

1.3. Gramáticas



Índice

Objetivos	3
Gramáticas	4
Definición	4
XML Schema	5
Ventajas de los XML Schema.....	7
Reglas gramáticas XML	7
Despedida	8
Resumen.....	8

Objetivos

- Conocer el significado de la gramática para los lenguajes de marcas.
- Conocer cómo se define una gramática.
- Conocer la gramática de HTML.
- Conocer la gramática de XML.

Gramáticas

Definición

Las gramáticas en los documentos de marcado de tipo electrónico como HTML, XML, etc., vienen conformadas por las **reglas de escritura** de este tipo de documentos más las restricciones e indicaciones de la DTD o XML Schema pertinente.

Todo documento de un lenguaje de marcas tiene en común una gramática que define el marcado permitido en esa clase, el marcado requerido y cómo debe ser utilizado dicho marcado en la instancia del documento.

Según el diccionario de la lengua española, gramática significa:

“Parte de la lingüística que estudia los elementos de una lengua, así como la forma en que estos se organizan y se combinan.”

DICCIONARIO DE LA LENGUA ESPAÑOLA

Sin embargo, en relación a los lenguajes de marcas, la gramática permite reconocer si la **representación de marcas de un documento específico es correcta y cumple las reglas** que permiten que este pueda ser interpretado por un programa externo y mostrado de manera correcta.

La gramática para documentos HTML, por ejemplo, es un **conjunto de reglas que indican el orden de los elementos del lenguaje**. Estos elementos del lenguaje se pueden dividir en dos grupos: terminales (las propias palabras de la lengua) y no terminales (todas las demás reglas gramaticales).

En HTML las palabras se corresponden con las etiquetas de marcado incrustadas y el texto en un documento.

El análisis que se realiza para un documento se basa en las reglas de sintaxis por las que se rige el documento, es decir, el lenguaje o el formato en el que está escrito. Todos los formatos que se pueden analizar deben tener una gramática determinista formada por un vocabulario y unas reglas de sintaxis. Esto se denomina **gramática libre de contexto**. Los lenguajes humanos no son de este tipo y, por tanto, no se pueden analizar con técnicas de análisis convencionales.

Reglas gramáticas para HTML

El estándar define esta gramática mediante la **DTD (Definición de Tipo de Documento)** que establece las reglas de formación del lenguaje formal, es decir, qué combinaciones de símbolos elementales son sintácticamente correctas.

En la DTD se identifica la estructura del documento, es decir, aquellos elementos que son necesarios en la elaboración de un documento o un grupo de documentos estructurados de manera similar. Contiene las reglas de dichos elementos: el nombre, su significado, dónde pueden ser utilizados y qué pueden contener.

La entidad **W3C** ha definido los estándares HTML y otros desde el lenguaje SGML a través de las DTD. La construcción de una DTD se revisará en otra lección.

Estas DTD se aproximan al HTML 4 DTD. El W3C recomienda utilizar las versiones autorizadas de estas DTD en sus identificadores de sistema, definidos al validar el contenido. Si es necesario utilizar estas DTD localmente se debe descargar uno de los archivos de esta versión.

Toda DTD debe tener un, y solo un, elemento raíz, también conocido como elemento documento. Este elemento raíz debe coincidir con el nombre que aparece a continuación del DOCTYPE.

Un documento DTD puede contener:

- Declaraciones de elementos.
- Declaraciones de atributos para un elemento.
- Declaraciones de entidades.
- Declaraciones de notaciones.
- Instrucciones de procesamiento.
- Comentarios.
- Referencias a entidades de parámetro.

XML Schema

Para definir la **gramática de los documentos XML** se puede usar DTD o XML Schema, que es una alternativa más potente a las DTD.

XML Schema permite escribir esquemas detallados para documentos XML utilizando la sintaxis estándar de XML. XML Schema describe la estructura de un documento XML. El lenguaje XML Schema también se denomina XML Schema Definition (XSD).

El objetivo de un XML Schema es definir los elementos que permiten construir un documento XML válido, igual que las DTD.

Un XML Schema define:

- Los elementos que pueden aparecer en un documento.
- Atributos que pueden aparecer en un documento.
- Qué elementos son elementos hijos.
- El orden de los elementos hijos.
- El número de elementos hijos.
- Si un elemento debe estar vacío o puede incluir texto.
- Los tipos de datos de sus elementos y atributos.
- Los valores por defecto y fijos para elementos y atributos.

