



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería en Informática



TFG del Grado en Ingeniería Informática

**Framework web de uso de
sistemas de machine learning.
Documentación Técnica**



Presentado por Javier Martínez Riberas
en Universidad de Burgos — 30 de agosto de 2017

Tutores: Dr. José Francisco Díez Pastor
Dr. César Ignacio García Osorio

Copyright (C) 2017 Javier Martínez Riberas. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included as a pdf file titled “LICENSE.pdf”.

Índice general

Índice general	I
Índice de figuras	III
Índice de tablas	IV
Apéndice A Plan de Proyecto Software	1
A.1. Introducción	1
A.2. Planificación temporal	1
A.3. Estudio de viabilidad	4
Apéndice B Especificación de Requisitos	7
B.1. Introducción	7
B.2. Objetivos generales	7
B.3. Catalogo de requisitos	7
B.4. Especificación de requisitos	8
Apéndice C Especificación de diseño	14
C.1. Introducción	14
C.2. Diseño de datos	14
C.3. Diseño arquitectónico	15
C.4. Diseño procedimental	17
C.5. Diseño de paquetes	18
Apéndice D Documentación técnica de programación	22
D.1. Introducción	22
D.2. Estructura de directorios	22
D.3. Manual del programador	24
D.4. Despliegue	39
D.5. Pruebas del sistema	40

<i>ÍNDICE GENERAL</i>	II
-----------------------	----

Apéndice E Documentación de usuario	41
E.1. Introducción	41
E.2. Requisitos de usuarios	41
E.3. Instalación	41
E.4. Manual del usuario	42
Bibliografía	44

Índice de figuras

B.1. Diagrama de casos de uso	13
C.1. Diagrama entidad relación	15
C.2. Model view controller	16
C.3. Arquitectura de la aplicación	17
C.4. Diagrama de secuencia	18
D.1. Clonación de los repositorios	26
D.2. Creación de entorno virtual e instalación de dependencias	28
D.3. Prueba de uso de pip freeze	29
D.4. Proceso de traducción	33
E.1. Punto de entrada a la aplicación	42
E.2. Pantalla de registro de usuario	42
E.3. Pantalla de inicio de sesión	43
E.4. Inicios de sesión alternativos	43
E.5. Cambio de las zonas accesibles al iniciar sesión	43
E.6. Pantalla de acceso a el servicio de clasificación	43

Índice de tablas

A.1. Costes de personal	4
A.2. Costes de material al mes	5
A.3. Dependencias del proyecto en produccion	6
A.4. Dependencias exclusivas del desarrollo	6
B.1. CU-01 Registro de usuario.	9
B.2. CU-02 Login.	9
B.3. CU-03 Logout.	10
B.4. CU-04 Clasificación de imagen.	10
B.5. CU-05 Subida de datasets.	11
B.6. CU-06 Re-entrenado de modelos.	12

Apéndice A

Plan de Proyecto Software

A.1. Introducción

La planificación es un punto importante en cualquier proyecto. Estimar el trabajo, el tiempo y el dinero que va a suponer la realización del proyecto aunque vaya a cambiar más tarde es interesante para saber si puede haber posibilidades de que sea viable. Para ello, debemos analizar cuidadosamente los componentes del proyecto. Con este análisis pretendemos conocer los requisitos del proyecto y pretendemos que mediante modificaciones siga sirviendo en un futuro.

A.2. Planificación temporal

En un principio se planteó seguir una metodología ágil, esta sería scrum ya que existía experiencia anterior. Por supuesto no se pudo usar completamente ya que no se tenía un equipo, no se hicieron reuniones diarias. . .

Se empezó a usar ZenHub como tablero kanban donde se situarían las tareas con sus costes. Este tipo de planificación con *milestones* e *issues* no ha sido seguido la mayor parte del tiempo ya que no era la metodología más adecuada. En labores de investigación que no se sabían si iban a ser usables más adelante las cuales no tenían ningún entregable claro se decidió no incluirlas a partir de cierto punto para no acumular demasiada información innecesaria en el repositorio.

En retrospectiva se ha deducido que una forma mejor de hacer este tipo de trabajos es llevar varios tableros kanban en varios repositorios: uno para todo el proyecto y otro para todo el informe del proyecto en vez de todo junto.

Sistema de *sprints*

El sistema de *sprints* no fue un constante a lo largo del desarrollo del proyecto, esto se debió a el hecho de que algunas cosas se hicieran antes de empezar el cuatrimestre por cuestiones de balanceamiento de la carga de trabajo.

Lo que esto supuso es que antes de empezar con el sistema de *sprints* se usase el tiempo de 1-3 *sprints* para aprender Flask, tras este periodo se llevó un sistema más estandarizado de *sprints*, estos se pueden ver en las *milestones* de GitHub y en las *issues*. Por problemas en cómo se movieron las *issues* por el tablero (a *QA* hasta la reunión del *sprint*), no se pudieron sacar gráficos automatizados suficientemente exactos, ya que acababan siendo un peldaño enorme en la reunión del *sprint*.

Tras una serie de semanas llevando este sistema, se tubo que abandonar por varios motivos. Los *sprints* empezaron a ser menos concretos y más basados en la investigación y desarrollo de la alternativa más atrayente, la carga de trabajo del resto de asignaturas se tenía en picos, propiciando *sprints* con mucho trabajo y otros sin casi nada y por último una serie de circunstancias personales hicieron que acabase por abandonar la documentación de cada *sprint online*.

Sprints antes del cuatrimestre

Antes del cuatrimestre como se ha comentado se siguieron una serie de *sprints*, estos se basaron en seguir el tutorial explore Flask. El *sprint* se llevo a cabo con éxito, terminando aquello que se tenía pensado hacer, una aplicación Flask.

Sprint 1

En este *sprint* se planteo el hecho de mover el trabajo anterior a una plataforma como Git y llevar el desarrollo más formalmente. También se empezó a investigar sobre subida de archivos y subida de archivos tipo *drag and drop*. La subida de archivos se consiguió con un poco de retraso pero el *drag and drop* no. Esto se debe a mi inexperiencia en desarrollo *front end*.

Sprint 2

En este *sprint* se empezó a subir el código a Heroku, comprobando que se podía desplegar en condiciones reales. Se intentó mejorar la interfaz gráfica.

Sprint 3

Se empezó a generar documentación fuera de código. Este *sprint* fue lento y no muy productivo ya que no se sabía cómo se quería desarrollar la documentación. En este punto se desarrolló como una comparativa entre todas las

alternativas hasta el momento, lo cual no ha sido la dirección que ha tomado el proyecto con el tiempo.

Sprint 4-6

En esta serie de *sprints* se planteó la elección del *framework* de *Machine Learning* que usar. La decisión se investigó más entre *sklearn*, *Theano* y *Tensorflow*, aunque de tener la opción ahora se hubiese elegido *Keras*.

También se empezó a implementar la integración, algo que no se pudo cumplir y que acabó resultando mucho más costoso de lo que se había pensado previamente.

Sprint 7

En este *sprint* se pretendía mejorar la aplicación y documentar mucho de los *sprints* anteriores. En este punto se empiezan a tener problemas con el seguimiento de las tareas de manera online.

Docker

Tras investigar sobre el tema se decidió poner algo de énfasis en el despliegue ya que ya se había intentado y llevado a cabo con *Heroku*. De los contenedores existentes se tubo que investigar y decir si usarlos y cual. La decisión fue *Docker* al ser la única alternativa que tenía proyectos en producción. En el momento de la elección era totalmente *open source*, algo que cambió con el tiempo y tras lo cuál hubo que hacer otra investigación de cómo influía el cambio. Resulto no influir apenas, puede ser que una vez maduren otras alternativas haya que intentar desacerse de *Docker*

Transfer learning

Tras investigar un poco sobre cómo se podía hacer la integración con *Tensorflow*, se vió que el *transfer learning* de un modelo como *inception v3* era suficientemente simple como para intentar llevarlo a cabo. Se desarrollaron los *scripts* para predecir clases a partir de imágenes.

Docker Compose y Swarm

Docker Compose y *Docker Swarm* fueron una de las razones por las que se eligió *Docker* en vez de otras alternativas, pero aún así la barrera de entrada a estas tecnologías estaba lo suficientemente alta como para que *Swarm* acabase siendo inalcanzable.

Concepto	Coste
Salario neto	1000
Retención IRPF (19 %)	360.53
Seguridad social (28,30 %)	537.00
Salario bruto	1897.53
4 meses tiempo parcial(3/4)	5692.59

Tabla A.1: Costes de personal

Microservicios

Los microservicios se empezaron a conocer gracias a *Docker*, ya que son una alternativa muy usada en este tipo de sistemas al tener una relación simbiótica. Cada uno beneficia al otro.

Escalabilidad y despliegue de páginas web

Por último se decidió centrar el proyecto en parte en la escalabilidad ya que en dos de los ejes ya estaba, y para implementarlo en el tercero sólo hacía falta cambiar de una imagen de la base de datos a otra. Este último cambio no se llevo a cabo porque, salvo en aplicaciones excesivamente grandes, el usar una base de datos normal es más beneficioso.

A.3. Estudio de viabilidad

En este apartado se estudiarán los costes en los que se incurren al desarrollar este proyecto.

Viabilidad económica

El proyecto incurre en distintos tipos de costes

Costes de personal

El proyecto se lleva a cabo por un desarrollador junior empleado a tiempo parcial (30h/semana) durante cuatro meses. Se considera el siguiente salario:

Costes de material: hardware y software

Como material podemos considerar lo mínimo necesario para llevar un proyecto así:

Un único coste puntual (ordenador portátil) que aproximamos en 600 y en la tabla se pondrá su coste amortizado contando con una amortización a 4 años. Se ha comprobado que internet está incluido en el alquiler de la oficina.

