



IES Sant Vicent Ferrer
Algemesí



< Llenguatges de {JSON} XML marques i HTML CSS JS Sistemes de Gestió d'Informació >

UNIDAD 1: INTRODUCCIÓN A LOS LENGUAJES DE MARCAS



Tabla de contenidos

1. PRESENTACIÓN DE LA UNIDAD	1
2. INTRODUCCIÓN A LOS LENGUAJES DE MARCAS. CLASIFICACIÓN.....	2
2.1. LOS DATOS	2
2.1.1. CARACTERÍSTICAS DE LOS DATOS.....	2
2.2. ALMACENAMIENTO DE DATOS EN ORDENADORES	3
2.2.1. DATOS BINARIOS	4
2.2.2. DATOS DE TEXTO	5
2.3. FICHEROS DE MARCAS	6
2.3.1. LAS MARCAS.....	8
2.3.2. CARACTERÍSTICAS DE LOS LENGUAJES DE MARCAS	9
2.3.3. CLASIFICACIÓN DE LOS LENGUAJES DE MARCAS	10
3. BIBLIOGRAFÍA Y REFERENCIAS.....	12

1. PRESENTACIÓN DE LA UNIDAD

Los lenguajes de marcas están en pleno auge. Apoyándose en el grande éxito que han representado las páginas web han ido apareciendo nuevas tecnologías basadas en lenguajes de marcas que permiten a los usuarios ya los programas conseguir resultados con los que antes sólo podía soñarse.

Los nuevos lenguajes de marcas intentan ir un poco más allá de la simple representación de datos que ofrecía HTML, y que eran ideales para las personas, a nuevos sistemas que puedan hacer que los programas puedan recuperar la información que se encuentra en los ordenadores y procesarla de forma automática.

Esta unidad muestra qué son los lenguajes de marcas, cuáles son las suyas características, qué evolución han sufrido y cuáles son sus usos más importantes.

Resultados de aprendizaje y criterios de evaluación

Al terminar esta unidad, el alumno:

- Reconoce las características de lenguajes de marcas analizando y interpretando fragmentos de código.
 - o Identifica las características generales de los lenguajes de marcas.
 - o Reconoce las ventajas que proporcionan en el tratamiento de la información.
 - o Clasifica los lenguajes de marcas e identifica los más relevantes.
 - o Diferencia los ámbitos de aplicación de los lenguajes de marcas.
 - o Reconoce la necesidad y los ámbitos específicos de aplicación de un lenguaje de marcas de propósito general.

2. INTRODUCCIÓN A LOS LENGUAJES DE MARCAS.

CLASIFICACIÓN

Una definición poco estricta de lo que es un ordenador podría ser que “es una máquina electrónica que recibe y procesa datos para convertirlos en información útil”.

Uno de los componentes básicos en un sistema informático son los datos que se puedan introducir y cómo lo hace este sistema para almacenarlos para su uso posteriormente o mostrarlos de nuevo.

Por tanto, una de las tareas básicas que hacen los ordenadores es almacenar la información que les proporcionamos para poder ser procesada posteriormente.

Esta información puede ser de muchos tipos diferentes (texto, imágenes, vídeos, música...) pero lo realmente importante será cómo lo almacena el ordenador para poder tratarla posteriormente de manera eficiente para generar más información.

2.1. LOS DATOS

Los datos son representaciones de aspectos del mundo real y se suelen recoger para realizar cálculos, mostrarlas, organizarlas, etc., con el objetivo de que posteriormente alguien pueda hacer algo: tomar decisiones, generar nuevos datos...

2.1.1. CARACTERÍSTICAS DE LOS DATOS

Entre las características interesantes sobre los datos destacan sobre todo tres aspectos:

- A quién van dirigidas
- La posibilidad de reutilizarlas
- Que se puedan compartir

Destinatario de datos

Los datos tendrán una forma u otra en función del destinatario al que vayan dirigidas:

- Datos destinados a los humanos: generalmente los datos destinados al humanos requerirán que tengan alguna estructura concreta, con unos formatos determinados, con textos decorados de alguna forma. Hi aparecerán títulos, caracteres en negrita, etc. Generalmente no es necesario conocer qué significado tienen los datos, puesto que la interpretación se deja al lector.
- Datos destinados a los programas: los programas generalmente no necesitan que los datos tengan información sobre cómo deben

representar, sino que basta con que sean fácilmente identificables, que quede claro de qué tipo son y que haya alguna forma de determinar qué significan para poder tratarlas automáticamente.

