

XSL y XSLT



Baltasar Fernández Manjón

Dpto. de Sistemas Informáticos y Programación,
Universidad Complutense de Madrid
Avda. Complutense s/n, 28040, Madrid, Spain.

<http://bogart.sip.ucm.es/~balta>

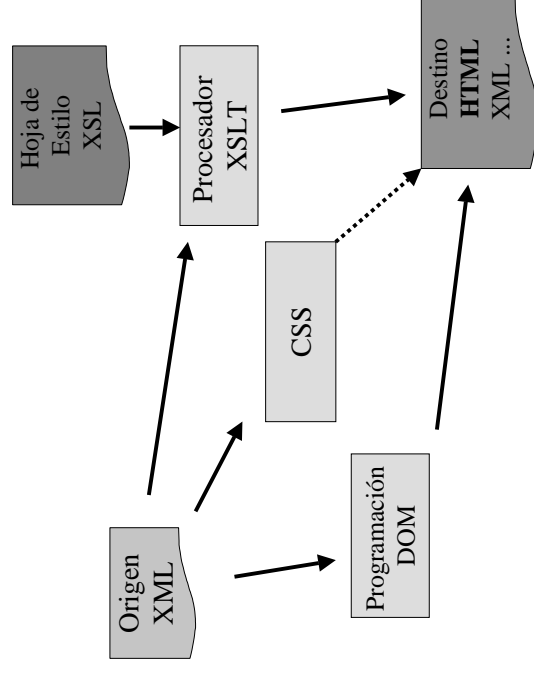
XSL - eXtensible Stylesheet Language

- ❑ Lenguaje de hojas de estilo para XML e implementado en XML
 - Lenguaje de hojas de estilo XSL
 - Incluye un lenguaje de transformaciones XSLT
 - Incluye un lenguaje de formateo XSLFO –XSL Formatting Objects
 - W3C Recommendation en octubre de 2001
- ❑ XSL define como se debe transformar un documento fuente para obtener un documento destino
 - p.e. De un documento XML obtener otro documento XML o WML o HTML
- ❑ Tipos de procesamiento
 - Extracción
 - ♦ Partes del documento original se copian al documento destino
 - Transformación
 - ♦ Ordenaciones, otras modificaciones
- ❑ Utiliza XPath que es un lenguaje que permite definir partes o patrones de un documento XML
 - XSL y XPath son una W3C Recommendation 16. Noviembre 1999

Procesadores y hojas de estilo

- El procesador XSLT es el programa que realiza la transformación
- El procesador utiliza una hoja de estilo XSLT para realizar la transformación.
 - ❑ La hoja de estilo es a su vez otro documento XML
- La hoja de estilo especifica como se deben transformar los elementos de los documentos XML originales para obtener el resultado deseado.
 - ❑ A partir de un árbol de documento origen obtiene un árbol destino
 - ❑ Por ejemplo, en la hoja de estilo se puede especificar:
 - que el resultado deseado es un fichero HTML y
 - como se deben transformar y marcar con etiquetas HTML los elementos del archivo origen.

Transformación de documentos XML



Hojas de estilo XSL

- El elemento `xsl:stylesheet` en una instancia de documento XML referencia una hoja de estilo XSL
 - En general los hijos del elemento `stylesheet` son plantillas (templates)
- Una plantilla especifica un patrón; la plantilla se aplica a los nodos del documento XML origen que coinciden con el patrón
- Se utiliza el espacio de nombres `http://www.w3.org/1999/XSL/Transform` asociado al prefijo `xsl`.
- El elemento raíz de una hoja de estilo XSLT es un elemento `xsl:stylesheet`, e.g.:

- `xsl:transform` es un sinónimo de `xsl:stylesheet`.