Concepto	Coste
Ordenador portátil	25
Alquiler de oficina	99
1 mes	124
4 meses	496

Tabla A.2: Costes de material al mes

Costes totales

El sumatorio de todos los costes es de 6188,59. Podríamos recortar más en ciertos puntos pero debido a que no se va a llevar a cabo como esta planteado aquí, esto solo es una aproximación del coste de oportunidad.

Beneficios

Si nuestro interés fuese vender el proyecto este no sería el proyecto que venderíamos, tendríamos que añadir medidas de tiempo computacional en cada operación.

Una vez consigamos calcular tiempos de computación podemos restringir a cada usuario una cantidad de tiempo. De esta manera podemos crear planes para cada usuario, podemos plantearnos hacer un plan gratuito y varios planes de pago según cantidad de tiempo computacional que se le permita usar al cliente.

Viabilidad legal

La licencia necesaria para nuestro proyecto debido a las dependencias que tiene tendrá que ser compatible con aquellas de las bibliotecas que hemos usado, la licencia más restrictiva que hemos usado es la de paramiko siendo LGPL (2.1 y 3.0)[2] , esta licencia esta pensada para librerías e incluye el siguiente párrafo:

5. A program that contains no derivative of any portion of the Library, but is designed to work with the Library by being compiled or linked with it, is called a "work that uses the Library". Such a work, in isolation, is not a derivative work of the Library, and therefore falls outside the scope of this License.

Por lo que las restricciones de esa licencia no se nos aplican. Esto quiere decir que podemos publicar nuestro código bajo la licencia que mejor nos parezca o incluso podríamos mantenerlo privado como un Secreto de negocio (Trade Secret) ya que proporcionamos SaaS (Software as a Service). Basandome en las licencias más comunes open source, ya que me parece interesante

Dependencia	Versión	Licencia
Alembic	0.9.3	MIT
Bcrypt	3.1.3	Apache 2.0
Flask	0.12.2	BSD
Flask-Babel	0.11.0	BSD
Flask-Bcrypt	0.7.1	BSD
Flask-Login	0.4.0	MIT
Flask-SQLAlchemy	2.2	BSD
Flask-Oauthlib	0.9.4	BSD
Flask-WTF	0.14.2	BSD
itsdangerous	0.24	BSD
Jinja2	2.9.6	BSD
Paramiko	2.2.1	LGPL-3.0
Psycpg2	2.7.1	LGPL-3.0
SQLAlchemy	1.1.11	MIT
Werkzeug	0.12.2	BSD
WTForms	2.1	BSD

Tabla A.3: Dependencias del proyecto en produccion

Dependencia	Versión	Licencia
codeclimate-test-reporter	0.2.3	MIT
coverage	4.4.1	Apache 2.0
Pytest	3.1.3	MIT

Tabla A.4: Dependencias exclusivas del desarrollo

el hecho de que otras personas puedan usar el código, y con ayuda de las recomendaciones de gnu[1], se ha decidido usar la GNU GPL 3.0.

Para la documentación se ha decidido que la mejor licencia para las necesidades era la GNU FDL (GNU Free Documentation License). La otra alternativa más interesante era Creative Commons.

Para completar la documentación sobre la viabilidad legal se adjunta una tabla con todas las dependencias del proyecto.

Apéndice B

Especificación de Requisitos

B.1. Introducción

Este apartado consistirá en una descripción de los propósitos y requerimientos de la aplicación.

B.2. Objetivos generales

El objetivo principal del proyecto es el crear una aplicación web, esta debe tener control de usuarios y capacidad de clasificar imágenes mediante una red neuronal. Por último debe ser capaz de re-entrenar el modelo de clasificación de imágenes.

La red neuronal a usar es inception v3 y el conjunto de datos es Imagenet.

Además se pide que todo esto se haga de una manera escalable, desplegable y con un enfoque que reduzca la deuda técnica.

B.3. Catalogo de requisitos

Requisitos funcionales

- **RF-1 Control de usuarios:** la aplicación tiene que tener capacidades para controlar usuarios.
 - **RF-1.1 Registro de usuarios:** el usuario debe poder registrarse en la aplicación con su correo.
 - **RF-1.2 Login:** el usuario debe poder entrar a la aplicación una vez esté registrado.
 - **RF-1.3 Logout:** el usuario debe poder finalizar una sesión.
 - **RF-1.4 Integración con Google:** en caso de que el usuario tenga cuenta de google podrá usarla para hacer el login.

- **RF-2 Uso de aprendizaje automático:** se debe permitir el uso de un sistema de aprendizaje automático basado en una red neuronal.
 - **RF-2.1 Clasificación de imágenes:** el usuario podrá clasificar imágenes individuales.
 - **RF-2.2 Re-entrenamiento del modelo:** se deberá permitir el re-entrenamiento de la red neuronal con distintos conjuntos de datos.
- **RF-3 Manejo de imágenes:** el usuario podrá subir imágenes para ser usadas por los algoritmos de aprendizaje automático.
 - **RF-3.1 Subir imagen individual:** se necesita poder subir imágenes individuales para su posterior clasificación.
 - **RF-3.2 Subir conjunto de imágenes:** se podrá subir un conjunto de imágenes con un formato determinado para poder re-entrenar los modelos.

Requisitos no funcionales

- **RNF-1 Usabilidad:** la aplicación debe ser intuitiva y con una interfaz sencilla.
- **RNF-2 Responsividad:** la interfaz tiene que cambiar de tamaño y dimensiones con respecto a la pantalla y su tamaño.
- **RNF-3 Escalabilidad:** la aplicación debe poder aumentar su rendimiento con el incremento de recursos hardware.
- **RNF-4 Seguridad:** Los datos sensibles, como las contraseñas, deben tratarse de forma segura. También se tendrá que tener las medidas básicas de seguridad contra los ataques a web más comunes.
- **RNF-5 Mantenibilidad:** la aplicación debe ser desarrollada de manera que permita escalabilidad a la hora de añadir características.
- **RNF-6 Internacionalización:** la aplicación deberá estar preparada para soportar varios idiomas.
- **RNF-7 Facilidad de despliegue:** se debe poder desplegar la aplicación en poco tiempo.

B.4. Especificación de requisitos

En esta sección se desarrollarán los casos de uso de la aplicación, cabe destacar que los actores del sistema corresponden a la misma persona en distintos momentos de su uso de la aplicación.

CU-01	Registro de usuario
Autor	Javier Martínez Riberas
Requisitos asociados	RF-1.1

CU-01	Registro de usuario
Descripción	Permite al usuario registrarse.
Precondición	Ninguna.
Acciones	<ol style="list-style-type: none"> 1. El usuario rellena el formulario de registro (email y contraseña). 2. Si esta configurada la confirmación por email se necesita confirmar para terminar el registro.
Postcondición	El usuario se registra en la base de datos.
Excepciones	<ul style="list-style-type: none"> ■ Fallo de la configuración de email. (Si no hay no se da)
Importancia	Alta

Tabla B.1: CU-01 Registro de usuario.

CU-02	Login
Autor	Javier Martínez Riberas
Requisitos asociados	RF-1.2, RF-1.4
Descripción	Permite al usuario autenticarse.
Precondición	El usuario debe de estar registrado.
Acciones	<ol style="list-style-type: none"> 1. El usuario rellena el formulario de login (email y contraseña). O el equivalente en google.
Postcondición	El usuario consigue una sesión.
Excepciones	<ul style="list-style-type: none"> ■ Fallo de rellenado del formulario.
Importancia	Alta

Tabla B.2: CU-02 Login.

CU-03	Logout
Autor	Javier Martínez Riberas
Requisitos asociados	RF-1.3
Descripción	El usuario se des-autentica.
Precondición	El usuario debe de estar autenticado.
Acciones	<ol style="list-style-type: none"> 1. El usuario hace click el botón de logout.
Postcondición	El usuario cierra una sesión, la cual se elimina.
Excepciones	Ninguna.
Importancia	Alta

Tabla B.3: CU-03 Logout.

CU-04	Clasificar imagen
Autor	Javier Martínez Riberas
Requisitos asociados	RF-2.1, RF-3.1
Descripción	Permite al usuario obtener la clase de una imagen.
Precondición	El usuario debe de estar autenticado.
Acciones	<ol style="list-style-type: none"> 1. El usuario sube una foto. 2. Opcional. El usuario elige un re-entrenamiento específico.
Postcondición	El usuario ve la clase más probable y otras dos probables.
Excepciones	<ul style="list-style-type: none"> ■ Fallo de formato de imagen.
Importancia	Alta

Tabla B.4: CU-04 Clasificación de imagen.

CU-05	Subida de datasets
Autor	Javier Martínez Riberas

CU-05	Subida de datasets
Requisitos asociados	RF-3.2
Descripción	Permite al usuario subir un conjunto de imágenes para re-entrenar la red neuronal.
Precondición	El usuario debe de estar autenticado.
Acciones	<ol style="list-style-type: none"> 1. El usuario elige un dataset en su ordenador para subir. 2. El dataset debe tener una estructura de carpetas equivalentes a las clases. 3. Cada carpeta equivalente a una clase debe contener imágenes en jpg de la clase en cuestión.
Postcondición	El usuario tras un periodo de tiempo tiene disponible un dataset para re-entrenar el modelo.
Excepciones	<ul style="list-style-type: none"> ■ Fallo a la hora de subir el zip.
Importancia	Media-Alta

Tabla B.5: CU-05 Subida de datasets.

