



Web Scrapping



Sergio González Barbe

Alicia Tugores Bahana

CIFP Francesc de Borja Moll

9-12-2020

<i>Definición del proyecto</i>	3
<i>Requerimientos</i>	3
<i>Programación</i>	3
<i>Lenguaje de Marcas</i>	3
<i>Entornos de Desarrollo</i>	4
<i>Sistemas informáticos</i>	4
<i>Bases de Datos</i>	4
<i>Requerimientos Transversales</i>	5
<i>Herramientas utilizadas</i>	5
<i>Pasos realizados y metodologías utilizadas</i>	5
<u>1º sprint</u> ➔	5
<u>2º sprint</u> ➔	6
<u>3º sprint</u> ➔	6
<i>Temporalización</i>	6
<i>Diagrama de Arquitectura</i>	7
<i>Conclusiones y posibles mejoras</i>	7

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.

Requerimientos

Programación

1. Reconoce la estructura de un programa informático, por eso, identifica y relaciona los elementos propios del lenguaje de programación utilizado.
2. Escribe y prueba programas sencillos, por eso reconoce y aplica los fundamentos de la programación estructurada y modular.
3. Escribe y depura código, por eso analiza y utiliza las estructuras de control del lenguaje.
4. Realiza operaciones de entrada y salida de información, utilizando procedimientos específicos del lenguaje y librerías de clases.
5. Escribe programas que manipulan información, por eso selecciona y utiliza estructuras avanzadas de datos.

Lenguaje de Marcas

1. Reconoce las características de lenguajes de marcas, analizando e interpretando fragmentos de código.
2. Utiliza lenguajes de marcas para la transmisión de información a través de la web analizando la estructura de los documentos e identificando sus elementos.
3. Establece mecanismos de validación para documentos HTML utilizando métodos para definir su sintaxis y estructura.

(añadido del Moodle por Adela):

La Web del proyecto ha de cumplir con los siguientes requerimientos:

- Ha de tener una página inicial nombrada por ejemplo: “index.html” o “inicio.html”.
- Ha de tener al menos un mínimo de tres páginas html enlazadas a la primera.

- Tienes que utilizar etiquetas semánticas para estructurar la web.
- Se han de utilizar CSS (mejor con archivos externos), que empleen los diferentes tipos de selectores explicados en clase (de etiqueta, de clase, de identificador, descendientes, combinatorios, pseudoclases).
- El diseño de la web ha de ser claro, limpio y que facilite la navegación. (Busca páginas donde expliquen las tendencias de diseño actuales o más recomendadas).

Entornos de Desarrollo

1. Reconoce los elementos y las herramientas que intervienen en el desarrollo de un programa informático y analiza sus características y las fases en que actúan hasta llegar a su puesta en funcionamiento.
2. Evalúa contornos integrados de desarrollo, y analiza sus características para editar código fuente y generar ejecutables.
3. Verifica el funcionamiento de programas, por eso diseña y realiza pruebas → módulo de Programación.
4. Optimiza código utilizando las herramientas disponibles en el contorno de desarrollo → módulo de Programación
5. Elabora documentación utilizando aplicaciones informáticas de propósito general.

Sistemas informáticos

1. Evalúa sistemas informáticos, e identifica sus componentes y sus características.
2. Instala sistemas operativos, después de planificar el proceso por el cual interpreta documentación técnica.
3. Gestiona la información del sistema aplicando medidas para asegurar la integridad de los datos, e identifica las estructuras de almacenaje.
4. Gestiona sistemas operativos utilizando comandos y herramientas gráficas y evalúa las necesidades del sistema.

Bases de Datos

1. Reconoce los elementos e las bases de datos analizando sus funciones y valorando la utilidad de los sistemas gestores.
2. Consulta la información almacenada en una base de datos empleando asistentes, herramientas gráficas y el lenguaje de manipulación de datos.
3. Modifica la información almacenada en la base de datos utilizando asistentes, herramientas gráficas y el lenguaje de manipulación de datos.
4. Desarrolla procedimientos de almacenaje evaluando y utilizando las sentencias del lenguaje incorporado en el sistema gestor de base de datos.

