

Graduado en Ingeniería Informática

Universidad Politécnica de Madrid

Escuela Técnica Superior de
Ingenieros Informáticos

TRABAJO FIN DE GRADO

Integración y puesta en valor de conjuntos de datos abiertos

Autor: Jesús Pérez Melero

Director: Raúl García Castro

MADRID, ENERO 2018

Índice de Contenido

1.- Descripción general del trabajo	3
2.- Lista de tareas	3
3.- Diagrama de Gantt	4
4.- Copia de la propuesta de trabajo	5
4.1.- <i>Resumen general del trabajo</i>	<i>5</i>
4.2.- <i>Lista de objetivos concretos</i>	<i>5</i>
4.3.- <i>Desglose de la dedicación total del trabajo en horas</i>	<i>5</i>
4.4.- <i>Conocimientos previos</i>	<i>5</i>
5.- Bibliografía	6

Índice de Ilustraciones

Ilustración 1 - Diagrama de Gantt.....	4
Ilustración 2 - Duración de cada tarea.....	4

Índice de Tablas

Tabla 1 - Desglose de la dedicación en horas.....	5
--	----------

1.- Descripción general del trabajo

El presente trabajo pretende realizar un sistema que tome como entrada distintos **conjuntos de datos abiertos** del portal de datos abiertos de la ciudad de **Madrid**.

La información contenida en dichos conjuntos de datos **se enriquecerá y modificará** en función de las necesidades oportunas (falta o exceso de información).

Posteriormente, se publicará siguiendo los principios de los datos enlazados, concretamente en el **formato RDF**. Los datos enlazados ^[1] constituyen un **método de publicación de información estructurada e interconectada**, que además de servir para nutrir páginas web para uso de los humanos, extiende su uso a la **lectura automática** por parte de las máquinas, siendo esto un pilar fundamental de la **Web Semántica**.

Los datos resultantes de dicho método serán la base de una aplicación que permita realizar distintos tipos de consultas sobre la ciudad de Madrid, la mayoría de ellas de índole **turístico y cultural**. Esta aplicación será presentada en el **Concurso de Datos Abiertos de la Ciudad de Madrid** ^[2], que tiene lugar el último trimestre del año 2017.

Por tanto, **los objetivos de este proyecto son:**

- Identificar conjuntos de datos útiles de la ciudad de Madrid.
- Enriquecer dicha información mediante el enlazado de datos, recalcando el valor añadido del enlazado, y publicándolos en formato RDF.
- Construir una aplicación que use dichos datos como parte nuclear.

2.- Lista de tareas

Tomando como referencia la propuesta de trabajo proporcionada por el profesor tutor, la lista de tareas en las cuales se dividirá la realización del presente trabajo son:

- **Estudio del dominio:** Se realizará un estudio previo del dominio de los datos enlazados, es decir, métodos de publicación, tipos de conjuntos de datos, análisis de conjuntos de datos, análisis de software adecuado para el tratamiento de datos enlazados, etc.
- **Análisis de sistemas relevantes al trabajo:** Se realizará un estudio de los sistemas que pueden ser útiles de cara a la realización de este trabajo.
- **Extracción de requisitos:** Los requisitos serán la base del sistema que se construya. Se definirá cuáles son y por qué deben considerarse como requisitos.
- **Diseño del sistema:** Una vez definidos los requisitos se procederá al diseño técnico del sistema, indicando qué herramientas (lenguajes, *IDEs*, *Frameworks*...) serán utilizados en la fase de implementación.
- **Implementación del sistema:** Se realizará la implementación del sistema siguiendo el diseño previo.
- **Aplicación y evaluación del sistema:** Una vez construido el sistema, este será evaluado utilizando un caso de uso para comprobar su efectividad, y acompañando dicha evaluación de un test de usabilidad para asegurar la calidad del sistema construido, de cara al usuario.
- **Escritura del documento y preparación de la presentación:** De forma paralela a todas las tareas, se redactará el documento que contendrá toda la información relativa al desarrollo del presente trabajo.

3.- Diagrama de Gantt

En base a las tareas definidas en el apartado anterior, y al desglose de horas proporcionado por el profesor tutor ([véase Desglose de horas](#)), se ha realizado el diagrama de Gantt que muestra la planificación inicial del desarrollo del proyecto.