Reutilización de los datos

Muy a menudo los datos querrán reutilizarse para poder realizar tareas diferentes. Uno error corriente suele ser almacenarlas específicamente para realizar una tarea concreta, ya que esto puede provocar que posteriormente sea mucho más complicado utilizarlas para realizar otras tareas.

Por tanto, es básico disponer de un sistema de almacenamiento que permita conseguir que los datos puedan ser reutilizados fácilmente y si puede que puedan ser reutilizadas tanto por las personas como por los programas.

Compartición de los datos

En el pasado, con los ordenadores centrales la información se generaba y se procesaba en el mismo sitio. Pero la aparición de los ordenadores personales, la eclosión de las redes y, sobre todo, el éxito de Internet, ha creado toda una serie de problemáticas que hasta el momento no existían: los datos generados en un sitio ahora pueden ser consumidas en un sitio totalmente diferente, tales como: en sistemas operativos totalmente distintos.

en máquinas que pueden funcionar de formas muy diversas.

Por tanto, en un sistema informático moderno se debe tener en cuenta esta posibilidad a la hora de almacenar datos. Existe la posibilidad de que estas datos sean compartidos y, por tanto, hay que almacenarlos de alguna manera que no tenga problemas por usarlas en sistemas diferentes.

2.2. ALMACENAMIENTO DE DATOS EN ORDENADORES

Los ordenadores almacenan la información en binario y, por tanto, toda la información que se puede almacenar siempre se representará en unos y ceros (1, 0). Esto hace que para representar cualquier tipo de datos (imágenes, vídeos, texto...) haya que hacer algún tipo de proceso que convierta los datos a una representación en formato binario.

Tradicionalmente en los ordenadores los datos se organizan de dos formas:

- Datos de texto
- Datos binarios

2.2.1. DATOS BINARIOS

Almacenar los datos de forma binaria es la manera natural de almacenar datos en ordenadores. Estrictamente hablando, los datos binarios están en el formato que utiliza el ordenador, ya que sólo son una tira de bits tras otro. Por tanto, normalmente, un ordenador no deberá hacer ningún proceso especial para almacenar y leer datos binarios.

Los datos en formato binario tienen una serie de características que los hacen ideales para los ordenadores:

- Generalmente están optimizadas para ocupar el espacio necesario.
- Los ordenadores las leen fácilmente.
- Pueden tener estructura.
- Es relativamente fácil añadir metadatos.

Si un programa desea usar los datos binarios normalmente sólo necesitará conocer el tamaño en bits y, sobre todo, conocer de qué forma hay almacenado la información.

Para almacenar un número basta con convertir este valor decimal a la su representación en binario y almacenarlo. Por ejemplo, el valor decimal 150 se representa con el valor binario 10010110.

Metadatos

Muy a menudo no se almacenan directamente los datos tal y como están sino que se procesan para optimizarlas, tales como almacenando información sobre el su contenido o aplicando procedimientos de optimización. Estas optimizaciones son transparentes por el usuario final, que visualizará los datos normalmente.

Datos estructurados

Los datos en la forma en que los generamos los humanos no están en un formato que facilite el tratamiento automático por parte de un ordenador. Por este motivo a menudo los datos que deben ser procesados por los ordenadores se convierten a alguno formato que sea más idóneo para el tratamiento. Lo más corriente es tratar los datos a fin de que tengan algún tipo de estructura.

Los tipos de datos estructurados son agrupaciones de otros tipos de datos (normalmente tipos más sencillos).

La forma más corriente de estructurar datos binarios suele ser tenerlos agrupados en registros que contienen la información repetitiva de un dato en concreto. Es

habitual que los lenguajes de programación tengan alguna forma de definir datos estructurados.

