

Bases de datos - Sesión 7: Datos y XML

XML

XML es un lenguaje de marcas (o etiquetas) que permite disponer un conjunto de elementos y atributos en formato de texto mediante una organización jerárquica de datos semiestructurados.

XML puede emplearse por sí solo (como se hará en esta práctica) o en combinación con documentos *DTD y/o *XSL.

Se puede explorar básicamente documentos XML mediante expresiones *XPath, y también transformarlos o hacer operaciones más elaboradas con *XSLT o *XQuery (aunque estas últimas opciones no se trabajarán en esta sesión por la limitación temporal).

XML se utiliza tanto para la organización de datos en archivos como para bases de datos que integran este formato.

Objetivos de la sesión

- Comprender la sintaxis y la organización de los documentos XML para poder interpretar su estructura y editarlos.
- Entender el funcionamiento de las expresiones XPath para poder hacer consultas básicas de los documentos XML.

Software necesario

Un archivo XML se puede visualizar y editar con un simple editor de texto, pero existen programas específicos para facilitar el trabajo con este tipo de archivos y ofrecer más posibilidades de consulta de los mismos como por ejemplo la evaluación de expresiones XPath.

El software necesario por esta práctica es XML Copy Editor, que se puede descargar para algunos sistemas operativo desde la dirección

<https://xml-copy-editor.sourceforge.io>

Estructura de un document XML

En varias fuentes se ofrece la posibilidad de descargar un conjunto de datos en formato XML, por ejemplo el portal de la Generalitat Valenciana permite descargar varios ficheros XML con datos extraídos para la comunidad:

https://dadesobertes.gva.es/dataset?res_format=XML

Un documento XML se puede ver y editar simplemente con cualquier editor de texto, o también con un programa específico como XML Copy Editor.

Con XML Copy Editor, abre el archivo **incidencia-de-interrupciones-voluntarias-del-embarazo-ive.xml** disponible en el Aula Virtual, observa la estructura jerárquica del documento XML mediante las opciones de contraer (-) y expandir (+) y refleja en un papel esta estructura de árbol.

Edición de un documento XML

En el menú XML del editor, sigue los pasos siguientes:

1. Desactiva la opción "Bloquear etiquetas".
2. A continuación, edita el archivo XML **incidencia-de-interrupciones-voluntarias-del-embarazo-ive.xml**.
3. Añade un nuevo nodo <Row> al final del documento con el contenido siguiente:

```
<Row>
  <PERIODO>2026</PERIODO>
  <COMUNITAT_VALENCIANA>0</COMUNITAT_VALENCIANA>
  <PROV_ALICANTE>0</PROV_ALICANTE>
  <PROV_CASTELLON>0</PROV_CASTELLON>
  <PROV_VALENCIA>0</PROV_VALENCIA>
  <DPTO_SALUD_NOM_USUARI>0</DPTO_SALUD_NOM_USUARI>
</Row>
```

4. Finalmente, comprueba que el fichero esté bien formado haciendo clic al icono de validación (signo de check o "Validate").

XPath: ejemplos e interpretación

XPath es uno de los componentes del estándar XSLT que permite navegar en un documento XML a través de sus elementos y atributos, seleccionar una parte de los datos así como hacer algunos cálculos simples.

Mediante la opción de XML "Evaluar XPath" explica el resultado que se obtiene con cada una de las siguientes expresiones XPath (después de evaluar cada XPath cierra si se ha abierto una nueva pestaña para evaluar siempre lo XPath sobre el archivo original)

Ejemplo: //Row

Muestra el contenido de todos los nodos "Row", que en este caso son todos los datos que aparecen al documento XML.

- 1) //PERIODO
- 2) /Rows
- 3) /Rows/Row
- 4) //DPTO_SALUD_VALENCIA_LAFE

- 5) //Row[1]
- 6) //Row[2]/PROV_VALENCIA
- 7) //Row[PERIODO="2011"]
- 8) //Row[PERIODO="2012"]/DPTO_SALUD_DENIA
- 9) //Row[PERIODO="2013"]/DPTO_SALUD_VALENCIA_HOSPITAL_GENERAL
- 10) //Row[last()]
- 11) //Row[PROV_ALICANTE>10]
- 12) //Row[DPTO_SALUD_XATIVA_ONTINYENT<3]
- 13) count(//Row)
- 14) count(//Row[PROV_CASTELLON<7])
- 15) sum(//Row/DPTO_SALUD_RIBERA)
- 16) //Row[1] | //Row[DPTO_SALUD_DENIA<6]
- 17) //Row[PERIODO="2014"]/PROV_VALENCIA +
//Row[PERIODO="2014"]/PROV_CASTELLON
- 18) 100 * //Row[PERIODO="2014"]/DPTO_SALUD_GANDIA div
//Row[PERIODO="2014"]/DPTO_SALUD_VALENCIA_HOSPITAL_GENERAL
- 19) //PERIODO/text()
- 20) ¿Qué pasaría al eliminar un año del fichero XML?

Trabajo con un segundo documento XML

Abre el archivo **lista-alojamientos.xml** disponible en el Aula Virtual y comprueba si está bien formado. Si no lo estuviera, corrige lo que sea necesario.

Añade al fichero **lista-alojamientos.xml** dos nuevas entradas <Camping>.

Después de hacer los cambios, comprueba que el documento esté bien formado mediante la validación XML.

XPath: ejercicios

Escribe la expresión XPath necesaria para cada caso siguiente:

- 21) Muestra todos los nombres de los campings.
- 22) Muestra todos los datos del primer camping.
- 23) Muestra todos los datos del último camping.
- 24) Muestra todos los datos del camping con ese nombre.
- 25) Muestra el email del camping ALANNIA COSTABLANCA.
- 26) Nombres de campings registrados antes de 2014.
- 27) Campings con categoría cinco estrellas.
- 28) Campings cuyo campo email no está vacío.



- 29) Total de campings listados.
- 30) Suma total de plazas entre todos los campings.
- 31) Multiplica las plazas de LA MARINA CAMPING & RESORT por 5.
- 32) Suma de plazas para campings dados de alta después de 2010.
- 33) Cuenta los campings con categoría "WEB".
- 34) ¿Cuántos campings hay en GUARDAMAR DEL SEGURA con más de 500 plazas?
- 35) ¿Cuál es el número total de alojamientos (campings y fincas) en el XML?
- 36) ¿Cuál es el número total de plazas disponibles solo en campings?
- 37) ¿Cuál es el número total de fincas registradas en el XML?
- 38) ¿Cuál es el número total de alojamientos entre campings y fincas?
- 39) ¿Qué porcentaje del total representan los campings y las fincas?

Enviar al final de la sesión un archivo de texto .txt que contenga:

- **la respuesta/explicación de los puntos 1 a 20**
- **la expresión XPath que permite obtener la respuesta a los puntos 21 a 39**
- **la consulta original tuya/vuestra con su solución mediante Xpath**
- **el texto del documento XML de alojamientos modificado por tú/vosotros**