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform" >
  plantilla
  ...
  plantilla
</xsl:stylesheet>
```

©Baltasar Fernández Manjón 5

Plantillas en las hojas de estilo

- Cuando se seleccionan nodos o se encuentran nodos que coinciden con el patrón, la plantilla indica al procesador XSLT como transformar dichos nodos para obtener la salida deseada
 - Las plantillas tienen la forma siguiente:
`<xsl:template match="patrón">`
`cuerpo de la plantilla`
`</xsl:template>`
- El patrón es una expresión Xpath que describe los nodos a los cuales se aplica la plantilla
 - El patrón `"/` coincide con el nodo raíz del documento
- El procesador analiza el documento origen buscando nodos que coincidan con el patrón y los reemplaza con el texto (o por otros elementos XSL) incluido en el cuerpo de la plantilla
- En el cuerpo puede haber fragmentos de documentos XML bien formados (que se pasa directamente al resultado) y marcado de XSLT (con prefijo `xsl`) para refinar el comportamiento de la transformación

©Baltasar Fernández Manjón 6

Documento XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml-stylesheet href="hojaestilo.xml" type="text/xsl" ?>
</libros>
</libro>
<titulo>Java2: iniciación y referencia</titulo>
<autor>Jesus Sánchez Allende, Baltasar Fernández</autor>
<descripcion>introducción a la programación en java</descripcion>
</libro>
</libro>
<titulo>Creación de sitios web con XML</titulo>
<autor>Michael Floyd</autor>
<descripcion>Desarrollo de sitios web, programación, XML</descripcion>
</libro>
</libros>
```

Instrucción de procesamiento que le asocia una hoja de estilo xsl

Hoja de estilo XSL para generar una tabla HTML

```
<?xml version="1.0" encoding="iso-8859-1"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform" >
<xsl:template match="/">
<html>
<head><title>XML traducido a HTML</title></head>
<body>
<table border="1" align="center">
<thead><th>Titulo libro</th><th>Autores</th></thead>
<tbody>
<!-- procesamiento de cada libro -->
<xsl:for-each select="libros/libro">
<tr>
<td>
<xsl:value-of select="/titulo" /> </td>
<td>
<xsl:value-of select="/autor" /></td>
</tr>
</xsl:for-each>
</tbody>
</table></body></html>
</xsl:template>
</xsl:stylesheet>
```

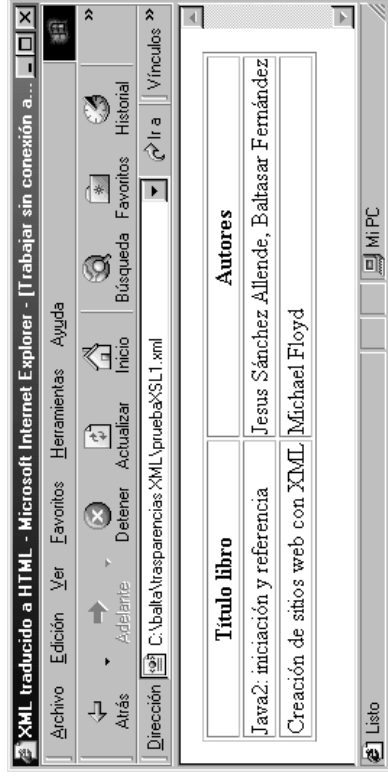
Fichero
hojaestilo.xml

©Baltasar Fernández Manjón 7

©Baltasar Fernández Manjón 8

Resultado

- En la generación del HTML se ha filtrado la descripción de los libros de modo que no aparece.



Espacios de nombres – XML namespaces

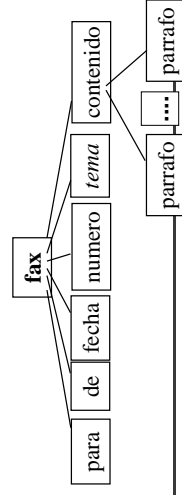
- Permiten hacer públicas las definiciones de estructuras.
 - Se utiliza un atributo especial: xmlns
 - El valor del atributo de xmlns debe ser una URI válida.
- El espacio de nombres define la estructura válida en el ámbito del elemento.
- Para usar tags de HTML en XML:

```
<html xmlns="http://www.w3.org/TR/REC-html40">  
... </html>
```
- Un documento XML puede referenciar varios espacios de nombres

```
<?xml version="1.0"?>  
<persona xmlns="http://mihost/miDTD"  
  xmlns:html="http://www.w3.org/TR/REC-html40">  
<html:p>  
<nombre>Baltasar Fernández</nombre>  
</html:p>  
</persona>
```

