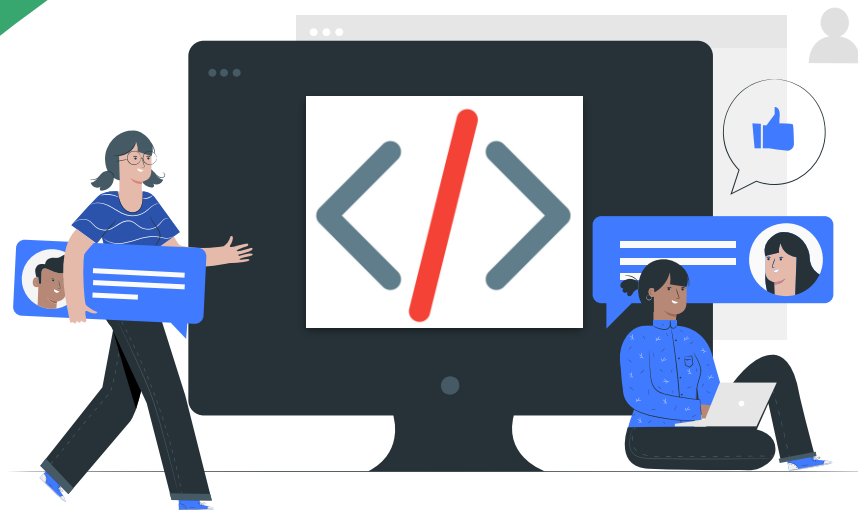


UT1. Introducció als llenguatges de marques.

LLENGUATGES DE MARQUES I SISTEMES DE GESTIÓ DE LA INFORMACIÓ – 1º DAM




Rubén Mayor Hernández



Index



- 
1. Concepte i avantatges
 2. Evolució dels llenguatges de marques
 3. Característiques
 4. Classificació
 5. Organitzacions i estàndars
 6. Llenguatges de marques. Etiquetes, elements i atributs
 7. Principals llenguatges de marques. HTML
 - 7.1 Evolució de HTML
 8. Principals llenguatges de marques. XML
 9. Principals llenguatges de marques. XHTML
 10. Principals llenguatges de marques. JSON
 11. Principals llenguatges de marques. RSS y Atom
 12. Evolució de CSS
 13. Eines d'edició
 14. Bibliografia

1. Concepte i avantatges

- Una marca és una senyal intercalada en un text, que delimita part d'aquest, i en alguns casos li aplica un determinat format (també hi ha marques que s'utilitzen amb altres propòsits).
- Les marques més comunes estan formades per una paraula que descriu la seua funció, delimitada pels símbols menor (<) i major (>), com per exemple <html>. Solen aparéixer per parelles, amb una marca d'inici i altra de tancament.

Por exemple:

<h1> Primera prova </h1>

<h2> Segona prova </h2>

(*) Activitat: Guarda el marcat anterior en un fitxer .txt y obri-ho amb un navegador web. Fes el mateix, però esta vegada guarda el fitxer en format .html.

Els llenguatges de marques (també anomenats llenguatges de marcat) són aquells que **combinen la informació**, generalment textual, que conté un document **amb marques** o anotacions relatives a l'estructura de les dades o de la forma de representar-los.



El llenguatge de marques és el que especifica quines seran les etiquetes possibles, on s'han de col·locar i el significat que tindrà cadascuna d'elles.

D'altra banda, cal tindre en compte que les pròpies etiquetes o marques **generalment no solen presentar-se a l'usuari final**, ja que aquest sol estar interessat en el propi contingut de el document.



1. Concepte i avantatges

- Els llenguatges de marques **no són equivalents als llenguatges de programació** tot i que es diguen també llenguatges. De fet, no hem d'utilitzar la paraula 'programar' quan ens referim als llenguatges de marques, ja que **no disposen dels elements típics** com ara variables, arrays, sentències de control, funcions, etc.
- No obstant això, els llenguatges de marques **es poden combinar** dins d'el mateix document amb altres llenguatges com Python o PHP, que sí que són llenguatges de programació, amb l'objectiu d'aportar funcionalitat i dinamisme a la pàgina web.
- Els llenguatges de marques **són interpretats per multitud de clients**, no només navegadors, per això parlarem d'**agent d'usuari** (user-agent) en lloc de navegador (cas d'aplicacions mòbils híbrides).

