

# T1 - INTRODUCCIÓN A LOS LENGUAJES DE MARCAS

Lenguaje de Marcas y Sistemas de Gestión de la  
Información

# **CONTENIDOS**

- 1.1. INTRODUCCIÓN A LOS LENGUAJES DE MARCAS**
- 1.2. CONTEXTO DE LOS LENGUAJES DE MARCAS**
- 1.3. LENGUAJES DE MARCAS: CARACTERÍSTICAS, ESTRUCTURA Y FORMATOS**
- 1.4. LENGUAJES DE MARCAS Y DIALECTOS MÁS COMUNES**

# Introducción

Los **lenguajes de marcas** existen prácticamente desde el comienzo de la informática. Cubren la necesidad de disponer de lenguajes sencillos para el intercambio de información entre distintos sistemas. Con la llegada de internet, algunos de los lenguajes de marcas se convirtieron en piedras angulares sobre las que construir la **World Wide Web**.

Existe un gran número de lenguajes de marcas distintos y, obviamente, no todos son igual de importantes. No obstante, comparten algunos aspectos comunes: fundamentos, orígenes, características básicas y elementos relacionados con su estructura. Es importante conocer qué son, cómo han evolucionado y, de aquellos que han tenido más éxito, qué pueden proporcionar.

# 1.1. Introducción a los lenguajes de marcas

Un lenguaje de marcas o lenguaje de marcado es aquel que permite representar información que contiene, además de los datos propiamente dichos, marcas o etiquetas que indican cómo se estructuran estos datos, su significado o cómo debe representarse desde un punto de vista gráfico o visual.

Los lenguajes de marcas no son lenguajes de programación. No contienen instrucciones ni implican acciones directas. Son reglas que organizan la información con el objetivo de darle una estructura uniforme y facilitar su procesamiento automático.

# 1.1. Introducción a los lenguajes de marcas

Una clasificación simplificada de los lenguajes de marcas podría ser la siguiente:

- *Lenguajes orientados a la presentación de información.*
- *Lenguajes orientados al almacenamiento y procesamiento de información.*

El ejemplo más representativo de un lenguaje de marcas orientado a la presentación de información es el lenguaje de marcas de hipertexto (HyperText Markup Language), conocido como HTML.

# 1.1. Introducción a los lenguajes de marcas

El etiquetado de HTML indica al programa que va a presentar en la pantalla la información (generalmente un navegador web o Web Browser) cómo debe hacer dicha presentación.

Es decir, proporciona junto a los datos que va a mostrar información referente a cómo han de mostrarse. Aunque ya existían previamente, a partir de la versión 5 de HTML se incluye un importante número de etiquetas con lo que se denomina información semántica. Estas etiquetas indican, además de cómo ha de mostrarse la información, qué significado tiene desde un punto de vista estructural. En la Unidad 2 se profundizará en el conocimiento de este tipo de etiquetado y las diferentes alternativas que proporciona.

# 1.1. Introducción a los lenguajes de marcas

Para ilustrar el lenguaje HTML, en el siguiente código se puede observar un texto delimitado por la etiqueta de apertura `<h1>` y la etiqueta de cierre correspondiente `</h1>`:

```
<h1>El origen de la vida</h1>
```

Estas etiquetas o marcas indican que el texto «marcado» es una cabecera del más alto nivel. A nivel visual, la etiqueta `<h1>` indica al navegador web que el texto debe representarse de una forma determinada. Los diseñadores de los navegadores, a partir del estándar publicado por el W3C, saben que la representación debe ser como se muestra en la siguiente imagen:

**El origen de la vida**

# 1.1. Introducción a los lenguajes de marcas

Por otra parte, la etiqueta `<h1>` indica que, semánticamente, el texto contenido es el título de un bloque de información del más alto nivel posible, como pueda ser el titular de una noticia en un periódico o el título de un capítulo de un libro.

HTML no es sino uno más de los múltiples lenguajes de marcas existentes. No existe un lenguaje de marcas universal, sino que pueden definirse tantos lenguajes de marcas como se desee, aunque probablemente ya exista algún lenguaje que proporcione la funcionalidad suficiente para casi cualquier necesidad.