CU-06	Re-entrenado de modelos
Autor	Javier Martínez Riberas
Requisitos asociados	RF-2.2, RF-3.2
Descripción	Permite al usuario re-entrenar la red neuronal.
Precondición	El usuario debe de estar autenticado y haber subido un dataset.
Acciones	<ol style="list-style-type: none"> 1. El usuario elige un dataset. 2. El usuario debe esperar una cantidad decente de tiempo (mínimo media hora) y tendrá el modelo disponible.
Postcondición	El usuario tras un periodo de tiempo tiene disponible un modelo.

CU-06	Re-entrenado de modelos
Excepciones	<ul style="list-style-type: none">Fallo del formato de las imágenes o de la estructura de archivos.
Importancia	Media-Alta

Tabla B.6: CU-06 Re-entrenado de modelos.

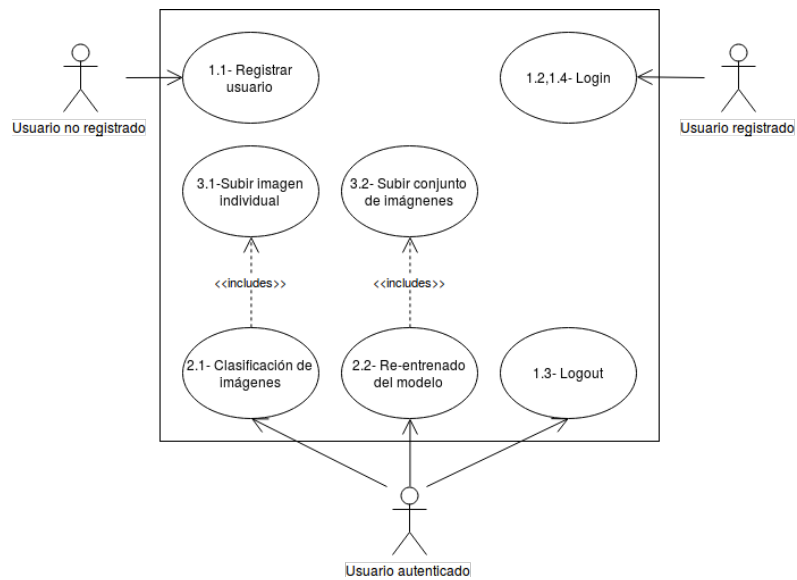


Figura B.1: Diagrama de casos de uso

Por último se expone el diagrama de casos de uso como resumen de la especificación de requisitos.

Especificación de diseño

C.1. Introducción

En este anexo se expone el diseño que se ha usado para llevar a cabo los objetivos anteriores. Como se manejan los datos, la arquitectura y modelos.

C.2. Diseño de datos

La aplicación cuenta con el modelado de los siguientes:

- **Usuario:** la entidad del usuario es una entidad simple la cual tiene un id auto-generado para poder identificar a cualquier usuario. También dispone de email, contraseña y confirmación del email. Por último dispone de un campo para posibles *tokens* de *oauth* de manera que se puedan añadir distintas maneras de registrar usuarios sin tener que cambiar constantemente el modelo de la base de datos.
- **Red neuronal:** la red neuronal solo tiene lo que podríamos considerar *primary key* en una base de datos y el objeto de bytes. Su identificador se compone de el nombre del usuario y de el nombre del conjunto de datos que se ha usado para entrenarlo.
- **Conjunto de datos:** el conjunto de datos se identifica mediante su nombre y del usuario que lo ha subido. Tiene un objeto de bytes asociado que guardamos debido a la lentitud del entrenamiento de una red neuronal y si este entrenamiento se interrumpe no queremos repetir la subida del conjunto de datos al servidor.

Dentro de esta configuración tentemos que tener en cuenta que sólo el modelo del usuario esta dentro de la base de datos de manera que el resto se controla por software y se guarda dentro de los volúmenes de *docker*

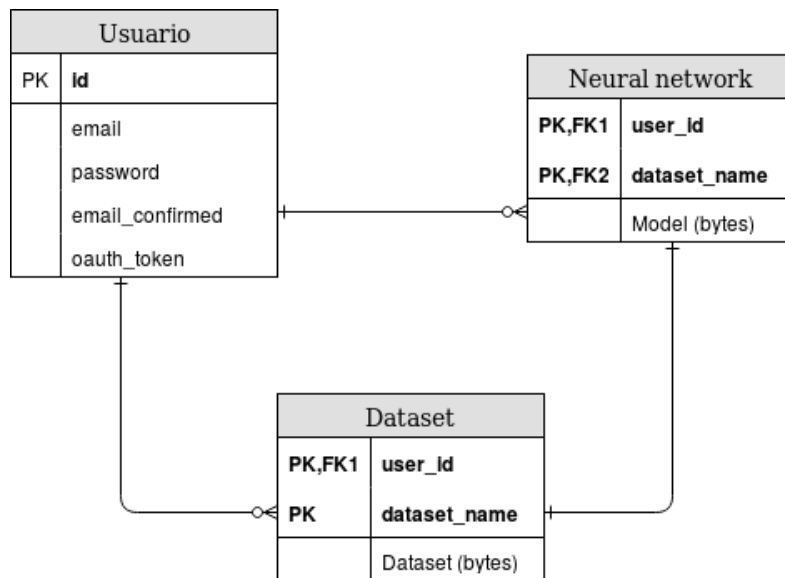


Figura C.1: Diagrama entidad relación

Paso de datos

El paso de datos entre los servicios se hace a través de SSH, esto se hace ya que nos proporciona seguridad. Otra opción es tener micro servidores de flask para exponer una *API Rest* que nos permita hacer las operaciones de una forma más elegante. Esto dificulta algo la configuración interna poniendo más peso en el equipo de *devops* pero nos permite tener un sistema mucho más desacoplado lo cual facilita el mantenimiento de ambas partes por separado. Si se quisiera tener seguridad tras ese cambio las comunicaciones se deberían hacer con SSL.

C.3. Diseño arquitectónico

La decisión de proporcionar el servicio como una página web ha condicionado la arquitectura. Al buscar la escalabilidad dentro del diseño se ha tenido que tener en cuenta para el diseño final. A continuación se exponen las partes más conocidas del diseño y el resultado final.

Model View Controller (MVC)

La parte de la aplicación tiene algo más de diseño en su interior. Para facilitar el desarrollo de futuros desarrolladores se ha decidido seguir la arquitectura ampliamente conocida como Modelo Vista Controlador o MVC.

De esta manera logramos separar la vista con la cual interaccionará el usuario, con la lógica de interacción con la aplicación y del modelo de nuestro

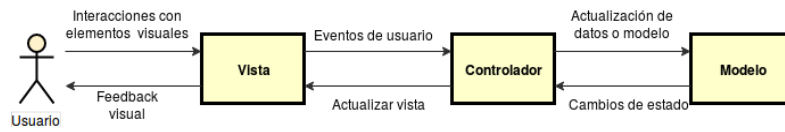


Figura C.2: Model view controller

negocio. Gracias a esta abstracción podemos ocultar toda la complejidad de cada capa a los desarrolladores de las otras dos en caso de que nuestro negocio crezca hasta ese punto.

En nuestra estructura de archivos las vistas están bajo `WhatAClass/WhatAClass/templates`, los controladores se encuentran en `WhatAClass/WhatAClass/blueprints` y el único modelo explícito es el de usuario, situado en `WhatAClass/WhatAClass/models`, el resto son estructuras lógicas sobre el sistema de archivos.

Builder

En la aplicación que hemos creado se ha usado un builder para permitir la creación de varias instancias de la aplicación de manera sencilla, ya que la aplicación es un agregado de varios compuestos. Además permite que se pueda escalar verticalmente (añadiendo más recursos en la misma máquina).

Esto también facilita la creación de la aplicación para programadores nuevos que desconozcan nuestro proyecto. Esto se debe a que los componentes están bien establecidos y un programador que no conozca toda la aplicación puede solo añadir un par de líneas en el método que corresponda, si las necesidades cambian mucho entre varios despliegues, lo suficiente para que no solo se tengan que hacer un par de cambios triviales, podríamos mantener varios métodos builder.

Iterator

Aunque no se ha implementado explícitamente python permite la iteración sobre elementos de una colección de manera transparente. Esto podría no considerarse como un patrón de diseño ya que está incorporado en el lenguaje.

Null Object

El objeto de comunicación con el servidor de E-mail hace la función de un objeto nulo cuando esta funcionalidad no está habilitada. Esto permite un cambio de comportamiento de manera simple sin condicionales excesivas en sitios inadecuados para tratar con el cambio en el comportamiento al estar habilitada la funcionalidad.

