Seminario sobre GeoKettle

Actividad 6

Enunciado del problema.

Se dispone de una URL en la web de la que se puede obtener información de localización, descripción y datos climáticos de estaciones meteorológicas de aficionados (Meteoclimatic). El URL del servidor del que se puede descargar la información en formato RSS (XML) es: http://www.meteoclimatic.net/feed/rss/ESAND. A continuación se muestra la primera parte del archivo para observar la estructura.

```
<?xml version="1.0" encoding="ISO-8859-15"?>
<rss version="2.0"   xmlns:content="http://purl.org/rss/1.0/modules/content/"
 <channel>
  <title>Meteoclimatic - RSS</title>   <link>http://meteoclimatic.net/</link>
  <language>es</language>   <ttl>60</ttl>
  <pubDate>Mon, 28 Sep 2015 07:59:18 +0000</pubDate>
  <image>    <title>Meteoclimatic - RSS</title>    <url>http://meteoclimatic.net/
  <docs>http://meteoclimatic.net/index/wp/rss_es.html</docs>

  <item>
   <title>Aguadulce-Roquetas de Mar (Almer&#237;a)</title>
   <link>http://www.meteoclimatic.net/perfil/ESAND0400000004720B</link>
   <pubDate>Mon, 28 Sep 2015 07:47:39 +0000</pubDate>
   <guid>b3b1888d3ed0ffa1449d050123e65c86</guid>
   <description>
       .....
   </description>
   <georss:point>36.8 -2.6</georss:point>
   <geo:Point>
    <geo:lat>36.8</geo:lat>
    <geo:long>-2.6</geo:long>
   </geo:Point>
  </item>
```

Se desea ***convertir esta información en un archivo CSV*** que preserve los atributos almacenados en el documento RSS (XML) y las coordenadas de las estaciones.

Solución basada en GeoKettle:

Consideramos la creación de un trabajo o job y también de una transformación. El trabajo consiste en un paso para establecer una conexión HTTP y recuperar el fichero RSS (XML) de meteoclimatic y posteriormente arranca la trasformación.

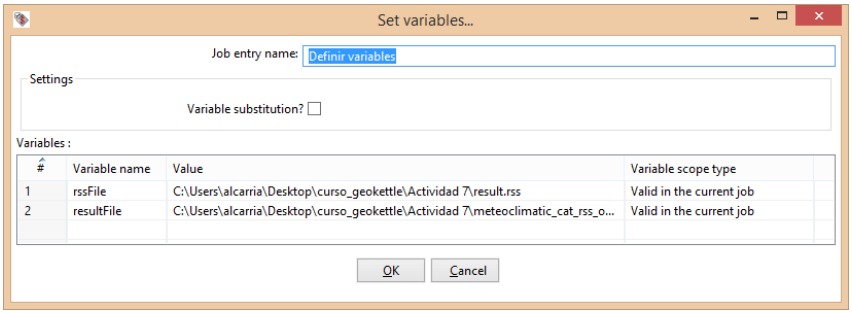
La transformación recibe del trabajo dos variables, que contienen el nombre y dirección del fichero a transformar y el nombre y dirección del fichero que se generará como resultado de la transformación.

La trasformación abre el fichero XML que se obtiene del servidor HTTP, selecciona los elementos del árbol XML con el patrón /rss/chanel/item y seleccione los atributos (en este caso alguno de los atributos del elemento item). Como último elemento de esta transformación es almacenar los datos en un archivo de texto separado por comas (CSV). Esta última etapa se podría haber sustituido por una cadena de 4 pasos para generar un archivo

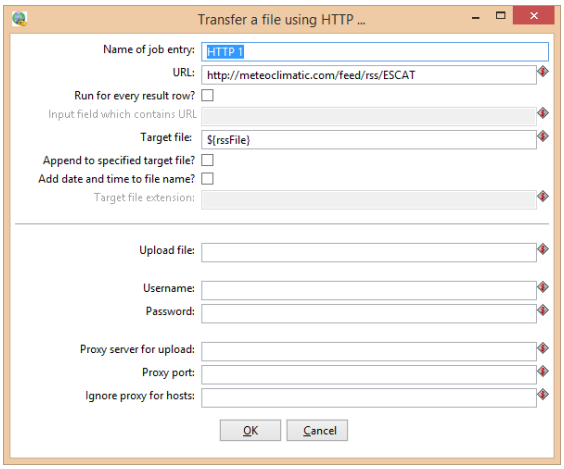
Shapefile en vez del archivo de texto (generar el atributo con la geometría a partir de la longitud y latitud, adaptar el tipo de dato para la geometría a formato Geometria, asignar el sistema de coordenadas (SRID de la geometría) y finalmente almacenar el contenido en el documento shapefile).