# 1.1. Introducción a los lenguajes de marcas

Por ejemplo, un lenguaje de marcas sencillo podría definirse de la siguiente manera:

- *Los elementos delimitados por la marca <#> son **títulos de películas** y deberán representarse en color azul y utilizando la tipografía Times New Roman con un tamaño de 16 puntos.*
- *Los elementos delimitados por el símbolo <%> son **nombres de director** y deberán representarse en color rojo y utilizando la tipografía Times New Roman con un tamaño de 14 puntos.*

Dado este formato, un documento creado con este nuevo lenguaje de marcas podría ser el siguiente:

```
<#>Tron<#>  
<%>Steven Lisberger<%>  
<#>Juegos de guerra<#>  
<%>John Badham<%>
```

# 1.1. Introducción a los lenguajes de marcas

Gracias a las marcas o etiquetas, un programa informático realizado según esta estructura puede saber cuál es el título y el director de una película de manera sencilla. También se puede saber cómo debe representarse la información contenida en el documento, ya que así se indica en la especificación del lenguaje.

Tron

Steven Lisberger

Juegos de guerra

John Badham

## 1.1. Introducción a los lenguajes de marcas

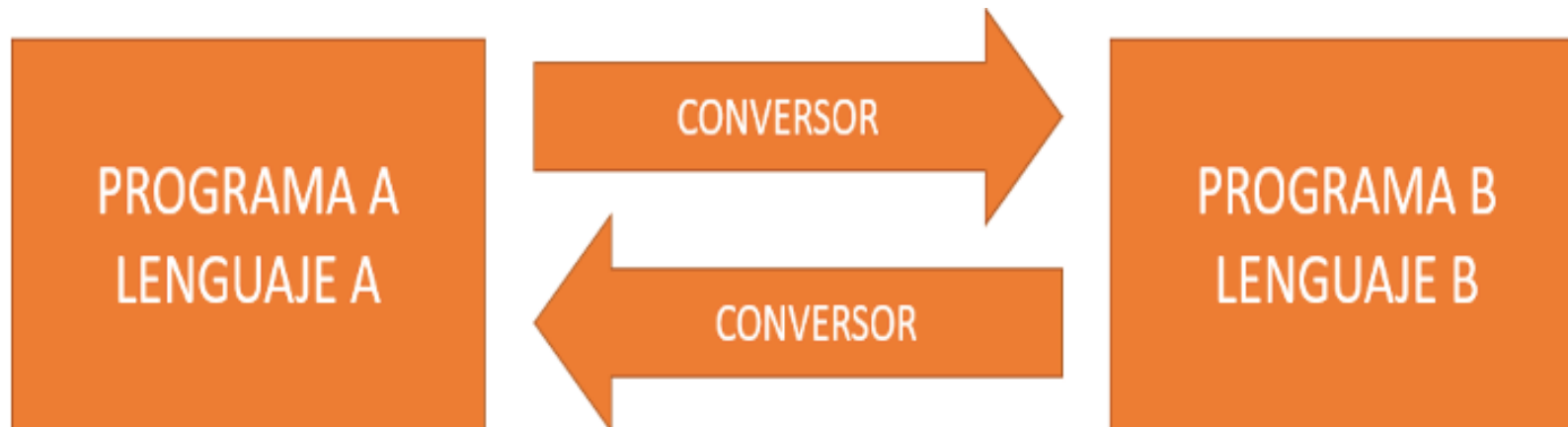
Se puede observar, por lo tanto, que la **definición** de un lenguaje de marcas está **al alcance de cualquiera** y no tiene por qué suponer mucho esfuerzo. El esfuerzo sería necesario en crear los **programas que leyesen e interpretasen las etiquetas correctamente**: habría que hacer tantos programas como lenguajes de marcas existentes.

Por fortuna, como se ha indicado con anterioridad, ya existe una serie de lenguajes de marcas capaces de cubrir prácticamente todas las necesidades posibles gracias a su flexibilidad. Además, no requieren la elaboración de sofisticados programas ni aplicaciones de procesamiento porque ya están creadas. Lenguajes como HTML y XML, junto con sus dialectos, están estandarizados y son capaces de satisfacer cualquier necesidad referente a cómo se deben estructurar los datos o cómo deben representarse.

## 1.2. Contexto de los lenguajes de marcas

El origen de los lenguajes de marcas se remonta prácticamente al origen de la computación moderna. En el mismo momento en que se comenzó a querer almacenar información para procesarla por un programa informático, se detectó la necesidad de dar «formato» a dicha información. Un dato etiquetado con una determinada marca facilita su proceso automático.