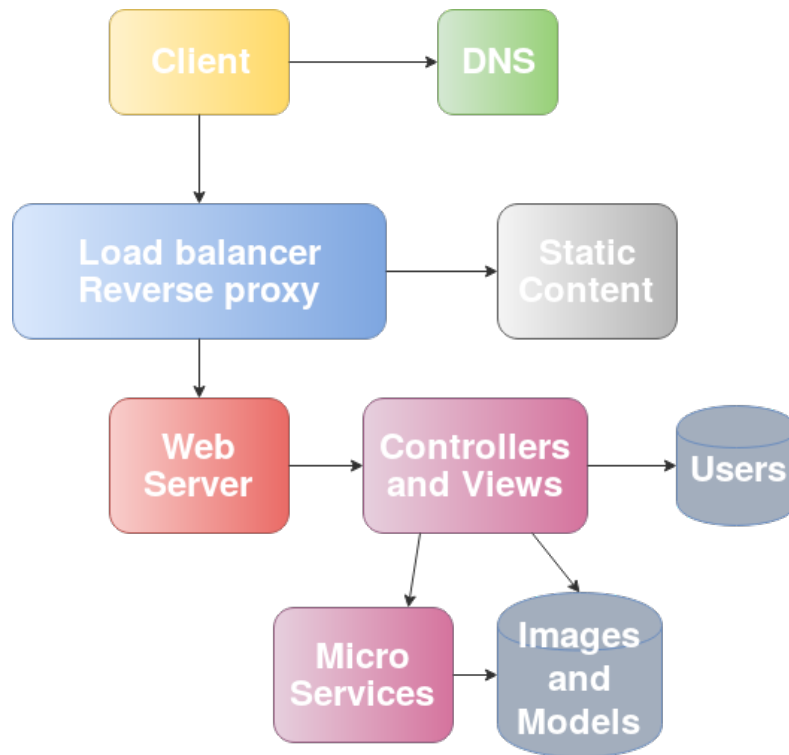


Figura C.3: Arquitectura de la aplicación

Arquitectura final

El diseño arquitectónico final es uno muy parecido a los que se usan en muchas páginas web para permitir mayor escalabilidad y picos de servicio sin caída del mismo. Simplemente se basa en permitir ejecutar múltiples instancias de los mismos objetos. Se podría ampliar la escalabilidad mediante *reverse proxies* en distintos puntos de la arquitectura y con un servicio de caché como *Redis* antes de las bases de datos para permitir mayor velocidad de acceso. Si quisiésemos replicar micro servicios deberíamos poner un *load balancer* o unirlos desde el que ya está.

C.4. Diseño procedimental

En este apartado se expondrá que flujo sigue la información de forma general cuando se usa la aplicación.

Como se puede observar la secuencia pasa por nginx para poder servir archivos estáticos a gran velocidad y para poder si se quisiese balancear la carga entre varios servidores de aplicación (uWSGI).

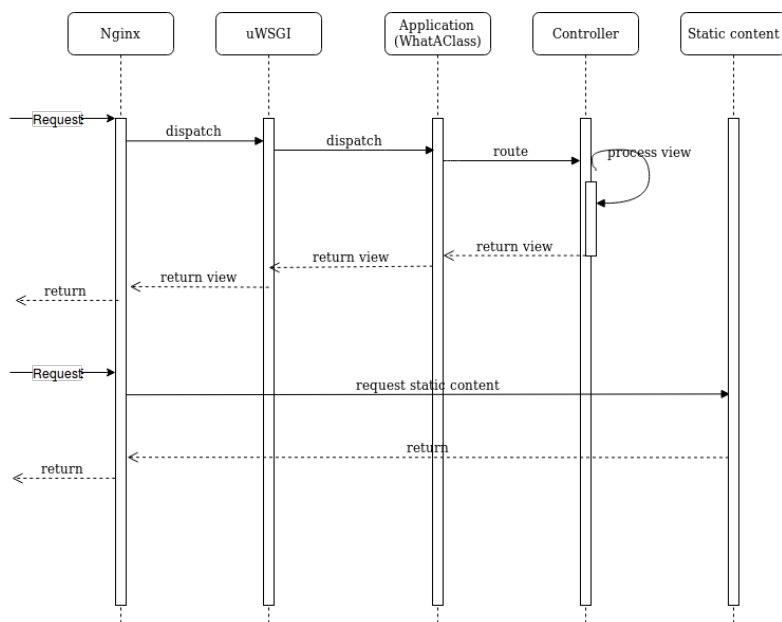


Figura C.4: Diagrama de secuencia

C.5. Diseño de paquetes

El diseño de la organización de las diferentes partes de la aplicación en paquetes, tradicionalmente, se puede hacer de dos maneras: descomposición funcional o divisional.

A continuación se pone un ejemplo de cada descomposición.

Descomposición funcional

```

yourapp/
  __init__.py
  static/
  templates/
    home/
    control_panel/
  views/
    __init__.py
    home.py
    control_panel.py
  models.py
  
```

Descomposición divisional

```
yourapp/  
  __init__.py  
  home/  
    __init__.py  
    views.py  
    static/  
    templates/  
  control_panel/  
    __init__.py  
    views.py  
    static/  
    templates/  
  models.py
```

Descomposición de la aplicación

La que se ha elegido en el proyecto es la funcional.

Como se puede ver, python no necesita una clase para guardar código, esto resulta en un diagrama de clases más simple que en otros lenguajes como java.

```
WhatAClass/  
  blueprints/  
    oauth/  
      __init__.py  
      google.py  
    __init__.py  
    index_blue.py  
    tensorflow_mng_blue.py  
    user_mng_blue.py  
  static/  
    css/  
    fonts/  
    js/  
    favicon.ico  
  templates/  
    tensorflow_mng/  
      predict.html  
      predicted.html  
      retrain.html  
      upload_ds.html  
    user_mng/
```

```
email/
  activate.html
  recover.html
login.html
other_logins.html
recover.html
reset.html
signup.html
error.html
index.html
js_upload.html
layout.html
macros.html
translations/
utils/
  __init__.py
  email.py
  __init__.py
app.py
controllers.py
extensions.py
forms.py
models.py
util.py
```

Aplicación

Como se puede ver la aplicación esta descompuesta en el builder (“app.py”), extensiones, formularios(“templates”), modelo, utilidades y controladores.

Los controladores están en la carpeta blueprints pero se incluyen dentro de “controllers.py” para reducir el acoplamiento.

Las utilidades tienen una configuración parecida a los controladores, las utilidades están en el “utils” y se importan con “util.py”.

Configuración

En la jerarquía de carpetas superior al código de la aplicación esta la carpeta “config” usada por el patrón builder para generar la aplicación.

En este momento solo se dispone de una configuración (default) esto se debe a que gracias al uso de *docker* y *docker compose* podemos cambiar la configuración mediante las variables de entorno que estos pueden modificar, esto simplifica las configuraciones situacionales, ya que de esta manera toda la

configuración que debamos cambiar podemos cambiarla en el archivo “docker-compose.yml”.

Tensorflow

En el microservicio de tensorflow no necesitamos una estructura compleja ya que la biblioteca es suficientemente completa como para poder usarla con scripts sencillos.

Apéndice *D*

Documentación técnica de programación

D.1. Introducción

El proyecto al estar descompuesto en microservicios tiene varias partes bien diferenciadas. Esto se refleja en todas las partes del proyecto.

D.2. Estructura de directorios

Aplicación web

La aplicación web tiene una estructura de directorios algo compleja, por lo que algunos elementos se han omitido del esquema.

WhatAClass/	<i>Principal localización para nuestro código fuente</i>
blueprints/	<i>Localización de los controladores</i>
oauth/	<i>Controladores de OAuth</i>
static/	<i>Archivos estáticos (bootstrap y favicon)</i>
templates/	<i>Plantillas de Jinja2 para generar html o javascript</i>
tensorflow_mng/	<i>Plantillas para tensorflow</i>
user_mng/	<i>Plantillas para el control de usuarios</i>
email/	<i>Plantillas para los email</i>
translations/	<i>Mensajes traducidos (compilación)</i>
utils/	<i>Utilidades</i>
email.py	<i>Utilidad de email</i>
app.py	<i>Builder</i>

APÉNDICE D. DOCUMENTACIÓN TÉCNICA DE PROGRAMACIÓN²³

extensions.py	<i>Extensiones de flask</i>
forms.py	<i>Formularios</i>
models.py	<i>Modelos</i>
Alembic/	<i>Versionado de la DB</i>
versions/	<i>Scripts de migración entre versiones de la DB</i>
babel/	<i>Mensajes con sus traducciones</i>
config/	<i>Configuración de la aplicación</i>
default.py	<i>Configuración por defecto</i>
docs/	<i>Documentación autogenerada</i>
report/	<i>Documentación manual</i>
resources/	<i>Recursos como claves rsa</i>
tests/	<i>Tests de la aplicación</i>
.coveragerc	<i>Que código se ignora por el cubrimiento</i>
.dockerignore	<i>Que ignora docker</i>
.gitignore	<i>Que ignora git</i>
.travis.yml	<i>Configuración de travis</i>
babel.cfg	<i>Configuración de babel (internacionalización)</i>
create_db.py	<i>Script de creación de la DB a partir del modelo</i>
docker-compose.yml	<i>Configuración de docker compose (solo webapp)</i>
Dockerfile	<i>Descripción de construcción del contenedor</i>
LICENSE	<i>Licencia</i>
messages.pot	<i>Mensajes traducidos</i>
Procfile	<i>Comandos para heroku</i>
README.md	<i>Readme para Github</i>
requirements.txt	<i>Requisitos para desarrollador y para VersionEye</i>
requirements-prod.txt	<i>Requisitos para despliegue</i>
run.py	<i>Punto de entrada para debug</i>
runtime.txt	<i>Establece la versión de python para heroku</i>
setup.cfg	<i>Alias para test</i>
setup.py	<i>Instalación de las dependencias</i>
start.py	<i>Punto de entrada sin configuración adicional</i>
test.py	<i>Punto de entrada para los test</i>
uwsgi.ini	<i>Configuración para que uwsgi encuentra la aplicación</i>
wsgi.py	<i>Punto de entrada para uwsgi</i>

Microservicio de tensorflow

dataset/	<i>Estructura de archivos en producción para facilitar los test</i>
graphs/	<i>Estructura de archivos de producción para los test</i>
images/	<i>Estructura de archivos de producción para los test</i>
test/	<i>Carpeta donde se guardan los test</i>

Dockerfile	<i>Descripción de construcción del contenedor</i>
fit_dataset.py	<i>Script para ajustar el modelo a un dataset</i>
id_rsa.pub	<i>Clave rsa para comunicación con la aplicación web</i>
inceptionv3*	<i>Modelo y clases</i>
maybe_start_retrain.py	<i>Script que comprueba si la petición ya se esta ejecutando o no</i>
run_inference.py	<i>Script para inferir la clase de una imagen</i>

Docker compose

Hay dos archivos que se han hecho fuera de las dos anteriores y estos son los docker-compose.yml que coordinan ambos servicios.

