

# **UD1. INTRODUCCIÓN A LOS LENGUAJES DE MARCAS**

Lenguajes de Marcas y Sistemas de Gestión de  
Información

1º Desarrollo de Aplicaciones Web – Presencial

Paco Sebastián

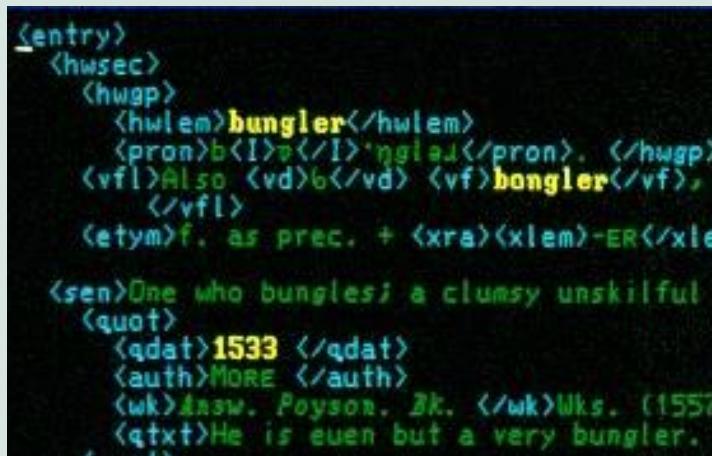
f.sebastianrebolla@edu.gva.es

# 0. OBJETIVOS DIDÁCTICOS

- Los objetivos de esta Unidad Didáctica son:
- Conocer la evolución de los lenguajes de marcas desde su aparición para dar contexto al estado del arte de los actuales lenguajes.
- Identificar rápidamente los diferentes lenguajes de marcas visualizando su apariencia sin renderizar.
- Identificar las características y ámbitos de aplicación de los lenguajes de marcas más comunes.
- Conocer los componentes básicos de los lenguajes de marcas.
- Identificar herramientas de edición de lenguajes de marcas.

# 1. INTRODUCCIÓN

- Algunos procesadores de texto optaron por guardar toda la información como texto, haciendo que las indicaciones de formato no se almacenan de manera binaria sino textual. Estas indicaciones son caracteres marcados de manera especial para que así un programa adecuado pueda traducir estos caracteres no como texto sino como operaciones que finalmente producirán mostrar el texto del documento de manera adecuada.
- La idea del marcado procede del inglés marking up término con el que se referían a la técnica de marcar manuscritos con lápiz de color para hacer anotaciones como por ejemplo la tipografía a emplear en las imprentas. Este mismo término se ha utilizado para los documentos de texto que contienen comandos o anotaciones.



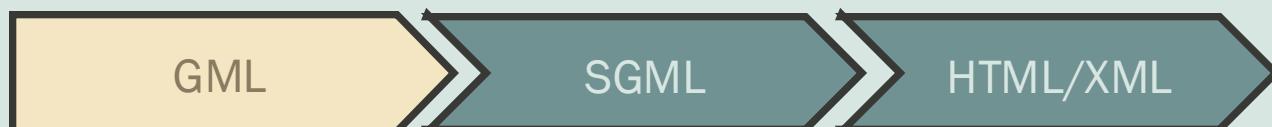
The image shows a screenshot of a computer screen displaying an XML document. The XML code is color-coded for syntax highlighting, with tags in blue and text in black. The document describes the word 'bungler'. It includes sections for headwords, pronunciation, etymology, and several examples of the word in use. The code uses various XML tags like <entry>, <hwsec>, <hwgrp>, <hwlem>, <pron>, <vfl>, <etym>, <sen>, <quot>, <qdat>, <auth>, <wk>, and <qtxt>. The word 'bungler' is highlighted in red in several places throughout the document.

```
<entry>
  <hwsec>
    <hwgrp>
      <hwlem>bungler</hwlem>
      <pron>b<I>v</I>nglər</pron>. </hwgrp>
      <vfl>Also <vd>b</vd> <vf>bongler</vf>,
      </vfl>
      <etym>f. as prec. + <xra><xlem>-ER</xlem></xra>
    </hwsec>
  <sen>One who bungles; a clumsy unskillful person.
  <quot>
    <qdat>1533 </qdat>
    <auth>MORE </auth>
    <wk>new. Poyson. 3k. </wk>Wks. (1557)
    <qtxt>He is even but a very bungler.
  </quot>
</entry>
```