El problema surge cuando cada programa utiliza su propio lenguaje, ya que obliga a crear un conversor entre cada par de programas con los que se desea compartir información.



## 1.2. Contexto de los lenguajes de marcas

A finales de los años sesenta y principios de los setenta del siglo XX, un investigador en la compañía IBM, desarrolló el que se considera el precursor de los lenguajes de marcas modernos: GML (Generalized Markup Language).

GML es el primer paso para conseguir unificar los formatos de representación de información. GML sienta las bases para definir los estándares de representación de información posteriores. GML se diseñó para utilizarse en procesadores de texto, de tal manera que, en el documento escrito con este lenguaje, se indica mediante unas etiquetas o marcas cómo está estructurado el contenido y así indicar que un texto es un encabezado y cuál es su nivel de importancia, una lista y sus elementos o una tabla y su contenido en filas, columnas y celdas, por citar algunos ejemplos.

## 1.2. Contexto de los lenguajes de marcas

Las marcas GML están precedidas de dos puntos (:) y seguidas de un punto (.) y tienen el aspecto que se muestra a continuación.

```
:hl.Capítulo 1: Introducción a los lenguajes de marcas  
:p.Características y origen de los lenguajes de marcas.  
:ol.  
:li.Introducción  
:li.Contexto  
:li.Tipos de lenguajes  
:eol.
```

## 1.2. Contexto de los lenguajes de marcas

Para que no dependiese de un ente privado, ya que GML había sido creado por IBM, con las implicaciones de todo tipo que esto puede tener (económicas, legales y técnicas). Este hecho dio lugar a la creación de lenguaje SGML.

SGML (Standard Generalized Markup Language) es el estándar creado en 1986 por la ISO (International Organization for Standardization). Esta organización internacional y no gubernamental está constituida por varias organizaciones de estandarización nacionales y tiene como objetivo crear estándares de ámbito internacional dentro de diversos campos.

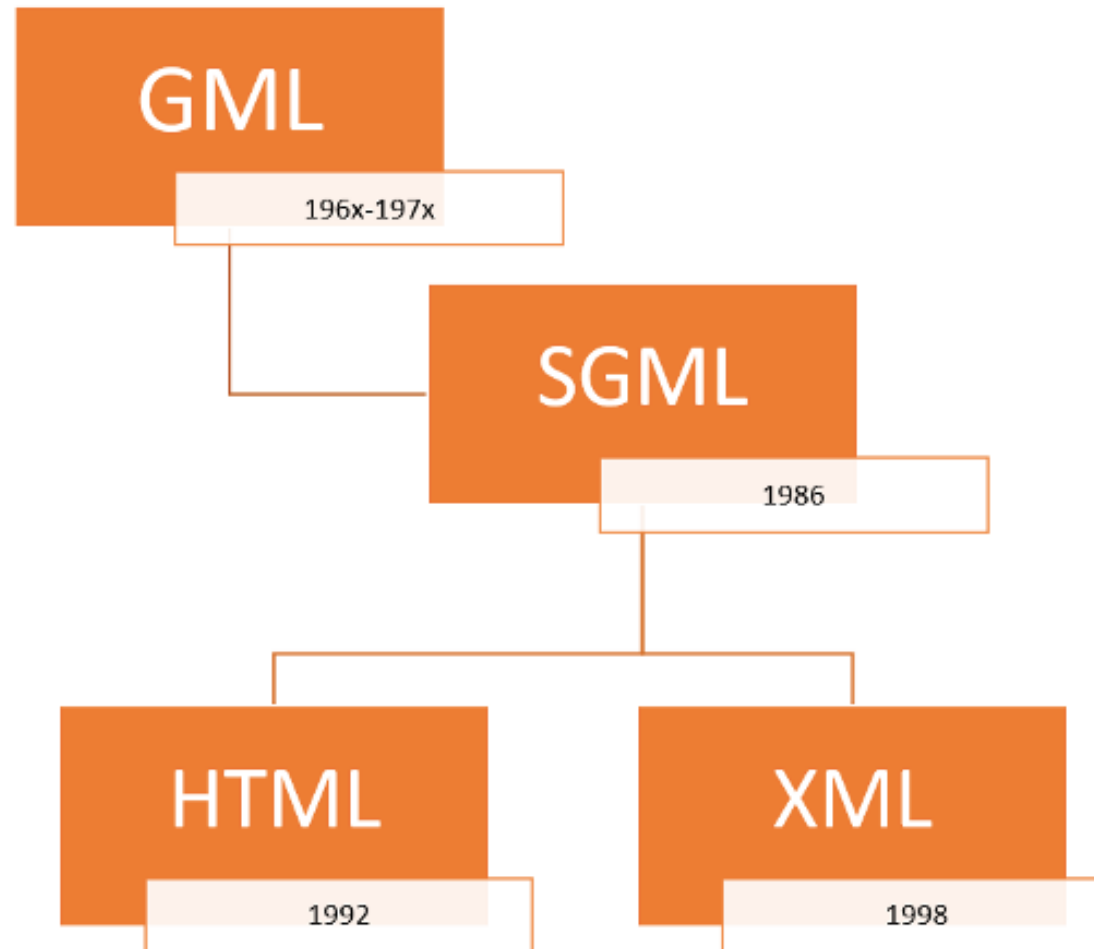
## 1.2. Contexto de los lenguajes de marcas

