

Tema 1. Introducció als llenguatges de marques

1.1. Origen

Quan ens comuniquem oralment, no només transmetem informació amb les paraules. Fem servir altres recursos per donar èmfasi o afegir nova informació. Mitjançant l'entonació, les pauses o determinats gestos afegeixen informació a les paraules que acompanyen. Tanmateix, quan fem servir el llenguatge escrit utilitzem altres recursos per afegir també nova informació: cursiva, signes de puntuació, subratllat, majúscules, etc. Quan apliquem negreta a una paraula, estem remarcant la seva importància.

En l'àmbit informàtic, l'exemple més conegut són les pàgines web, un llenguatge de marques que ens permet donar format i transformar la informació que es mostra al nostre navegador web.

Les pàgines web fan servir el llenguatge de marques HTML. Aquest llenguatge permet modificar el títol d'una pàgina perquè es mostri més gran, modificar el color de la font, la disposició dels elements de text dintre de la pàgina, etc. És a dir permet aplicar un format determinat a la informació.

Una **marca** és una senyal que delimita un text, donant-li informació al navegador de com interpretar i renderitzar aquest text al navegador. Les marques més comuns estan formades per una paraula delimitada pels símbols "<" i ">". Per exemple <html> o <h2>.

Activitat 1

Escriu el següent text en un document anomenat activitat1.txt, a continuació guarda'l amb extensió .html i obre ambdós documents amb el navegador. Comprova com canvia el format de text:

```
<h1>Hola mundo pequeño</h1>
<h3>Hola mundo todavía más pequeño </h3>
```

NOTA: Feu una captura de pantalla i deseu-la com activitat1_inicials.jpg.

Els llenguatges de marques no es consideren llenguatges de programació i per tant no podem dir que estem programant, ja que aquests llenguatges no disposen de sentències de control, funcions, etc. Els llenguatges de marques sí que es podem combinar amb altres llenguatges de programació- Java, Javascript, PHP, etc.- per mostrar la informació per pantalla d'una forma més ordenada i amigable.

1.2. Cronologia dels llenguatges de marques

- **GML** va ser creat a finals dels 60 per IBM com a solució a la falta de compatibilitat dels documents electrònics que s'enviaven entre els diferents mainframes de l'època.
- GML va passar a formar part de l'estàndar ISO i es va convertir en **SGML** (ISO 8879) el 1986. SGML es va definir com a metallenguatge, és a dir,

permet crear altres llenguatges de marques.

- L'any 1996 va sorgir el primer esborrany de l'especificació del **XML**. La primera versió de XML (**XML 1.0**) es va definir l'any 1998. XML és deriva de SGML i es considera un **metallenguatge**.
- La versió XML 1.1 va sortir al febrer de 2004. Conté algunes característiques amb usos molt específics i no està gaire implementada. Es recomana només si volem fer servir les seves característiques especials.
- **HTML** o HyperText Markup Language és el llenguatge de marques estandar dissenyat per a ser mostrat al navegador i va ser creat a finals de 1990 per Tim Berners-Lee. HTML també deriva de SGML, però no és un metallenguatge. L'última versió de HTML, HTML5 va sortir el 2014. Apart dels llenguatges abans esmenats, n'hi ha d'altres basats en XML

La següent imatge ens relaciona els diferents llenguatges de la família GML.