## 2. GML

- Golfarb era un investigador de IBM que propuso ideas para que los documentos de texto que incluyeran la posibilidad de marcar el formato del mismo. Al final ayudó a realizar el lenguaje GML de IBM el cual puso los cimientos del futuro SGML (padre de HTML y XML) ideado por el propio Goldfarb y padre de la mayoría de lenguajes de marcas actuales.

```
:h0. El reino de los animales
:h1. Mamíferos
:p. Los mamíferos (:hp1.Mammalia:ehp1.) son una clase de
vertebrados :hp2.amniotas homeotermos:ehp2. que poseen
glándulas mamarias productoras de leche con las que
alimentan a las crías
:h1. Aves
:p. Las aves son animales vertebrados, de sangre
caliente, que caminan, saltan o se mantienen solo sobre
las extremidades posteriores
```



# 3. TeX y LaTeX

- En la década de los 70 Donald Knuth creó el lenguaje TeX para producir documentos científicos utilizando una tipografía y capacidades que fueron iguales en cualquier computadora.
- El éxito de TeX produjo numerosos derivados de los cuales el más popular es LaTeX. LaTeX fue definido en 1984 por Leslie Lamport (premio Turing 2003), aunque luego ha sido numerosas veces revisado. La idea es que los científicos se centren en el contenido y no en la presentación.

```
\documentclass[12pt]{article}
\usepackage{amsmath}
\title{\textbf{Ejemplo}}
\begin{document}
Este es el texto ejemplo de \LaTeX{}.
Con datos en \emph{cursiva} o \textbf{negrita}.
Ejemplo de f\textit{ormula}
\begin{align}
E &= mc^2
\end{align}
\end{document}
```



Este es el texto ejemplo de \LaTeX{} Con datos en *cursiva* o **negrita**. Ejemplo de formula

$$E = mc^2 \quad (1)$$

# 4. RTF

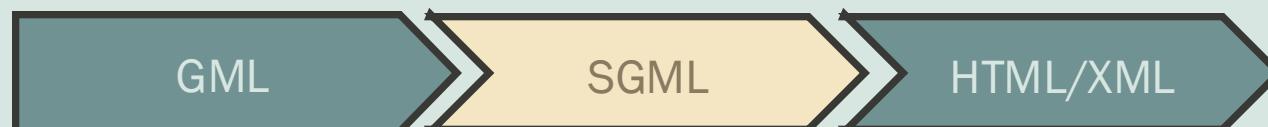
- RTF es el acrónimo de Rich Text Format, por miembros del equipo de desarrollo de Microsoft Word.
- Actualmente se trata de un formato aceptado para documentos de texto que cuenten con información sobre el estilo del texto. Se usa mucho como formato de intercambio entre diferentes procesadores por su potencia.
- Su éxito procede a que las indicaciones de formato son potentes y son más legibles para las personas que el formato nativo de los procesadores de textos, aunque es, como lenguaje de marcado, uno de los más crípticos.

```
{\rtf\ansicpg1252\deff0\deflang3082
{\fonttbl
{\f0\fcharset0\froman Times New Roman}
{\f1\fcharset0\fswiss Arial Black}
}
{\pard \f1\fs48
El reino de los animales
\par}
{\pard \f1\fs40
Mamíferos
\par}
{\pard \f0\fs25
Los mamíferos (\b Mammalia) son una clase de
vertebrados \i amniotas homeotermos que poseen
glándulas mamarias productoras de leche con las que
alimentan a las crías
\par}
{\pard \f1\fs40
Aves
\par}
{\pard \f0\fs25
Las aves son animales vertebrados, de sangre caliente,
que caminan, saltan o se mantienen solo sobre las
extremidades posteriores
\par}
}
```

