

UT1. Els llenguatges de marques

UT1. Els llenguatges de marques

1.1. Introducció

1.2. SGML. L'origen

1.3. Característiques dels llenguatges de marques

1.4. Classificació dels llenguatges de marques

1.5. Organitzacions i estàndards

1.6. Introducció als principals llenguatges de marques

1.6.1. HTML

1.6.2. XML

1.1 Introducció

Les persones utilitzem en el nostre llenguatge parlat certs gestos, diferent entonació, pauses, etc. que acompanyen el contingut que volem transmetre.

De manera semblant al llenguatge escrit utilitzem mides de lletra, negreta, numeració, vinyetes, taules, colors, etc. que també acompanyen la informació perquè sigui més fàcil d'entendre.

En definitiva **estem aplicant un determinat format a la informació que transmetem.**

1.1 Introducció

En el cas dels **documents** que intercanviem a través d'**Internet**, com ara les **pàgines web**, són els llenguatges de marques els que ens permeten aplicar aquest format.

Un document que contingui exclusivament text és perfectament llegible per nosaltres, encara que evidentment, tediós i inadequat per publicar pàgines web. Si li apliquem format mitjançant un **llenguatge de marques** com ara **HTML**, obtenim un arxiu també llegible però més difícil d'interpretar.

D'aquesta tasca se n'encarregarà el navegador o millor dit l'**agent d'usuari**, que interpreta les marques de format i les aplica convenientment al text per donar lloc a una pàgina web, que serà molt més agradable de llegir que el text original.

1.1 Introducció

Una "**marca**" és un senyal col·locat dins d'un text, per tal de delimitar una part del mateix i en molts casos, aplicar-li un determinat format (encara que hi ha marques amb altres propòsits).

Les marques més comunes estan formades per una paraula que descriu la seva **funció** tancada entre els símbols menor que < i més gran que > com <html>.

És molt habitual que apareguin per parelles, una de començament i una altra de fi.

Per exemple:

```
<h1> Hola, aquest text apareix més gran </h1>  
<h3> Aquest text apareix més petit, adéu </h3>
```

1.1 Introducció

Els llenguatges de marques no són equivalents als llenguatges de programació encara que es anomenin també llenguatges.

De fet, no hem d'utilitzar la paraula “programar” quan ens referim a llenguatges de marques, ja que no disposen dels elements típics com variables, arrays, sentències de control, funcions, etc.

No obstant això, els llenguatges de marques es poden combinar dins del mateix document, amb altres llenguatges com JavaScript o PHP, que sí que són llenguatges de programació, amb el objectiu d'aportar funcionalitat i dinamisme a la pàgina web.

1.1 Introducció

**llenguatges
de marques**



**llenguatges
de programació**



1.1 Introducció

Un altre aspecte important a tenir en compte quan parlem de llenguatges de marques és el destinatari de la informació.

Potser el més habitual, és un usuari final utilitzant un navegador web al pe de casa, però hem de considerar la resta d'opcions que van en augment, altres destinataris podrien ser: usuaris en dispositius mòbils, usuaris amb deficiències visuals o motrius, usuaris d'avançada edat, un perifèric com la impressora, els robots dels cercadors, etc.

Per aquesta raó, és més correcte utilitzar el terme general, **agent d'usuari** (user-agent) en lloc de navegador.

La presentació de la mateixa pàgina web per a cadascun d'aquests usuaris ha de ser lògicament molt diferent, així per exemple, un text en negreta es pot representar respectivament, per caràcters amb més gruix, per un volum més alt al sintetitzador de veu, per més punts al paper, etc.

1.1 Introducció

La qüestió és que **el llenguatge de marques ha de ser independent del destinatari final**, és el intèrpret del llenguatge qui s'encarrega de representar les marques de la manera adequada.

HTML per exemple, no especifica a les etiquetes com seran representades més tard pel navegador.

Aquesta és una de les raons per la qual podem trobar certes **diferències en la visualització d'una mateixa pàgina**, per part de diferents navegadors.