En realidad, SGML no es un lenguaje como tal, ya que no especifica qué marcas deben utilizarse. SGML indica cómo deben ser las marcas de un lenguaje. Esto significa que **SGML** puede ser el origen de un lenguaje de marcas, lo que le convierte en lo que se denomina **metalenguaje**. **HTML** y **XML**, por ejemplo, son lenguajes de marcas creados a partir del estándar SGML, aunque hay que aclarar que la versión 5 de HTML tiene sus propias reglas y ya no está basado en este estándar.



## 1.2. Contexto de los lenguajes de marcas

Cronología de los lenguajes de marcas más usados.



## 1.3. Lenguajes de marcas: características, estructura y formatos

Existen multitud de lenguajes de marcas. No existe, por lo tanto, una estructura común o un único formato. Por definición, cuando se habla de un lenguaje de marcas, se está haciendo referencia a la información añadida mediante marcas a los datos propiamente dichos del documento.

Cada lenguaje tiene su propia estructura y formato. No obstante, con el objetivo de ilustrar, se presenta a continuación parte de la estructura de uno de los lenguajes de marcas más importantes, que no es otro que XML.

# 1.3. Lenguajes de marcas: características, estructura y formatos

La estructura parcial y simplificada de un documento XML es la siguiente:

- Cada elemento de un documento XML está precedido por una etiqueta de apertura compuesta por un nombre y rodeada por los símbolos < y >.
- Cada elemento de un documento XML está seguido por una etiqueta de cierre, compuesta por el mismo nombre de la etiqueta de apertura y rodeada por los símbolos </y>.
- El contenido de un elemento de un documento XML puede ser un texto, o uno o más elementos.
- Solo existe un elemento raíz.
- Un atributo está compuesto de un nombre del atributo, el símbolo = y el valor del atributo entrecomillado.
- Un elemento puede tener entre 0 y n atributos.

## 1.3. Lenguajes de marcas: características, estructura y formatos

Un ejemplo de un documento XML de acuerdo con esta estructura podría estar almacenado en un fichero denominado misjuegos.xml sería como se muestra a continuación.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<videojuegos>
  <videojuego multijugador="si">
    <titulo>Call Of Duty</titulo>
    <plataforma>Play Station 5</plataforma>
  </videojuego>
  <videojuego multijugador="si">
    <titulo>Fortnite</titulo>
    <plataforma>XBox One </plataforma>
  </videojuego>
  <videojuego multijugador="no">
    <titulo>Manic Miner</titulo>
    <plataforma>AmstradCPC</plataforma>
  </videojuego>
</videojuegos>
```

## 1.3. Lenguajes de marcas: características, estructura y formatos

Por su parte, HTML, siendo también un lenguaje surgido a raíz de SGML, no es un **metalenguaje** como XML. Tiene también una estructura bien definida (similar a la de XML, pero no idéntica), aunque, al tratarse de un lenguaje concreto, tiene definidos con precisión los **nombres de etiquetas y de los atributos**.

En el siguiente ejemplo se muestra una sección de un documento HTML. Las etiquetas `<table>`, `<tr>` y `<td>` están definidas en la especificación del lenguaje HTML, así como el atributo `class`.

