

# Seminario sobre GeoKettle

## Actividad 10

### Enunciado del problema.

Se dispone de un listado de direcciones postales que geocodificar. Este listado está almacenado en un fichero de texto, de tal forma que cada línea es una dirección.

```
1 direccion
2 PASEO DE YESERIAS, 33, , 28005, MADRID
3 AVDA DOCTRO FEDERICO RUBIO Y GALI, 61, , 28040, MADRID
4 BALTAR, 5, A PASTORIZA, 27246, LUGO
5 BARRIO DE COTARELO. MARCO DE ALVARE, A PASTORIZA, 27287, LUGO
6 MESON VIEJO - MARCO ALTO, A PASTORIZA, 27287, LUGO
7 CTRA NACIONAL 120 PKM 511 (ABRENCE), A POBRA DO BROLLON, 27000, LUGO
8 BARRIO LEBARIO, S/N, ABADIANO, 48220, VIZCAYA
9 RESIDENCIA ATXARTE - BO KANTARRIKETAS, 66, ABADIANO, 48220, VIZCAYA
10 TOKI ALAI - CTRA DURANGO A VITORIA, ABADIANO ,VIZCAYA
11 C.E.I.P. DE ABADIN, ABADIN, 27730, LUGO
12 C/ GENERAL FRANCO, 29, ABADIN, 27730, LUGO
13 ABAMIA, ABAMIA (CORAO), 33556, ASTURIAS
```

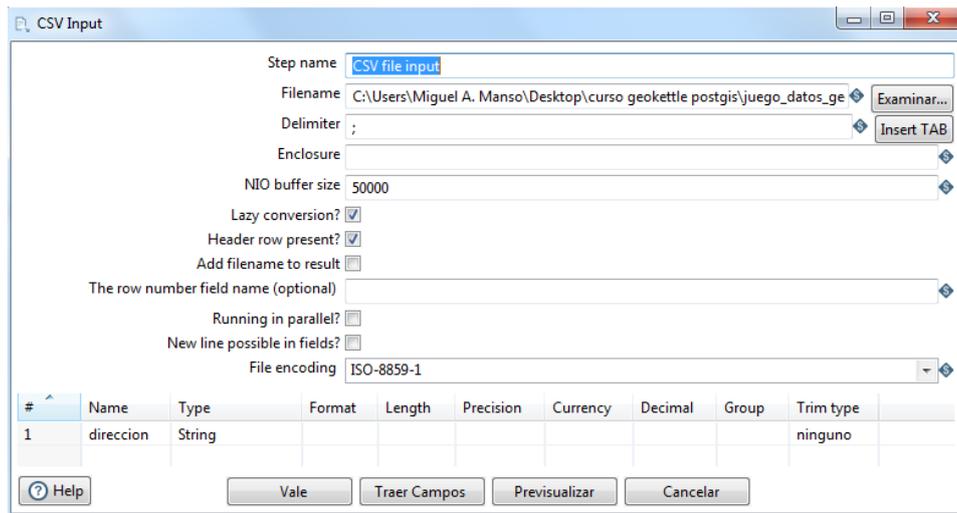
Se desea **generar** una transformación que utilice el servicio de geocodificación de Google (<http://maps.googleapis.com/maps/api/geocode/json?address=...&sensor=false>) para obtener las coordenadas de las direcciones.

### Solución basada en GeoKettle:

Se generará una transformación que obtenga las direcciones de un archivo de texto tipo CSV, se codifiquen los caracteres para remitirlos por una petición GET a un servicio web con cliente HTTP y se obtenga un documento JSON. Seguidamente se procesa los nodos de éste documento y se selecciona la información de interés: Dirección normalizada, coordenadas y nivel geográfico de la misma. Esta información finalmente se almacena en un archivo de texto.