Exemple: Compatibilitat dels navegadors amb HTML5 html5test.com

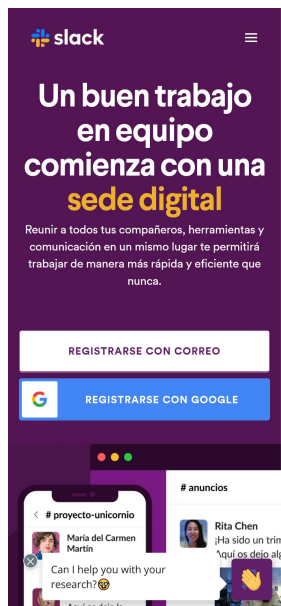
1.1 Introducció

D'altra banda, per independitzar encara més la representació de la pàgina web de la seva contingut, es va crear **CSS**, que serveix per definir estils.

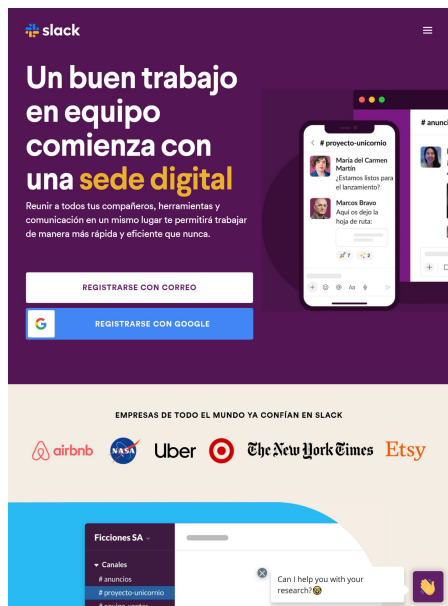
Mitjançant CSS podem especificar amb més precisió i eficàcia la **representació de la informació**, per cada intèrpret i per a diferents suports, com a monitors, dispositius mòbils, paper, veu, etc.

Donat l'auge dels dispositius mòbils, moltes pàgines presenten diferents versions adaptades al dispositiu que utilitza l'usuari.

1. Introducció



Mòbil



Tablet



Portàtil

www.slack.com

1.2. SGML. L'origen

Als 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 carència important era la separació entre estructura i aspecte del document.

IBM, empresa pionera en investigació en informàtica i electrònica va intentar resoldre aquests problemes a través d'un llenguatge de marques anomenat **GML (Generalized Markup Language)**.

GML independitza el document del dispositiu que l'utilitzarà, usant marques genèriques. D'altra banda, **GML incorpora marques descriptives per a l'estructura del document** que permeten distingir el text, de les llistes, les taules, etc. El mateix document pot, llavors, ser utilitzat per diversos dispositius, simplement especificant un perfil per a cada un.

1.2. SGML. L'origen

El 1986 GML va passar a mans d'**ISO** i es va convertir en **SGML** (ISO 8879), **Standard Generalized Markup Language**, programari lliure i de codi obert.

És important tenir en compte que SGML no és estrictament un llenguatge sinó un **metallenguatge**, és a dir, un conjunt de normes que permeten crear altres llenguatges de marques.

Això es fa definint un vocabulari o conjunt d'elements a utilitzar, i una gramàtica o conjunt de regles que regeixen l'ús dels elements i els seus atributs.

SGML, per tant, és un metallenguatge que permet definir llenguatges de marcatge. HTML per exemple, és un dels llenguatges creats a partir de SGML.

1.2. SGML. L'origen

- **Avantatges** de SGML: Reutilització de les dades, integritat i control sobre les dades, portabilitat, adaptabilitat.
- **Inconvenients** de SGML: Alta complexitat

Un **document SGML** consta de 2 parts:

- El **pròleg**: conté l'estructura.
 - La **declaració**: indica que el document és SGML i alguns paràmetres.
 - La **definició de tipus de document (DTD)**: indica la sintaxi particular del llenguatge creat.
- La **instància de document**: conté les dades.

