

UNIVERSIDAD REY JUAN CARLOS

Master en Data Science

Obtención de datos



# **RED DE METRO, CERCANIAS Y TRANVIA DE MADRID**

Autor: Alejandro Pérez Barreiro

Móstoles, Diciembre 2016

## **Índice de contenidos**

1. Objetivo de la práctica.
2. Obtención de datos.
3. Descripción del código.
4. Ontologías.
5. Diagrama de clases.

## 1. Objetivo de la práctica.

El objetivo de esta práctica será proporcionar los datos de las paradas de Metro, Metro Ligero y Cercanías de Madrid en formato RDF para su uso posterior en la obtención de rutas accesibles.

## 2. Obtención de datos

Para proporcionar los datos con mayor detalle vamos a extraerlos de dos fuentes distintas para después combinarlos obteniendo así más información sobre cada parada/estación.

El primer lugar del que extraeremos los datos será de la pagina de datos abiertos del Consorcio Regional de Transportes de Madrid (CRTM) [<http://datos.crtm.es/>] en formato GTFS, habrá que obtener los datos de las estaciones (contenidas en el fichero *stops.txt*) de los transportes METRO, METRO Ligero y cercanías RENFE. En el siguiente enlace se puede ver el significado de cada campo que aparece en el fichero stops.txt <https://developers.google.com/transit/gtfs/reference?hl=es-419>

El resto de información se obtendrá haciendo un scrap sobre la página <http://www.crtm.es/tu-transportepublico.aspx> con el fin de obtener el número de línea, el orden de cada parada y el tipo de transporte de cada parada.

En el caso de que por una parada pasen varias líneas o distintos tipos de transporte esta para aparecerá en los datos tantas veces como esto ocurra, es decir, si por una parada pasan dos líneas de metro y una de metro ligero esa para aparecerá tres veces en los datos.

## 3. Descripción del código

Los pasos a seguir para obtener los datos en formato RDF se encuentran recogidos en el fichero README.txt adjunto en el zip.

Los scripts fueron desarrollados en Ubuntu, por tanto no se recomienda usar Windows para la ejecución de los mismos ya que podría dar errores.

## 4. Ontologías

En la siguiente tabla se encuentran las ontologías analizadas y en verde las que he utilizado:

Ontology	Namespace	Homepage
Dbpedia ontology	<a href="http://dbpedia.org/ontology/">http://dbpedia.org/ontology/</a>	<a href="http://dublincore.org/">http://dublincore.org/</a>
The Friend Of A Friend (FOAF)	<a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a>	<a href="http://www.foaf-project.org/">http://www.foaf-project.org/</a>
Geonames ontology	<a href="http://www.w3.org/2003/01/geo/wgs84_pos#">http://www.w3.org/2003/01/geo/wgs84_pos#</a>	<a href="http://www.geonames.org/">http://www.geonames.org/</a>
Daml ontologies	<a href="http://www.daml.org/2003/05/subway/subway-ont">http://www.daml.org/2003/05/subway/subway-ont</a>	<a href="http://www.daml.org/ontologies/396">http://www.daml.org/ontologies/396</a>
The Dublin Core (DC)	<a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a>	<a href="http://dublincore.org/">http://dublincore.org/</a>

## 5. Diagrama de clases