# 5. SGML

- Se trata de una mejora muy notable del lenguaje de GML que estandarizaba el lenguaje de marcado y que fue definida finalmente por ISO como estándar mundial en documentos de texto con etiquetas de marcado. Su responsable fue Charles Goldfarb.
- Su importancia radica en que es el padre del lenguaje XML y la base sobre la que se sostiene el lenguaje HTML, dos de los lenguajes de marcas más populares de la historia
- En realidad no es un lenguaje con unas etiquetas concretas, sino que se trata de un lenguaje que sirve para definir lenguajes. Entre los lenguajes definidos mediante SGML, sin duda HTML es el más popular.

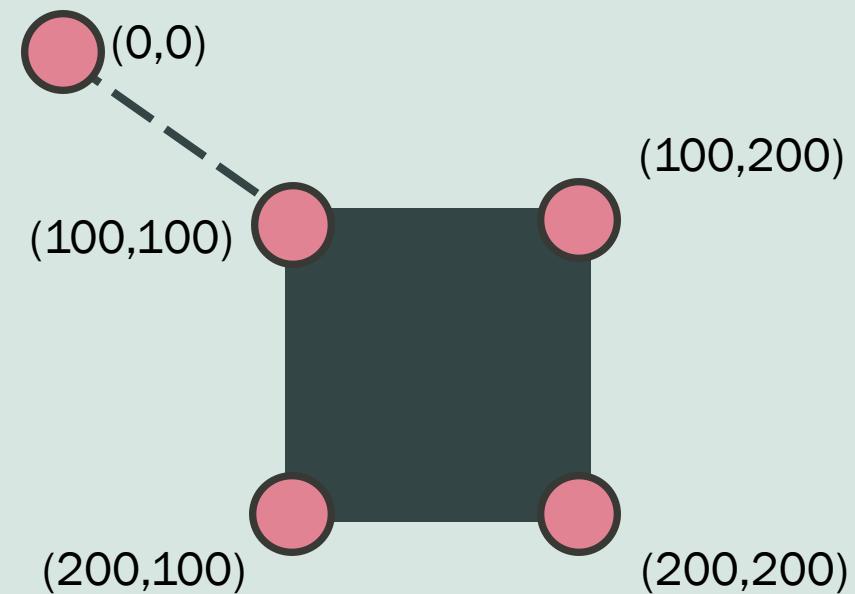
```
<articulo>
    <titulo1>El reino de los animales</titulo1>
    <titulo2>Mamíferos</titulo2>
    <normal>Los mamíferos (<negrita>Mammalia</negrita>)
    <titulo2>Aves</titulo2>
    <normal>Las aves son animales vertebrados, de sangre
</articulo>
```



# 6. PostScript

- Se trata de un lenguaje de descripción de páginas. De hecho es el más popular para este fin, siendo el lenguaje más utilizado por los sistemas de impresión de alta gama.
- Es en realidad todo un lenguaje de programación que indica la forma en que se debe mostrar la información que puede incluir texto y el tipo de letra del mismo, píxeles individuales y formas vectoriales (líneas, curvas). Sus posibilidades son muy amplias.