1.2. SGML. L'origen

document HTML 4.01

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
<HTML>
  <HEAD>
    <TITLE>My first HTML document</TITLE>
  </HEAD>
  <BODY>
    <P>Hello world!
  </BODY>
</HTML>
```

DTD HTML 4.01

```
<!--===== The Anchor Element =====>
<!ENTITY % Shape "(rect|circle|poly|default)">
<!ENTITY % Coords "CDATA" -- comma-separated list of lengths -->

<!ELEMENT A - - (%inline;)* -(A) -- anchor -->
<!ATTLIST A
  %attrs;
  charset %Charset; #IMPLIED -- %coreattrs, %i18n, %events --
  type %ContentType; #IMPLIED -- char encoding of linked resource --
  name CDATA #IMPLIED -- advisory content type --
  href %URI; #IMPLIED -- named link end --
  hreflang %LanguageCode; #IMPLIED -- URI for linked resource --
  rel %LinkTypes; #IMPLIED -- language code --
  rev %LinkTypes; #IMPLIED -- forward link types --
  accesskey %Character; #IMPLIED -- reverse link types --
  shape %Shape; #IMPLIED -- accessibility key character --
  coords %Coords; #IMPLIED -- for use with client-side image maps --
  tabindex NUMBER #IMPLIED -- for use with client-side image maps --
  onfocus %Script; #IMPLIED -- position in tabbing order --
  onblur %Script; #IMPLIED -- the element got the focus --
  > -- the element lost the focus --
```

1.3. Característiques dels llenguatges de marques

TEXT PLA

Els fitxers de text pla són aquells que estan compostos únicament per caràcters de text, a diferència dels fitxers binaris que poden contenir imatges, so, fitxers comprimits, programes compilats, etc.

Aquests caràcters es poden codificar amb diferents codis segons l'idioma o alfabet que calgui, per exemple: ASCII, ISO-8859-15, UTF-8.

Un dels principals avantatges dels fitxers de text pla és que poden ser interpretats directament per un simple editor de text, a diferència dels binaris que necessiten programari específic (visors multimèdia, descompressors, compiladors, etc.)

Aquesta característica fa que els documents siguin **independents del sistema operatiu** o programa amb què van ser creats, això facilita la interoperabilitat, que constitueix una important avantatge per a l'**intercanvi d'informació** a Internet.

1.3. Característiques dels llenguatges de marques

COMPACTESA

Les instruccions de marcatge es barregen amb el propi contingut, per exemple, `<h2>Contingut</h2>` .

El codi entre claus angulars com `<h2>`, són instruccions de marcatge, també anomenades **etiquetes**. Aquesta etiqueta en concret és una etiqueta de presentació, indica que el text comprès ha de tenir el format assignat a la capçalera núm. 2. El text entre les marques és el propi contingut del document.

INDEPENDÈNCIA DEL DISPOSITIU FINAL

El mateix document pot ser interpretat de diferents formes segons el dispositiu final, així tindrem diferents resultats si es fa servir un dispositiu mòbil, un ordinador de sobretaula o una impressora.

