

OBJETIVO DE LA PRÁCTICA

La práctica de la asignatura tiene como objetivo proporcionar datos en formato RDF de los transportes METRO, METRO Ligero y Cercanías RENFE, para su uso posterior en la obtención de rutas accesibles.

RESUMEN DE LA PRÁCTICA

Partiendo de la fuente de datos abiertos del Consorcio Regional de Transportes de Madrid (CRTM) [1], en formato GTFS [2], habrá que obtener los datos de las estaciones (contenidas en el fichero *stops.txt*) de los transportes METRO, METRO Ligero y cercanías RENFE.

Dado que en estos ficheros GTFS, las estaciones no están ordenadas según el itinerario de la línea, ni en ellos aparece la línea a la que pertenece la estación, será necesario obtener dicho orden y dicha pertenencia a partir de los datos de las líneas de cada medio de transporte, accesibles en el Consorcio Regional de Transportes de Madrid [3].

REALIZACIÓN DE LA PRÁCTICA

El CRTM proporciona, en formato abierto, datos de los medios de transporte de la Comunidad de Madrid. Uno de los formatos que ofrece es el de la especificación de los *feeds* de transporte de Google (GTFS). Para obtener estos ficheros GTFS, **se accederá manualmente a la página de datos abiertos del consorcio, y se descargarán (también manualmente) los ficheros de Metro, Metro Ligero y Cercanías, en su correspondiente enlace GTFS, para extraer el fichero *stops.txt* de cada medio de transporte.**

Dado que estos ficheros no contienen el orden del itinerario de las líneas, ni las líneas a las que pertenece cada estación, es necesario buscar esta información realizando un *scraper* en [3]. Será necesario acceder al enlace Líneas de cada medio de transporte especificado con anterioridad. Una vez allí, habrá que recorrer todas las líneas, accediendo a su contenido, con el fin de obtener la ordenación de las estaciones.

La integración de estas dos fuentes de datos se realizará en un fichero de texto en formato CSV.

Posteriormente, será necesario elaborar un diagrama conceptual con las entidades y las relaciones entre ellas, que soporte la información del dominio de este problema. Finalmente, se llevará a cabo la creación de un grafo RDF utilizando *rdflib*. El resultado de este último paso será un fichero RDF/XML.

Se pide:

- Código del *scraper* desarrollado para obtener la información. **Hay que entregar los ficheros .py con el código fuente, no un notebook. Preparar el fichero comprimido que entreguéis con los ficheros stops.txt para que al ejecutar el código dichos ficheros estén disponibles y no haya que copiarlos en vuestro directorio para la ejecución.**
- Fichero de texto, en formato CSV, con información del medio de transporte, la línea y el conjunto de estaciones en el orden marcado por el itinerario de cada línea. **Este fichero debe ser único y contener la información integrada de los tres medios de transporte.** La cabecera de dicho fichero debe incluir los siguientes campos (en cursiva los propios de GTFS):

transportmean_name,line_number,order_number,stop_id,stop_code,stop_name,stop_desc,stop_lat,stop_lon,zone_id,stop_url,location_type,parent_station,stop_timezone,wheelchair_boarding

- *transportmean_name*: indicará el nombre del medio de transporte. Los valores posibles son: METRO, ML, CR.
- *line_number*: indicará el número de la línea a la que pertenece la estación.

OBTENCIÓN DE DATOS

Curso 2019-2020



- `order_number`: indicará el número de la línea concatenado con el número de secuencia de la estación en la línea, según el itinerario.

- Diagrama conceptual del dominio del problema. **Diagrama de clases UML**.
- Fichero RDF, en formato RDF/XML, generado con la información procedente del fichero anterior.
- Código de creación del fichero RDF. **Entregar el fichero fuente .py**.

FECHA DE ENTREGA DE LA PRÁCTICA

La fecha tope de entrega de la práctica será el **día 15 de enero de 2020** a través del Aula Virtual, en un enlace disponible para dicha entrega.

La práctica se realizará en grupos de 4 ó 5 personas.

FORMATO DE ENTREGA DE LA PRÁCTICA

La entrega será un fichero ZIP con el nombre formado por el primer apellido concatenado de todos los miembros del grupo de trabajo. Por ejemplo, si Paloma y Almudena forman un grupo el nombre del fichero sería "CaceresSierra.zip". Este fichero ZIP deberá tener el siguiente contenido:

- Código fuente en Python del *scraper* desarrollado y el programa que integra los datos de las dos fuentes.
- Ficheros de datos *stops.txt*.
- Fichero CSV.
- Fichero RDF/XML.

ENLACES DE INTERÉS PARA LA REALIZACIÓN DE LA PRÁCTICA

- [1]. Datos abiertos del Consorcio Regional de Transporte de Madrid (CRTM), <http://datos.crtm.es/>
- [2]. Google Transit Feed Specification (GTFS), <https://developers.google.com/transit/gtfs/reference?hl=es-419>
- [3]. Información de los medios de transporte en el CRTM, <http://www.crtm.es/tu-transporte-publico.aspx>