Para ello, se han tenido en cuenta las siguientes consideraciones:

- La fecha de comienzo del trabajo es el 18 de septiembre.
- La fecha de finalización (estimada) del trabajo es el 18 de diciembre.
- Cada día se dedicarán aproximadamente 4 horas a la realización del trabajo.

Tomando dichos datos, se tienen 360 (90 días * 4 horas / día) horas de trabajo útiles.

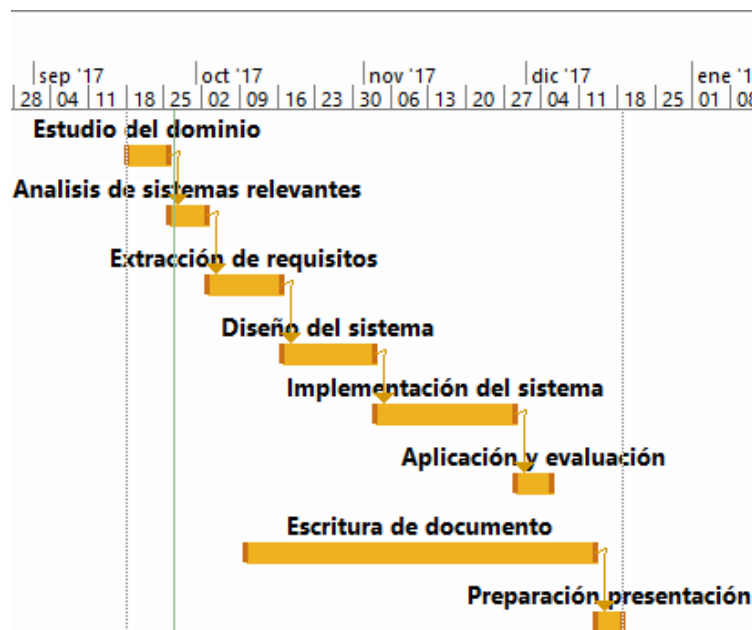


Ilustración 1 - Diagrama de Gantt

Nombre de tarea	Duración	Comienzo	Fin
Estudio del dominio	6 días	lun 18/09/17	lun 25/09/17
Analisis de sistemas	5 días	mar 26/09/17	lun 02/10/17
Extracción de requisitos	10 días	mar 03/10/17	lun 16/10/17
Diseño del sistema	13 días	mar 17/10/17	jue 02/11/17
Implementación del sistema	18 días	vie 03/11/17	mar 28/11/17
Aplicación y evaluación	5 días	mié 29/11/17	mar 05/12/17
Escritura de documento	47 días	mar 10/10/17	mié 13/12/17
Preparación presentación	3 días	jue 14/12/17	lun 18/12/17

Ilustración 2 - Duración de cada tarea

4.- Copia de la propuesta de trabajo

4.1.- Resumen general del trabajo

Desarrollo de un sistema que tome como entrada distintos conjuntos de datos abiertos y que los publique siguiendo los principios de los datos enlazados, en el formato **RDF**, **añadiendo valor a los mismos**. Junto con los datos generados se implementará una aplicación que los use en algún dominio concreto con el **objetivo de presentarla a algún datathon**.

4.2.- Lista de objetivos concretos

El trabajo realizado se puede dividir en distintos sub-objetivos que el alumno tendrá que ir cumpliendo secuencialmente.

- Realización de un **estado de la cuestión** en el tema de integración de datos abiertos.
- Elaboración de un **documento de especificación de requisitos** del sistema.
- **Desarrollo del sistema**, partiendo de los requisitos identificados.
- **Aplicación y evaluación del sistema** en un caso de uso concreto.

Cada uno de los objetivos mencionados anteriormente llevará asociado un entregable que servirá de entrada a la tarea de escritura del Trabajo Fin de Grado.

4.3.- Desglose de la dedicación total del trabajo en horas

Tarea	Horas asociadas
Estudio del dominio	25
Análisis de sistemas relevantes al trabajo	29
Extracción de requisitos	65
Diseño del sistema	60
Implementación del sistema	70
Aplicación y evaluación del sistema	25
Escritura del documento y preparación de la presentación	50

Tabla 1 - Desglose de la dedicación en horas

4.4.- Conocimientos previos

Haber cursado la asignatura de Web Semántica y Linked Data.

5.- Bibliografía

[1] – Wikipedia (2017). *Datos enlazados*. [Enlace](#)

[2] – MediaLab Prado (2017). *Datatón Ciudad de Madrid 2017*. [Enlace](#)