Uno esta en la raíz de ambas estructuras, este hace la construcción del servicio contando con los fuentes.

El otro que esta en docker-compose-hub/docker-compose.yml, y este usa los contenedores disponibles en dockerhub en vez de construirlos desde los fuentes.

D.3. Manual del programador

En la siguiente sección se expondrá como se puede continuar desarrollando la aplicación. Primero preparando el entorno de desarrollo y el código del proyecto, cómo ejecutar tanto el proyecto como los test y cómo servir el proyecto.

Entorno de desarrollo

Los siguientes elementos son necesarios para trabajar en el proyecto de forma efectiva:

- Python 3.5 o 3.6
- Git
- Gestor de paquetes
- Entorno virtual
- IDE o editor de texto
- Docker

Python

El lenguaje de programación más popular en este momento, tanto para web, como para aplicaciones de empresa es Python [3].

El proyecto se inició con Python 3.5 aunque se ha mantenido la compatibilidad con Python 3.6. Debido a algunos cambios que se han hecho en

Python 3.6 con los diccionarios en Python se recomienda esta versión, ya que proporciona un aumento de rendimiento.

Se puede obtener esta versión (3.6) en [?]. Hay que elegir correctamente la versión y sistema operativo. Después de seguir el manual de instalación tendremos Python instalado.

Git

La forma más simple y útil de descargar y trabajar con el código fuente es a través de un sistema de control de versiones como es Git.

Este programa permite movernos a través de la evolución que ha tenido el código a lo largo del tiempo.

Una vez lo tengamos instalado tenemos dos posibilidades: usar un soporte con GUI o usar un soporte en línea de comandos, por supuesto hay algunos programas que lo instalan automáticamente.

De la primera opción están el proporcionado por Git en algunas instalaciones, los plugin o utilidades para IDEs o un programa especializado como GitKraken.

La segunda opción es la que se recomienda porque ayuda a comprender el funcionamiento de Git de una manera más clara.

Obtención código fuente (Clonación del repositorio)

Con cualquier herramienta básica de git se puede clonar un repositorio, en algunas la opción se llama importar. En esta sección se expondrá cómo hacerlo en línea de comandos (en windows git bash). Primero abriremos la terminal o su equivalente.

```
$ mkdir <DIR>
$ cd <DIR>
$ git clone https://github.com/Jazriel/tfg
$ cd tfg
$ git clone https://github.com/Jazriel/WhatAClass
```

Gestor de paquetes

El gestor de paquetes es una utilidad considerada necesaria para facilitar la instalación de paquetes (en este contexto significa un conjunto de software), ya que no solo instalamos el código fuente que pedimos sino que también instalamos las dependencias de dicho código.

Las principales herramientas son pip, setuptools y wheel

Si hemos instalado Python 2 con la versión 2.7.9+ o Python 3 con la versión 3.4+ ya tienen instalado pip y setuptools.

```

Terminal - jack@jack-ISM: ~/dir/tfg/WhatAClass
File Edit View Terminal Tabs Help
-rw-r--r-- 1 jack jack 14 dic 12 2016 .xscreensaver
-rw----- 1 jack jack 82 ago 19 03:25 .xsession-errors
-rw----- 1 jack jack 272 ago 18 20:20 .xsession-errors.old
jack@jack-ISM:~$ cd algo
bash: cd: algo: No such file or directory
jack@jack-ISM:~$ mkdir dir
jack@jack-ISM:~$ cd dir
jack@jack-ISM:~/dir$ git clone https://github.com/Jazriel/tfg
Cloning into 'tfg'...
remote: Counting objects: 51, done.
remote: Total 51 (delta 0), reused 0 (delta 0), pack-reused 51
Unpacking objects: 100% (51/51), done.
Checking connectivity... done.
jack@jack-ISM:~/dir$ cd tfg/WhatAClass
jack@jack-ISM:~/dir/tfg/WhatAClass$ git clone https://github.com/Jazriel/WhatAClass
Cloning into 'WhatAClass'...
remote: Counting objects: 1158, done.
remote: Compressing objects: 100% (42/42), done.
remote: Total 1158 (delta 2), reused 22 (delta 2), pack-reused 1114
Receiving objects: 100% (1158/1158), 8.41 MiB | 2.81 MiB/s, done.
Resolving deltas: 100% (662/662), done.
Checking connectivity... done.
jack@jack-ISM:~/dir/tfg/WhatAClass$

```

Figura D.1: Clonación de los repositorios

En linux, ya que el sistema operativo tiene muchas veces tanto la versión 2.x y 3.x de Python para usar pip tendremos que usar pip2 o pip3 para saber que versión específica se esta usando.

Uso básico de pip

Para instalar un proyecto específico se hace con:

```
$ pip install <PROJECT>
```

Algunas opciones adicionales son las versiones específicas o compatibles:

```
$ pip install <PROJECT>==<SPECIFIC>
$ pip install <PROJECT>~=<COMPATIBLE>
```

Para actualizar un proyecto:

```
$ pip install --upgrade <PROJECT>
```

Por último cabe destacar que para un conjunto de dependencias la manera más simple de instalar todas es tenerlas en un archivo requirements.txt, para instalar a partir de ese archivo se usa el siguiente comando:

```
$ pip install -r requirements.txt
```

Para obtener este archivo el desarrollador anterior debe haber generado ese archivo usando pip, esto descargará las dependencias que tiene el entorno desde el que se ha generado.

```
$ pip freeze
```

Entorno virtual

Usar un entorno virtual no es una necesidad a la hora de desarrollar programas en Python pero se recomienda encarecidamente debido a que eso nos permite encapsular nuestras dependencias. La excepción a la regla anterior es el caso de desarrollar dos programas con dependencias a las mismas bibliotecas en diferentes versiones.

Algunos de los gestores de entornos virtuales más usados son venv, virtualenv y conda. Hay que tener en cuenta que los IDEs necesitan configuración adicional para trabajar con entornos virtuales.

Las características principales de las opciones anteriores son:

- **Venv**: viene incluido en la biblioteca standard de Python, no soporta Python 2.x.
- **Virtualenv**: es la biblioteca más recomendada por los principales tutoriales de Python y se encuentran disponibles envoltorios que facilitan o implementan funcionalidades adicionales, soporta Python 2.6+ y 3.3+.
- **Conda**: no solo es un gestor de entornos virtuales sino que también gestiona paquetes, soporta también 2.6+ y 3.3+.

Creación de un entorno virtual

Para este ejemplo pondremos como ejemplos a venv y virtualenv. Las palabras que están entre ‘<’ y ‘>’ son variables.

Venv

Para crear el entorno:

```
$ python3 -m venv <DIR>
```

Para activarlo:

```
$ source <DIR>/bin/activate
```

Para desactivarlo:

```
$ deactivate
```

Virtualenv

Virtualenv al no estar incluido en Python deberemos instalarlo, para ello se recomienda tener un gestor de paquetes como pip instalado antes.

Para instalar virtualenv valdrá con:

```
$ pip install virtualenv
```

```

Terminal - jack@jack-ISM: ~/dir/tfg/WhatAClass
File Edit View Terminal Tabs Help
jack@jack-ISM:~/dir/tfg$ cd WhatAClass/
jack@jack-ISM:~/dir/tfg/WhatAClass$ virtualenv -p /usr/bin/python3.5 venv
Running virtualenv with interpreter /usr/bin/python3.5
Using base prefix '/usr'
New python executable in /home/jack/dir/tfg/WhatAClass/venv/bin/python3.5
Also creating executable in /home/jack/dir/tfg/WhatAClass/venv/bin/python
Installing setuptools, pip, wheel...done.
jack@jack-ISM:~/dir/tfg/WhatAClass$ source venv/bin/activate
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass$ pip install -r requirements.txt
Collecting alembic==0.9.5 (from -r requirements.txt (line 1))
  Downloading alembic-0.9.5.tar.gz (990kB)
    100% |#####| 993kB 1.0MB/s
Collecting bcrypt==3.1.3 (from -r requirements.txt (line 2))
  Downloading bcrypt-3.1.3-cp35-cp35m-manylinux1_x86_64.whl (54kB)
    100% |#####| 61kB 3.0MB/s
Collecting codeclimate-test-reporter==0.2.3 (from -r requirements.txt (line 3))
  Downloading codeclimate-test-reporter-0.2.3.tar.gz
Collecting coverage==4.4.1 (from -r requirements.txt (line 4))
  Downloading coverage-4.4.1-cp35-cp35m-manylinux1_x86_64.whl (196kB)
    100% |#####| 204kB 4.4MB/s

```

Figura D.2: Creación de entorno virtual e instalación de dependencias

Para crear el entorno:

```
$ virtualenv -p /usr/bin/python<VERSION> <DIR>
```

Para activarlo: `source <DIR>/bin/activate`

```
$ source <DIR>/bin/activate
```

Para desactivarlo:

```
$ deactivate
```

Entorno virtual e instalación de las dependencias

Para este ejemplo crearemos un entorno virtual e instalaremos las dependencias de la parte de la aplicación web. Se continuará desde el punto donde lo hemos dejado en la parte de obtención del código fuente.