La gramática en los XML Schema está escrita también en XML, por lo que:

1. No hay necesidad de aprender un lenguaje nuevo.
2. Se puede utilizar un editor de XML para editar el XML Schema.
3. Se puede utilizar un procesador de XML para procesar un XML Schema.
4. Se puede manipular un XML Schema utilizando XML DOM.
5. Se puede transformar un XML Schema utilizando XSLT.

Cuando se transmite información desde un emisor a un receptor es necesario que ambas partes conozcan el mismo protocolo de comunicación. Con XML Schema el emisor puede describir los datos de manera que el receptor lo entienda.

Una fecha como: "03-11-2006", dependiendo del país, puede ser interpretada como "3 de noviembre" y en otros países como "11 de marzo". Con lo cual, al definir un elemento de la siguiente manera:

```
<fechaInicio type="date">2006-03-11</fechaInicio>
```

Se asegura un mutuo entendimiento sobre el contenido, porque el tipo de dato "date" requiere el formato "YYYY-MM-DD".

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://dominio.com">

<xs:element name="Contacto">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="nombre" type="xs:string"/>
      <xs:element name="dni" type="xs:string"/>
      <xs:element name="telefono" type="xs:string"/>
      <xs:element name="descripcion" type="xs:string"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>

</xs:schema>
```

Ejemplo XML Schema.

Ventajas de los XML Schema

Ventajas

Las ventajas de los XML Schema son:

- Los XML Schema son **extensibles**.
- Los XML Schema son **más ricos semánticamente que las DTD**.
- Los XML Schema están **escritos siguiendo la sintaxis estándar de XML**.
- Los XML Schema soportan **tipos de datos**.
- Los XML Schema soportan **namespaces**.

La característica más importante de los XML Schema es que soportan tipos de datos. Los tipos de datos permiten:

- Describir qué **elementos están permitidos**.
- **Validar** si los datos son correctos.
- Trabajar con los datos de una base de datos.
- Definir **Facets** (restricciones).
- Definir **patrones** de datos (formatos).
- Convertir datos considerando diferentes tipos de datos.

Reglas gramáticas XML

Además de las reglas definidas en cada XML Schema, los XML deben cumplir las siguientes normas para que puedan ser entendibles por sistemas externos:

1. Empezar con la declaración XML.
2. Tener un único elemento raíz.
3. Las etiquetas de apertura deben tener sus correspondientes etiquetas de cierre.
4. Los elementos son "case-sensitive" (sensibles a mayúsculas y minúsculas).
5. Todos los elementos deben cerrarse.
6. Todos los elementos tienen que estar adecuadamente anidados.
7. Los atributos deben estar entre comillas.
8. Se deben utilizar las entidades para utilizar caracteres especiales.

```
<?xml version="1.0"?>
<nota>
<a>Pedro</a>
<de>María</de>
<cabecera>A su atención</cabecera>
<cuerpo>¡En breve nos vemos en la cita prevista!</cuerpo>
</nota>
```

Despedida

Resumen

Has finalizado esta unidad, veamos los aspectos más importantes que hemos tratado.

En todo lenguaje necesitamos de unas reglas para transmitir la información deseada. Estas reglas son las que definen la estructura del propio lenguaje.

En esta unidad hemos visto las reglas gramáticas que tenemos que cumplir en los documentos XML y HTML.

En XML estas reglas están definidas por la estructura definida en su diseño y además podemos añadir las normas que definamos nosotros en la construcción de un DTD o XML Schema.

```
<xml>::= <elemento> FIN_TOKENS
<elemento>::= <inicio_elemento> <atributos> <cierres>
<inicio_elemento>::= '<' TOKEN_IDENTIFICADOR
<atributos>::= <atributo><atributos> | vacío
<atributo>::= TOKEN_IDENTIFICADOR '='
TOKEN_CADENA
<cierres>::= '>' | '>' <mas_elementos> <fin_elemento>
<mas_elementos>::= <elemento> <mas_elementos> |
vacío
<fin_elemento>::= '</' TOKEN_IDENTIFICADOR '>'
```

Gramática XML simple.

En el caso de HTML las normas ya nos vienen fijadas por la propia DTD establecida por la organización W3C y las propias de los lenguajes XML del que procede.