1.3. Característiques dels llenguatges de marques

ESPECIALITZACIÓ

Inicialment els llenguatges de marques es van idear per visualitzar documents de text, però progressivament **s'han començat a utilitzar en moltes altres àrees** com ara gràfics vectorials, sindicació de continguts, notació científica, interfícies d'usuari, síntesi de veu, etc.

FLEXIBILITAT

Els llenguatges de marques es poden combinar al mateix fitxer amb altres llenguatges, com HTML amb PHP i JavaScript. Fins i tot hi ha etiquetes específiques com a `<script>`.

XML ha permès que es puguin combinar diversos llenguatges de marques diferents en un mateix fitxer, com en el cas de XHTML amb MathML i SVG.

1.4. Classificació dels llenguatges de marques

Normalment els llenguatges de marques se solen classificar en tres tipus, atenent el tipus de marques que utilitzen:

- **De presentació:** Indiquen el format del text o tipografia, sense especificar-ne l'estructura, per exemple augmentar la mida de la font, centrar o canviar a negreta.

Aquesta categoria inclou els llenguatges de procediment que agrupen diverses marques de presentació en una macro. Per exemple, per formatar un títol, hi ha d'haver una sèrie de directives immediatament abans del text indicant: mida de lletra 16p, font Arial, negreta. Just després del títol hi ha d'haver etiquetes inverses que anul·lin el format, per continuar amb el text normal. El programari que representa el document ha d'interpretar el codi en el mateix ordre a que apareix.

Els processadors de text i en general les aplicacions d'edició professional utilitzen aquest tipus de marcatge.

Exemples: nroff, troff, RTF, TeX, Docbook (derivats de SGML)

1.4. Classificació dels llenguatges de marques

- **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.

XML és un metallenguatge expressament dissenyat per generar marcatge descriptiu i els llenguatges derivats de XML amb aquest propòsit són: EBML, RDF, XFML, OWL i XTM.

Encara que XML emmagatzema informació de tot tipus, els altres tenen contingut específic.

Aquests llenguatges creen documents amb estructura en arbre que emmagatzemen informació, per això són bases de dades, però no usen taules ni respecten les regles d'integritat pròpies de les BD Relacionals, per això se'ls anomena **bases de dades semiestructurades**.

Exemples: ASN.1, YAML, EBML, RDF, XFML, OWL, XTM (derivats d'XML)

- **Híbrid:** Llenguatges que contenen marques dels dos tipus anteriors indistintament.

Exemples: HTML (derivat de SGML), XHTML, WML (derivats d'XML)

1.4. Classificació dels llenguatges de marques

Una classificació de llenguatges de marques atenent a la seva funcionalitat:

- **Per a crear documentació electrònica:**

- RTF, TeX, troff, nroff
- ASN.1, EBML, Y AML
- Wikitexto, DocBook, LinuxDoc

