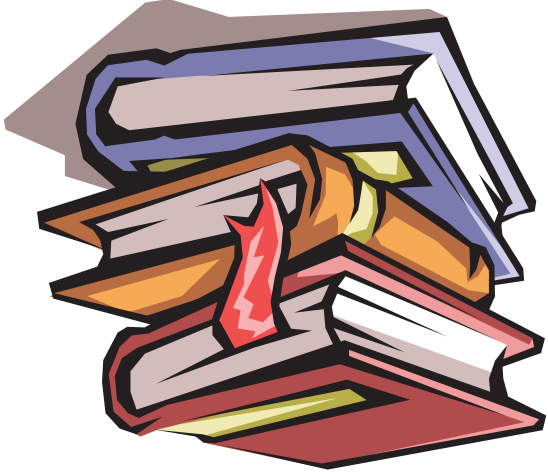


# (I) Introducción: La evolución hacia XML



# Guion

- Introducción a los lenguajes de marcas
- La evolución hacia XML

– SGML

– HTML y XHTML

– XML

– XHTML

# El documento digital

- La diferencia entre bases de datos con información estructurada y las bases textuales se ha diluido, La representación o codificación de información – estructurada y no estructurada – ha aceptado al lenguaje XML como un estándar
- Utilizamos el mismo “lenguaje” o “formato”, XML, para:
  - Codificar y transferir metadatos (información estructurada)
  - Codificar y transferir contenidos (información no estructurada)
  - ¿cuál ha sido la evolución?

## El documento digital - Formatos

- En la actualidad existen distintos formatos disponibles:
  - Basados en lenguajes de marcas: SGML/HTML/XML, a los que se llama 'la opción semántica'.
  - Formatos de imagen - TIFF
  - Formatos de réplica o portables - PDF
  - Formatos para el intercambio - MIF, RTF, ...
  - y un largo etcétera...

## Paradoja de la compatibilidad

- La existencia de distintos formatos originó la llamada 'paradoja de la compatibilidad':  
*“A medida que el número de materiales disponibles en soporte digital aumentaba, también se hacían mayores las dificultades para acceder a los mismos”*
- Para solucionar este problema, se comenzó a trabajar a favor de la normalización de formatos
- La normalización nos lleva... a XML

# El camino hacia la normalización

☀ ¿el camino hacia la normalización...

• 1974 - SGML

• 1985 - ODA/ODIF

• 1994 - HTML

• 1996 - XML

☀ o hacia la

desnormalización?

☀ Distintos formatos para distintas fases del ciclo de producción

- Los formatos se complementan
- La evolución ha favorecido a los formatos:
  - simples,
  - con mayores posibilidades de interacción entre aplicaciones e intercambio de información.
- El tratamiento de formatos forma ya parte de la GED y de todas las aplicaciones y sistemas de gestión (incluyendo bibliotecarias)

# SGML Structured Markup Language

- Creado en 1969 en *IBM* por *Charles Goldbarb*
- En 1986 se convierte en norma ISO 8879-1986
- Se basa en el 'marcado descriptivo' o 'generalizado', que indica cómo se deben intercalar marcas en un documento para diferenciar sus componentes estructurales.
- Las marcas no señalan cómo se debe presentar el documento en pantalla ni cómo se tiene que formatear el documento al imprimirlo.
- Es un Metalenguaje

# SGML Structured Markup Language

- SGML no propone un conjunto de marcas predefinidas, sino:

– la sintaxis que debe utilizarse para definir un conjunto de marcas aplicables a los documentos de un mismo tipo (aplicación)

– la forma en la que se debe intercalar estas marcas en los documentos

– los juegos de caracteres que se pueden utilizar en los documentos (ISO 646 e ISO 10646)



# SGML Structured Markup Language

- Los elementos que pueden aparecer en un tipo de documento, sus características y el orden en el que deben escribirse se definen en un documento aparte llamado DTD (Document type definition)
- Los documentos se consideran 'instancias' de un tipo de documento específico que define su estructura válida

- Un documento SGML siempre debe cumplir las restricciones que se indican en su DTD

- Un documento SGML debe incluir una referencia a la DTD a partir de la cual se ha definido