1. Concepte i avantatges

- El llenguatge de marques ha de ser **independent del destinatari final**. És **l'interpret** del llenguatge qui s'encarrega de representar les marques de la forma adequada. Per tant, es poden trobar **discrepàncies** en la forma de representar les etiquetes pels diferents navegadors. Per exemple, en aquest enllaç podràs comprovar quines versions de navegadors suporten l'element <svg> o el tipus de camp per a dates:

http://www.w3schools.com/html/html5_svg.asp

https://www.w3schools.com/TAGS/att_input_type_date.asp

- **SVG** significa Scalable Vector Graphics i defineix un llenguatge basat en XML per a la construcció de gràfics vectorials 2D amb multitud d' efectes i característiques avançades. SVG és una recomanació del W3, el contingut de cada fitxer és XML i no un format binari tancat.

1. Concepte i avantatges

- Per **independitzar la representació de la pàgina web del seu contingut**, es va crear el llenguatge **CSS** (Cascade Style Sheet), mitjançant el qual es pot especificar amb major precisió i eficàcia la representació de la informació, per a cada intèrpret i per a diferents suports, com a monitors , dispositius mòbils, paper, etc.
- A causa de l'auge dels dispositius mòbils, moltes pàgines web presenten diferents **versions adaptades** al dispositiu que utilitza l'usuari. Per exemple: <http://iespacomolla.es>

1. Concepte i avantatges

Ejemplos:

- XML, XHTML, HTM o HTML5, representen les dades, la informació a representar. Alguns llenguatges basats en XML com MathML, NewsML o CML defineixen conjunts concrets de tags destinats a un fi específic com són el llenguatge matemàtic, notícies electròniques o notació química.
- CSS o XSL defineixen la presentació que tindran les dades en diversos mitjans (HTML i documents XML, respectivament).
- SVG defineix la representació de gràfics dins de qualsevol document.
- SMIL permet integrar àudio, vídeo, imatges, text o qualsevol altre contingut multimèdia.
- XAML defineix interfícies gràfiques per a aplicacions del S.O. Windows.
- LaTeX, sistema de composició de textos per documents d'alta qualitat, s'utilitza molt en ambients científics.

2. Evolució dels llenguatges de marques

- En els anys 60 les empreses de publicació i maneig de documents electrònics tenien el problema de **manca de compatibilitat** entre aplicacions. El problema existent era que cada aplicació utilitzava les seves pròpies marques per descriure els diferents elements, això **impedia l'intercanvi de documents entre plataformes**. Una altra mancança important era la separació entre estructura i aspecte del document.
- IBM va intentar primerament resoldre el problema mitjançant el **llenguatge GML**, i més tard la ISO (International Organization for Standardization) el va convertir en **SGML**, que és un **metallenguatge** i constitueix un conjunt de normes que permeten crear altres llenguatges de marques.
- **HTML** va ser creat a partir de SGML.

3. Característiques

Text pla	<ul style="list-style-type: none">- Compostos únicament per caràcters de text, en contraposició als arxius binaris, que poden contenir imatges, so, fitxers comprimits ...- Els caràcters de text es poden codificar amb diferents codis (ASCII, UTF-8 ...)- Poden ser oberts directament per un editor de text, mentre que els binaris necessiten programari específic.- S'independitza els fitxers de sistema operatiu, facilitant la interoperabilitat (essencial en les comunicacions per internet).
Compacitat	<ul style="list-style-type: none">- Les instruccions de marcat es barregen amb el propi contingut: HTML: <h2> Contingut </ h2> LaTeX: \section{Título} Wikitexto: == Título ==
Independència de l'agent d'usuari	<ul style="list-style-type: none">- El mateix document pot ser interpretat de forma diferent pel dispositiu final.
Especialització	<ul style="list-style-type: none">- Els llenguatges de marques s'utilitzen en multitud d'àrees, com gràfics vectorials, sindicació de continguts, notació científica, interfícies d'usuari, síntesi de veu, etc.
Flexibilitat	<ul style="list-style-type: none">- Els llenguatges de marques es poden combinar amb altres llenguatges, en el mateix arxiu, com HTML amb Python, JavaScript o PHP.

