplicación.

4.1 Arquitectura general

Este proyecto consiste en una aplicación web Django en la que se mostrarán los resultados de un análisis producido en el back-end de la aplicación.

El back-end o lado del servidor ha sido desarrollado en Python mediante el framework web Django. En esta parte del código se ha desarrollado una API REST que devolverá en función de los recursos a los cuales se acceda, los resultados analizados en el momento de la petición o almacenados en la base de datos, del proyecto FOSS requerido por el usuario. Además permitirá la obtención de los datos globales de la aplicación, así como los correspondientes a un usuario y aquellos relacionados con una determinada inada inforamción legal.

De los resultados devueltos por el servidor se servirá el código del navegador o front-end. Este código realizará funciones de visualización de los datos obtenidos. Mediante el framework web AngularJS se realizará una SPA que mostrará dos secciones en las que visualizar los datos, una en la que analizar un proyecto y otra en la que visualizar las estadísticas totales almacenadas

en la base de datos, además de poder filtrarlas.

Para una mejor comprensión de la aplicación, se desarrollará en mayor profundidad el funcionamiento de la misma en los siguientes apartados.

4.2 Back-end

En este apartado se explicarán las características de la API REST desarrollada en el servidor Django.

Para implementar esta aplicación y esta API REST, se puso en funcionamiento primero un proyecto Django sobre el cual se desarrollará el back-end de la aplicación.

Desde este servidor Django se sirven los archivos correspondientes a la página principal y en este caso única de la aplicación, además del código correspondiente al funcionamiento de la página. También en este servidor se puede acceder a cuatro recursos diferentes correspondientes a la API REST ya mencionada.

Estos recursos son los siguientes:

- **/analyze:** Accediendo a este recurso y enviando en el cuerpo de la petición HTTP, la url de un proyecto, se obtienen los datos correspondientes al análisis del proyecto al que corresponde la url. Entre estos datos se encuentran los ficheros que contiene y los contribuidores, las licencias y los copyrights presentes en ellos.
- **/all:** Mediante el acceso a este recurso se devolverán en la respuesta todos los datos contenidos en la base de datos que hagan referencia al total de líneas que ha desarrollado cada contribuidor, los lenguajes utilizados y los copyrights y licencias utilizados en los distintos ficheros almacenados.
- **/user:** Al realizar una petición HTTP sobre este recurso con el nombre de un usuario en su cuerpo, se devolverán los lenguajes utilizados por este usuario, los proyectos en los que ha participado y las licencias y copyrights que ha utilizado en sus ficheros.
- **/legal:** Al acceder a este recurso, se podrán enviar una licencia o un copyright por los cual se quieran filtrar los ficheros almacenados en la base de datos, devolviendo en la respuesta todos los ficheros que contengan esta información legal.

4.2.1 Análisis de proyectos

A continuación se describirá el proceso de análisis de los distintos proyectos así como la implementación de dicho análisis.

Proceso

El análisis se realizará al acceder al recurso `/analyze` siempre y cuando no haya sido almacenado previamente en la base de datos, y la fecha en la que fue almacenado sea más reciente que la fecha de última modificación del proyecto.

Si se cumplen estas dos condiciones, en la vista correspondiente al recurso `/analyze` se utilizará la librería `GitPython` para clonar el repositorio del proyecto a analizar, para posteriormente utilizar la herramienta `Scancode Toolkit` sobre la carpeta en la que este proyecto fue clonado.

Mediante esta herramienta finalmente se obtendrá la información legal de cada fichero, y mediante la librería `GitPython` y la API de Github se obtendrá otra información como pueden ser los contribuidores y las contribuciones de cada uno de ellos.

Implementación

Para implementar este proceso se hará uso de una función que con la herramienta `Scancode Toolkit`, recorrerá todos los archivos contenidos en la carpeta para obtener de cada uno de ellos sus licencias y copyrights. Esta herramienta utiliza una lista de licencias y reglas que le sirven como patrón para comparar el contenido de cada fichero en busca de las licencias a las que correspondan dichas reglas.

Además de `Scancode`, se utilizará como se especificó anteriormente, la API de Github, que devuelve un objeto JSON con la información requerida, en este caso los contribuidores e información general del proyecto.

Por otro lado, también se utilizan `GitPython` para obtener las líneas equivalentes al comando `git blame` de Git, que serán posteriormente parseadas para obtener el autor de cada una y la biblioteca `Pygments` que a partir del contenido de cada fichero, determinará el lenguaje en que está programado en base a su sintaxis.

