



# Web Scrapping



Sergio González Barbe

Alicia Tugores Bahana

CIFP Francesc de Borja Moll

9-12-2020

<i>Definición del proyecto</i> .....	3
<i>Herramientas utilizadas</i> .....	3
<i>Librerías utilizadas</i> .....	3
<i>GIR WORK FLOW</i> .....	3
<i>Pasos realizados y metodologías utilizadas</i> .....	4
<u>1º sprint</u> →	
<u>2º sprint</u> →	
<u>3º sprint</u> →	
<i>Temporalización</i> .....	4
<i>Diagrama de Componentes</i> .....	5
<i>Conclusiones y posibles mejoras</i> .....	6
<i>Dificultades encontradas</i> .....	6

## Definición del proyecto

Rick realiza una súper fiesta con muchos invitados (tanto humanos como aliens) y los invitados han de elegir y reservar el UFO, el pack de bienvenida y el menú. Cada pareja de proyecto debía elegir 1 de los 3 servicios. Nosotros elegimos los menús.

Nuestra tarea se resume en hacer una página que contenga la información de varios menús para luego hacer un programa cuya función sea scrapear la información de dicha página para devolverla en un documento json y crear una base de datos.

- Construir el sistema que realiza web scraping sobre las webs de los productos y servicios.
- Construir un sitio web con el servicio que prefieras (reserva de UFO, reserva de packs de bienvenida o reserva de menús) y poblarlo con los datos que pretendes “rick-scrapear”.
- Almacenar los datos en una base de datos MongoDB o en un repo en la nube - GitHub hace el truco- con los documentos JSON en un esquema específico que indiquen los socios del grupo de segundo de Dual.

## Herramientas utilizadas

**VSCode** → Es el IDE que hemos utilizado para desarrollar el código fuente. (Conventional commits, live Share, git graph, Python).

**Git** → Hemos utilizado Git como sistema de versionado de código para compartir y trabajar sobre nuestra aplicación y para mantener un registro de los cambios realizados.

**Clockify** → Para documentar las horas empleadas en cada parte del proyecto.

**CMD** → Es la terminal utilizada para instalar las librerías de Python utilizadas en el proyecto y para los comandos de Git.

**Mongodb** → Para crear la base de datos.

**GitHub** → Hemos utilizado esta plataforma para almacenar nuestro proyecto en la nube y además hemos utilizado la rama de github pages para hostear nuestra web en la red.

## Librerías utilizadas

- Pymongo → Para conectar con la base de datos en Mongodb.
- Request → Para hacer peticiones al servidor.
- Dnspython → Para realizar consultas y transferencias (para solucionar error de pymongo)

## GIT WORK FLOW

Hemos utilizado 4 ramas: Main, donde hemos almacenado el proyecto una vez ya terminado (versión 1.0), Feature-s, donde Sergio ha estado trabajando las modificaciones del proyecto y Feature-a es donde Alicia ha estado trabajando. Para terminar, tenemos una última rama; Developer, donde hemos estado fusionando (haciendo merge) las ramas feature-s y feature-a.

## Pasos realizados y metodologías utilizadas

Hemos utilizado una combinación de pair-programming y el marco de trabajo Scrum. Cada día que tocaba ponerse con el proyecto hacíamos una daily para exponer los problemas que nos pudiéramos encontrar e intentábamos solucionarlas. También hemos realizado sprints para cada módulo del proyecto.

1º sprint → Crear y diseñar las páginas web y todo su contenido. (17-nov/24-nov)

- Creamos las páginas utilizando HTML y le dimos diseño con CSS.
- Buscamos y creamos el contenido de todos los menús guiándonos por los requisitos que nos pidió el alumno de segundo curso.
- Utilizamos git pages para hostear los documentos HTML.