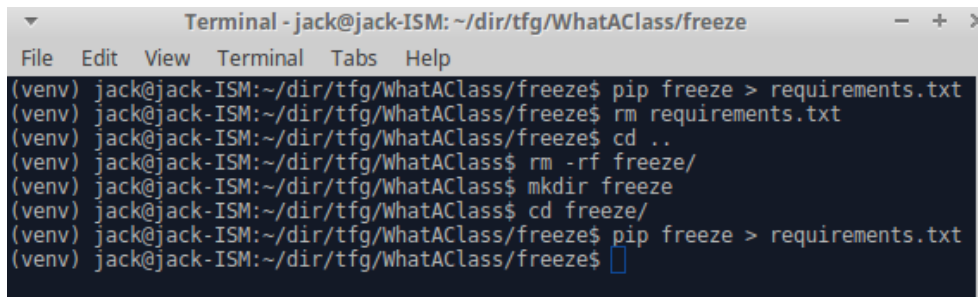
```

$ cd WhatAClass
$ virtualenv -p /usr/bin/python<VERSION> <DIR>
$ source <DIR>/bin/activate
$ pip install -r requirements.txt

```

Otra forma de instalar las dependencias es usando el archivo ‘setup.py’, este archivo esta pensado para poder usar pip como sistema de distribución del proyecto. No se ha visto necesario el subir el archivo a pip ya que no proporciona un servicio directo, sino que ayuda a proporcionar servicios.

Para comprobar como funciona pip freeze para la actualización de dependencias, se expondrá a continuación un ejemplo ejecutado como continuación del anterior.



```

Terminal - jack@jack-ISM: ~/dir/tfg/WhatAClass/freeze
File Edit View Terminal Tabs Help
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass/freeze$ pip freeze > requirements.txt
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass/freeze$ rm requirements.txt
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass$ rm -rf freeze/
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass$ mkdir freeze
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass$ cd freeze/
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass/freeze$ pip freeze > requirements.txt
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass/freeze$

```

Figura D.3: Prueba de uso de pip freeze

```

$ mkdir freeze
$ cd freeze
$ pip freeze > requirements.txt

```

Para actualizar las dependencias, como ya se ha comentado, solo debemos ejecutar este último comando y luego transcribirlo a ‘setup.py’.

IDE o editor de texto

Se recomienda usar un IDE aunque un editor de texto y una línea de comandos pueden hacer la misma función. En caso de usar un IDE hay que buscar la configuración para que pueda trabajar con el entorno virtual que hayamos creado o vayamos a crear.

Si no estamos usando un IDE simplemente debemos entrar en el entorno virtual antes de hacer nada relacionado con la ejecución del proyecto.

Importar el proyecto desde Pycharm

Por la separación del proyecto en microservicios se recomienda el no trabajar con un solo proyecto, se recomienda importar cada servicio en un proyecto y así mantenerlos separados, esto facilita el uso de entornos virtuales.

1. Abrir PyCharm
2. Desde la ventana de bienvenida, hacer click en abrir y abrir la localización tfg/WhatAClass
3. Esperar a que PyCharm lea todos los archivos
4. Ir abriendo los cuadros de dialogo y comprobando su importancia

PyCharm ya está preparado para trabajar con git, pip, entornos virtuales y docker, pero muchas veces hace falta configurar estas herramientas. Para configurarlas los cuadros de dialogo son muy útiles ya que proporcionan una buena cantidad de ayuda, facilitando la instalación de dependencias no instaladas, permitiendo tener docker como opción de ejecución predeterminada, llevando el control de versiones por ti y avisando si algo no sale como se espera.

Por último tenemos que añadir como *script* de depuración ‘run.py’ y como *script* de prueba ‘test.py’ o la carpeta de ‘test/’ como objetivo.

Añadir características

Para añadir características al programa debemos, en primer momento plantearnos que funcionalidad proveen, si cuadra dentro de una de las categorías que tenemos definidas (home, control de usuarios y tensorflow) o no.

Si cuadra con alguna de las categorías definidas nos será suficiente con añadir más rutas disponibles dentro de la blueprint de esa sección. Si no cuadra, lo que tendremos que hacer será crear esa blueprint que vayamos a implementar.

Sea cual sea la opción que elijamos acabaremos creando funciones decoradas con ‘@route’ o ‘@blueprint.route’. Estas funciones serán nuestros controladores, lo que devuelvan será lo que vea el cliente, principalmente deberíamos devolver la función ‘render_template’ invocada con la plantilla correspondiente. Para la internacionalización debemos importar de ‘flask_babel’ la función ‘gettext’, normalmente se importa como ‘_’ de manera que el texto internacionalizable pasa antes por esa función.

Como ejemplo de la mayoría de conocimientos requeridos para implementar una funcionalidad nueva se expone la función de login, excesivamente comentada, a continuación.

```

1  from flask import (Blueprint, flash, render_template,
2                        url_for, redirect)
3  from flask_login import current_user
4
5  from flask_babel import gettext as _
6
7  from WhatAClass.forms import LoginForm
8  from WhatAClass.models import User
9
10 user_mng = Blueprint('user_mng', __name__)
11
12 # El decorador que da una ruta de acceso
13 @user_mng.route('/login', methods=['GET', 'POST'])
14 def login():
15
16     # Ruta desde la carpeta templates hasta el archivo
17     # de la vista
18     view = 'user_mng/login.html'
19
20     # Usuario actual
21     if current_user.is_authenticated:
22         # Función _ para permitir internacionalización

```

```

23         flash(_('There is a logged in user already.'))
24         return redirect(url_for('index.base'))
25
26     form = LoginForm()
27
28     # Si el formulario esta siendo subido (POST)
29     if form.validate_on_submit():
30
31         # Hacemos una petición a la base de datos
32         user = User.query.filter_by(
33             email=form.email.data).first()
34
35         # Intentamos hacer login al usuario
36         if not check_and_login(
37             form.password.data, user):
38
39             return render_template(view, form=form)
40
41         flash(_('Logged in successfully.'))
42         return redirect(url_for('index.base'))
43
44     # Salida de una petición GET
45     return render_template(view, form=form)

```

Como se puede ver, lo que en otras tecnologías es una clase, lo que aquí es la *blueprint*, esto hace que no sean tan útiles los diagramas uml.

Si se dese a usar git para integrar el cambio, hacer un fork del proyecto y luego una pull request sería lo más adecuado.

Compilación

Python es un lenguaje interpretado y no necesita compilación previa, pero algunas partes de nuestro proyecto sí que pueden necesitar ser compiladas. De momento solo las traducciones necesitan ser compiladas.

Babel

Para traducir con Babel, como ya se ha introducido anteriormente, usaremos gettext y `_`. Una vez todo nuestro texto este en llamadas a esas funciones solo nos queda extraerlo para traducirlo, generar el catálogo del idioma y compilar ese catálogo para que sea de uso más eficiente.

A continuación se expone como se lleva a cabo este proceso:

```
$ pybabel extract -F babel.cfg -o messages.pot WhatAClass/
```

Este primer comando lo que lleva a cabo es la extracción de todos los mensajes a ser traducidos al archivo ‘messages.pot’. Debemos abrir ese archivo con un editor de texto o nuestra IDE para añadir las traducciones correspondientes. Una vez estén añadidas es hora de generar el catálogo del idioma correspondiente.

```
$ pybabel init -i messages.pot
                  -d WhatAClass/translations
                  -l es
```

Con esto estamos especificando que la entrada es messages.pot, la salida estará en WhatAClass/translations y el lenguaje es el español, para otros idiomas habría que buscar sus extensiones. Por último se compilan las traducciones para que babel pueda usarlas de manera eficiente

```
$ pybabel compile -d WhatAClass/translations
```

Docker

Como ya se ha hablado Docker se usa para facilitar el despliegue en cualquier servidor que lo tenga instalado. Para poder disfrutar las ventajas que nos da el uso de este servicio de virtualización debemos configurar el contenedor cómo si fuese un sistema operativo normal y corriente. Esto se lleva a cabo en un archivo llamado ‘Dockerfile’.

Cómo no sería rentable invertir el tiempo de instalar un sistema operativo en un contenedor, la opción que nos da Docker es el equivalente en git a hacer un fork, replica todo el sistema que pidamos y a partir de el podemos usar el sistema como queramos. Cómo se hace esto se verá a continuación explicado en las siguientes secciones.

Antes de poder ejecutar un contenedor debemos construirlo, para ello debemos situarnos en la carpeta que contenga nuestro proyecto, tras eso ejecutaremos el comando siguiente:

```
$ docker build . [-t <name>[:<version>]]
```

Si por cualquier razón no estuviese el ‘Dockerfile’ en la carpeta raíz del proyecto deberíamos cambiar el ‘.’ por la ruta hasta el directorio.