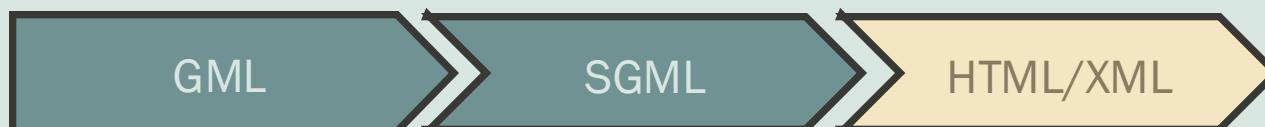
```
%colocar el cursor  
100 100 moveto  
%dibuja cuadrado  
100 200 lineto  
200 200 lineto  
200 100 lineto  
100 100 lineto  
%relleno  
stroke
```



# 7. HTML

- Tim Bernes Lee utilizó SGML para definir un nuevo lenguaje de etiquetas que llamó Hypertext Markup Language (lenguaje de marcado de hipertexto) para crear documentos transportables a través de Internet en los que fuera posible el hipertexto.
- Inicialmente estos documentos se veían con ayuda de intérpretes de texto (como por ejemplo el Lynx de Unix) que simplemente colorearían el texto y remarcaban el hipertexto. Después el software se mejoró y aparecieron navegadores con capacidad más gráfica para mostrar formatos más avanzados y visuales.

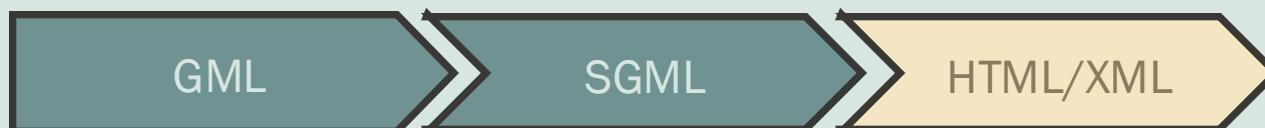
```
<!DOCTYPE html>
<html lang="es">
<head>
    <meta charset="UTF-8">
    <title>Document</title>
</head>
<body>
    <h1>El reino de los animales</h1>
    <h2>Mamíferos</h2>
    <p>Los mamíferos (<strong>Mammalia</strong>) son una clase de vertebrados
    <h2>Aves</h2>
    <p>Las aves son animales vertebrados, de sangre caliente, que caminan, sa
</body>
</html>
```



# 8. XML

- Se trata de un subconjunto de SGML ideado para mejorar el propio SGML y con él definir lenguajes de marcado con sintaxis más estricta, pero más comprensible.
- Actualmente está siendo sobrepasado en la mayoría de sus usos por JSON

```
<?xml version="1.0" encoding="UTF-8"?>
<nombre>Jorge</nombre>
<apellido1>Sánchez</apellido1>
<dirección>
    <calle>C/ Falsa nº 0</calle>
    <localidad>Palencia</localidad>
    < código_Postal>34001</ código_Postal>
    <pais>España</pais>
</dirección>
<teléfonos>
    <teléfono tipo="fijo">999 999 999</teléfono>
    <teléfono tipo="móvil">666 666 666</teléfono>
</teléfono>
```



# 9. Markdown

- Fue creado por John Gruber con la ayuda de Aaron Shwartz. La pretensión de este lenguaje es definir unas normas muy sencillas para crear documentos parecidos a los que se crean mediante el lenguaje HTML.
- Ha tenido un éxito muy notable, especialmente desde que fue adoptado por sitios tan populares como GitHub, Reddit o StackExchange para que los usuarios publicaran contenido con formato.

```
# El reino de los animales
## Mamíferos
Los mamíferos (**Mammalia**) son una clase de vertebrados *amniotas homeotermos*. que poseen gl
## Aves
Las aves son animales vertebrados, de sangre caliente, que caminan, saltan o se mantienen solo
```

# 10. JSON

- Abreviatura de JavaScript Object Notation, Se trata de una notación de datos procedente del lenguaje JavaScript estándar.
- En el año 2002 se le apoyaba desde muchos de los navegadores y su fama ha sido tal que ahora se ha convertido en una notación independiente de JavaScript que compite claramente con XML en funcionalidad.
- En JSON, el texto se divide en datos y metadatos. De manera que el símbolo de los dos puntos separa el metadato del dato. Por otro lado, los símbolos de clave y claudátor permiten agrupar de diversas formas los datos.

```
{  
  "nombre": "Jorge",  
  "apellido1": "Sánchez",  
  "dirección": {  
    "calle": "C/ Falsa nº 0",  
    "localidad": "Palencia",  
    "código Postal": 34001,  
    "país": "España"  
  },  
  "teléfonos": [  
    {  
      "tipo": "fijo",  
      "número": "999 999 999"  
    },  
    {  
      "tipo": "móvil",  
      "number": "666 666 666"  
    }  
  ]  
}
```