Otras consideraciones

Para el correcto funcionamiento de la aplicación fueron necesarias ciertas modificaciones sobre la herramienta Scancode Toolkit:

- El funcionamiento por defecto de Scancode Toolkit carga la lista de reglas y licencias en cada análisis. En este caso, al tratarse de una aplicación web, esta funcionalidad no es viable, por lo tanto se tuvo que modificar el funcionamiento de la herramienta para que la carga de licencias se realice nada más arrancar la aplicación web, almacenando después la lista en el servidor, y evitando así que se tengan que realizar cargas posteriores.
- Para facilitar el análisis completo de los proyectos se decidió leer el contenido de los distintos ficheros al comenzar el proceso de análisis, evitando así, las distintas lecturas que habría que realizar por defecto en el funcionamiento de las herramientas utilizadas. De esta forma, para evitar estas múltiples lecturas, se modificó la herramienta Scancode Toolkit para poder analizar el contenido de cada fichero directamente en lugar de leerlo dentro del código de esta herramienta.

4.2.2 Procesado de estadísticas

En esta sección se describirán los procesos de obtención y envío de los datos estadísticos de los proyectos, además de su filtrado.

Tablas de la base de datos

La obtención de los datos de la base de datos de MySQL se realizará mediante el uso de los modelos que nos da la funcionalidad de Django. De esta manera, al acceder a los recursos `/all`, `/user` o `/legal`, se realizarán las interacciones necesarias con la base de datos gracias a la sencilla interfaz que nos proporcionan estos modelos.

Es necesario destacar que al tratarse de una base de datos relacional, se tendrá que acceder a las distintas tablas a las cuales pertenecen los datos de estas relaciones, antes de elaborar la respuesta.

La disposición de las tablas presentes en la base de datos, además de las distintas relaciones entre ellas se puede observar en la figura 4.1.

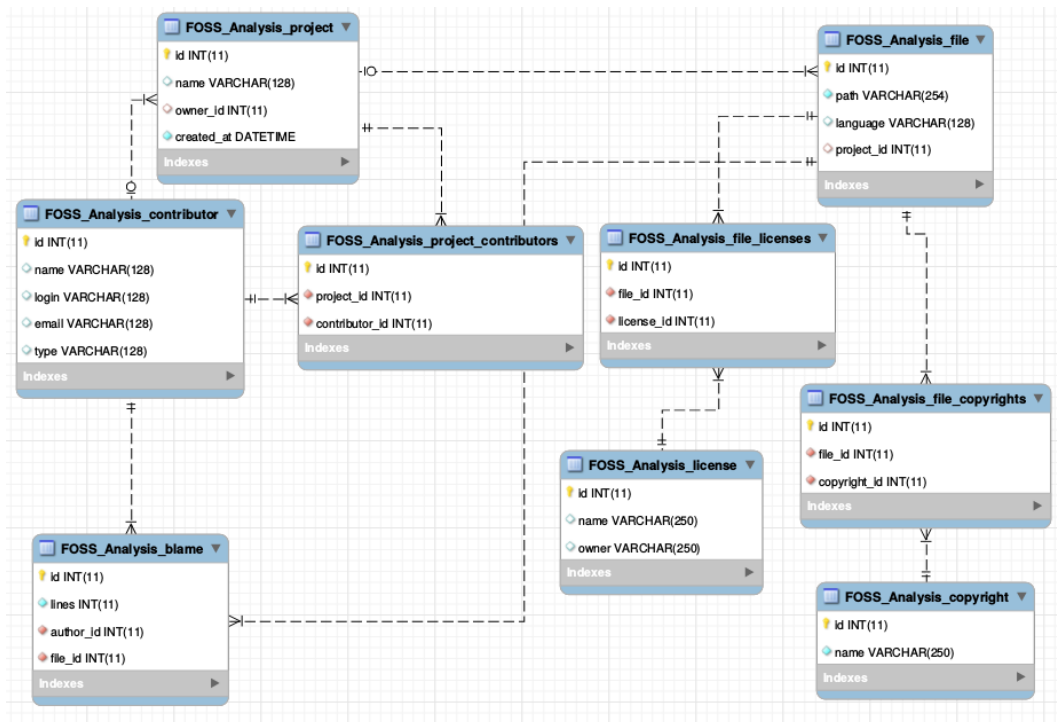


Figura 4.1: Distribución de la base de datos

En esta figura observamos las siguientes tablas de la base de datos:

- **Project:** En esta tabla contiene el nombre y la fecha de almacenamiento del proyecto, además del dueño y los contribuidores del proyecto mediante unas relaciones de Foreign Key y Many to Many respectivamente con la tabla de Contributor.
- **Contributor:** Donde se almacena el nombre, login e email del contribuidor, teniendo además el par login, email como clave única de los contribuidores, de forma que no podrán almacenarse dos contribuidores cuyos campos de login y email sean ambos iguales.
- **File:** Para almacenar los ficheros contenidos en cada proyecto utilizaremos esta tabla, donde se almacenan el path absoluto donde ha sido analizado el fichero y el lenguaje en que este está programado. Además de esto, esta tabla estará relacionada con las tablas Project, Copyright y License para determinar a que proyecto pertenece el archivo y que copyrights y licencias están contenidas en el.
- **Blame:** En la tabla Blame se almacenarán las contribuciones por cada fichero que ha realizado cada usuario. Esto se realizará almacenando las líneas de código que cada con-

tribuidor ha realizado, guardando el id correspondiente con el contribuidor en la tabla Contributor en el campo author, y así mismo con el fichero en el campo file.

