

RETO1 – Por Gerard Luque Oliver

OBJETIVO DEL PROGRAMA:

Este programa tiene como objetivo leer un archivo XML llamado “noticias.xml” con una estructura de datos específica. El programa extrae y muestra:

- *El título del canal.*
- *Los derechos de autor.*
- *El título de cada noticia.*
- *El título de la imagen de la noticia y su atributo (media:title).*
- *El creador de la noticia.*

Se utiliza Java con la biblioteca estándar para el manejo de XML (DOM).

1. Creación del DocumentBuilderFactory y DocumentBuilder

```
DocumentBuilderFactory factory = DocumentBuilderFactory.newInstance();
DocumentBuilder builder = factory.newDocumentBuilder();
Document document = builder.parse(new File("noticias.xml"));
```

- **DocumentBuilderFactory:** Es una fábrica que permite crear un analizador de documentos XML. Usamos newInstance() para obtener una instancia de esta fábrica.
- **DocumentBuilder:** Se crea a partir de la fábrica (factory) y permite analizar el archivo XML. Esto se logra con el método newDocumentBuilder().
- **parse:** Este método analiza el archivo XML especificado (noticias.xml) y devuelve un objeto Document que representa la estructura del XML en memoria.

2. Normalización del Documento

```
document.getDocumentElement().normalize();
```

- La normalización organiza el XML de manera estándar, eliminando espacios en blanco innecesarios y combinando nodos de texto adyacentes. Esto facilita el análisis del contenido.

3. Acceso a la Raíz del Documento

```
Element root = document.getDocumentElement();
NodeList channel = root.getElementsByTagName("channel");
```

- **getDocumentElement():** Obtiene el elemento raíz del archivo XML (en este caso <rss>).
- **getElementsByTagName("channel"):** Encuentra y devuelve todos los elementos <channel> dentro del documento. El resultado es NodeList una colección de nodos que se puede recorrer.

4. Recorrido de los Canales

```
for (int i = 0; i < channel.getLength(); i++) {
    System.out.println("-----> CANAL <-----");
    Element channelElement = (Element) channel.item(i);
```

```

String tituloChannel =
channelElement.getElementsByTagName("title").item(0).getTextContent();
System.out.println("Titulo del Canal: " + tituloChannel);
String copyright =
channelElement.getElementsByTagName("copyright").item(0).getTextContent();
System.out.println("Copyright: " + copyright);
}

```

- Se recorre cada canal dentro del NodeList.
- **channel.item(i)**: Obtiene el elemento específico de la lista de canales.
- **getElementsByTagName("title")**: Busca el nodo <title> dentro del canal .item(0) accede al primer nodo encontrado.
- **getTextContent()**: Extrae el texto del nodo, como el título del canal o el copyright.

5. Acceso a las Noticias

```

for (int j = 0; j < item.getLength(); j++) {
    System.out.println("-----> NOTÍCIA " + (j + 1) + " <-----");
    Element itemElement = (Element) item.item(j);
    String tituloNoticia = itemElement.getElementsByTagName("title").item(0).getTextContent();
    System.out.println("\tEl título de la noticia es: " + tituloNoticia);
}

```

- **getElementsByTagName("item")**: Encuentra todos los elementos <item> dentro del canal actual.
- **For anidado**: Se recorre cada noticia dentro de la lista.
- **item(j)**: Obtiene el nodo <item> actual.
- **getTextContent()**: Extrae el título de la noticia.

6. Procesamiento del medio:título

```

NodeList mediaTitleList = itemElement.getElementsByTagName("media:title");
if (mediaTitleList.getLength() > 0) { // Verificar si media:title existe
    Element mediaTitleElement = (Element) mediaTitleList.item(0);
    if (mediaTitleElement != null) { // Asegurar que el elemento no sea null
        String tituloMedia = mediaTitleElement.getTextContent();
        System.out.println("\tEl título de media:title es: " + tituloMedia);

        // Atributo de media:title
        String atributoMedia = mediaTitleElement.getAttribute("type");
        System.out.println("\tEl atributo de media:title es: " + atributoMedia);
    } else {
        System.out.println("\t----> media:title no tiene contenido <----");
    }
} else {
    System.out.println("\t----> No hay media:title en este item <-----");
}
}

```

- **getElementsByTagName("media:title")**: Busca todos los elementos <media:title> en la noticia.
- **if (mediaTitleList.getLength() > 0)**: Verifica si existe al menos un nodo <media:title>.
- **getAttribute("type")**: Extrae el valor del atributo type dentro del nodo <media:title>.

7. Procesamiento del dc:creator

```
NodeList creador = itemElement.getElementsByTagName("dc:creator");
if (creador.getLength() > 0) {
    Element creatorElement = (Element) creador.item(0);
    if (creatorElement != null) {
        String creator = creatorElement.getTextContent();
        System.out.println("\tEl creador de la noticia es: " + creator);
    }
} else {
    System.out.println("\t----> No hay dc:creator en este item <----");
}
```

- **getElementsByTagName("dc:creator")**: Encuentra los nodos <dc:creator> en la noticia.
- **if (creador.getLength() > 0)**: Verifica si hay al menos un creador definido.
- **getTextContent()**: Extrae el texto del nudo <dc:creator> y lo imprime.

8. Impresión del resultado

A lo largo del código, se imprime información relevante:

1. El título del canal.
2. Los derechos de autor.
3. El título de la noticia.
4. El título de la imagen de la noticia (media:title) y su atributo type.
5. El creador de la noticia (dc:creator).

Cada noticia y su información se separan con líneas visuales para una presentación clara.

```
-----> CANAL <-----
Titulo del Canal: Mercados // expansion
Copyright: (c) 2024, Unidad Editorial, S.A.
*****
-----> NOTICIA 1 <-----
    El título de la noticia es: El Ibex se sostiene con Grifols y los bancos
    El titulo de media:title es: CRÓNICA DE BOLSA
    El atributo de media:title es: html
    El creador de la noticia es: Alejandro Sánchez
=====
-----> NOTICIA 2 <-----
    El título de la noticia es: Los riesgos a vigilar en el rally del bitcoin
    El titulo de media:title es: CRIPTOMONEDAS
    El atributo de media:title es: html
    El creador de la noticia es: Alejandro Sánchez
=====
-----> NOTICIA 3 <-----
    El título de la noticia es: Kepler corta las alas en el Ibex a CaixaBank
    El titulo de media:title es: MERCADOS
    El atributo de media:title es: html
    El creador de la noticia es: Alejandro Sánchez
=====
```