# 11. TIPOS DE LENGUAJES DE MARCAS

- Orientados a la presentación. En ellos los metadatos permiten indicar el formato en el que se debe presentar el texto. Es el caso de RTF, en el que sus etiquetas especifican tipos de letra, tamaños de página, colores, etc. Las primeras versiones de HTML también se consideran así, ya que incluían etiquetas como fuente mediante la cual se especificaba el formato de fuente.
- Orientados a la descripción. En ellos las marcas especiales permiten dar significado al texto pero no indican cómo se debe presentar en pantalla lo mismo. Sería el caso de XML (o de SGML), JSON, Markdown y de las versiones actuales de HTML. En estos lenguajes simplemente se indica el significado del contenido: si el texto es un título, un párrafo normal, un pie de ilustración, una dirección postal etc.
- Orientados a procedimientos. Se trata de documentos en los que el texto marcado, se interpreta como órdenes a seguir, y así el archivo en realidad contiene instrucciones a realizar con el texto (girarle, convertirle en una fórmula, realizar una suma, etc.). Es el caso de LaTeX o PostScript.

# 12. CARACTERÍSTICAS Y COMPONENTES

- Etiquetas: Las etiquetas se escriben cerradas entre ángulos, es decir < y >. Normalmente, se utilizan dos etiquetas: una de inicio y otra de fin para indicar que ha terminado el efecto que queríamos presentar. La única diferencia entre ambas es que la de cierre lleva una barra inclinada "/" antes del código.
- Jerarquía: La posición de las marcas a este tipo de lenguajes es muy importante. Si una marca está contenida dentro de otra se dice que pertenece a su jerarquía. La jerarquía es de mucha importancia tanto en los Lenguajes de marcas orientados a la presentación como a los orientados a la descripción.

```
<?xml version="1.0" encoding="iso-8859-1" standalone="yes"?>
<!DOCTYPE biblioteca >
<biblioteca>
    <ejemplar tipo_ejem="libro" titulo="XML practico">
        <tipo> <libro isbn="978-2-7460-4958-1" edicion="1" </libro> </tipo>
        <autor nombre="Sebastien Lecomte"></autor>
        <autor nombre="Thierry Boulanger"></autor>
        <autor nombre="Angel Belinchon Calleja" funcion="traductor"></autor>
        <prestado lector="Pepito Grillo">
            <fecha_pres dia="13" mes="mar" año="2009"></fecha_pres>
            <fecha_devol dia="21" mes="jun" año="2009"></fecha_devol>
        </prestado>
    </ejemplar>
</biblioteca>
```