A continuación mostramos el trabajo o Job:

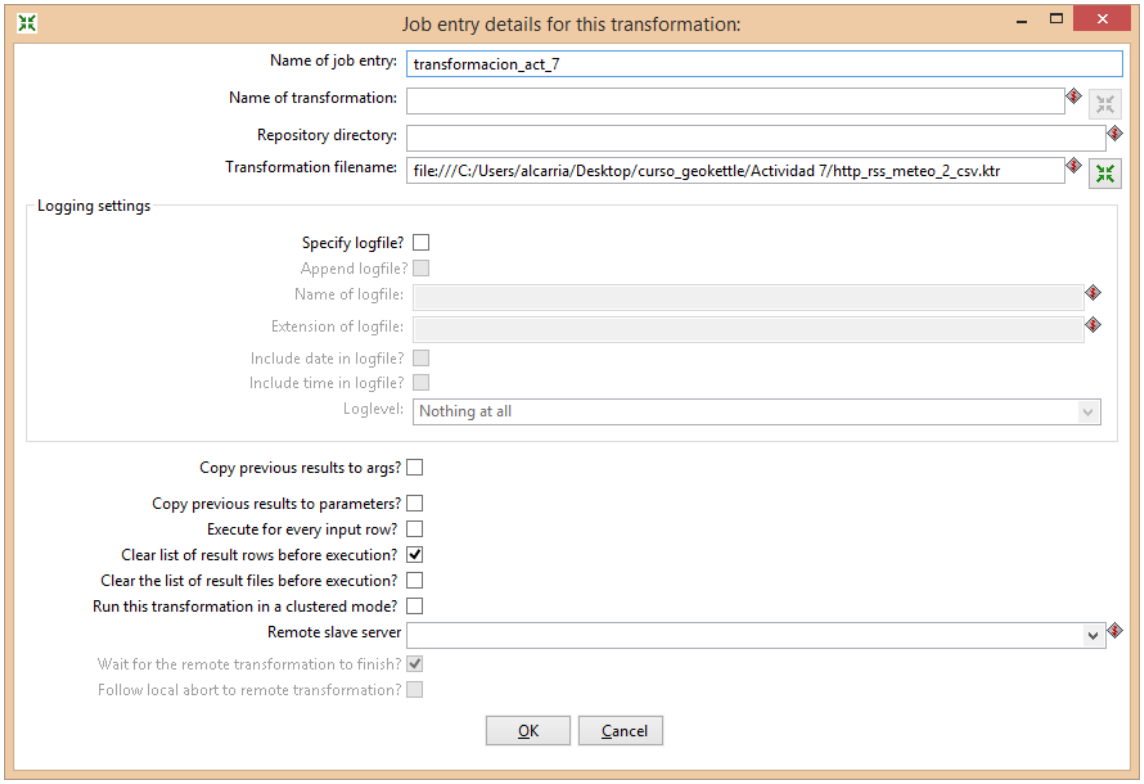
El paso de definir variables debe contener la dirección del archivo que almacenará el XML resultado de la consulta HTTP y la dirección del archivo que se generará como resultado de la transformación:



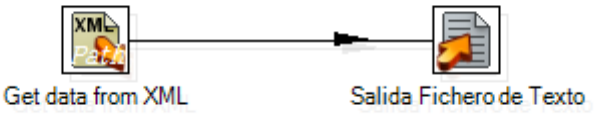
El paso de HTTP se define con la siguiente información:



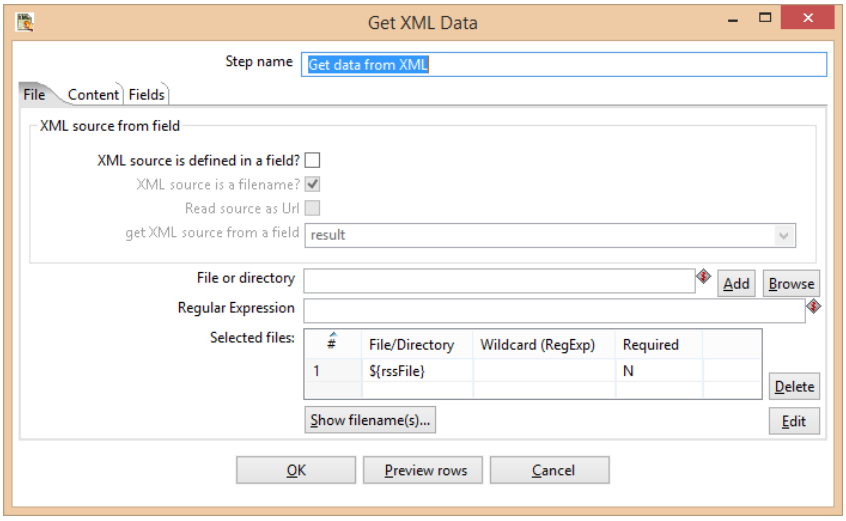
El último paso del trabajo invoca a la transformación:



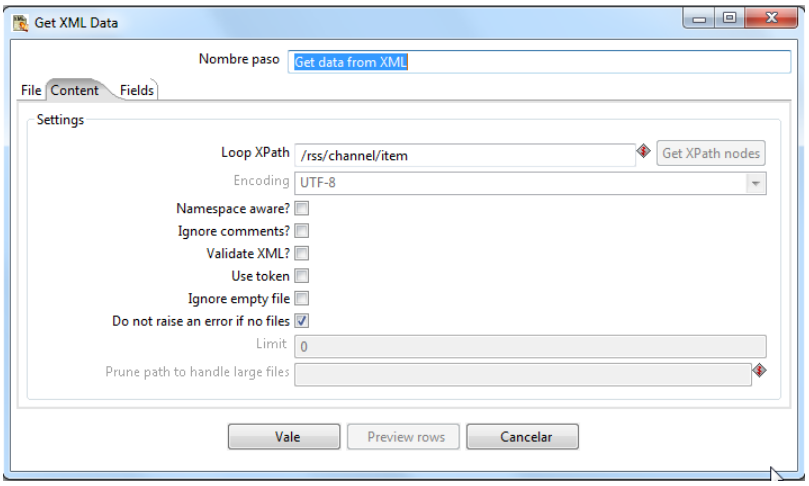
La transformación invocada es muy simple, su estructura se muestra ne la siguiente figura:



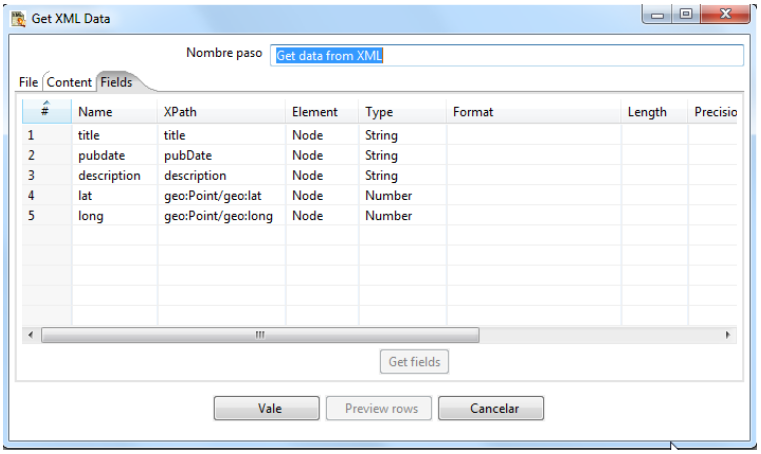
El primer paso procesa el documento XML con los datos de las estaciones meteorológicas. Se selecciona como fuente de datos la variable definida como ${rssFile}



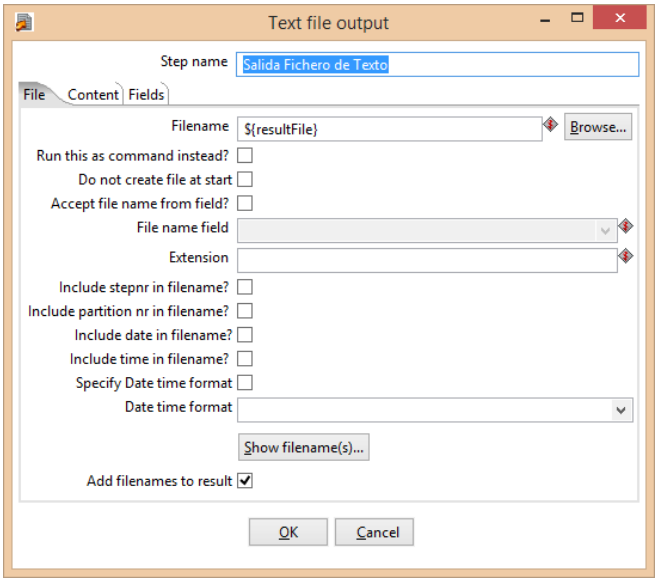
También hay que indicar en ***Content*** el patrón de árbol de etiquetas XML a utilizar (/rss/chanel/ítem). Como se puede apreciar no se puede usar la herramienta GetXpathNodes que en otras ocasiones resulta tan útil.



En tercer lugar hay que indicar en la solapa ***Fields*** los nodos, atributos y tipos de datos se pretende recuperar. Tampoco se puede usar la herramienta Get Fields que nos hubiera ayudado a entender la estructura del documento. Seleccionaremos algunos de los atributos como se muestra en la siguiente figura.



Finamente la etapa de almacenamiento de los datos en un archivo CSV. Se selecciona un paso de tipo Salida de Fichero de Texto, y se establece el archivo donde almacenar los datos como la variable que hemos definido como ${resultFile}.



También los atributos que se incluirán en él, así como el separador entre atributos, si se incluyen las comillas para texto, si se incluyen cabecera con los nombres de los atributos, etc.