2º sprint → Desarrollar los primeros archivo Python del proyecto. (24-nov/05-dic)

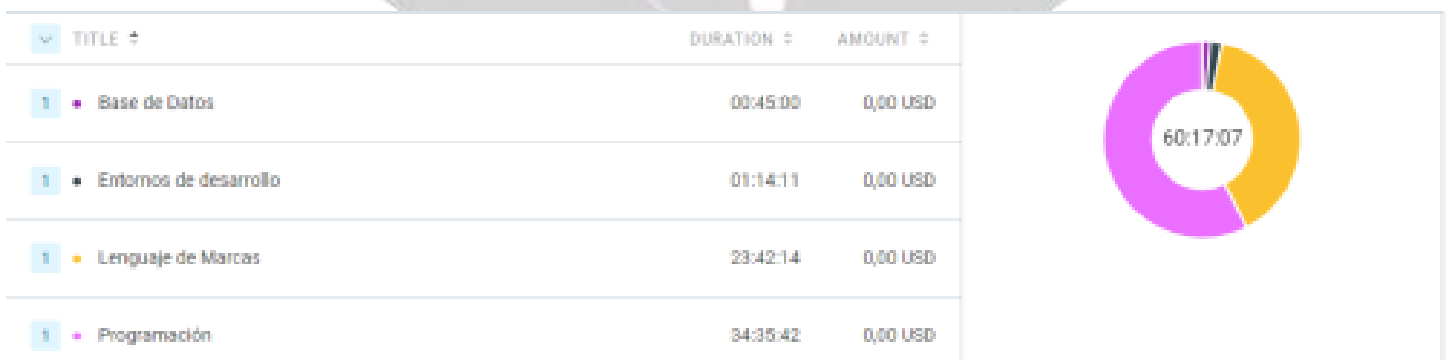
- Un archivo que almacenara los links los cuales contienen la información que deseamos extraer.
- Un documento que extrajera dicha información.

3º sprint → Desarrollar la resta de archivos Python. (05-dic/09-dic)

- Un archivo que almacenara la información scrapeada en diccionarios.
- Un archivo que conectara con MongoDB para crear una bbdd y subir los documentos.

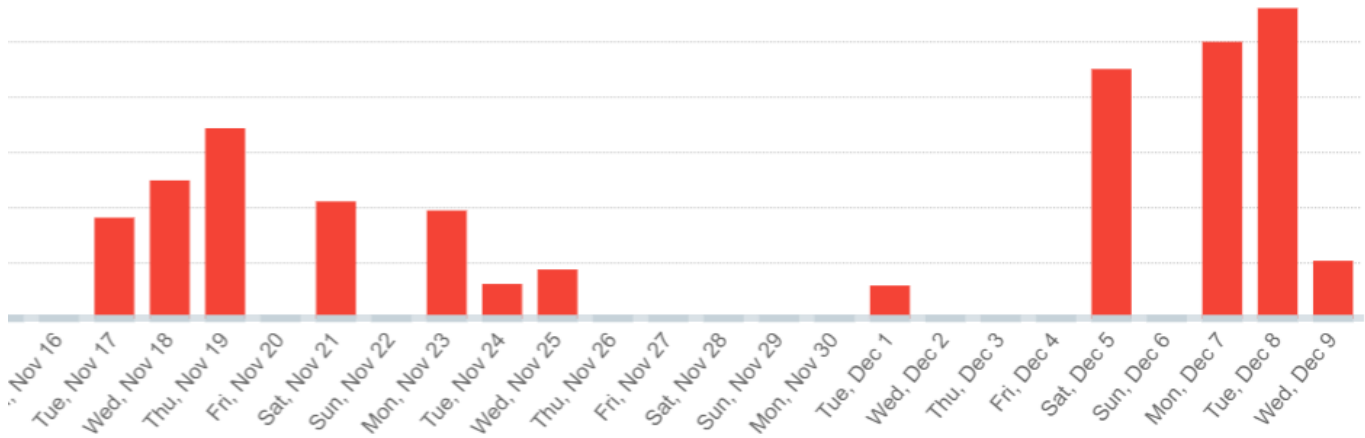
## Temporalización

Hemos utilizado la herramienta Clockify para documentar las horas empleadas en el proyecto separadas por módulos.