4. Classificació. Tipus de marques.

1. De procediment i de presentació:

Indiquen el format de text o tipografia, sense especificar la seua estructura. Exemples: nroff, troff, RTF, TeX, Docbook, MS Word ...

2. Descriptiu, estructural o semàntic:

Indiquen les diferents parts en què s'estructura el document, però sense especificar com s'han de representar ni en quin ordre. Exemples: XML (i els seus derivats EBML, RDF, XFML, OWL, XTM), YAML ...

3. Híbrid:

Són una combinació dels dos tipus anteriors. Exemples: HTML, XHTML, WML ...

4. Classificació. Segons la funcionalitat

Per a crear documentació electrònica	<ul style="list-style-type: none">• RTF, TeX, troff, nroff• ASN.1, EBML, YAML• Wikitext, DocBook, LinuxDoc
Tecnologies d'internet	<ul style="list-style-type: none">• HTML, XHTML, WML (web)• GladeXML, Xforms, XAML (formularis/interfícies d'usuari)• RSS, Atom (sindicació de continguts)• WSDL, SOAP, UDDI (serveis web)• XMPP (missatgeria instantània)
De propòsit específic	<ul style="list-style-type: none">• MathML, CML (matemàtiques)• SSML, SRGS, VoiceXML (síntesi de veu)• MusicXML (partitures de música)• SVG, VML, X3D (gràfics vectorials)• SMIL (arxius multimèdia)• XLL (enllaços)• XLINK (associació de recursos)• XML Base (URI bàsic)• XPOINTER (localització de recursos)• XSLT (transformació de documents)• XTM (mapes percentuals)• RDF, XFML, OWL, XMP (catalogació i classificació de documents)• GML (informació geogràfica)• OFX (intercanvi d'informació financera)• ebXML (comerç electrònic)• XML Dsig, XML Enc, SAML, XACML, XKMS, XrML (seguretat)• Xinclude (inclusió d'arxius)

5. Organitzacions i estàndards

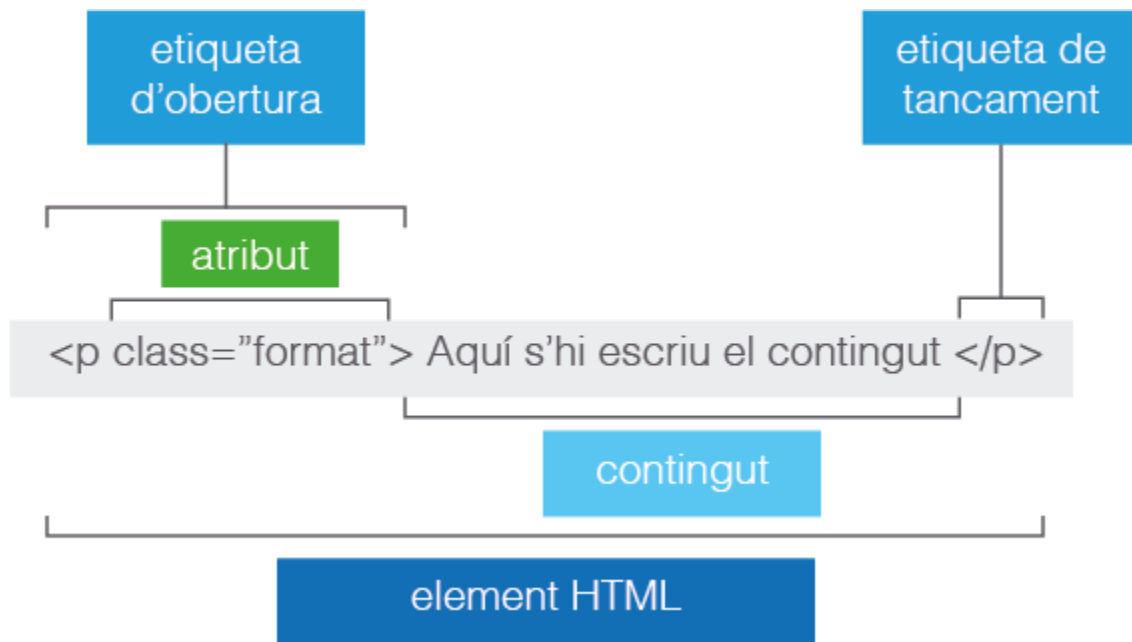
- Es pot definir l'estandardització com el procés d'**especificació de normes**, per garantir el correcte funcionament d'elements construïts de forma independent.
- Per a la **definició d'aquestes normes** hi ha organismes internacionals, nacionals i fins i tot organitzacions privades. Les més importants són **W3C**, **ISO** i **Open Source**.
- El **World Wide Web Consortium** (W3C) és una comunitat internacional que desenvolupa estàndards que assegurin el creixement de la web a llarg termini. És imprescindible el consens amb les empreses involucrades com els navegadors, cercadors, desenvolupadors web i fabricants de dispositius mòbils (Microsoft, Apple, Google ...)