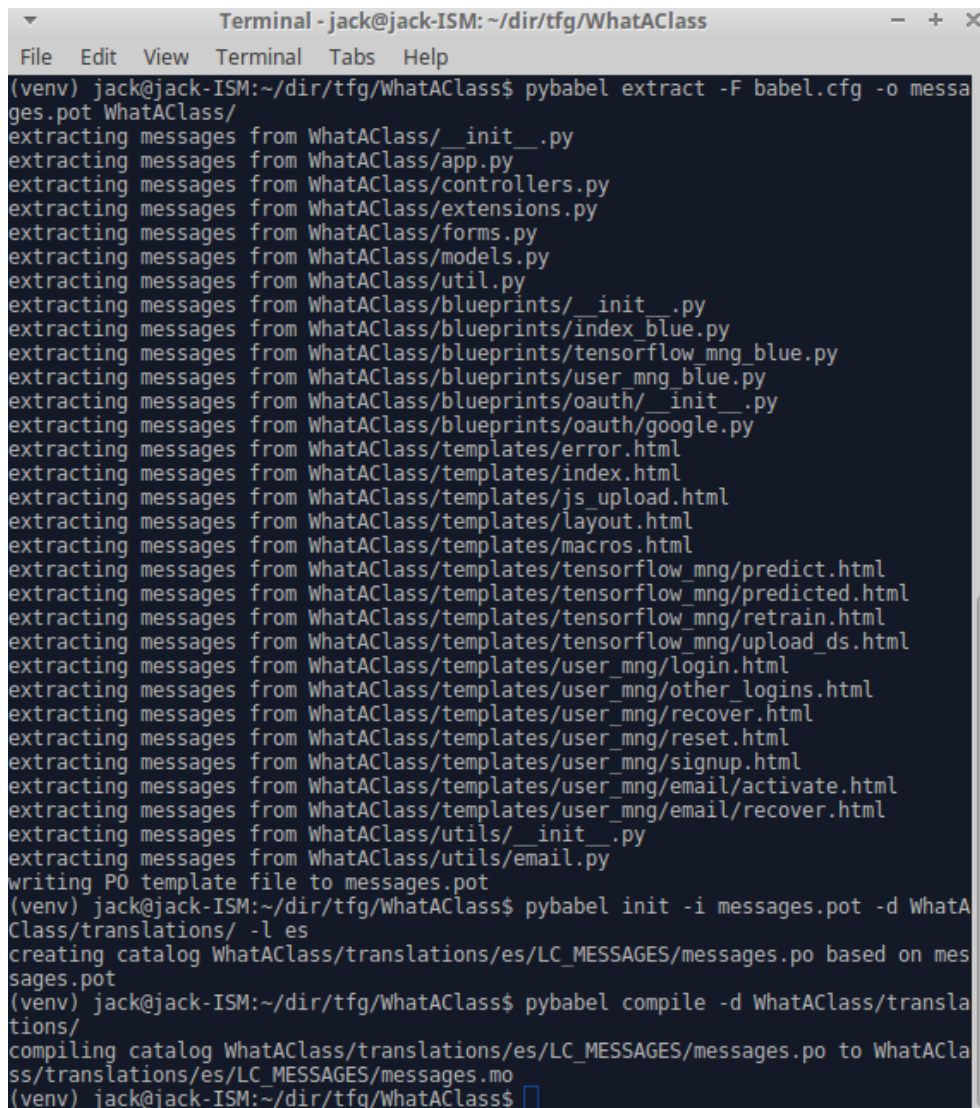
Para lanzar un contenedor y ponerlo en ejecución se puede hacer con el siguiente comando:

```
$ docker run -i -t <image_id>
```

O el siguiente si queremos usar el nombre del contenedor:

```
$ docker run -i -t <image_name>:<version>
```

Si no nos acordásemos de que imágenes o que nombres tienen, ya que es difícil de acordarnos tanto de las ids como de los nombres auto-generados podemos usar:



```

Terminal - jack@jack-ISM: ~/dir/tfg/WhatAClass
File Edit View Terminal Tabs Help
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass$ pybabel extract -F babel.cfg -o messages.pot WhatAClass/
extracting messages from WhatAClass/__init__.py
extracting messages from WhatAClass/app.py
extracting messages from WhatAClass/controllers.py
extracting messages from WhatAClass/extensions.py
extracting messages from WhatAClass/forms.py
extracting messages from WhatAClass/models.py
extracting messages from WhatAClass/util.py
extracting messages from WhatAClass/blueprints/__init__.py
extracting messages from WhatAClass/blueprints/index_blue.py
extracting messages from WhatAClass/blueprints/tensorflow_mng_blue.py
extracting messages from WhatAClass/blueprints/user_mng_blue.py
extracting messages from WhatAClass/blueprints/oauth/__init__.py
extracting messages from WhatAClass/blueprints/oauth/google.py
extracting messages from WhatAClass/templates/error.html
extracting messages from WhatAClass/templates/index.html
extracting messages from WhatAClass/templates/js_upload.html
extracting messages from WhatAClass/templates/layout.html
extracting messages from WhatAClass/templates/macros.html
extracting messages from WhatAClass/templates/tensorflow_mng/predict.html
extracting messages from WhatAClass/templates/tensorflow_mng/predicted.html
extracting messages from WhatAClass/templates/tensorflow_mng/retrain.html
extracting messages from WhatAClass/templates/tensorflow_mng/upload_ds.html
extracting messages from WhatAClass/templates/user_mng/login.html
extracting messages from WhatAClass/templates/user_mng/other_logins.html
extracting messages from WhatAClass/templates/user_mng/recover.html
extracting messages from WhatAClass/templates/user_mng/reset.html
extracting messages from WhatAClass/templates/user_mng/signup.html
extracting messages from WhatAClass/templates/user_mng/email/activate.html
extracting messages from WhatAClass/templates/user_mng/email/recover.html
extracting messages from WhatAClass/utis/__init__.py
extracting messages from WhatAClass/utis/email.py
writing PO template file to messages.pot
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass$ pybabel init -i messages.pot -d WhatAClass/translations/ -l es
creating catalog WhatAClass/translations/es/LC_MESSAGES/messages.po based on messages.pot
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass$ pybabel compile -d WhatAClass/translations/
compiling catalog WhatAClass/translations/es/LC_MESSAGES/messages.po to WhatAClass/translations/es/LC_MESSAGES/messages.mo
(venv) jack@jack-ISM:~/dir/tfg/WhatAClass$

```

Figura D.4: Proceso de traducción

```
$ docker images
```

Para todos estos comandos existen muchas variaciones pero para su uso básico esta información es suficiente.

Aplicación web

En el siguiente ejemplo se muestra un archivo Dockerfile, específicamente el de la aplicación web, lo que hace cada línea se verá comentado en el propio código.

Dado que necesita usar una base de datos que no se lanza en el mismo contenedor no podemos usar este archivo de forma individual. Si quisiésemos poder usarlo de forma individual podríamos configurar un volumen, que es la unidad de almacenamiento persistente de docker, para almacenar una base de datos SQLite, esto se podría añadir en los archivos de configuración para que funcionase.

```
1 FROM tiangolo/uwsgi-nginx-flask:flask-python3.5
2 # De que <usuario>/<imagen>:<version> estamos 'heredando'
3 # Es un link a dockerhub de manera implícita.
4
5 MAINTAINER Javier Martinez "javymartinez@gmail.com"
6 # Quién mantiene la imagen.
7
8 WORKDIR /app
9 # Directorio de trabajo ,
10 # el equivalente a cd /app de linux
11
12 # Copiamos el código fuente
13 COPY ./WhatAClass ./WhatAClass
14
15 # Copiamos la parte estática a una carpeta
16 # especialmente configurada para que nginx
17 # pueda servirlo más rápidamente
18 COPY ./WhatAClass/static ./static
19 # Copy es un comando que copia desde nuestra maquina y
20 # directorio hasta el contenedor y directorio de trabajo.
21
22 RUN mkdir -p ./static/images
23 # Podemos ejecutar comandos del sistema operativo
24 # con run, en este caso creamos un directorio.
25
26 # {...
27 # ...}
28
```

```

29 # Añadimos una variable de estado con env
30 ENV PYTHONPATH ‘$PYTHONPATH:
31     /WhatAClass:
32     /WhatAClass/WhatAClass’
33
34 # Instalamos las dependencias
35 RUN pip3 install —editable .

```

Tensorflow

El contenedor de tensorflow es heredado de una instalación de tensorflow, a partir de ese punto añadimos un daemon de ssh.

En este archivo tenemos un par de comandos distintos a los del otro que explicaremos a continuación.

```

1
2 FROM gcr.io/tensorflow/tensorflow:latest-devel-py3
3 # Imagen de tensorflow y con un link explicito a grc.io
4
5 RUN apt-get update && apt-get install -y openssh-server
6 # Instalar el daemon de ssh
7
8 # {...
9 # ...}
10
11 EXPOSE 22
12 # Exponemos un puerto para que se pueda
13 # acceder a el más facilmente
14
15 CMD ["/usr/sbin/sshd", "-D"]
16 # CMD es parecido a run, pero en algunos casos
17 # simplifica la operación.

```

Docker compose

Para coordinar ambos contenedores en el momento que queramos ejecutar la aplicación debemos usar un servicio de orquestación, como ya estamos usando docker, la opción más fácil y, de momento, profesional es docker compose.

La forma de usar docker compose es con un archivo ‘docker-compose.yml’. En el se especifican todas las configuraciones adicionales que se necesiten. A continuación se mostraran un par de ejemplos del mismo.

```

1 # Versión del archivo docker compose

```

```

2 # para permitir retrocompatibilidad
3 version: '3'
4
5 # Servicios que se desplegarán
6 services:
7   # Nombre arbitrario del servicio
8   web:
9     # Directorio de construcción
10    build: ./
11    # Puertos
12    ports:
13      - "80:80"
14    # Redes internas a las que puede acceder
15    networks:
16      - backend
17    # Servicios de los que depende, no funciona
18    # suficientemente bien.
19    depends_on:
20      - database
21    # Variables de entorno
22    environment:
23      # Como se puede ver tras el @ accedemos a
24      # database como una dirección de red, esto
25      # se debe a que su dirección interna es el
26      # nombre de su servicio.
27      - DATABASE_URL=postgresql://postgres@database #{...}
28    database:
29      # Imagen que usamos ya que no la construimos.
30      image: "postgres"
31      networks:
32        - backend
33 # Redes existentes
34 networks:
35   backend:

```

‘depends_on’ no funciona de forma correcta ya que no espera a ninguna señal del servicio del que depende. Por lo que los servicios no tienen porque estar preparados para recibir clientes, pudiendo causar errores. Esto llega a ser un problema sobre todo en los sistemas que pueden tardar mucho en encenderse como las bases de datos.