- **Tecnologies d'Internet:**

- HTML, XHTML, WML (pàgines web)
- GladeXML, XForms, XAML (formularis/interfícies d'usuari)
- RSS, Atom (sindicació de continguts)
- WSDL, SOAP, UDDI (serveis web)
- XMPP (missatgeria instantània)

1.4. Classificació dels llenguatges de marques

- **De propòsit específic:**

- MathML, CML (fòrmules matemàtiques)
- SSML, SRGS, VoiceXML (síntesi de veu)
- MusicXML (partitures de música)
- SVG, VML, X3D (gràfiques vectorials)
- SMIL (arxius multimèdia)
- XLL (enllaços):
 - XLINK (associació de recursos)
 - XML Base (URI bàsic)
 - XPOINTER (localització de recursos)
- XSL T (transformació de documents)
- XTM (mapes conceptuals)
- RDF, XFML, OWL, XMP (catalogació i classificació de documents, metadades)
- GML (informació geogràfica)
- OFX (intercanvi d'informació financiera)
- ebXML (comerç electrònic)
- XML Dsig, XML Ene, SAML, XACML, XKMS, XrML (seguretat)
- Xlnclude (inclusió d'arxius)

1.5. Organitzacions i estàndards

Podem definir la **normalització** o **estandardització** com el procés d'especificació de normes, per garantir el funcionament correcte d'elements construïts de forma independent.

Aplicat al context dels llenguatges de marques, seria, per exemple, el desenvolupament de pàgines web atenent les especificacions oficials del llenguatge utilitzat.

Per definir aquestes normes hi ha organismes internacionals, nacionals fins i tot organitzacions privades. Algunes organitzacions importants **W3C** i **ISO**.

Segons el mateix W3C:

"El World Wide Web Consortium (W3C) és una comunitat internacional que desenvolupa estàndards que assegurin el creixement de la Web a llarg termini."

1.5. Organitzacions i estàndards

Podem definir la **normalització** o **estandardització** com el procés d'especificació de normes, per garantir el funcionament correcte d'elements construïts de forma independent.

Aplicat al context dels llenguatges de marques, seria, per exemple, el desenvolupament de pàgines web atenent les especificacions oficials del llenguatge utilitzat.

Per definir aquestes normes hi ha organismes internacionals, nacionals fins i tot organitzacions privades. Algunes organitzacions importants **W3C** i **ISO**.

Segons el mateix W3C:



"El World Wide Web Consortium (W3C) és una comunitat internacional que desenvolupa estàndards que assegurin el creixement de la Web a llarg termini."

1.5. Organitzacions i estàndards

El W3C rep ingressos de les quotes dels seus membres, de beques de recerca, subvencions i donacions privades. Per tant, **no és una empresa amb finalitats lucratives** sinó una comunitat heterogènia formada per diferents organismes que en són membres que són els que fabriquen principalment els estàndards.

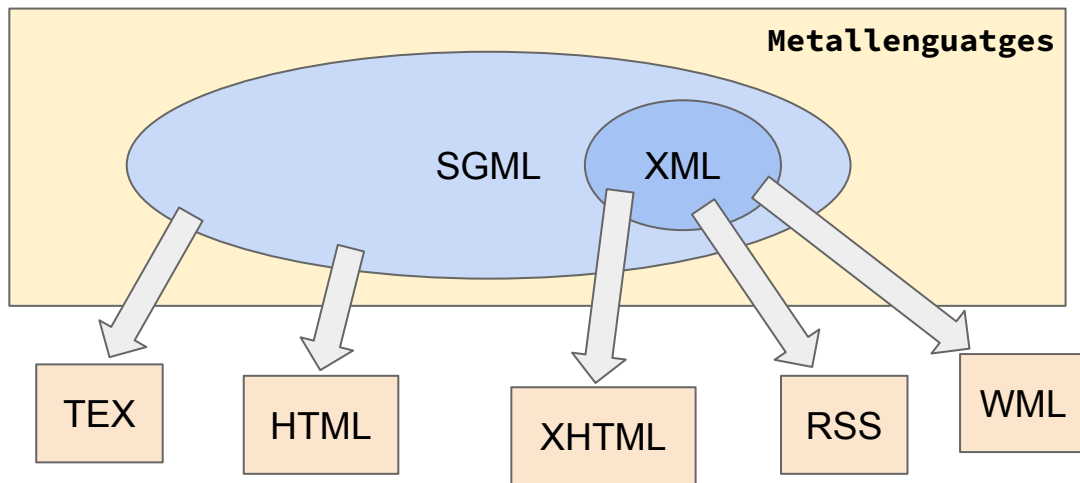