Resultado

- XPath es un lenguaje para direccionar, navegar a o seleccionar (referenciar) partes (p.e. elementos, atributos) de un documento XML
 - Se utiliza en XSLT y XPointer.
- XPath permite especificar expresiones basadas en el modelo de árbol (modelo de datos) de un documento
 - Documento XML como árbol de nodos
- Define un lenguaje de expresiones (patrones) para referenciar partes de un documento XML basándose en su estructura
 - Expresiones de camino o rutas de ubicación (path expressions) dentro del árbol



Modelo de datos de XPath

- Un documento XML se considera un árbol ordenado de nodos
 - Existe un nodo raíz
 - Tanto el nodo raíz como los nodos elemento tienen una lista ordenada de nodos hijos
- Existen varios tipos de nodos
 - Nodos raíz
 - Nodos de elementos
 - Nodos de atributos
 - Nodos de texto
 - Nodos de comentario
 - Nodos de instrucciones de procesamiento
 - Nodos de espacios de nombres
- Cada nodo tiene un valor asociado que es una cadena de caracteres
 - El valor puede ser parte del nodo
 - El valor puede obtenerse a partir de los valores de los nodos hijo
- Paradoja
 - Un elemento es el padre de los nodos atributo que contiene pero esos nodos atributo no son hijos del elemento

Funcionamiento de XPath

- El nodo de contexto define el punto de origen para la evaluación de la ruta de localización que puede estar formada por uno o mas pasos de localización
- El eje define la dirección que se toma en la estructura jerárquica a partir del nodo de contexto
 - Inicialmente se devolverían como resultado todos los nodos del tipo elemento situados a partir del nodo de contexto siguiendo dicho eje
- La condición de nodo permite refinar o filtrar el conjunto de nodos resultado a sólo aquellos para los que dicha condición sea cierta
- Los predicados permiten refinar aún mas el conjunto de nodos resultante de la aplicación del contexto-eje-condición
- Una vez ejecutado el paso de localización, y si no hay mas pasos, se devuelve el conjunto de nodos resultado. Si hay mas pasos de localización cada uno de los nodos resultado se utilizan como nodos de contexto para el siguiente paso de localización

13

©Baltasar Fernández Manjón

Expresiones XPath

- La principal construcción XPath es la expresión
- Cuando se evalúa una expresión se obtiene como resultado un objeto cuyo tipo puede ser
 - Conjunto de nodos o nodeset (colección desordenada de nodos no repetidos)
 - boolean (true o false)
 - number (un número real)
 - String (una secuencia de caracteres)
- Las expresiones se evalúan con respecto a un contexto
- En casos simples una expresión XPath tiene una apariencia similar a una especificación de ruta en Unix, donde los subdirectorios son los elementos anidados:
 - "/" es el elemento raíz del documento
 - Las expresiones pueden ser absolutas (relativas a la raíz) o relativas a un nodo de contexto

14

©Baltasar Fernández Manjón

Rutas o caminos de localización

- El principal tipo de expresiones en XPath para referenciar nodos o partes de un documento
 - Devuelve un conjunto de nodos (nodeset)
- Una ruta de localización está compuesta por:
 - Un nodo contexto (punto origen)
 - Un eje (dirección de movimiento)
 - Un destino determinado por uno o varios pasos de localización
- Una ruta de localización puede ser absoluta (respecto al nodo raíz) del documento, o relativa (respecto al nodo de contexto).
 - ruta absoluta: /paso/paso/paso/...../paso
 - ruta relativa: paso/paso/paso/...../paso