Los datos estructurados facilitarán que las aplicaciones puedan tratarlas de modo automático.

Problemas

Los datos binarios también tienen problemas asociados a la hora de ser compartidas, ya que en el mundo moderno los datos deben compartirse en máquinas en las que no han sido generadas, que pueden tener sistemas operativos diferentes, pueden ser máquinas totalmente distintas, etc. Los datos binarios optimizados para el sistema en el que se generan no siempre serán bien entendidos por los demás sistemas.

2.2.2. DATOS DE TEXTO

Para solucionar el problema que sea necesario recurrir a programas específicos para recuperar los datos que hay en un archivo una posibilidad es hacer lo más obvio, hacer lo mismo que han hecho los humanos durante siglos. Los humanos al escribir ya están utilizando una codificación y, por tanto, si se utiliza la misma codificación tendremos los datos en un formato fácil de utilizar y entender que no tendrá problemas para ser leído por los programas.

Los archivos de texto almacenan la información letra por letra de una forma similar a cómo lo haría un humano al escribir. Esto hace que se esté generando una información que se podrá leer de la misma forma que se lee un documento de paper.

Para un ordenador no hay mucha diferencia a la hora de almacenar los archivos de texto o archivos binarios, ya que los archivos de texto también son tiras de bits. La diferencia es que esta vez los bits están agrupados de forma estándar y conocida: un código de caracteres.

Códigos de caracteres

Representar los datos en un ordenador en forma de texto implica que para poder representar una palabra cualquiera en el ordenador previamente tendrá que ser codificada para que pueda ser representada en binario (recordemos que los ordenadores sólo pueden representar datos en binario). Esta codificación solo consistir en determinar una cantidad de bits predefinida para marcar un carácter y posteriormente se asocia un valor numérico a cada uno de sus caracteres.

La equivalencia entre los caracteres y sus valores numéricos no puede hacerse de forma aleatoria, ya que se estaría creando el mismo problema que hay con las

datos binarios. Si se quiere conseguir que los datos se puedan leer en diferentes sistemas hay que seguir algún tipo de norma conocida por todos. Por este motivo aparecieron los estándares de codificación de caracteres.

Compartición de información

Gracias al uso de estándares de códigos de caracteres la información en forma de texto es más fácilmente compartida que la información binaria, ya que los códigos de caracteres que utilizan los sistemas para representar el texto son conocidos y pueden ser implementados libremente.

Por tanto, almacenar los datos en formato de texto aporta dos grandes ventajas:

- Las pueden utilizar una gran cantidad de programas que ya existen (editores de texto, navegadores, etc.).
- Pueden leerlos por humanos.

Con uno de los programas más simples que hay, un editor de texto, se puede crear un documento que se podrá compartir con cualquier persona que entienda el idioma en el que ha sido escrito. Y como todos los sistemas operativos llevan de serie programas capaces de cargar archivos, si se envía el archivo a alguien éste no tendrá ningún problema para interpretar los datos cuando los reciba.

Problemas

Generar información en archivos de texto también tiene algunos problemas:

- Ocupan más espacio en el disco que los binarios.
- Existen múltiples códigos de caracteres diferentes.
- La forma en que los tratan los diferentes sistemas operativos.
- Falta de estructuración de los datos.

Pero a pesar de los problemas, éstos son mucho menos importantes que los que tenemos para compartir archivos binarios. Por tanto los archivos de texto son la manera más sencilla de asegurarnos de que podemos compartir la información que hay con otras personas.

2.3. FICHEROS DE MARCAS

Se puede decir que los archivos de marcas son una manera diferente de almacenar información en ordenadores que se añade a las formas de almacenar la información por medio de archivos binarios o archivos de texto. El objetivo principal de los

archivos de marcas es intentar recoger las mejores características de los archivos de texto y binarios.

Los archivos de marcas toman como base los archivos de texto para aprovecharse de las características más interesantes de este tipo de archivos:

- La facilidad de creación y lectura.
- El cumplimiento de estándares de almacenamiento definidos y públicos.