Se ha solucionado añadiendo una espera de hasta 25 segundos en la aplicación web para la base de datos.

Otro ejemplo de docker compose algo más complejo, pero casi idéntico:

```

1 version: '3'

```

```

2 services:
3   web:
4     # {...}
5     environment:
6       - DATABASE_URL=postgresql:// # {...}
7       # Se puede cambiar el despliegue de forma simple
8       # mediante variables de entorno, si se mantiene la
9       # configuración.
10      - WORKER_HOST_NAME=worker
11      - WORKER_PORT=22
12      - WORKER_USER=root
13      - WORKER_PASSWORD=screencast
14      - TENSORFLOW=True
15     volumes:
16       # Donde se monta el volumen
17       # <nombre_del_volumen>:<directorio donde se monta>
18       - images:/app/static/images
19
20   database:
21     # {...}
22   worker:
23     # {...}
24 networks:
25   backend:
26 # Volúmenes compartidos
27 volumes:
28 # Volumen con nombre
29 images:

```

Una vez tenemos el archivo ‘docker-compose.yml’ podemos usar los siguientes comandos para hacer la construcción de los contenedores y para desplegarlos:

```

$ docker-compose build
$ docker-compose up

```

Este proceso es lento y a no ser que queramos probar como se integran los servicios no conviene usar este método de depuración.

Integración continua

En el repositorio se han configurado ciertos servicios para facilitar el trabajo sobre el proyecto, de manera que si hay algún tipo de problema sea lo más fácil de ver posible.

A continuación se describe cada servicio y porqué se ha integrado.

Travis CI

Travis CI (Continuos Integration) es una herramienta en la nube que ejecuta tu proyecto como tu le digas, normalmente esto supone la ejecución de test. Es gratuita para proyectos Open Source y funciona automáticamente cada vez que se realiza un *commit*.

La forma de configurar este servicio es con el fichero `‘.travis.yml’`, el siguiente ejemplo ilustra cómo se ha usado en nuestro proyecto en concreto.

```

1 language: python
2 python:
3   - "3.5"
4   - "3.6"
5
6 install: "pip install -r requirements.txt"
7
8 addons:
9   codeclimate:
10    repo_token: d115b35656471f {...}
11
12
13 notifications:
14   slack:
15     rooms:
16       - javiermartineztf:Ij6ci {...} # tools
17
18 script:
19   - coverage run test.py
20
21 after_success:
22   - codeclimate-test-reporter --token d115b3565 {...}

```

Language sirve para decir en que lenguaje estamos trabajando y como trabajamos en python decimos en que versiones se ha de probar que funciona el proyecto.

Install le dice a Travis que debe ejecutar antes de script que normalmente son los test.

Addons son complementos, en este caso el de CodeClimate, para proporcionar el cubrimiento que hacen los test en comparación con nuestro código.

Notifications es a que servicios debe notificar, en este caso notifica a slack.

Script es el comando para ejecutar los test, como se puede ver lo envolvemos en coverage para que también halle en recubrimiento.

After_success es la directiva que dice que se ejecuta si los test pasan, en este caso envía el cubrimiento a CodeClimate.

CodeClimate

CodeClimate es una herramienta que realiza revisiones de código automáticamente, y guarda una serie de historiales a lo largo del tiempo.

Se configura con un archivo `.codeclimate.yml` pero como no se ha necesitado modificar la configuración básica no se ha creado. El sistema de puntuación usado es GPA, más propio de Estados Unidos, que va del 0 al 4, y que en las notas parciales usa un sistema de puntuación con una escala ordinal, A siendo la mejor.

VersionEye

VersionEye es una herramienta de monitorización de versiones, de manera que vigila las dependencias del proyecto y avisa cuando alguna de ellas queda desactualizada o tiene problemas de seguridad.

También lleva a cabo una comprobación de la compatibilidad de las licencias entre la del proyecto propio y la de las dependencias.

Slack

Slack es un servicio de comunicación entre trabajadores, principalmente sirve para comunicación en equipos, pero en este proyecto ha tenido un uso un tanto diferente. Se han configurado el resto de servicios para que reportasen en un canal concreto, de esta manera la información resumida del proyecto estaba disponible en un solo lugar, en vez de tener que estar buscándola entre diferentes servicios.

D.4. Despliegue

Para hacer un despliegue en un lugar en el que no este disponible el código fuente la manera más simple es instalar docker y ejecutar un comando de los dos siguientes:

```
$ wget 'URL' -O docker-compose.yml && docker-compose up
```

```
$ curl 'URL' >> docker-compose.yml && docker-compose up
```

Siendo `'https://raw.githubusercontent.com/Jazriel/TFG/master/docker-compose-hub/docker-compose.yml'` la url.

Si ya hubiésemos el archivo alojado en esa url, bastaría con hacer `'docker-compose up'` en el mismo directorio.

D.5. Pruebas del sistema

Las pruebas se pueden ejecutar con `test.py` que lanza los tests de la carpeta `tests/`. si queremos ver el recubrimiento de el codigo podemos usar una herramienta como `coverage` y ejecutar `test.py` a traves de esa herramienta.

Documentación de usuario

E.1. Introducción

En este anexo mostramos en que sistemas se puede usar la aplicación y bajo qué condiciones de manera que la puedan usar los clientes.

E.2. Requisitos de usuarios

El único requisito que necesita un usuario que quiera usar nuestro sistema es un navegador. Necesitamos *javascript* y *cookies* activados, las *cookies* son para mantener la sesión.

Navegadores con los que se ha comprobado:

1. Firefox 53
2. Chrome 57

E.3. Instalación

Al proporcionar nuestro servicio como una página web no necesitamos que el cliente o usuario instale la aplicación.

Si de todas las maneras se hubiese decidido que cada usuario tiene la aplicación en su ordenador se pueden seguir las siguientes instrucciones.

Primero se debe comprobar que el ordenador permita la virtualización.

E.4. Manual del usuario

El manual de usuario es más simple que otras aplicaciones, esto se debe a que el enfoque del proyecto no ha sido tanto a tener una aplicación con muchas funcionalidades sino a cómo hacerla.

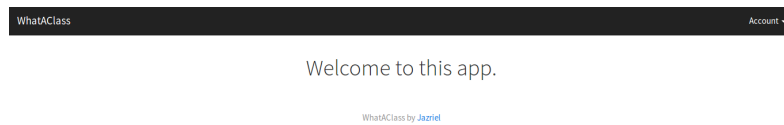


Figura E.1: Punto de entrada a la aplicación

The image shows a registration form titled 'Sign up!'. It has two input fields: 'Email' with the placeholder 'email@company.domain' and 'Password'. Below the fields is a blue button labeled 'Sign up'.

Figura E.2: Pantalla de registro de usuario

The image shows a login form titled 'Login'. It has two input fields: 'Email' with the placeholder 'email@company.domain' and 'Password'. Below the fields is a blue button labeled 'Login'.

Figura E.3: Pantalla de inicio de sesión

El usuario generalmente llegará a la aplicación por su index, o índice.

Para usar la aplicación a partir de ese punto necesita registrarse o iniciar sesión con una de las cuentas alternativas.

Tras registrarse como usuario o tripulante, si esta preparado el correo, se envía un correo electrónico a su dirección personal.

Una vez se confirma la recepción del correo, se activa la cuenta, permitiendo así usar servicios que antes no estaban disponibles.

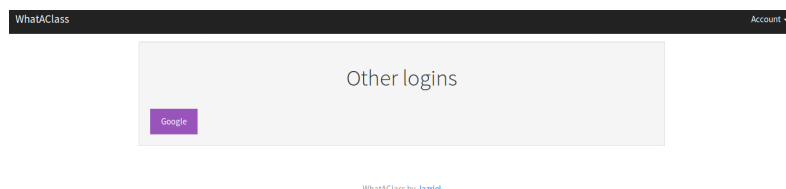


Figura E.4: Inicios de sesión alternativos

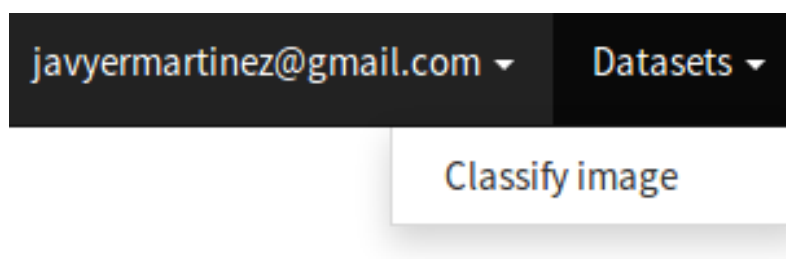


Figura E.5: Cambio de las zonas accesibles al iniciar sesión

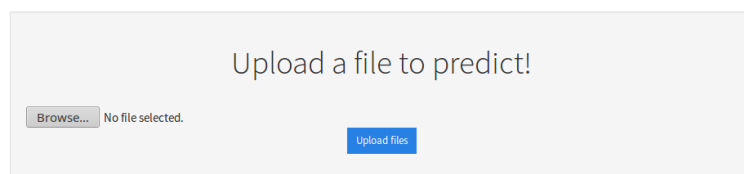


Figura E.6: Pantalla de acceso a el servicio de clasificación

Bibliografía

- [1] License recommendations.
- [2] Gnu lesser general public license, 1999.
- [3] The top programming languages 2017, 2017.