15

©Baltasar Fernández Manjón

Paso de localización

- Un paso en una ruta de localización tiene tres partes:
 - El eje de movimiento.
 - La condición de nodo que especifica el tipo de los nodos seleccionados a través del eje.
 - Cero o más predicados que permite refinar el conjunto de nodos seleccionados.
- Sintaxis no abreviada:
 - Eje :: Condición
 - Eje :: Condición [Predicado] [Predicado] ... [Predicado]

16

©Baltasar Fernández Manjón

- Definición de un conjunto de nodos relativa a un nodo

○ Valores

- self:: El nodo de contexto en sí.
- child:: Los hijos del nodo de contexto.
- parent:: El padre del nodo de contexto.
- descendant:: Los descendientes del nodo de contexto.
- ancestor:: Los antecesores del nodo de contexto.
- following-siblings::
- preceding-siblings::
 - Hermanos posteriores y anteriores del nodo de contexto.
- following::
- preceding::
 - Nodos anteriores y siguientes al nodo de contexto en el orden del documento (sin incluir antecesores ni descendientes respectivamente)
- attribute:: Atributos del nodo de contexto.
- namespace:: Espacio de nombres.
- descendant-or-self::
- ancestor-or-self::
 - Los descendientes (antecesores) del nodo de contexto y el propio nodo de contexto.

17

©Baltasar Fernández Manjón

- Un nombre (posiblemente cualificado)
 - El asterisco * permite referir a cualquier elemento o atributo
- Condiciones sobre el tipo de nodo (devuelven true si el nodo es del tipo correspondiente)
 - text().
 - comment().
 - processing-instruction().
 - node().

/child::* → selecciona todos los nodos elemento hijos del nodo raíz (en este caso el nodo documento)

child::párrafo → selecciona los nodos párrafo hijos del nodo de contexto

child::text() → selecciona los nodos de texto hijos del nodo de contexto

child::node() → selecciona todos los nodos hijos del nodo contexto (y no sólo los nodos elemento hijos)

attribute::* → selecciona todos los atributos del nodo de contexto

attribute::moneda → selecciona el atributo moneda del nodo contexto

child::*[child::párrafo] → selecciona todos los nodos párrafo "nietos" del contexto

18

©Baltasar Fernández Manjón

- Definición de un conjunto de nodos relativa a un nodo

○ Valores

- self:: El nodo de contexto en sí.
- child:: Los hijos del nodo de contexto.
- parent:: El padre del nodo de contexto.
- descendant:: Los descendientes del nodo de contexto.
- ancestor:: Los antecesores del nodo de contexto.
- following-siblings::
- preceding-siblings::
 - Hermanos posteriores y anteriores del nodo de contexto.
- following::
- preceding::
 - Nodos anteriores y siguientes al nodo de contexto en el orden del documento (sin incluir antecesores ni descendientes respectivamente)
- attribute:: Atributos del nodo de contexto.
- namespace:: Espacio de nombres.
- descendant-or-self::
- ancestor-or-self::
 - Los descendientes (antecesores) del nodo de contexto y el propio nodo de contexto.

17

©Baltasar Fernández Manjón

Predicados

- Es una expresión que se utiliza para filtrar un conjunto de nodos resultado
 - Se seleccionan los nodos resultado que cumplen el predicado
 - Los predicados se escriben entre corchetes
 - En los predicados se utilizan expresiones y funciones
 - Habitualmente el orden de los predicados afecta al conjunto de nodos resultante

child::párrafo[position()=1] → selecciona el primer nodo párrafo hijo del nodo de contexto

child::párrafo[position()=last()] → selecciona el último nodo párrafo hijo del nodo de contexto

child::libro[child::precio[@attribute=moneda]] → selecciona todos los nodos hijos precio hijos del nodo libro del nodo contexto que posean un atributo moneda

child::libro[child::precio[@attribute=moneda]] [position()=2] → selecciona el segundo de los nodos hijos precio hijos del nodo libro del nodo contexto que posean un atributo moneda

child::libro[child::precio [position()=2] [attribute=moneda]] → selecciona el segundo de los nodos hijos precio hijos del nodo libro del nodo contexto si posee un atributo moneda