(*) Consulteu: <http://www.evolutionoftheweb.com>

6. LM. Etiquetes, elements i atributs

- Una etiqueta (tag) és un text que va entre els símbols **menor que** i **major que** < >. Hi ha etiquetes d'obertura i de tancament. <Nom> </ nom>
- Els elements **representen estructures** mitjançant les quals s'organitzarà el contingut de el document o accions que es desencadenen quan el programa navegador interpreta el document. Consten de l'etiqueta **d'obertura**, l'etiqueta de **tancament** i de tot allò que **es troba entre les dues**. Alguns elements no tenen contingut, se'ls denomina **elements buits** i no han de portar etiqueta de tancament (etiqueta única).
- Un atribut és un **parell nom-valor** que es troba dins de l'etiqueta d'obertura d'un element, i indiquen les **propietats** que poden portar associades els elements.

6. LM. Etiquetes, elements i atributs



Exemple

```
<domicili>
...
  <propietari>
    <tractament>Mrs.</tractament>
    <nom>Mary</nom>
    <cognoms>McGoon</cognoms>
  </propietari>
  <carrer>1401 Main Street</carrer>
  <ciutat estat="NC">Anytown</ciutat>
  <codi_postal>34829</codi_postal>
</domicili>
```

En l'exemple anterior, l'element "propietari" conté tres elements fills: "tractament", "nom" i "cognoms". A més, "estat" es un atribut de l'element "ciutat"

7. Principals llenguatges de marques. HTML

- Llenguatge destinat a la **creació de pàgines web**, que ens permet combinar text i contingut multimèdia.
- Disposa **d'enllaços** o **hipervincles**, per connectar amb altres pàgines amb un contingut relacionat amb la pàgina actual.
- Creat a principis dels 90, estava pensat per a un **ús divulgatiu**.
- Ha evolucionat **de forma anàrquica**, donant lloc a deficiències compensades per altres llenguatges com JavaScript i CSS.
- Els diferents navegadors realitzen una implementació independent de l'estàndard, per això pot haver **discrepàncies** en la representació.
- Tot i que és possible crear documents HTML amb editors de text, també hi ha aplicacions específiques per facilitar la creació de codi.

7.1 Evolució del llenguatge HTML

- HTML1. (HYPERTEXT MARKUP LANGUAGE) Va ser la primera versió, creada per **Tim-Berners-Lee en 1991**.
- HTML2. La segona versió, va aparèixer el 1994 i es va acabar el 1996 amb l'aparició de HTML3. No va aportar grans canvis.
- HTML3.0. És la versió que en realitat planteja les bases de les següents versions d'HTML. Les regles i el funcionament d'aquesta versió estan donades per la "W3C". Va aparèixer en 1996 i afegeix moltes possibilitat, com **taules**, **applets**, **scripts**, **posicionament** de el text al voltant d'imatges, entre d'altres.

