



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería en Informática



TFG del Grado en Ingeniería Informática

Sistema de reconocimiento
automático en arqueobotánica
Documentación Técnica



Presentado por Jaime Sagüillo Revilla
en Universidad de Burgos — 16 de marzo de 2017
Tutor: Álgvar Arnaiz González, José Francisco Díez
Pastor y Virginia Ahedo García

Índice general

Índice general	I
Índice de figuras	III
Índice de tablas	IV
Apéndice A Planificación	1
A.1. Introducción	1
A.2. Planificación temporal	6
A.3. Estudio de viabilidad	6
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	7
Apéndice C Especificación de diseño	8
C.1. Introducción	8
C.2. Diseño de datos	8
C.3. Diseño procedimental	8
C.4. Diseño arquitectónico	8
Apéndice D Documentación técnica de programación	9
D.1. Introducción	9
D.2. Estructura de directorios	9
D.3. Manual del programador	9
D.4. Compilación, instalación y ejecución del proyecto	9
D.5. Pruebas del sistema	9

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

Apéndice E Documentación de usuario	10
E.1. Introducción	10
E.2. Requisitos de usuarios	10
E.3. Instalación	10
E.4. Manual del usuario	10
Bibliografía	11

Índice de figuras

A.1. Burndown del <i>sprint</i> 0	2
A.2. Burndown del <i>sprint</i> 1	3
A.3. Burndown del <i>sprint</i> 2	4
A.4. Burndown del <i>sprint</i> 3	4
A.5. Burndown del <i>sprint</i> 4	5

Índice de tablas

Planificación

A.1. Introducción

Para llevar a cabo este proyecto vamos a aplicar una metodología llamada Scrum. Scrum es una metodología de desarrollo software ágil, es decir, durante cada *sprint*¹ se asignarán unas determinadas tareas a cumplimentar, con un producto como consecuencia de estas tareas. Al final de cada *sprint* se realizará una reunión junto a los tutores para validar los avances realizados y determinar las tareas a realizar durante el siguiente *sprint*.

Sprint 0

Estas son las tareas a realizar durante este *sprint 0*:

- Probar L^AT_EX.
- Gestor de tareas/versiones: *Github* y *Zenhub*.
- Instalar *Anaconda* y *Jupyter*.
- Leer los artículos propuestos por los tutores.
- Comenzar a probar algunos algoritmos de binarización.

Como se puede ver las tareas a realizar son básicas, puesto que es el *sprint 0* y es un *sprint* de mera adaptación al entorno de trabajo. La única tarea que supone un esfuerzo de comprensión mayor es la lectura de los artículos propuestos sobre trabajos relacionados o con una problemática similar a la nuestra. A continuación, en la figura A.1, se muestra el diagrama *burndown* de este *sprint*.

¹*Sprint*: es el período en el cual se lleva a cabo el trabajo en sí.[1]

Figura A.1: Burndown del *sprint* 0

Sprint 1

Estas son las tareas a realizar durante esta *sprint* 1:

- Documentar lo realizado durante el *sprint* 0.
- Documentar lo que se irá realizando durante este *sprint* 1.
- Continuar probando con algoritmos de procesamiento de imágenes.
- Probar una aproximación con clasificadores al problema.

Puesto que en el *sprint* anterior no se documentó lo realizado, durante este se pretende documentar todo lo realizado durante el *sprint* anterior y este. Además de continuar probando con algoritmos de procesamiento de imágenes y comenzar a probar con la aproximación al problema mediante clasificadores.

En este *sprint* me vi desbordado de trabajo debido a la subestimación del esfuerzo a empeñar en las distintas tareas. No siendo capaz de comenzar a probar una aproximación con clasificadores. Por ello la tarea «Probar una aproximación con clasificadores al problema» se vio movida al siguiente *sprint*.

A continuación, en la figura A.2, se muestra el diagrama *burndown* de este *sprint*. El cual tiene dicho aspecto debido a que muchas de las tareas se trabajaron de manera paralela, no siendo acabadas hasta el final del *sprint*. Y, además, algunas de las tareas no fueron cerradas cuando se debió, aspecto que se corregirá en los siguientes *sprints*.

Sprint 2

Estas son las tareas a realizar durante este *sprint* 2:

Figura A.2: Burndown del *sprint* 1

- Probar una aproximación con clasificadores al problema.
- Aplicación del método "Non maximum suppression" sobre el clasificado.

Puesto que la aproximación mediante reconocimiento de imágenes no reflejaba unos resultados muy positivos, durante la reunión mantenida con los tutores se decidió el uso de una técnica distinta. Nos referimos a la utilización de un clasificador, junto a un descriptor visual.