Requerimientos Transversales

- Desarrollar la creatividad y el espíritu de innovación para resolver los desafíos que se presenten en los procesos y en la organización del trabajo y de la vida personal.
- Tomar decisiones fundamentales analizando las variables implicadas, integrando saberes de diferente ámbito y aceptando los riesgos y la posibilidad de equivocación, para afrontar y resolver diferentes situaciones, problemas o contingencias.

- Desarrollar técnicas de liderazgo, motivación, supervisión y comunicación en contextos de trabajo en grupo, para facilitar la organización y la coordinación e equipos de trabajo.
- Aplicar estrategias y técnicas de comunicación adaptadas a los contenidos que se vayan transmitiendo, así como la finalidad y las características de los receptores, para asegurar la eficacia en los procesos de comunicación.
- Identificar y proponer las acciones profesionales necesarias para dar respuesta a la accesibilidad universal y al diseño para todos.
- Reconocer los derechos y los deberes como agente activo en la sociedad, teniendo en cuenta el marco legal que regula las condiciones sociales y laborales, para participar en la ciudadanía democrática.
- Analizar y valorar la participación, el respeto, la tolerancia y la igualdad de oportunidades, por hacer efectivo el principio de igualdad entre mujeres y hombres.

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.

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.

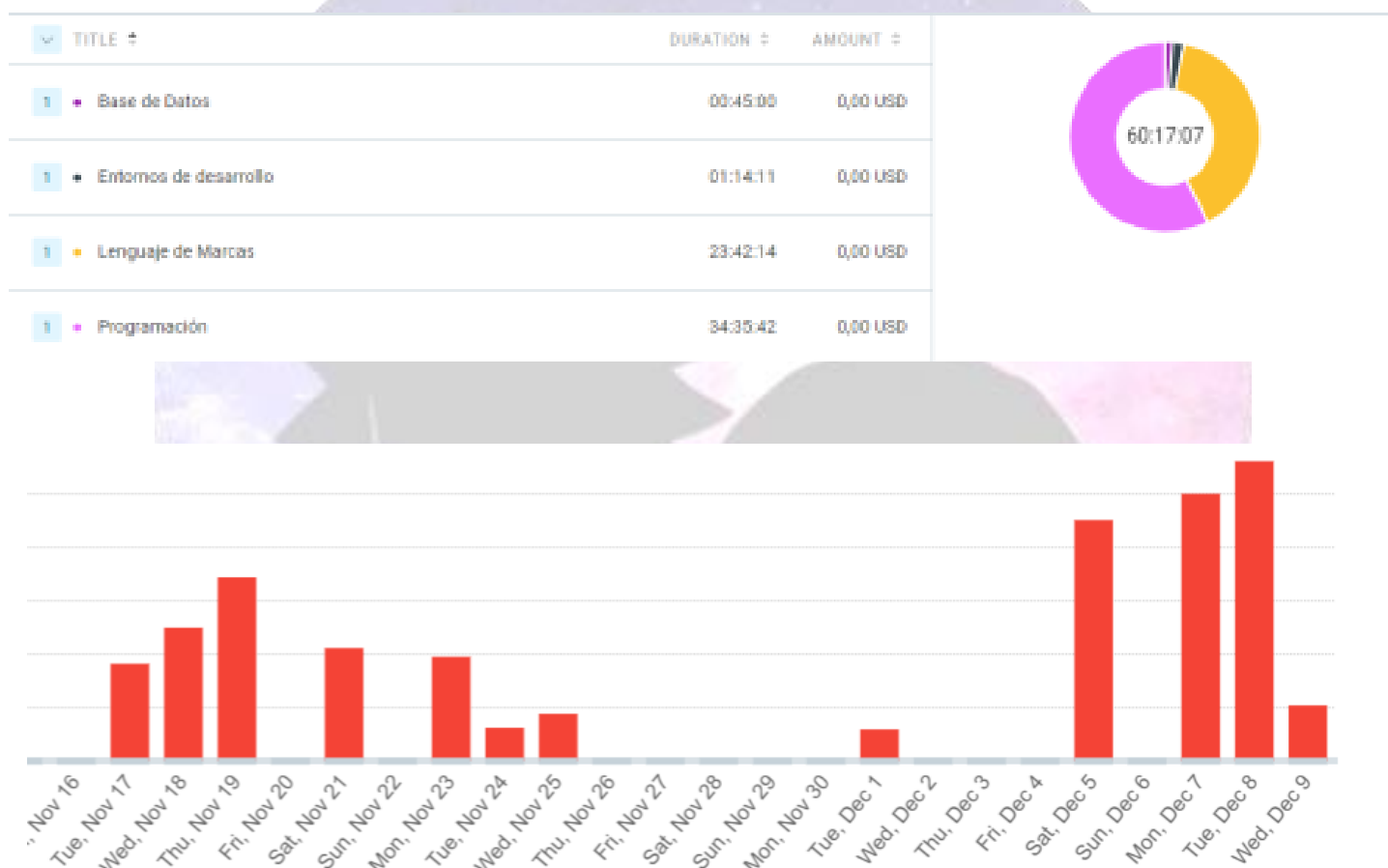
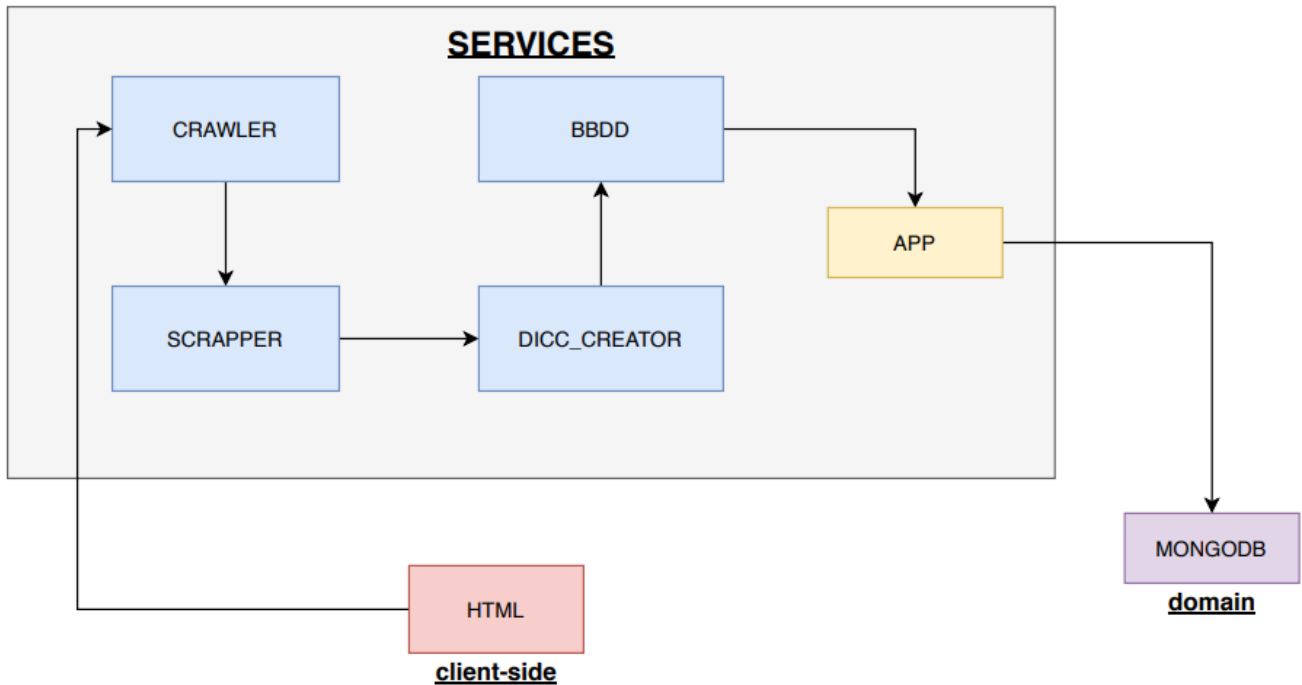


Diagrama de Arquitectura



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 mejoras? Probablemente muchas, nuestro código podría ser más eficiente y de mejor calidad.