```
<!DOCTYPE html>
<table class="fondoGris">
  <tr>
    <td>Elemento 1</td>
    <td>Elemento 2</td>
  </tr>
  <tr>
    <td>Elemento 3</td>
    <td>Elemento 4</td>
  </tr>
</table>
</html>
```

---

Elemento 1	Elemento 2
Elemento 3	Elemento 4

# 1.3. Lenguajes de marcas: características, estructura y formatos

Las **características fundamentales** de los **lenguajes de marcas** son las siguientes:

- **Independencia.** Su uso no debe depender de ninguna plataforma hardware ni sistema software concreto.
- **Almacenados en texto plano.** El documento está compuesto únicamente por caracteres de texto.
- **Flexibilidad.** Los lenguajes de marcas deben ser lo suficientemente flexibles para poder utilizarse en diferentes contextos.
- **Compactibilidad.** Las marcas y el contenido se encuentran en el mismo documento.
- **Facilidad de procesamiento.** El formato debe ser diseñado de tal manera que el procesamiento automático sea sencillo.

## 1.4. Lenguajes de marcas y dialectos más comunes

Actualmente, los lenguajes de marcas más habituales tienen como objetivos principales la **representación de información** (HTML y XHTML) o el **almacenamiento e intercambio de información** (XML y sus dialectos, y el formato de intercambio JSON).

Vamos a presentar algunos de los lenguajes de marcas y dialectos de uso más común en la actualidad, pero no son los únicos.

# 1.4. Lenguajes de marcas y dialectos más comunes

## 1.4.1 HTML

Es el lenguaje de marcas más conocido y utilizado, y el W3C se encarga de su mantenimiento y estandarización. La primera versión de **HTML5** se publicó en octubre de 2014 y supuso un cambio radical y muy exitoso en el lenguaje.

Su objetivo principal es **proporcionar un mecanismo para diseñar páginas web** que van a ser representadas en un navegador, ya sea este ejecutado en un ordenador, en una videoconsola o en un dispositivo móvil.

Versión	Año
HTML	1991
HTML 2.0	1995
HTML 3.2	1997
HTML 4.01	2000
HTML5	2014



## 1.4. Lenguajes de marcas y dialectos más comunes

### 1.4.1 HTML

El éxito de **HTML5** es indudable. Los diseñadores gráficos cuentan con una capacidad ilimitada de elementos de maquetación gracias al uso conjunto con las **hojas de estilo en cascada (CSS)**.

También los programadores de aplicaciones web han podido aprovechar las mejoras que se han ido incorporando al lenguaje de marcado **HTML** para programar aplicaciones más funcionales y consistentes.

La extensión de los ficheros **HTML** suele ser **.html** o **.htm**

# 1.4. Lenguajes de marcas y dialectos más comunes

## 1.4.2 XML

XML (lenguaje de marcado extensible o **Extensible Markup Language**) es el lenguaje de intercambio de información entre sistemas por excelencia. Un documento XML puede servir para almacenar la configuración de una aplicación, los mejores marcadores de un videojuego, el diseño de una interfaz de usuario o la información de un gráfico vectorial.

La primera versión de XML salió a la luz en 1998. Es, al igual que HTML, una especificación del W3C y se define como «de propósito general». Esto significa que es un **metalenguaje** al venir **definido por reglas** y no por etiquetas concretas.

La última versión vigente de XML es la 1.1.

Debido a su naturaleza flexible y no específica, XML no requiere de revisiones ni mejoras frecuentes, ya que alberga en sí mismo la capacidad suficiente para adaptarse a las necesidades presentes y futuras.

La extensión de los ficheros XML suele ser .xml.

## 1.4. Lenguajes de marcas y dialectos más comunes

### 1.4.3 XHTML

El lenguaje de marcas **XHTML** (**eXtensible HyperText Markup Language**) se puede definir de manera simplificada como la versión en formato **XML** del lenguaje de marcas **HTML**. Se presentó en el año 2000 como una recomendación del W3C.

**XHTML** tiene la misma relación con **XML** que **HTML** con **SGML**. Tanto **HTML** como **XHTML** surgen a partir de un **metalenguaje** y ambos son lenguajes de marcado para el **desarrollo de páginas web**. La diferencia fundamental es que **XHTML** tiene una sintaxis más estricta y es más formal que **HTML**.

## 1.4. Lenguajes de marcas y dialectos más comunes

### 1.4.3 XHTML

Como consecuencia negativa, una página **HTML mal construida** es difícil de procesar por un **programa informático**, ya que estos no son tan flexibles como las personas. Si un programa «se inventa» qué es la **cantidad de un producto** en un listado y **qué es el precio**, las consecuencias pueden ser imprevisibles.

**HTML** tiene una sintaxis definida por un estándar, tal y como se ha citado previamente, pero son los navegadores los programas que procesan los documentos y los representan en pantalla. Los navegadores, cuando encuentran un **documento HTML mal construido** (nombres de etiquetas incorrectas, atributos inexistentes, errores en la estructura del documento, etc.), **interpretan** («se inventan») lo que quería decir el autor de la página.

## 1.4. Lenguajes de marcas y dialectos más comunes

### 1.4.3 XHTML

La mayoría de las veces aciertan y este funcionamiento ha sido una gran contribución al éxito de la web: a más sencillez, más uso. Por otra parte, si la interpretación es errónea, el autor podrá ver en su navegador que la página está mal construida o, como mal mayor, la persona que acceda a la página verá una página mal maquetada.

La extensión de los ficheros XHTML suele ser `.xhtml`.

## 1.4. Lenguajes de marcas y dialectos más comunes

### 1.4.4 JSON

JSON es el acrónimo de JavaScript Object Notation («notación de objetos de JavaScript»).

En realidad, es el formato de texto en el que JavaScript representa sus objetos, incluidos los arrays. Pese a esa naturaleza original tan específica, JSON ha abandonado su vínculo con JavaScript y se ha universalizado como alternativa a XML para realizar almacenamiento e intercambio de datos entre sistemas independientemente de la plataforma y del lenguaje de programación. Casi todos los lenguajes que se usan en la actualidad disponen de analizadores JSON.

# 1.4. Lenguajes de marcas y dialectos más comunes

## 1.4.4 JSON

En el siguiente código se puede observar un ejemplo de fichero en formato JSON.