7.1 Evolució del llenguatge HTML

- HTML4. Aquesta és la versió més comuna d'HTML (en concret, HTML4.01). Va aparèixer el 1998 i proposa l'ús de **marcs** (que divideixen una pàgina web en diverses parts), **taules** més complexes, millores en les formes, etc. Més important encara, aquesta versió permet per primera vegada utilitzar **fulls d'estil** del famós CSS.
- HTML5. És l'última versió. Porta moltes millores com la possibilitat d'incloure fàcilment **vídeos**, millorar el contingut, noves característiques per als **formularis**, etc. Aquesta és la versió que descriurem durant el curs. L'estàndard està completat des de l'any 2014.

8. Principals llenguatges de marques. XML

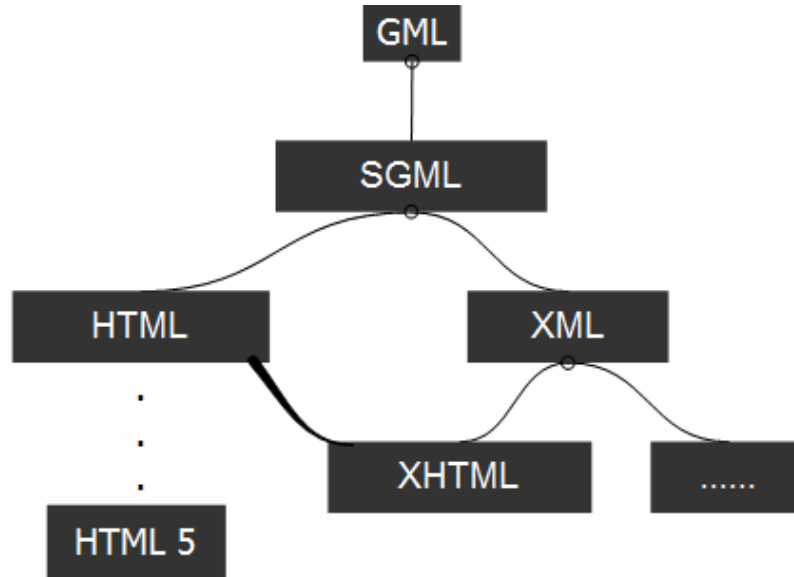
- Simplificació i adaptació del SGML que permet definir llenguatges específics. XHTML, SVG, MathML, RSS, i altres, es basen en XML.
- **Extensible**: es poden definir noves etiquetes.
- **Versàtil**: separa contingut, estructura i presentació.
- **Estructurat**: es poden modelar dades a qualsevol nivell de complexitat.
- **Validable**: cada document es pot validar enfront d'un DTD / Schema.
- **Obert**: independent d'empreses, sistemes operatius, llenguatges de programació o entorns de desenvolupament.
- **Senzill**: fàcil d'aprendre i d'utilitzar.

XML no s'utilitza només a Internet, sinó que és un estàndard consolidat per a l'intercanvi d'informació estructurada entre diferents plataformes.

Es pot usar en bases de dades lleugeres, editors de text, fulls de càlcul, transaccions comercials i en general on es necessita emmagatzemar informació sense les restriccions d'un SGBD relacional.



Llenguatges de marques



9. Principals llenguatges de marques. XHTML

- *(eXtensible HyperText Markup Language) Es pot definir de manera simplificada com la versió en format XML del llenguatge HTML. Any 2003 com a recomanació de W3C.*
- És una fórmula més estricta i formal que HTML.
- HTML té una sèrie de característiques que li impedeixen ser **document XML ben format**, com que l'etiqueta **
** no té etiqueta de tancament, En XML totes les etiquetes s'han de tancar, per tant en XHTML seria **
**.
- Una pàgina HTML mal construïda és difícil de processar pel navegador, i "s'inventa" el que voldria dir l'autor de la pàgina.

