

Intro Express a XML

Susanna Allés Torrent
Training School Menorca
16-24 Septiembre 2014

XML (1)

- ❖ XML = Extensible Markup Language
- ❖ XML es un heredero, más sencillo y eficaz, de su antecesor SGML (Standard Generalized Markup Language)
- ❖ Se publicó en febrero de 1998 por el W3C Consortium.
- ❖ La documentación oficial está disponible en: <http://www.w3.org/XML/>
- ❖ Última versión: [XML 1.0 \(5ª edición\)](#)



XML (2)

- ❖ Es un estándar web
- ❖ Es “extensible” y compatible con la mayoría de lenguajes web (tipos de documento, de metadatos, de modelización, de presentación, de programación, de protocolos...)
- ❖ Es el más usado para estructurar informaciones que pretendan durar en el tiempo y ser interoperables con otras aplicaciones, plataformas, etc.
- ❖ A partir de las directrices XML, se creó [XHTML 1.0](#) (enero 2000), adaptación de HTML 4 (aunque ahora ya en la versión HTML 5).

XML (3)

- ❖ simple, flexible y “human-readable”
- ❖ Soporta diferentes sistemas de escritura; utiliza el estándar Unicode, por defecto, el marcado es UTF-8.
- ❖ ampliamente utilizado (soporte y herramientas)
- ❖ [ISO](#) estándar (International Organization for Standardization)
- ❖ Independiente de cualquier software
- ❖ Facilita el movimiento de datos (interoperabilidad)
- ❖ Separa contenido de presentación
- ❖ Enorme cantidad de tutoriales en línea, entre los cuales: [Introducción a XML elaborada por TEI](#)

XML (4): Un document XML

- ❖ tiene que estar Bien formado, es decir estar expresado correctamente según la gramática de XML
- ❖ tiene que ser Válido  Document is well formed.
 - Para ello, necesita un modelo que fije la estructura básica y las posibilidades de anidación del documento. Las DTD o esquemas (RelaxNG, W3CSchema) establecen la sintaxis que debe tener el documento XML. Si este está formado según sus reglas, será válido.  Document is valid.

XML (5)

- ❖ Noción de marcado = “Markup” (en.), “Balisage” (fr.), “Codifica” (it.) → aislar una porción de texto con un significado semántico específico.
- ❖ ‘Tag’ = ‘Etiqueta’
 - `<etiqueta>concepto</etiqueta>`
 - `<titulo>Introduccion a XML</titulo>`
 - `<autor>Susanna Allés</autor>`
 - `<fecha>25 de junio de 2014</fecha>`
- ❖ La estructura consiste en la anidación sin límites de unos elementos al interior de otros, construyendo así una especie de árbol genealógico. Por eso, necesitamos siempre un solo elemento raíz del que cuelguen todos los otros.

Véase: documento_xml_basico.xml

XML (6): estructura de una etiqueta

Elemento: son las unidades de información semántica compuestas de una marca de apertura `<elemento>` y de cierre `</elemento>`. El nombre del elemento constituye su identificador general. Puede contener texto, otros elementos o estar vacío (`<elemento/>`).

Atributo: añaden una característica adicional al elemento; su orden no es relevante; deben siempre llevar un valor.

Valor: rellena el valor o contenido del atributo.

```
<elemento atributo="valor">contenido</elemento>
```

```
<date when="0972">972</date>
```

```
<div type="prologo">El prólogo... </date>
```

XML (7): Algunas reglas

- ❖ el nombre del elemento puede solo contener letras y números, sin acentos ni espacios, y solo algunos signos: - _ .
- ❖ el nombre no puede empezar por un número, ni por un signo, ni por “xml”
- ❖ XML es sensible a las mayúsculas y las minúsculas!

XML (8): Otros elementos

- ❖ Entidades: caracteres especiales que no deben ser interpretados como marcado.

Por ejemplo, determinados signos deben ser codificados de una manera especial:

`<` = `lt`

`>` = `gt`

`&` = `amp`

- ❖ Comentarios: se indican con

`<!-- el comentario -->`