Debido a que los archivos de texto siempre están almacenados en algún código de caracteres conocido (ASCII, UTF-8, etc.) se consigue que puedan ser transportados y leídos en cualquier plataforma, sistema operativo o programa que pueda interpretar estos códigos de caracteres. Por tanto, los lenguajes de marcas se aprovecharán de esta característica, al estar basados en el formato de texto.

Además, también tendrán la ventaja de que podrán ser abiertos y creados con los programas de edición de texto estándar.

Los archivos de marcas, por tanto, se aprovechan de una de las grandes ventajas de los archivos de texto sobre los archivos binarios, ya que estos últimos requieren ser abiertos con un programa específico que pueda interpretar su formato.

Pero los archivos de marcas no sólo se intentan aprovechar de las características de los archivos de texto sino que también intentan conseguir las características más interesantes de los archivos binarios, como:

- La incorporación de metadatos.
- La definición de la estructura de los datos.

Esto hace que los lenguajes de marcas adquieran una de las características más interesantes de los archivos binarios, que es la posibilidad de incorporar información sobre los datos –metadatos– pero intentando que afecte lo menos posible la legibilidad del documento.

Gracias a las ventajas que ofrecen los lenguajes de marcas, éstas se han convertido rápidamente en una de las maneras habituales de representar datos y se pueden encontrarse continuamente en las tareas habituales con ordenadores:

- Internet (la Web), basada totalmente en los lenguajes de marcas.
- Muchos de los programas de ordenador que utiliza habitualmente utilizan en algún momento alguna u otra forma de algún lenguaje de marcas por a almacenar sus datos de configuración o de resultados.

2.3.1. LAS MARCAS

Las marcas son una serie de códigos que se incorporan a los documentos electrónicos para determinar el formato, la forma en que deben imprimirse, la estructura de los datos, etc. Por tanto, son anotaciones que se incorporan a los datos pero que no son un partido formal.

Las marcas, por tanto, deben ser fácilmente distinguibles del texto normal (por su posición, porque siguen algún tipo de sintaxis, etc.). Las marcas más usadas son las que están formadas por textos descriptivos y están rodeadas de los símbolos de “más pequeño” (<) y “mayor” (>) y normalmente suele haber una en el principio y una al final:

```
<nom>Manel Puig Garcia</nom>
```

Estas marcas pueden ser imbricadas para indicar estructuras de datos:

```
<persona>
  <nom>Manel Puig Garcia</nom>
  <nom>Pere González Puigdevall</nom>
  <nom>Maria Pous Canadell</nom>
</persona>
```

Pero hay otras muchas formas de marcas. Otra idea consiste en encontrar alguna combinación de caracteres que salga raramente en el lenguaje habitual. El TeX utiliza las barras invertidas para indicar el inicio de las marcas

```
\section{Persones}
\begin{itemize}
\item Manel Puig Garcia
\item Pere González Puigdevall
\item Maria Pous Canadell
\end{itemize}
```

Otros lenguajes de marcas utilizan caracteres no habituales en determinadas posiciones para indicar que son marcas. Por ejemplo los caracteres “=” en la primera posición de una línea se utilizan para indicar que el texto es un título de apartado y el * por las listas de puntos:

```
= Persones =
* Manel Puig Garcia
* Pere González Puigdevall
* Maria Pous Canadell
```

La idea general es que es necesario que las marcas sean fácilmente identificables por poder aprovechar las ventajas que ofrecen los lenguajes de marcas.

2.3.2. CARACTERÍSTICAS DE LOS LENGUAJES DE MARCAS

Los lenguajes de marcas son una manera de codificar un documento de texto de forma que por medio de las mismas se incorpora información relativa a cómo debe representar el texto, sobre qué estructura tienen los datos que contiene, etc.

Los lenguajes de marcas han destacado por una serie de características que los han convertido en los tipos de lenguajes más usados en la informática actual para almacenar y representar los datos. Entre las características más interesantes que ofrecen los lenguajes de marcas existen:

- Que se basan en el texto plano.
- Que permiten utilizar metadatos.
- Que son fáciles de interpretar y procesar.
- Que son fáciles de crear y lo suficientemente flexibles para representar datos muy diversas.