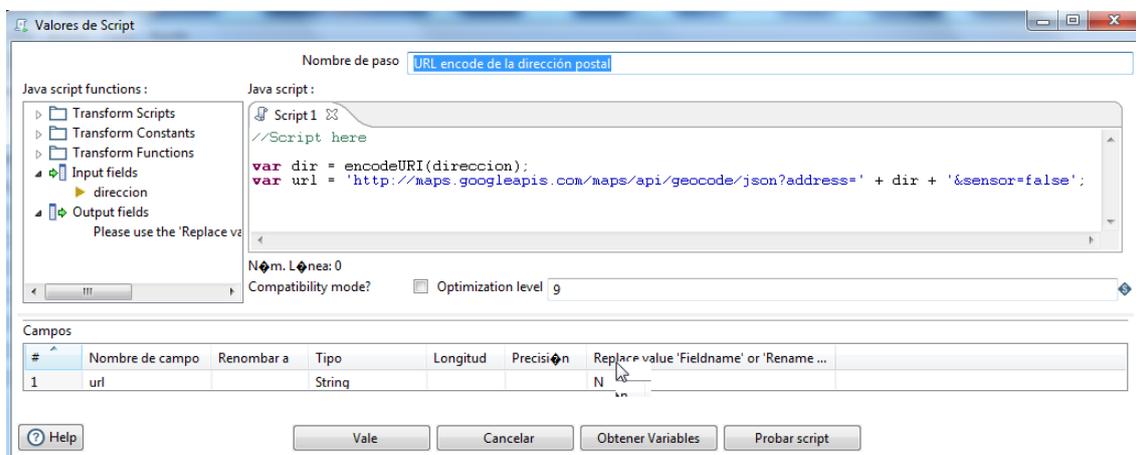


La primera etapa de la transformación es una entrada de tipo CSV file, vinculando la entrada con el documento que contiene las direcciones.

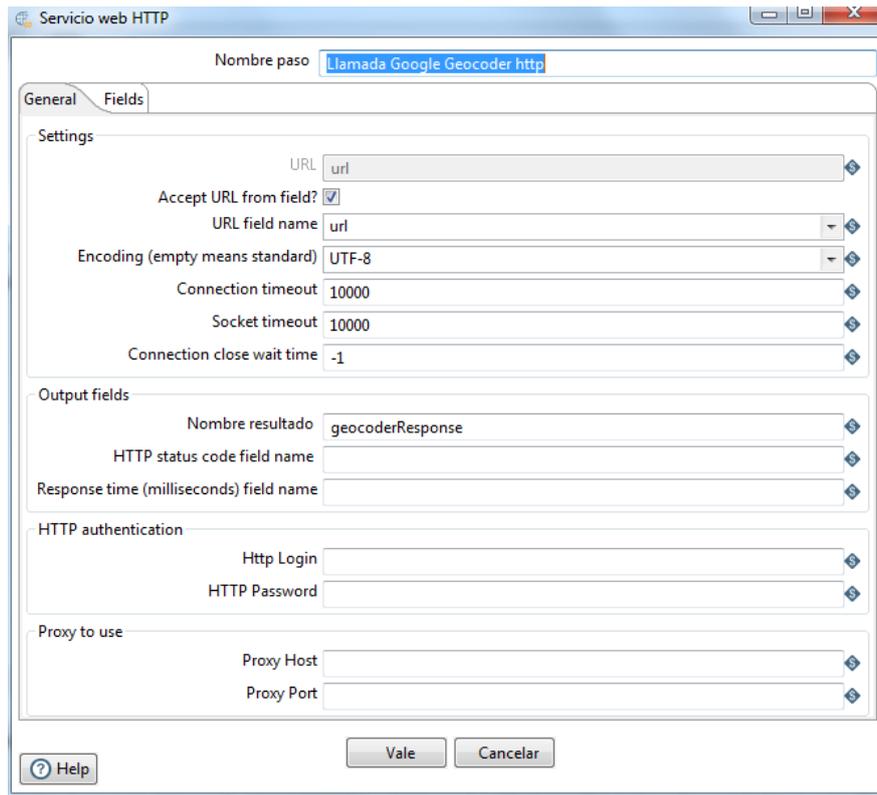


El nombre del atributo con la dirección lo denominaremos *direccion* y el tipo *String*. Es importante indicar que no se cierren o interprete el texto entre comillas si no lo está inicialmente y si en las direcciones aparecen las comas para separar elementos de la misma, se use como elemento delimitador otro que no sea la coma.

La siguiente etapa de tipo, Valores de Script, va a usar una función de JavaScript para codificar la dirección en formato que se pueda enviar por una URL. Se usa la función `encodeURIComponent()` y como se ve en la figura lo que hacemos es construir la url de las peticiones con una parte fija, la url del servicio de Google, la dirección codificada y finalmente otro parámetro fijo.



