

Arquitecturas Distribuidas

Tema 3. III. SGML

I. SGML

1. HTML como “aplicación SGML”
2. Resumen sintaxis SGML

Motivación

- Los S.D. son heterogéneos: diversidad de dispositivos, aplicaciones, sistemas operativos, lenguajes de programación, etc.
- ¿Cómo expresamos los datos de manera que puedan ser interpretados sin ambigüedad y pérdida de datos por sistemas heterogéneos?
 - Una solución muy extendida es mediante lenguajes de marcas estructurados.
 - Las marcas dotan de semántica (significado) a los datos

HTML de nuevo

- Hasta ahora:
 - Sabemos como definir etiquetas y atributos.
 - Poseemos un concepto general de la estructura de una página.
 - Conocemos ciertos elementos de HTML y la base general de su jerarquía (BLOQUE vs TEXTO).
 - Observamos la potencialidad de HTML para transmitir información hiper-enlazada.

¿Cómo se define/especifica HTML?

- Pero, a nivel sintáctico:
 - ¿Cómo identificar qué elementos pueden estar presentes dentro de otros? ¿Y el orden?
 - ¿Cómo conocer la sintaxis correcta de los atributos?
 - ¿Un elemento debe usar etiquetado de cierre?
- HTML debe definirse formalmente, de modo que en su especificación podamos encontrar rápidas respuestas a dichas preguntas. Para definir HTML se utiliza **SGML**, del que, formalmente, HTML es una aplicación

SGML (I)

- Standard Generalized Markup Language
- SGML es un **metalenguaje** que permite **definir lenguajes de marcas**
- **Document Type Definition (DTD)**: define el tipo de documento, los marcas usadas, la relación entre ellas y su significado.
- El DTD es la plantilla sobre la cual el contenido marcado puede ser entendido, interpretado y validado: “Manual del lenguaje de marcas”.
- **SGML proporciona una sintaxis para declarar DTDs**, es decir, lenguajes de marcas

SGML (II)

- Características que lo hacen apropiado para S.D.:
 - Marcas descriptivas: anota los datos en términos de su estructura o semántica y NO en función de lo que se debe hacer con ellos => Separa los datos de la estructura y por implicación de su representación
 - Proporciona un medio para asociar un documento marcado con un *tipo de documento* (DTD)
- A partir de SGML se definen diversos lenguajes de marcas.
 - Entre ellos, el más usado: HTML
- SGML es demasiado complejo y ambiguo en ciertos aspectos: XML se desarrolla para solucionar estos problemas