- **Copyright:** Donde se almacena el nombre del copyright y a quien pertenece.
- **License:** Donde se guardan los nombres de las distintas licencias.

Queries

Una vez descrito el modo en que están almacenados los datos en las distintas tablas de MySQL, procedemos a describir las interacciones que se han de realizar con la base de datos para poder obtener y enviar los distintos datos.

Estadísticas de proyecto Para la obtención de los datos globales de cada proyecto accedemos al recurso /analyze en cuya vista se realizará una query sobre la tabla de proyectos en la que se traerán además los datos relacionados con la tabla de contribuidores.

Tras realizar esta query se procesarán los datos para que sean serializables y para que tengan un formato JSON aceptable para el front-end programado en JavaScript.

Estadísticas totales En el caso de la obtención del total de los datos almacenados en la base de datos, se realizarán tres queries distintas sobre la tabla File, de las que se obtendrán los datos relacionados de las licencias y copyrights de todos los ficheros, además de los datos del lenguaje de cada uno de ellos. Por último, se hará una petición sobre la tabla Blame que devolverá el número de líneas totales realizadas por cada autor en los distintos ficheros.

Sobre estos datos se realizará una suma mediante el método count de SQL para cada valor diferente, gracias al método de la interfaz que aporta Django con el mismo nombre. Obteniendo así el número total de líneas realizadas por cada autor y el número total de copyrights, licencias y lenguajes diferentes.

Por último se procesarán los datos para adaptarlos al formato JSON, y además al formato de los datos que tendrán que ser introducidos en los gráficos de D3.

Estadísticas de usuario Para obtener los datos concretos de un usuario, sobre la tabla Blame se realizarán tres peticiones para obtener el número total de ficheros programados en cada lenguaje, y que contienen cada tipo de copyright y licencia, filtrándolos por el usuario determinado.

Además se harán dos queries en las que se obtendrán todos los nombres de proyectos en los que el usuario participa como dueño o como colaborador del mismo.

Este filtrado se realizará mediante el método `where` del lenguaje SQL o en este caso, del método `filter` de la API de Django models.

Al igual que en el caso anterior, se realizará un procesado final antes de enviar la respuesta con los datos obtenidos, filtrados y sumados.

Estadísticas legales Por último se obtendrán todos los ficheros filtrados por licencias o por copyrights.

En este caso se realizará un filtrado como en el caso anterior, pero con la diferencia de que en este caso se filtrará de manera no estricta, es decir, valdrá con introducir una palabra contenida en los copyrights o licencias del fichero, para que este sea devuelto en la respuesta.

4.3 Front-end

A continuación se pasará a explicar el funcionamiento y diseño de la parte del cliente de la aplicación.

Como ya se ha comentado antes, esta sección del proyecto está implementada con el framework AngularJS, el cual aporta una estructura al código que se pasará a explicar a continuación.

4.3.1 Estructura AngularJS

El framework Angular nos permite s permite dividir nuestro código por módulos, los cuales permitirán diferenciar de manera adecuada las distintas funcionalidades que pueda tener nuestra web. Los módulos que forman esta aplicación son los siguientes:

- **Main:** Permite la selección de las distintas secciones de la página sin necesidad de recargar o acceder a una nueva url, convirtiendo así la página en una Single Page Application.
- **Análisis:** Accediendo a la API implementada en el servidor Django, obtiene los datos analizados del proyecto, y los almacena de modo que puedan ser usados por la plantilla HTML correspondiente a este módulo y sus gráficos.

- **Statistics:** En este módulo al igual que en el de Analysis se obtendrán los datos de la API del servidor, y se mostrarán en tres plantillas diferentes ya sean datos totales, de usuario o legales.
- **Navbar:** Es el módulo que permite controlar el funcionamiento de la barra de navegación, permitiendo mostrar las secciones que contiene la página.
- **Chart:** El módulo de Chart se crearán los gráficos que nos servirán para representar los datos de los módulos de Analysis y Statistics. Por tanto estos dos módulos tendrán una dependencia con el de Chart.

Todos estos módulos se inyectan en las dependencias del módulo principal de la aplicación, que será el módulo app.

4.3.2 Funcionamiento

En los distintos módulos se establecerán una serie de directivas Angular que permitirán el intercambio de los distintos parciales o plantillas HTML, gracias a la colocación del método `ng-view` en un elemento del DOM. Mediante este método, en este elemento se colocarán los contenidos de las plantillas referenciadas dentro de cada directiva.