19

©Baltasar Fernández Manjón

Sintaxis abreviada

- Las expresiones Xpath se pueden simplificar utilizando una sintaxis abreviada
 - El eje child:: se considera eje por defecto y se puede omitir
 - El eje attribute:: se puede abreviar a @
 - El eje self:: se abrevia a un punto .
 - El eje parent:: se abrevia a dos puntos consecutivos ..
 - El eje descendant-or-self se abrevia a //
 - [position(=número)] se abrevia a [número]
 - El nodo raíz se abrevia a /

/ * → selecciona todos los nodos elemento hijos del nodo raíz (en este caso el nodo documento)

párrafo → selecciona los nodos párrafo hijos del nodo de contexto

text() → selecciona los nodos de texto hijos del nodo de contexto

node() → selecciona todos los nodos hijos del nodo contexto (y no sólo los nodos elemento hijos)

@ * → selecciona todos los atributos del nodo de contexto

párrafo[1] → selecciona el primer nodo párrafo hijo del nodo de contexto

libro/precio[@moneda] → selecciona todos los nodos hijos precio hijos del nodo libro del nodo contexto que posean un atributo moneda

20

©Baltasar Fernández Manjón

Ejemplo

```
<?xml version="1.0" encoding="iso-8859-1"?>
<?xml-stylesheet href="hojaestilo2.xsl" type="text/xsl" ?>
<libros>
  <libro>
    <titulo>Java2: iniciación y referencia</titulo>
    <autor>Jesus Sanchez Allende, Baltasar Fernández</autor>
    <descripcion>introducción a la programación en Java</descripcion>
    <precio moneda="euros">24</precio>
  </libro>
  <libro>
    <titulo>Creación de sitios web con XML</titulo>
    <autor>Michael Floyd</autor>
    <descripcion>Desarrollo de sitios web, programación, XML</descripcion>
    <precio moneda="pesetas">5000</precio>
  </libro>
</libros>
```

©Baltasar Fernández Manjón

25

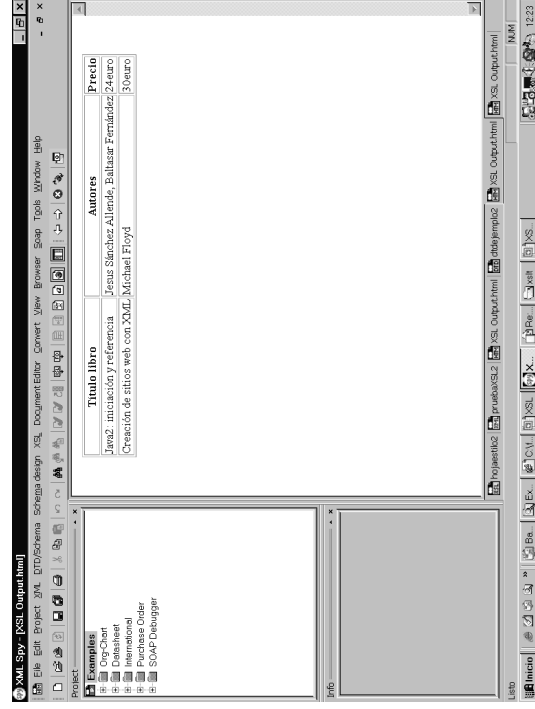
Ejemplo

```
<?xml version="1.0" encoding="iso-8859-1"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform" >
<xsl:template match="/">
<html><head><title>XML traducido a HTML</title></head><body>
<table border="1" align="center">
<thead><th>Titulo libro</th><th>Autores</th><th>Precio</th></thead>
<tbody>
<!-- procesamiento de cada libro -->
<xsl:for-each select="libros/libro" >
  <tr>
    <td>
      <xsl:value-of select="./titulo" />
    </td>
    <td>
      <xsl:value-of select="./autor" />
    </td>
    <td>
      <xsl:value-of select="./precio" />
    </td>
  </tr>
</tbody>
</table></body></html></xsl:template></xsl:stylesheet>
```

©Baltasar Fernández Manjón

26

Resultado



©Baltasar Fernández Manjón

27