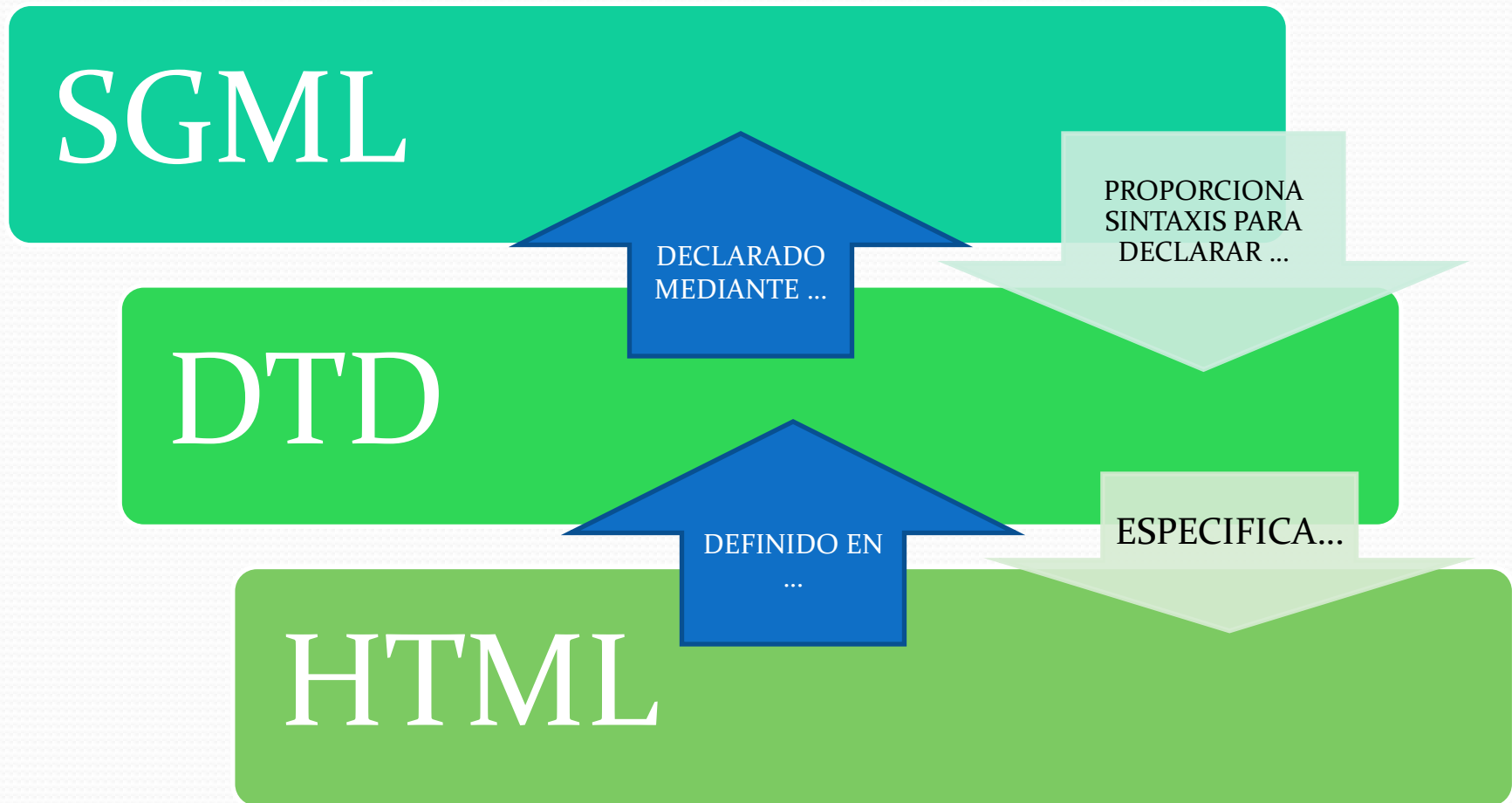
HTML como aplicación SGML

- SGML es anterior a HTML, y se ha usado para definir otros lenguajes de marcación.
- SGML proporciona la sintaxis para declarar DTDs, es decir, especificar lenguajes de marcas.
- La declaración formal de lenguajes de marcación permite:
 - Saber si un documento es “conforme” a la declaración del lenguaje al que pertenece.
 - Procesado automático por una aplicación.

HTML como aplicación SGML

- Cada lenguaje definido con SGML se dice que es una “aplicación SGML”. → HTML es una aplicación SGML.
- A cada aplicación SGML le corresponde un *Document Type Definition (DTD)* del lenguaje:
 - *Reglas permitidas*
 - *Reglas requeridas*
 - *Distinción entre marcas y texto del documento*

Relación HTML-DTD-SGML



HTML como aplicación SGML

```
<!doctype HTML PUBLIC "-//W3C/DTD  
HTML 3.2 Final//EN">
```

Indicación de la aplicación SGML utilizada.

```
<html>  
<!-- Ejemplo 1 -->  
<!-- Esto son comentarios -->  
  
<head>  
  <title>Titulo de la pagina</title>  
</head>  
  
<body>  
  Contenido de la pagina.  
</body>  
  
</html>
```

Documento debe ser conforme a ese lenguaje.

Resumen sintaxis SGML

```
<!DOCTYPE nombre [reglas]>
```

```
<!ELEMENT nombre_etiqueta marcación (contenido)>
```

marcación: “-” (obligatoria) “o” (opcional)

contenido: #PCDATA y otros elementos

- **Cardinal** (? * +)
- **Ordenación** (, | &)

```
<!ATTLIST nombre_etiqueta  
          (nombre_atributo tipo valor_por_defecto \n)+>
```

```
<!ENTITY % nombre "estructura">
```

Ejemplos sintaxis SGML

<!ELEMENT UL - - (LI)+>

<!ELEMENT DL - - (DT|DD)+>

<!ELEMENT FORM - - (%block;|SCRIPT)+ -(FORM)>

<!ELEMENT IMG - O EMPTY>

<!ENTITY % fontstyle "TT | I | B | BIG | SMALL">

<!ELEMENT LINK - O EMPTY -- a media-independent link -->

<!ATTLIST LINK

%attrs; -- %coreattrs, %i18n, %events --
charset %Charset; #IMPLIED -- char encoding of linked resource --
href %URI; #IMPLIED -- URI for linked resource --
hreflang %LanguageCode; #IMPLIED -- language code --
type %ContentType; #IMPLIED -- advisory content type --
rel %LinkTypes; #IMPLIED -- forward link types --
rev %LinkTypes; #IMPLIED -- reverse link types --
media %MediaDesc; #IMPLIED -- for rendering on these media --
>

Referencias y bibliografía

- Tutoriales de HTML en la Web (¡¡hay cientos!!)
- En inglés:
 - <http://www.w3schools.com/html>

Referencias y bibliografía

- Especificaciones formales HTML:
 - Disponibles desde la Web de la asignatura:
 - <http://ait.upct.es/asignaturas/ad/manuales>
 - “Especificación de Referencia para el HTML 3.2”, *Recomendación del W3C*, 14-ene-1997
 - “Especificación HTML 4.01(*HTML 4.01 Specification*)”, *Recomendación del W3C*, 24 de diciembre de 1999

Referencias y bibliografía

- Tutoriales de SGML en la Web
 - En inglés:
 - <http://ait.upct.es/asignaturas/ad/manuales/SGML/Gentle-Introduction-To-SGML.html>, “A gentle introduction to SGML”.
 - <http://www.w3.org/TR/REC-html40/intro/sgmltut.html> → Tutorial del W3C.