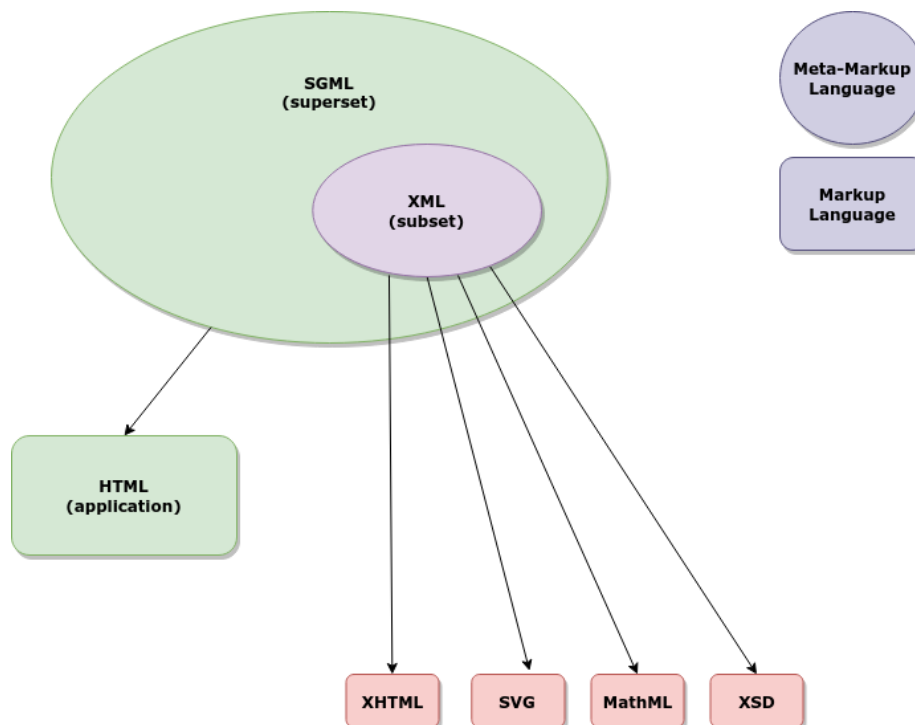


Figure 1: Llenguatges família GML

1.3. Característiques dels llenguatges de marques

Els llenguatges de marques tenen les següents característiques:

- **Text pla.** Els llenguatges de marques fan servir fitxers de text pla. Això vol dir que poden ser oberts per qualsevol editor de text a qualsevol sistema operatiu, ja sigui amb la línia d'ordres (shell) o amb una aplicació d'escriptori (p. ex. notepad).
- **Compactness (compactibilitat).** Les instruccions de marcatge estan barrejades amb el propi contingut, com per exemple a l'Activitat 1.
- **Independència del dispositiu final.** El mateix document es pot interpretar de diferent forma depenent de la resolució del dispositiu: mòvil, PC, tablet.
- **Especialització.** Els llenguatges de marques han evolucionat i és fan servir en moltes àrees, no només en la visualització de documents. Tenim llenguatges per a gràfics vectorials (SVG), matemàtiques (MathML), posicionament GPS (GPX), sindicació de continguts (RSS), ePub, síntesi de veu, etc.
- **Flexibilitat.** Els llenguatges de marques es poden combinar en el mateix arxiu amb altres llenguatges, com per exemple HTML amb llenguatges com PHP o Javascript. En aquest cas es fan servir etiquetes especials per indicar quan comença i termina el codi de programació. XHTML permet combinar MathML i SVG en un mateix document.

1.4. Classificació dels llenguatges de marques

Podem classificar els llenguatges de marques segons els tipus de marques que utilitzen:

- **Presentació.** Aquests llenguatges especifiquen característiques com el tamany de la font, centrar el text o canviar una paraula a negreta o cursiva. Els procesadors de text i en general les aplicacions d'edició professional fan servir aquest marcatge.
- **Descriptiu, estructural o semàntic.** En aquest cas els llenguatges especifiquen com s'estructura el document, sense especificar com s'han de presentar a la pantalla. Aquests llenguatges creen documents amb **estructures d'arbre** que emmagatzemen la informació. Per tant podríem parlar de bases de dades. No obstant això, els llenguatges de marques no tenen algunes de les característiques de les BBDD com les taules o les regles d'integritat, i també se les anomena **Bases de Dades semiestructurades**.
- **Híbrid.** Aquests llenguatges contenen marques dels dos tipus anteriors.

Activitat 2

Busca a Internet diferents exemples d'aquests tipus de llenguatges de marques. Indica la extensió de l'arxiu, el tipus i per a què serveix.