Tota organització d'estàndards pretén desenvolupar normes que siguin d'ampli seguiment per part de la comunitat, per a això **és imprescindible el consens** amb les empreses involucrades com els navegadors, cercadors, desenvolupadors web i fabricants de dispositius mòbils.

Entre els seus membres hi ha les principals empreses del sector com Microsoft, Apple o Google entre d'altres.

A més de produir estàndards, la Comunitat W3C ha creat programari de codi obert, sent el més conegut el **validador W3C**, que ens serà d'utilitat amb HTML, CSS, mobileOK i altres tecnologies.

1.6. Introducció als principals llenguatges de marques

L'origen dels llenguatges de marques com ja sabem és SGML, del qual se'n deriven directament alguns llenguatges com HTML. També es va crear a partir d'ell per simplificació XML, un altre metallenguatge més fàcil d'utilitzar i entendre. A partir d'XML s'han creat molts llenguatges com XHTML, RSS i un llarg etc.



1.6.1. HTML

HTML és un llenguatge destinat a la creació de **pàgines web**, que ens permet barrejar a un mateix document, text i contingut multimèdia.

A més, disposa d'una potent eina que són els **enllaços** o hipervincles, per **connectar amb altres pàgines** el contingut està relacionat la pàgina actual. Això suposa un avantatge important davant dels documents impresos (llibres, revistes, etc.)

HTML va ser creat per Tim Berners Lee a principis dels 90 amb objectius divulgatius. No va pensar que la web arribés a ser tan potent, de manera que inicialment no podia donar suport a totes les funcions que més tard hauria de fer. Aquestes mancances s'han anat resolent mitjançant la incorporació successiva de modificacions i nous elements, són les diferents versions d'HTML.

Aquesta **evolució poc planificada** i de vegades anàrquica d'HTML, ha suposat una sèrie d'inconvenients i deficiències, que han estat superats amb la introducció d'altres tecnologies capaços de millorar el funcionament i l'organització dels llocs web.

1.6.1. HTML

Exemples són CSS, JavaScript i els complements de navegador.

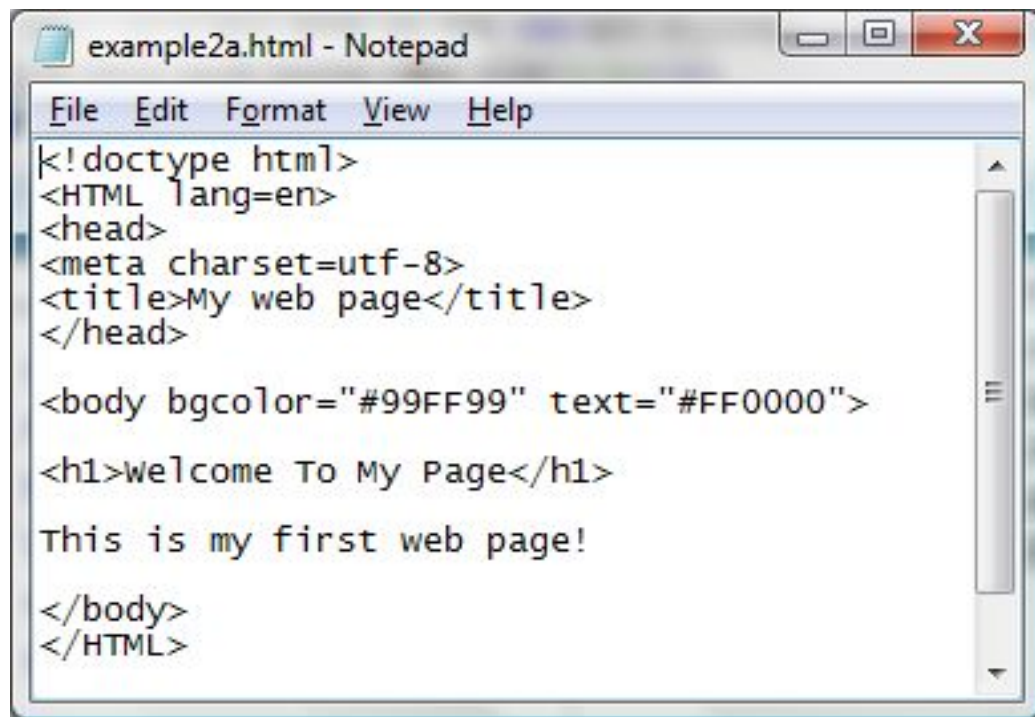
Un altre dels problemes que té HTML, especialment quan es combina amb CSS, ha estat la **diferent interpretació que fan els navegadors** de la mateixa pàgina web, a causa del diferent ritme en què els navegadors incorporen les novetats.

Això ha provocat l'aparició de hacks, que són petits trossos de codi per aconseguir un comportament homogeni als navegadors.

A més del navegador necessari per veure els resultats de la nostra feina, en necessitem una altra eina capaç de crear la pàgina. Un arxiu HTML és text pla, per la qual cosa per escriure a HTML necessitem un **simple editor de textos**.

Hi ha aplicacions específiques per a la creació de pàgines web, que ens ofereixen moltes funcionalitats per augmentar la nostra productivitat.

1.6.1. HTML