# SGML Structured Markup Language

- *SGML* tiene capacidad hipertexto basadas en la norma *HyTime*

- Los enlaces *SGML* unen un elemento origen con un elemento destino

- El elemento origen tendrá un atributo de tipo *IDREF* (por convención se llama *link* o *xref* a este atributo)

- El elemento destino tendrá un atributo de tipo *ID* (por convención se llama *target*)

# SGML Structured Markup Language

Ejemplo...

```
<nota  
  link="nota_1">Comentarios  
  de Rodríguez Sáez</nota>  
<textonota target="nota_1">La  
  edición de 1994 contiene  
  errores de  
  interpretación...</textonota>
```

# SGML Structured Markup Language

- Para imprimir un documento SGML o verlo en pantalla, es necesario aplicarle un formato
- El formato indicará cómo se debe formatear cada uno de los elementos que forman el documento: tipo de letra, márgenes, etc.
- Se han diseñado dos normas que indican cómo se deben crear hojas de estilo para documentos SGML:
  - DSSSL (Document Style Semantic and Specification Language)
  - FOSI (Format Output Specification Interface)

# SGML Structured Markup Language

- En el éxito del lenguaje SGML ha jugado un papel trascendental dos hechos:

- La decisión del DoD (Department of Defense) de los EUA de adoptar SGML en su proyecto CALS (Computer-aided Acquisition and Logistic Support)
- El desarrollo del lenguaje HTML y la ‘explosión’ del World Wide Web

- En los dos últimos años, XML ha devuelto el interés por SGML, llevando éste a XML

# HTML Hypertext Markup Language

- HTML es una aplicación SGML para codificar documentos y distribuirlos en el World Wide Web
- HTML define un conjunto limitado de marcas que se pueden intercalar en los documentos
- Comparte las ventajas de SGML: multipataforma, fácilmente procesable, etc.
- Ha alcanzado un gran éxito, hasta el punto de llegar a igualarse edición electrónica con edición HTML

# HTML - Limitaciones

- La simplicidad de HTML acarrea limitaciones:
  - Falta de fórmulas de compresión asociadas al formato (problema ancho de banda)
  - Posibilidades de formato muy limitadas
  - Falta de mecanismos de acceso: búsqueda texto completo, tablas de contenidos, etc.
  - Escasa capacidad expresiva del lenguaje
  - Disponemos de browsers y tecnologías (Java, lenguajes de script, etc.) muy potentes cuya capacidad está siendo infrutilizada por las limitaciones del formato HTML

## dHTML – HTML dinámico

- La evolución de HTML ha estado condicionada por la presión ejercida por los fabricantes de los browsers más utilizados

- dHTML hace referencia a unas características soportadas por la versión 4 de los browsers de Netscape y Microsoft

- Algunas de estas características se tomaron del borrador disponible para la versión 4 del formato HTML



## dHTML – HTML dinámico

- dHTML permite:
  - Control sobre los elementos que conforman una página HTML mediante el DOM (Document Object Model)
  - Posibilidad de formatear los elementos de la página
  - Posibilidad de modificar el formato de los elementos en respuesta a acciones realizadas por el usuario
  - Control de la posición de un elemento u objeto dentro de la página
  - Capacidad de cambiar el contenido de la página una vez ésta ha sido descargada por el navegador

## dHTML - HTML dinámico

- Problemas de dHTML

- Orientado a la presentación de los documentos y a aspectos 'visuales'
- No representa el contenido semántico ni la lógica de la información que contiene el documento
- Las 'implementaciones' de Microsoft y de Netscape no son idénticas: las mismas hojas de estilo son interpretadas de forma distinta por los navegadores, problemas en soporte a lenguajes de script, controles ActiveX, etc.

