



LABORATORIO DE PROCESAMIENTO DE INFORMACIÓN EN APLICACIONES TELEMÁTICAS

Práctica 4.

Curso 2021/2022

Contenido

1	Objetivos	1
2	Descripción.....	1
3	Realización	1
3.1	Especificaciones de la aplicación	1
3.2	Transformación de información y generación del documento de resultados	2
3.3	Restricciones de implementación	2
4	Material disponible	3
5	Entrega.....	3

1 Objetivos

El objetivo principal de esta práctica es la familiarización con la tecnología **GSON Streaming** y desarrollo de un analizador de documentos JSON basado en dicha tecnología.

Como objetivo secundario se pretende capacitar en el diseño de algoritmos eficientes que permitan la extracción y transformación de información a partir de fuentes de contenidos estructurados.

2 Descripción

En esta práctica se pretende ampliar la funcionalidad de la herramienta de búsqueda sobre el [Portal de Datos Abiertos del Ayuntamiento de Madrid](http://datos.madrid.es) (<http://datos.madrid.es>) desarrollada en la práctica 3, dotándole de la capacidad de extraer información sobre los recursos del Portal que se encuentren asociados a una determinada categoría (`concept`), accediendo para ello, a los `conjuntos de datos` (`dataset`) en formato **JSON**.

Con objeto de alcanzar los objetivos especificados, se deberá partir de una copia del código desarrollado en la **P3** y completarlo para que, mediante un **analizador GSON Streaming**, se procese la información de cada `conjunto de datos` pertinentes y se genere un documento XML válido con respecto al esquema de documento **ResultadosBusquedaP4.xsd**, con información de los recursos asociados a alguna de las `categorías` pertinentes.

3 Realización

3.1 Especificaciones de la aplicación

La aplicación (**P4_JSON**) deberá recibir los mismos argumentos que la práctica anteriores, es decir:

- (ARG0) Ruta al documento `catalog.xml`.
- (ARG1) Código de la categoría de la que se desea información.
- (ARG2) Ruta al documento XML de salida.

La aplicación deberá realizar los siguientes pasos:

1. *Verificación y validación de los argumentos de entrada, tal y como se describió en el apartado 3.1.1 de la práctica anterior.*
2. *Obtención de los `concepts` y `datasets` pertinentes, tal y como se describió en el apartado 3.1.2 de la práctica anterior*
3. **Transformación de la información y generación del documento de resultados.** Este será el paso que deberá realizar en esta práctica y que se describe en el siguiente apartado.

3.2 Transformación de información y generación del documento de resultados

Al igual que hizo en la práctica 3, deberá generar un documento XML (ARG2) válido con respecto a `ResultadosBusquedaP4.xsd`. Este documento es idéntico al `ResultadosBusquedaP3.xsd` pero añadiendo el elemento `resources`, que es el que se obtendrá en la realización de esta práctica.

A partir de la obtención de los `dataset` pertinentes, deberá analizar los ficheros `.json` indicados en el atributo `id` de cada `dataset` para realizar la serialización a XML de la información adecuada y poderla incorporar al documento XML resultante de acuerdo al esquema `ResultadosBusquedaP4.xsd`. De cada `dataset` se añadirán como máximo cinco `resource`.

El procedimiento de análisis, extracción y transformación de los ficheros `.json` indicados en los `dataset` para generar el elemento `resources` deberá realizarse a través de un analizador **GSON Streaming** implementado mediante la clase `JSONDatasetParser`

Se usará la API `gson streaming` versión 2.9.0. La documentación sobre esta API se encuentra disponible en la página <http://code.google.com/p/google-gson/>. El fichero `jar` y los `javadoc` pueden descargarse en:

<https://repo1.maven.org/maven2/com/google/code/gson/gson/2.9.0/>.

3.3 Restricciones de implementación

La aplicación deberá implementarse mediante, al menos, las siguientes clases `java` pertenecientes al paquete `piat.opendatasearch`:

- `P4_JSON`: clase inicial de la aplicación, que será idéntica a la clase `P3_SAX` pero añadiendo el código necesario para cumplir con las especificaciones de esta práctica. Deberá cumplir las mismas restricciones de implementación que la práctica 3.
- `ParserCatalogo`: la realizada en la práctica 3.
- `ManejadorXML`: la realizada en la práctica 3.
- `JSONDatasetParser`: parser JSON que utiliza el API GSON Streaming. Debe usar colecciones para almacenar la información pertinente durante el análisis de cada objeto o array del documento JSON. No serán válidas soluciones basadas en tomas de decisiones una vez analizado el contenido de todo el documento. Tampoco será válido usar las API GSON Binding o Tree. El resultado del análisis, almacenado en colecciones, estará accesible a través de métodos públicos que se invocarán desde la clase `P4_JSON`.

Se aplicará el resto de las restricciones de la práctica 3.

4 Material disponible

En el Moodle de la asignatura encontrará el siguiente material para la realización de la práctica:

- Esqueleto de la clase principal: `P4_JSON`.
- Esqueleto de la clase `JSONDatasetParser`.
- Esquema del documento XML que debe generar la aplicación como resultado de la búsqueda: `ResultadosBusquedaP4.xsd`.

5 Entrega

Con anterioridad al 16 de mayo de 2022 a las 23:55 horas, deberá haberse entregado en Moodle, dentro de "Espacio para la entrega de la práctica 4", los siguientes ficheros (sin comprimir):

- Los ficheros Java desarrollados con el nombre del alumno en formato Javadoc (dentro del tag `@author`) antes de la definición de la clase.
- Un fichero XML con la salida de la aplicación al ejecutarla con el código de catálogo 018.