

Metodología de la construcción de las distancias y tiempos de educación

El proceso que se ha seguido para la construcción de las matrices de distancias y tiempos para los centros educativos más apropiados a cada municipio se puede definir en tres etapas:

- 1) Tipificación de centros educativos y asignación a municipios
- 2) Llamada a Google Distance Matrix API

1. Tipificación de los centros educativos y asignación a municipios

Se usará como base de datos los [centros educativos de la Comunidad de Madrid](#). En este estudio, se quiere dividir los centros educativos disponibles en diversas categorías:

Por categoría:

- PREINF: corresponde a preinfantil. Dentro del csv, corresponde a 'ZNAS.CN.' y 'PRE'.
- INF: corresponde a infantil. Dentro del csv, corresponde a 'EEI', 'INF' y 'EIPR'.
- PRI: corresponde a primaria. Dentro del csv, corresponde a 'PRI'.
- SEC: corresponde a secundaria. Dentro del csv, corresponde a 'IES', 'SEC' y 'ESO'.

Por titularidad:

- PUB: corresponde a público
- PRIV: corresponde a privado. Dentro del csv, corresponde a 'PRIVADO' y 'PRIVADO CONCERTADO'. Al haber una alta oferta de estos, se limitarán a 3 por categoría por municipio como máximo.

Además, se filtrarán a los centros cuya situación sea 'ALTA' y del que tengamos coordenadas. Con esto, construiremos un JSON que almacene por municipio los centros educativos correspondientes por categoría.

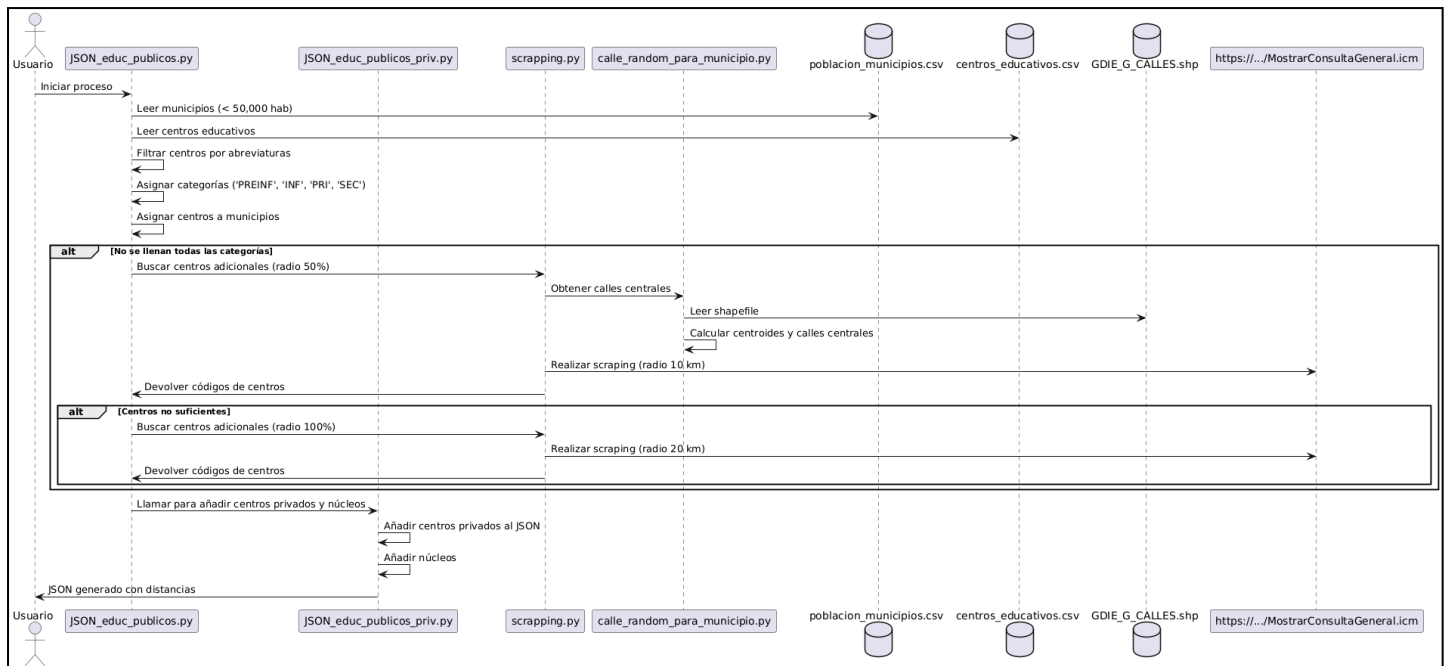
```
"28866": {
  "CENTROS_EDUCATIVOS": {
    "PUB": {
      "PREINF": [
        28056321
      ],
      "INF": [
        28003407,
        28072284
      ],
      "PRI": [
        28003407
      ],
      "SEC": [
        28046315
      ]
    },
    "PRIV": {
      "PREINF": [],
      "INF": [
        28003419,
        28042371,
        28070548
      ],
      "PRI": [
        28003419,
        28042371
      ],
      "SEC": [
        28003419,
        28042371
      ]
    }
  },
  "NUCLEOS": [
    660101
  ]
},
```

¿Qué metodología hemos utilizado para construir este JSON?

En primer lugar, se ha asignado automáticamente los centros educativos a los municipios a los que pertenecen, tanto privados como públicos, con las consideraciones mencionadas anteriormente.

Luego, al quedar municipios sin todas las categorías de centros educativos públicos con al menos una instancia, se realizó un script que automatiza la búsqueda de ellos usando el [buscador de colegios de la Comunidad de Madrid](#). Usaremos la calle más central del municipio y un número al azar de ellas (sin él, la búsqueda no funciona). Intentaremos primero una búsqueda con el radio de búsqueda al 50%, es decir, 10km. Si queda al menos una categoría sin rellenar de los centros educativos públicos, se repetirá el proceso pero para 100%, es decir, 20km, sin modificar las categorías ya rellenadas.

Todo este proceso puede ser resumido con el siguiente diagrama de secuencia:



2. Llamada a la Distance Matrix API de Google

Una vez construido el JSON, tendremos el número de elementos necesario para completar nuestra matriz. A cada núcleo le asociaremos los centros educativos de su municipio y usaremos el centroide de cada núcleo como referencia.

Entonces, se hará la llamada a la API donde los orígenes serán los núcleos de un municipio y los destinos los diferentes centros educativos. Se debe de tener en cuenta las restricciones que esta tiene:

- En una misma llamada los elementos máximos (en nuestro caso número de núcleos x número de centros educativos) no pueden superar 100.
- El número de orígenes (núcleos) y destinos (centros educativos) no debe de superar los 25.

Por ello, haremos *batches* en las llamadas para cumplir con estos requisitos.

Aparte, se puede utilizar el argumento *departure_time* para seleccionar a qué hora va a ocurrir la búsqueda, siendo siempre una fecha futura. En el caso que no se proporcione este argumento, la distancia y tiempo que devuelve la llamada son sin tráfico, calculados por la API teniendo en cuenta únicamente las velocidades límites de las carreteras involucradas en la ruta. Si este argumento es dado, nos devolverá las métricas teniendo en cuenta el tráfico por un lado, y las velocidades máximas de las carreteras por otro.