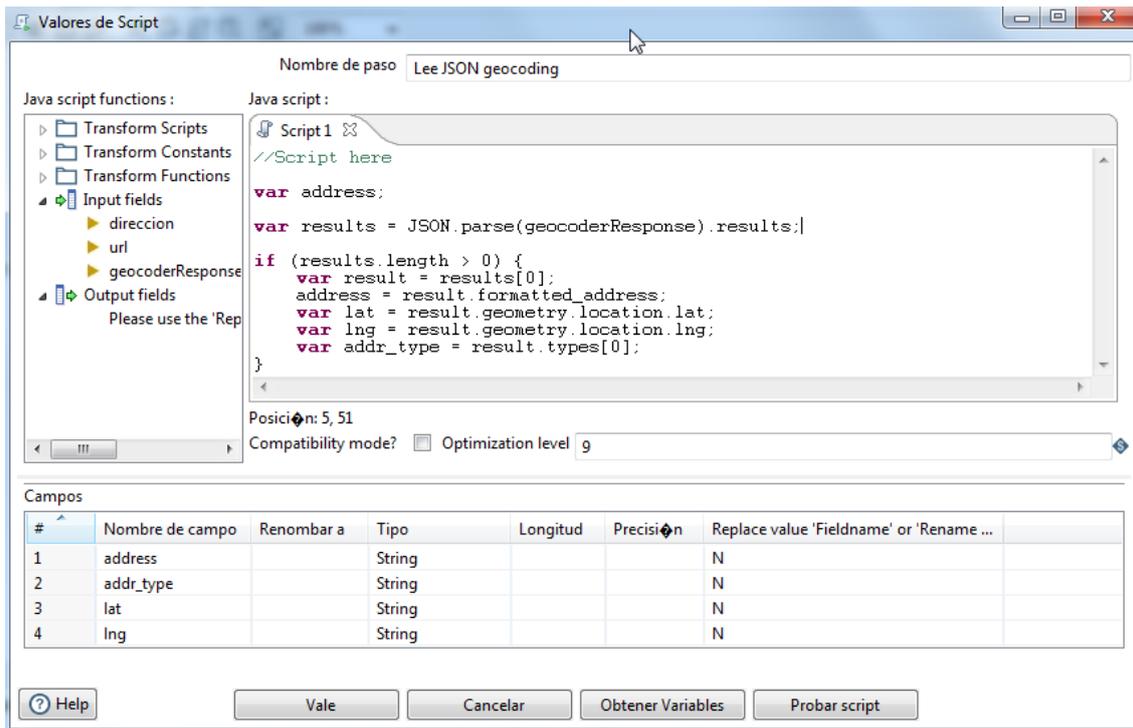
El tercer paso es la petición al servicio web con un paso (Servicio web HTTP), en el que se usa como url la generada en el paso anterior. También se da nombre al resultado (documento JSON), para que lo pueda procesar el siguiente paso (`geocoderResponse`).



El resultado de la consulta es un documento JSON con una estructura como la que se muestra en la siguiente figura.

```
{
  "results" : [
    {
      "address_components" : [
        {
          "long_name" : "1600",
          "short_name" : "1600",
          "types" : [ "street_number" ]
        }
      ],
      "formatted_address" : "1600 Amphitheatre Pkwy, Mountain View, CA 94043, EE. UU.",
      "geometry" : {
        "location" : {
          "lat" : 37.4220352,
          "lng" : -122.0841244
        },
        "location_type" : "ROOFTOP",
        "viewport" : {
          "northeast" : {
            "lat" : 37.42338418029149,
            "lng" : -122.0827754197085
          },
          "southwest" : {
            "lat" : 37.4206862197085,
            "lng" : -122.0854733802915
          }
        }
      },
      "place_id" : "ChIJ2eUgeAK6j4ARbn5u_wAGqWA",
      "types" : [ "street_address" ]
    }
  ],
  "status" : "OK"
}
```

De éste documento nos interesa, la dirección normalizada (formatted\_address), el tipo (types), y las coordenadas geográficas (lng, lat). Para poder extraer esta información se incluye otra etapa de tipo Valores de Script y se procesa geocoderResponse con JavaScript.



En último lugar, se genera una etapa de tipo Salida archivo de texto, para guardar, tanto la dirección original, como la normalizada, el tipo o nivel de geocodificación y las coordenadas.