```
{
  "asunto": "Lenguaje de Marcas",
  "publicacion": "2021",
  "editorial": "Desconocido",
  "lenguajes": [
    {
      "nombre": "HTML",
      "uso": "Web"
    },
    {
      "nombre": "XML",
      "uso": "Intercambio de datos"
    },
    {
      "nombre": "JSON",
      "uso": "Intercambio de datos"
    }
  ]
}
```

## 1.4. Lenguajes de marcas y dialectos más comunes

### 1.4.4 JSON

**JSON** es más ligero y cuenta con una sintaxis más sencilla que **XML**, pero por el contrario no dispone de las herramientas y riquezas que sí tiene **XML**. El debate entre **JSON** y **XML** no debe existir: hay contextos en los que **JSON** es más apropiado que **XML** y viceversa.

De hecho, es frecuente encontrar ambas alternativas conviviendo en un mismo sistema.

La extensión de los ficheros **JSON** suele ser .json.



## 1.4. Lenguajes de marcas y dialectos más comunes

### 1.4.5 SVG

SVG (Scalable Vector Graphics o gráficos vectoriales escalables) es un dialecto de XML. Se utiliza para representar gráficos vectoriales de dos dimensiones.

La extensión de los ficheros SVG suele ser `.svg` o `.svgz`.

### 1.4.6 MathML

MathML (Mathematical Markup Language o lenguaje de marcado matemático) es un dialecto de XML. Se utiliza para representar notaciones matemáticas.

### 1.4.7 NewsML

NewsML (News Markup Language o lenguaje de marcado de noticias) es un dialecto de XML. Se utiliza para representar noticias electrónicas.

## 1.4. Lenguajes de marcas y dialectos más comunes

### 1.4.8 FXML

FXML es un dialecto de XML. Se emplea para representar las interfaces de usuario en la tecnología JavaFX del lenguaje de programación Java.

### 1.4.9 XAML

XAML (eXtensible Application Markup Language o lenguaje de marcado extensible de aplicación) es un dialecto de XML. Es el lenguaje en el que se definen las interfaces de usuario en las aplicaciones para el sistema operativo Windows desarrolladas con la tecnología de Microsoft.

### 1.4.10 ANDROID

En las aplicaciones Android, la forma habitual de diseñar las interfaces de usuario es mediante ficheros escritos en un dialecto de XML. En estos ficheros se definen los esquemas de distribución, así como los componentes y sus características que van a permitir al usuario interactuar con la aplicación. **EXISTEN MUCHOS MÁS LENGUAJES DE MARCAS...**

# INSTALACIÓN WEBSTORM

<https://www.jetbrains.com/es-es/webstorm/>