9. Principals llenguatges de marques. XHTML

- Amb XHTML és possible crear webs amb mateix etiquetatge i mateixa funcionalitat que amb HTML, però garantint estructura i sintaxi correcta.
- L'extensió sol ser .xhtml.
- Exemple de començament d' arxiu:

```
<?xml version="1.0"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<title>Ejemplo</title>
</head>
```

10. Principals llenguatges de marques. JSON

- Acrònic de **JavaScript Object Notation**, és un format amb el qual Javascript representa els seus arrays.
- S'ha universalitzat com a alternativa a XML per a emmagatzematge i intercanvi de dades.
- Gairebé tots els llenguatges disposen d'analitzadors JSON.
- És més lleuger i compta amb sintaxi més senzilla però menys rica que XML.
- L'extensió sol ser .json

JSON

```
{  
  "siblings": [  
    {"firstName": "Anna", "lastName": "Clayton"},  
    {"firstName": "Alex", "lastName": "Clayton"}  
  ]  
}
```

XML

```
<siblings>  
  <sibling>  
    <firstName>Anna</firstName>  
    <lastName>Clayton</lastName>  
  </sibling>  
  <sibling>  
    <firstName>Alex</firstName>  
    <lastName>Clayton</lastName>  
  </sibling>  
</siblings>
```

11. Principals LM. RSS y Atom.

- Són dialectes de XML.
- RSS (Really Simple Syndication).
- Per a la distribució de continguts fonts als quals els usuaris s'han subscrit.
- Atom és una alternativa després de RSS.
- Document en format RSS de AEMET:

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0">
  <channel>
    <title>AEMET - Avisos de fenómenos meteorológicos adversos -
Extremadura</title>
    <link>http://www.aemet.es/es/eltiempo/prediccion/avisos?k=ext</link>
    <description>Avisos de fenómenos meteorológicos adversos según el Plan
meteoalerta de la Agencia Estatal de Meteorología (AEMET) para Extremadura.</
description>
    <language>es</language>
```

12. Evolució del llenguatge CSS

Tot i que CSS no es considera un llenguatge de marques, sinó un llenguatge de disseny, cal introduir-lo dins del temari per donar forma a les nostres pràctiques. Les versions existents són la següents:

- **CSS1**. El 1996, va aparèixer la primera versió de CSS. En ella s'estableixen les bases d'aquest llenguatge que permeten millores de presentació a les pàgines web, com ara colors, marges, fonts, entre d'altres.

12. Evolució del llenguatge CSS

- **CSS2**. Va aparèixer el 1999 i va ser completat per CSS 2.1, aquesta nova versió de CSS afegeix nombroses opcions. Ara es poden utilitzar tècniques de posicionament molt precises que permeten veure els elements en el lloc desitjat de la pàgina.
- **CSS3**. Aquesta és l'última versió, que afegeix característiques molt esperades, com ara vores arrodonides, degradats, ombres entre d'altres. Després entendre HTML, ens centrarem en els estils amb CSS 3.

13. Eines d'edició

Eines amb l' edició de pàgines web i creació de documents XML. Igual que per a tot tipus de programari hi ha eines online i offline que s'han de descarregar i instal·lar.

- **Bloc de notes.** Editor de text que permet crear documents de text plànel. Extensió .txt i permet codificació ASCII, Unicode, UTF-8.
- **Gedit.** Editor de text per a l' entorn GNOME. Se sol utilitzar com a bloc de notes i com a entorn de desenvolupament.
- **Notepad++.** Potent editor de text amb llicència GPL. Esta escrito em C++. Permet seleccionar multitud de llenguatges de programació. Permet ressaltat de sintaxi, numeració de línies...

13. Eines d'edició

- **XML Copy Editor.** Editor de XML que permet crear i comprovar de manera molt senzilla si el document està bé format.
- **Atom.** Editor de text Open Source macOS, Linux, y Windows amb suport per a múltiples plug-in escrits en Node.js (que amplien les seves capacitats) i control de versions Git integrat, que pot utilitzar-se com a IDE per a un siní de llenguatges de programació.

14. Bibliografía

- [1] José Luis Comesaña. Lenguaje de marcas. Desarrollo de Aplicaciones Web, Sin editorial, 2011.
- [2] Fernando Paniagua Martín. Lenguaje de marcas y sistemas de gestión de información, Paraninfo 2021.
- [3] Juan Carlos Moreno, Sergio Luis González. Lenguaje de marcas y sistemas de gestión de información, Sintesis 2018.

A continuació oferim en ordre alfabètic els autors que han fet aportacions a aquest document:

- Silvia Amorós Hernández
- Sandra Deltell Torres
- Ana Martínez
- José Más Davó
- Vicente Peñataro Sánchez
- José Javier Segura Martínez