Una vez mostrada la plantilla correspondiente a cada directiva, se permitirá mediante distintos formularios, la realización de peticiones de datos a la base de datos tal y como se ha explicado anteriormente.

Tras obtener la respuesta a las peticiones, se almacenará en campos del elemento `$scope` de Angular, permitiendo mediante el Two Way Data Binding mencionado anteriormente, la actualización y representación de los datos en cada plantilla.

Cabe mencionar, que en el caso de los gráficos de la biblioteca D3 es necesario actualizar el gráfico con los nuevos datos. Este proceso no es automático, por tanto se ha hecho uso del método `$watch` para añadir un manejador al evento de actualización de los datos, así cuanto este evento se produzca, se llamará a una función que generará el gráfico de nuevo y actualizado con los nuevos datos.

4.4 Adaptación de la información

Debido a las distintas tecnologías utilizadas en el back-end y en el front-end, donde se utilizan Python y JavaScript respectivamente, se producen ciertas incompatibilidades entre los formatos de uno y otro extremo.

Estos conflictos no suponen un problema a la hora de realizar ambas partes por separado, pero es necesario atajarlos ya que entre uno y otro punto se produce un intercambio de información.

En este caso, al ser JavaScript, en el front-end, el consumidor de la información, será necesario que antes de su envío, la información sea procesada para darle un formato JSON. Este formato implica tener una colección de elementos con una clave y un valor asociado a ella.

Este formato permitirá una adecuada serialización desde el servidor al cliente para el correcto envío de la respuesta, y que en el front-end no se requiera ningún procesamiento especial de la respuesta, más que el de un elemento JSON propio de JavaScript, y su posterior adaptación para su representación.

La adaptación de la información supone un elemento clave también para considerar la API desarrollada como una API REST, puesto que esto conlleva una homogeneidad entre los recursos y el formato JSON permite dar esta homogeneidad a las respuestas de estos recursos.

Capítulo 5

Resultados

Capítulo 6

Conclusiones

6.1 Consecución de objetivos

Esta sección es la sección espejo de las dos primeras del capítulo de objetivos, donde se planteaba el objetivo general y se elaboraban los específicos.

Es aquí donde hay que debatir qué se ha conseguido y qué no. Cuando algo no se ha conseguido, se ha de justificar, en términos de qué problemas se han encontrado y qué medidas se han tomado para mitigar esos problemas.

6.2 Aplicación de lo aprendido

Aquí viene lo que has aprendido durante el Grado/Máster y que has aplicado en el TFG/TFM. Una buena idea es poner las asignaturas más relacionadas y comentar en un párrafo los conocimientos y habilidades puestos en práctica.

1. a

2. b

6.3 Lecciones aprendidas

Aquí viene lo que has aprendido en el Trabajo Fin de Grado/Máster.

1. a

2. b

6.4 Trabajos futuros

Ningún software se termina, así que aquí vienen ideas y funcionalidades que estaría bien tener implementadas en el futuro.

Es un apartado que sirve para dar ideas de cara a futuros TFGs/TFM.

6.5 Valoración personal

Finalmente (y de manera opcional), hay gente que se anima a dar su punto de vista sobre el proyecto, lo que ha aprendido, lo que le gustaría haber aprendido, las tecnologías utilizadas y demás.

Apéndice A

Manual de usuario

Bibliografía

- [1] AngularJS. Web oficial de angularjs. <https://angularjs.org/>.
- [2] D3.js. Web oficial de la librería d3.js. <https://d3js.org/>.
- [3] Django. Web oficial de django project. <https://www.djangoproject.com/>.
- [4] FreeSoftwareFoundation. Web oficial de la free software foundation. <http://www.fsf.org/>.
- [5] GitPythonDevelopers. Repositorio en github de la librería gitpython. <https://github.com/gitpython-developers/GitPython>.
- [6] HTML5Doctor. Web de html5 doctor. <http://html5doctor.com/>.
- [7] MozillaDeveloperNetwork. Web oficial de mdn. <https://developer.mozilla.org/es/docs/Web/JavaScript>.
- [8] nexB. Repositorio en github de la herramienta scancode toolkit. <https://github.com/nexB/scancode-toolkit>.
- [9] OpenSourceInitiative. Web oficial de la open source initiative. <https://opensource.org/>.
- [10] Oracle. Web oficial de mysql. <https://www.mysql.com/>.
- [11] Python. Web oficial de python programming language. <https://www.python.org/>.
- [12] J. Serra. Relational databases vs non-relational databases. <http://www.jamesserra.com/archive/2015/08/relational-databases-vs-non-relational-databases/>.
- [13] TwitterBootstrap. Web oficial de bootstrap. <http://getbootstrap.com/>.