```
example2a.html - Notepad
File Edit Format View Help
<!doctype html>
<HTML lang=en>
<head>
<meta charset=utf-8>
<title>My web page</title>
</head>

<body bgcolor="#99FF99" text="#FF0000">

<h1>welcome To My Page</h1>

This is my first web page!

</body>
</HTML>
```

1.6.2. XML

És una simplificació i adaptació de SGML que permet definir llenguatges específics. Per lo tant, XML no és un llenguatge en particular, sinó una manera de definir llenguatges per a diferents necessitats, és a dir, lo que hem anomenat un metallenguatge.

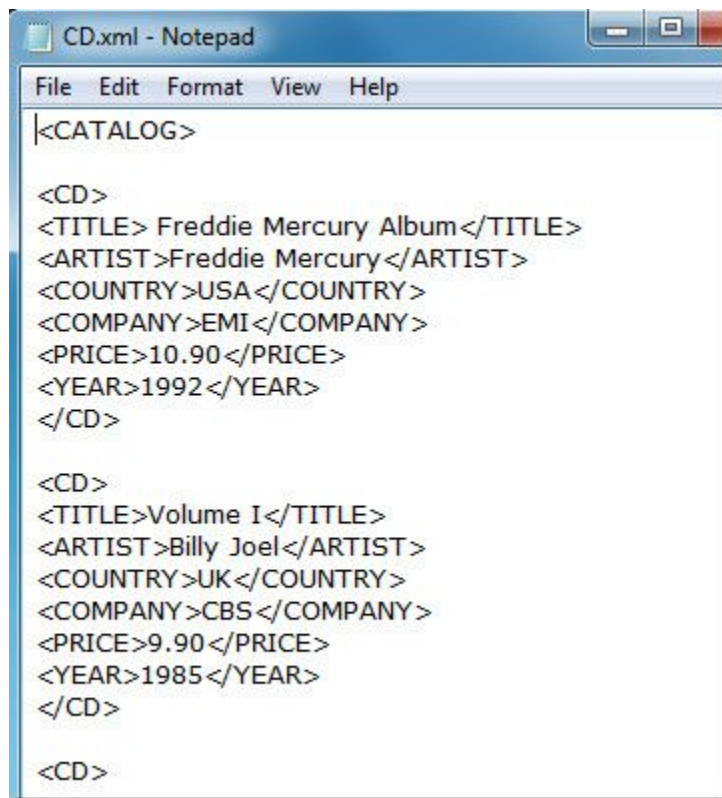
Alguns dels llenguatges que es basen en XML per a la seva definició són XHTML, SVG, MathML, RSS, etc.

Com a característiques podem citar:

- **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 davant d'un DTD/Schema
- **Obert:** independent d'empreses, sistemes operatius, llenguatges de programació o entorns de desenvolupament.
- **Senzill:** fàcil d'aprendre i fer servir.

XML no s'utilitza només a Internet, sinó que s'està convertint en un estàndard per a l'**intercanvi d'informació** estructurada entre diferents plataformes. Es pot fer servir en bases de dades lleugeres, editors de text, fulls de càlcul, transaccions comercials i en general on cal emmagatzemar informació sense les restriccions d'un SGBD Relacional.

1.6.2. XML



```
CD.xml - Notepad
File Edit Format View Help
<CATALOG>

  <CD>
    <TITLE> Freddie Mercury Album</TITLE>
    <ARTIST>Freddie Mercury</ARTIST>
    <COUNTRY>USA</COUNTRY>
    <COMPANY>EMI</COMPANY>
    <PRICE>10.90</PRICE>
    <YEAR>1992</YEAR>
  </CD>

  <CD>
    <TITLE>Volume I</TITLE>
    <ARTIST>Billy Joel</ARTIST>
    <COUNTRY>UK</COUNTRY>
    <COMPANY>CBS</COMPANY>
    <PRICE>9.90</PRICE>
    <YEAR>1985</YEAR>
  </CD>

  <CD>
```