



XML

Una breve introducción



eXtensible Markup Language (XML)

- <http://www.w3.org/XML/>
- Estándar para el intercambio de información estructurada entre distintas plataformas
- Utilizado para
 - ☐ Docx, odt,...
 - ☐ Servicios web: SOAP
 - ☐ ...
 - ☐ Bases de datos (eXists,...)

Estructura de un documento XML

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Edit_Mensaje SYSTEM "Edit_Mensaje.dtd">
```

Encabezado

Puede incluir información sobre el esquema o DTD que se utiliza

```
<Edit_Mensaje>
  <Mensaje>
    <Remitente>
      <Nombre>Nombre del remitente</Nombre>
      <Mail> Correo del remitente </Mail>
    </Remitente>
    <Destinatario>
      <Nombre>Nombre del destinatario</Nombre>
      <Mail>Correo del destinatario</Mail>
    </Destinatario>
    <Texto>
      <Asunto>
        Este es mi documento con una estructura muy sencilla no contiene atributos ni entidades...
      </Asunto>
      <Parrafo>
        Este es mi documento con una estructura muy sencilla no contiene atributos ni entidades...
      </Parrafo>
    </Texto>
  </Mensaje>
</Edit_Mensaje>
```

Cuerpo



Validez de un documento XML

■ Documento bien formado

- ☐ Las etiquetas deben estar balanceadas
 - `<formato>MP3</formato>`
- ☐ Campos vacíos finalizan con “/”
 - `<comentarios/>`
- ☐ ...

■ DTD o XML Schema

- ☐ Permiten definir las etiquetas y su estructura



Como procesar un documento XML

■ Document Object Model (DOM)

- <http://www.w3.org/DOM/>
- Se construye un árbol con el contenido del documento XML
- XPATH (<http://www.w3.org/TR/xpath/>) para navegar el árbol
- Costoso – requiere mucha memoria

■ Simple API for XML (SAX)

- <http://www.saxproject.org>
- Se procesan las etiquetas – más eficiente



SAX

■ Métodos

- ☐ *procesarComienzoDocumento*
- ☐ *procesarComienzoElemento*
- ☐ *procesarTexto*
 - Tratar el texto de un elemento
- ☐ *procesarFinalElemento*
- ☐ *procesarFinalDocumento*



JDOM

- Librería para el proceso de documentos XML
 - Soporte DOM y SAX
 - <http://www.jdom.org>