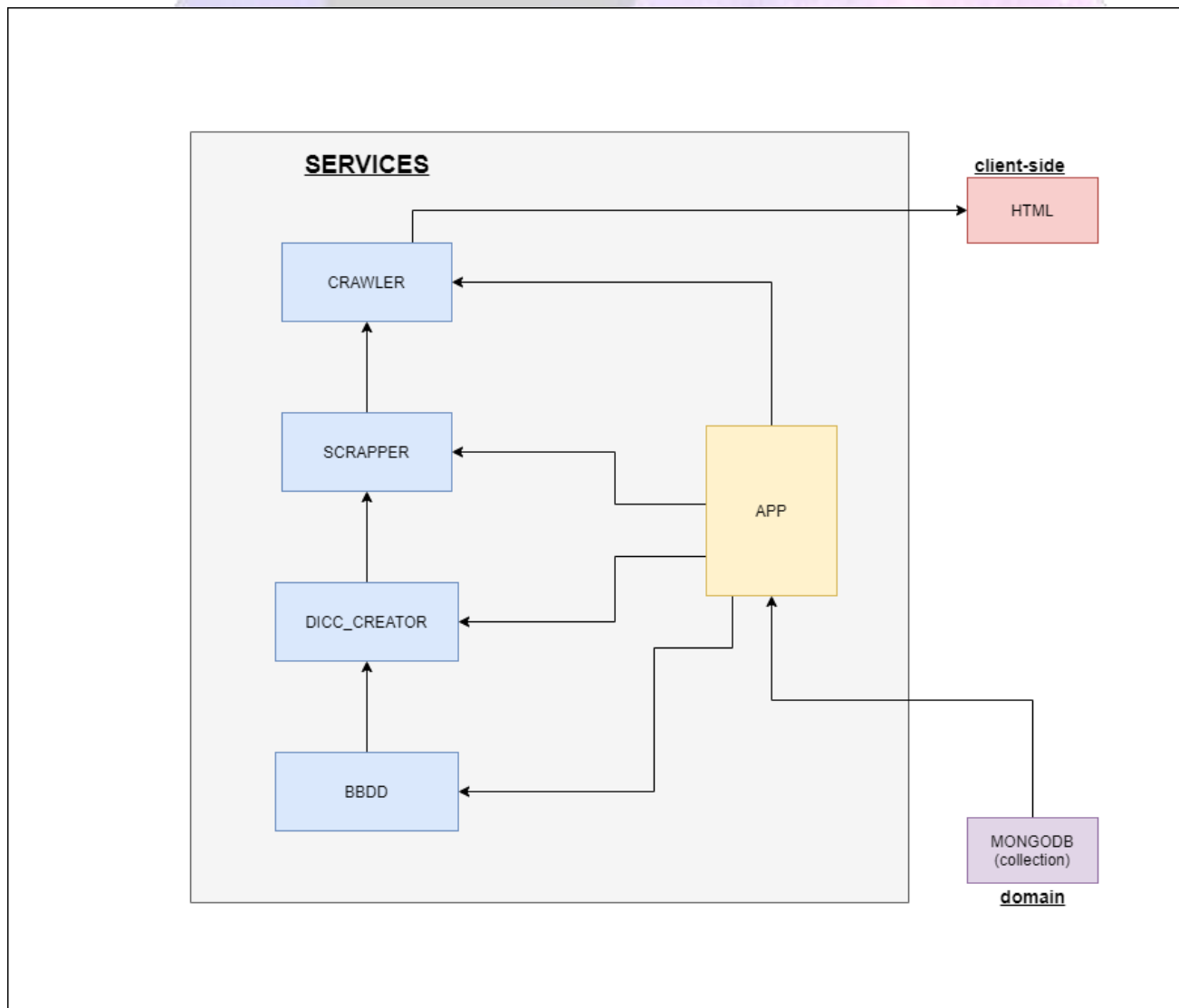




## WEB SCRAPPING



## Diagrama de Componentes



## Conclusiones y posibles mejoras

Creemos que hemos hecho un gran trabajo, dado que a pesar que no tenemos el mejor código, hemos estado programando con Pair-Programming (Ping Pong Programming) y los dos hemos aprendido por igual y hemos visto la evolución del proyecto.

¿Posibles mejores? Probablemente muchas, nuestro código podría ser más eficiente y de mejor calidad.

## Dificultades encontradas

Algunas de las dificultades que nos hemos encontrado a la hora de realizar nuestro proyecto han sido:

- Las exceptions
- Marcadores @
- Entrelazar los módulos
- Importaciones de variables y funciones (las rutas)