# XML: eXtensible Markup Language

- Su desarrollo comienza en septiembre de 1996 dirigido por el W3C y con la participación de importantes empresas: Microsoft, IBM, Sun, Novell, ArborText, H-P etc.
- El propósito es:
  - diseñar un lenguaje de marcas optimizado para el WWW
  - unir la simplicidad de HTML con la capacidad expresiva de SGML
- Versión 1.0 ratificada en diciembre de 1997

# XML: eXtensible Markup Language

- Áreas de aplicación:
  - Representación y distribución de documentos e información textual
  - Intercambio de datos e información estructurada a través de Internet y el WWW
  - Integración de datos procedentes de fuentes heterogéneas
- Elimina la 'barrera' entre información estructurada e información textual

# XML: eXtensible Markup Language

- XML
  - Perfil de SGML
    - No especifica etiquetas, sino cómo deben definirse conjuntos de etiquetas aplicables a un tipo de documento
  - Modelo de hiperenlaces complejo (múltiples destinos, hijos y relativos, etc.)
- HTML
  - Aplicación de SGML
    - Conjunto limitado de etiquetas y un único tipo de documento
  - Modelo de hiperenlaces simple (unidireccionales y hijos)

# XML: eXtensible Markup Language

- XML
  - Gran capacidad para procesar documentos, el browser es una plataforma para el desarrollo de aplicaciones
  - Fin de la guerra de los navegadores y etiquetas propietarias
- HTML
  - Escasa capacidad de procesamiento, el browser es un mero visor de páginas
  - El problema de la 'no compatibilidad' y las diferencias entre browsers ha alcanzado un punto en el que la solución es difícil

# XML: eXtensible Markup Language

- XML
  - Mayor simplicidad
  - Eliminación de las etiquetas opcionales
  - Un documento no debe ser validado obligatoriamente, basta con que esté 'bien formado'
  - XML es un documento válido y bien formado.
- SGML
  - Gran complejidad que dificulta su tratamiento e implementación
  - Validez requerida: todos los documentos deben cumplir las restricciones de su DTD

# XML: eXtensible Markup Language

- XML
  - La simplicidad de XML hace más fácil el desarrollo de aplicaciones de bajo coste
  - Amplio soporte de la industria informática y más áreas de aplicación
- SGML
  - Su complejidad hace que las aplicaciones informáticas para procesar SGML sean muy costosas
  - Escaso impacto al margen de sectores muy específicos: editoriales, doc.técnica



# XML: eXtensible Markup Language

- XML
  - Compatibilidad e integración con HTML (soporte a CSS y data-islands)
  - Formato y estilos fáciles de aplicar mediante:
    - CSS
    - XSL
- SGML
  - No hay una HTML definida
  - Formato y estilos relativamente complejos (norma DSSSL)

# XML: eXtensible Markup Language

- XML - Aplicaciones
  - RDF - Resource Description Framework
  - Catalogación de recursos Internet
  - CDF - Channel Description Format
  - Envío de información a través de push
  - OSD - Open Software Description
  - Descripción de packs de actualización de software
  - OFX - Open Financial Exchange
  - Intercambio de datos financieros para aplicaciones contabilidad doméstica

# XML: eXtensible Markup Language

- XML - Aplicaciones
  - XER - XML Encoding Rules
  - Codificar mensajes ASN.1 para Z39.50, ILL, etc.
  - MARTIF - Machine Readable Terminology Interchange Format
  - Codificar registros terminológicos
  - TMX - Translation Memory eXchange
  - Codificar memorias de traducción
  - DOCBOK
  - DTD para manuales y guías técnicas documentación de software

## ¿El futuro de XML?

- Gartner Group señala a XML como un medio óptimo para el intercambio y la integración de datos en torno a arquitecturas Internet y http
- Compañías antagónicas trabajaron a favor de XML (Microsoft, IBM, Sun, etc.)
- Los fabricantes de bases de datos (Oracle, Informix, Software AG) se volcaron con XML

# XHTML

- W3c lo hace público en agosto de 1999
- Reformular HTML 4.0, utilizando los mismos elementos y atributos, pero con sintaxis XML
- Uso obligatorio de etiquetas de inicio y de fin, p.ej. `<p>...</p>`
- Sintaxis XML para elementos vacíos:
- `<hr />` en lugar de `<hr>`
- Valores de atributos entrecomillados (números también)
- Nombres de elementos y atributos deben escribirse en minúsculas

# XHTML

- Se proponen tres DTDs para documentos HTML
  - Estricto – para documentos nuevos
  - Transición – para documentos ya disponibles
  - Para documentos con marcos (frames)
- Soporte a espacios de nombres “namespaces”
- Los elementos se agrupan en módulos: xhtml1-tables.mod, xhtml1-form.mod, xhtml1-linking.mod, etc.
- Facilitar el acceso al Web desde otros dispositivos: móviles, TV, etc.