# 12. CARACTERÍSTICAS Y COMPONENTES

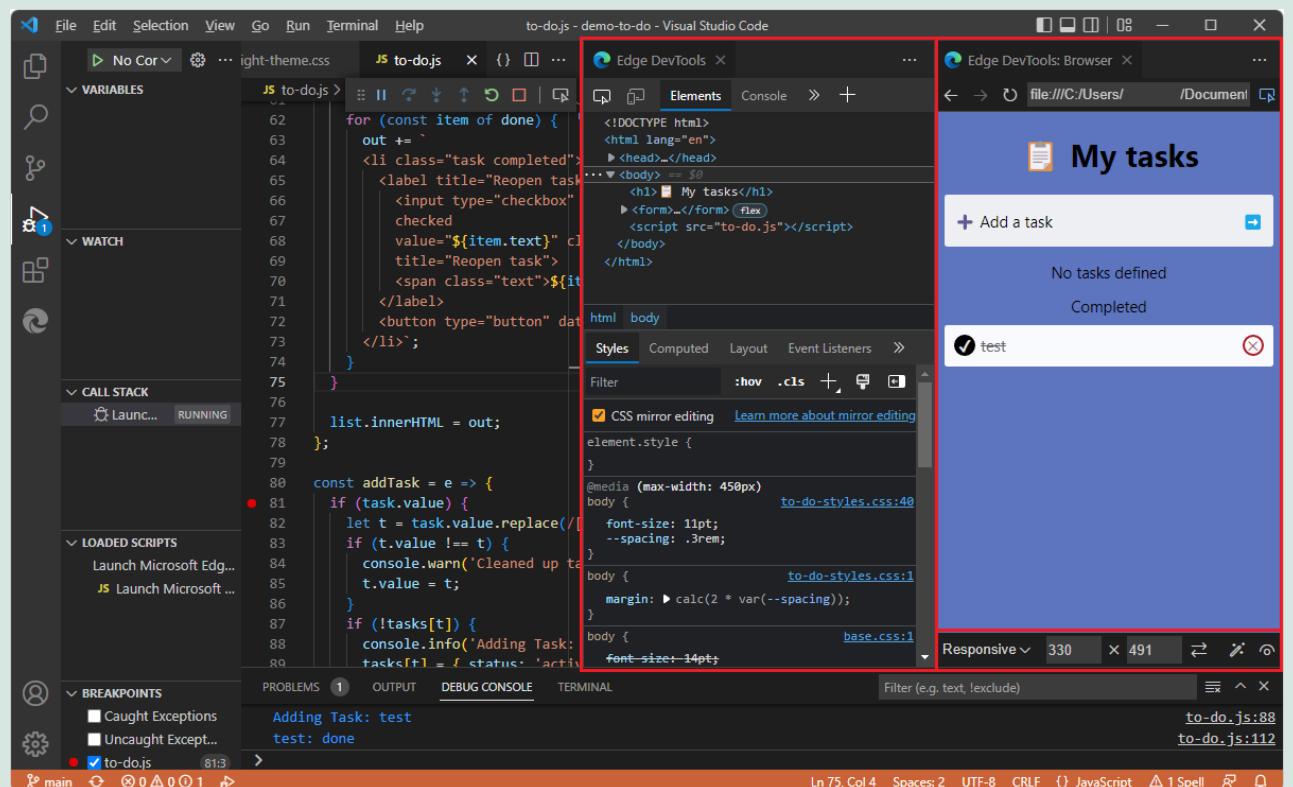
- Atributos: Permiten añadir propiedades a los elementos de un documento. Los atributos no pueden organizarse en ninguna jerarquía, no pueden contener ningún otro elemento o atributo y no reflejan ninguna estructura lógica.

```
<tipo> <libro isbn="978-2-7460-4958-1" edición="1" </libro> </tipo>
```

- Estándar: Para poder transmitir la información, es fundamental hablar el mismo idioma. Para las marcas pasa lo mismo, hay que establecer un estándar para interpretar las marcas. En los lenguajes de marcas modernos que parten de SGML como HTML y XML los estándar están establecidos por la World Wide Web Consortium (W3C).
- Facturación: Algunos lenguajes como XML, necesitan especificar cuál es exactamente la relación que debe verificarse entre los diferentes elementos presentes en el documento. Esta relación entre elementos se especifica en un documento externo o definición .

# 13. HERRAMIENTAS DE TRABAJO

- Las herramientas de edición de un lenguaje de marcas son normalmente los editores de texto. En la actualidad la mayor parte de editores de código tienen herramientas para hacer más fácil el trabajo con cada tipo de lenguaje.
- El herramientamiento de Visual Studio Code de Microsoft permite trabajar con mucho tipo de Lenguajes. Es el herramientamiento recomendado por esta asignatura.



→ Fer la pràctica 1.1 ←