Debido a que todavía no se poseían suficientes imágenes para el estudio del problema mediante esta técnica, lo que se decidió es aplicarla sobre otro problema de características similares, como es el reconocimiento de caras en imágenes. Con unos resultados bastante positivos debido a distintos razonamientos explicados en la Memoria, sección de Aspectos relevantes del proyecto.

A continuación, en la figura A.3, se muestra el diagrama *burndown* de este *sprint*.

Sprint 3

Estas son las tareas a realizar durante este *sprint* 3:

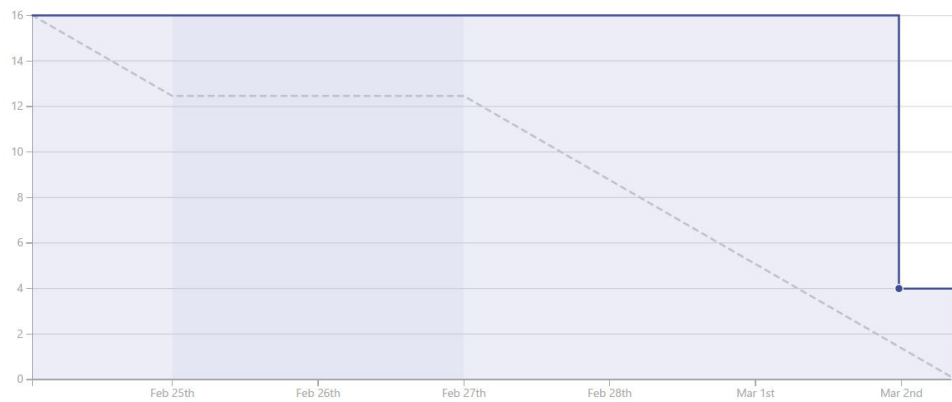
- Reorganizar los *Jupyter Notebooks*.
- Probar distintos clasificadores y métricas.
- Enviar fotos rotadas al clasificador.

Durante este *sprint*, primero, se reorganizó la estructura del proyecto. Aportando mucho más orden y claridad a nuestro proyecto. Después, se introdujeron múltiples clasificadores y métricas, los cuales introduciré en mayor

Figura A.3: Burndown del *sprint* 2

medida en la memoria, como *Random Forest* o *Gradient tree boosting*. Por último, se enviaron imágenes rotadas al clasificador, con el fin de poder analizar una posible problemática.

A continuación, en la figura A.4, se muestra el diagrama *burndown* de este *sprint*.

Figura A.4: Burndown del *sprint* 3

***Sprint* 4**

Estas son las tareas a realizar durante este *sprint* 4:

- Implementación de *Data Augmentation* en nuestro conjunto de entrenamiento.
- Implementación de controles de usuario.

Durante este *sprint* se aplicó en nuestro conjunto de entrenamiento la técnica *Data augmentation*. Esta técnica nos permitió aumentar el tamaño de nuestro conjunto de entrenamiento enormemente.

Además, se realizó un *notebook*², con controles de usuario, los cuales nos permiten escoger entre clasificadores, imágenes y probabilidades. Permitiendo la continua interacción entre el usuario y la clasificación de una imagen, sin la necesidad de modificar el código por parte del usuario del notebook para cambiar entre las distintas opciones.

A continuación, en la figura A.5, se muestra el diagrama *burndown* de este *sprint*.



Figura A.5: Burndown del *sprint* 4

Sprint 5

Estas son las tareas a realizar durante este *sprint* 4:

- Implementar un *file chooser*
- Añadir más clasificadores.
- Correcciones en la documentación.
- Estudiar como implementar un etiquetador de imágenes.

²Siempre que nos referimos a un *notebook*, a lo que nos referimos es a un *Jupyter Notebook*

A.2. Planificación temporal

A.3. Estudio de viabilidad

Viabilidad económica

Viabilidad legal

Apéndice B

Especificación de Requisitos

- B.1. Introducción
- B.2. Objetivos generales
- B.3. Catalogo de requisitos
- B.4. Especificación de requisitos

Especificación de diseño

- C.1. Introducción
- C.2. Diseño de datos
- C.3. Diseño procedimental
- C.4. Diseño arquitectónico

Documentación técnica de programación

- D.1. Introducción
- D.2. Estructura de directorios
- D.3. Manual del programador
- D.4. Compilación, instalación y ejecución del proyecto
- D.5. Pruebas del sistema

Apéndice E

Documentación de usuario

- E.1. Introducción**
- E.2. Requisitos de usuarios**
- E.3. Instalación**
- E.4. Manual del usuario**

Bibliografía

- [1] Wikipedia. Scrum — Wikipedia, the free encyclopedia, 2017. [Online; accedido 12-Marzo-2017].