Basados en texto plano

Los lenguajes de marcas se basan en texto plano sin formato. Estos caracteres pueden estar codificados en distintos códigos de caracteres: ASCII, ISO-8859-1, UTF 8, etc

Una de las ventajas que intentan aportar los lenguajes de marcas es que se pueden interpretar directamente y esto sólo es posible si utilizamos el formato de texto, puesto que los binarios requieren un programa para interpretarlos. Pero en más tienen la ventaja de que son independientes de la plataforma, del sistema operativo o del programa.

El hecho de que estén basados en formato de texto hace que sean fáciles de crear y de modificarlo porque sólo requieren un simple editor de textos.

Uso de metadatos

Las marcas se intercalan entre el contenido del documento, por lo que generalmente estas etiquetas suelen ser descriptivas de qué es lo que indica el contenido de los datos que contienen.

Estas marcas son la forma en que se añaden los metadatos a los documentos de texto y cómo se consiguen superar las limitaciones del formato de texto y conseguir algunas de las ventajas de los archivos binarios.

Facilidad de proceso

Los lenguajes de marcas permiten que el procesamiento de los datos que contengan pueda ser automatizado de alguna manera, ya que el archivo contiene la estructura de los datos que contiene.

El hecho de incluir la estructura permitirá que un programa pueda interpretar cada una de los datos de un archivo de marcas para representarlo o tratarlo convenientemente, ya que muestran la estructura de los datos que contienen.

Posteriormente un programa podrá interpretar gracias a las marcas qué es el que significa cada uno de los datos del documento.

Facilidad de creación y representación de datos diversos

A pesar de que fueron pensados para contener datos de texto, los lenguajes de marcas han demostrado que son capaces de contener datos de muchos tipos diferentes.

Actualmente se están utilizando archivos de marcas para representar imágenes vectoriales, fórmulas matemáticas, crear páginas web, ejecutar funciones remotas por medio de servicios web, representar música o sonidos, etc.

Y sin importar qué tipo de datos se representan siempre habrá la posibilidad de crear estos archivos desde un editor de texto básico.

2.3.3. CLASIFICACIÓN DE LOS LENGUAJES DE MARCAS

Generalmente se acepta que tenemos dos grandes grupos basándonos en cuál es el objetivo básico del lenguaje de marcas:

- Lenguajes procedimentales y de presentación, orientados a especificar cómo representar la información.
- Lenguajes descriptivos o semánticos: orientados a describir la estructura de los datos que contiene.

Procedimentales y de presentación

En estos lenguajes lo que se hace es indicar de qué manera se debe realizar la presentación de los datos. Ya sea por medio de información para el diseño (marcar

negritas, títulos, etc.) o de procedimientos que debe hacer el software de representación. El ejemplo más popular de estos lenguajes es el HTML. En estos casos los documentos pueden servirnos para determinar de cuál modo se mostrará el documento a quien lo lee.

Descriptivos o semánticos

En estos lenguajes se describe qué estructura lógica tiene el documento ignorando cómo será representada en los programas. Sólo se ponen marcas con el objetivo de definir las partes que dan estructura al documento. El ejemplo más importante es el XML pero hay otro que está teniendo mucho de soporte, como por ejemplo JSON.

Sistema de etiquetado

Tanto si el sistema es descriptivo como de presentación, las marcas no han sido colocadas de cualquier modo sino que se ha ido siguiendo un sistema determinado. A menudo las marcas envuelven el contenido que queremos que tenga un significado o que sea representado de forma determinada. No se pueden colocar las marcas de cualquier manera, ya que una de las cosas a evitar son posibles malinterpretaciones.

Por eso, además de definir las marcas que se pondrán, los lenguajes de marcas definen unas reglas de uso que especifican cómo deben ser las marcas, en qué condiciones se permite utilizarlas ya veces incluso qué significan.

3. BIBLIOGRAFÍA Y REFERENCIAS

Sala, Javier. (2023) Lenguajes de marcas y sistemas de gestión de información.
Instituto Abierto de Cataluña