

Introducció: Llenguatge de marques

RA1

MP0373: Llenguatge de marques i
Desenvolupament d'aplicacions web

Joel Vizcaíno

Índex

1. Resultats d'aprenentatge i criteris d'avaluació
2. Introducció al llenguatge de marques.
3. Exemple pseudocodi llenguatge de marques.
4. Tipus de llenguatges de marca.
5. Exemples tipus de llenguatges de marca.
6. Àmbits d'aplicació
7. Introducció al XML
 - 7.1 Estructura d'un document XML
 - 7.2 Elements
 - 7.3 Atributs
 - 7.4 Espais en blanc
 - 7.5 Documents XML ben formats
 - 7.6 Conflictes amb els noms
8. Referències

1. Resultats d'aprenentatge i criteris d'avaluació

Resultat d'aprenentatge:

1. Reconeix les característiques de llenguatges de marques analitzant i interpretant fragments de codi.

Criteris d'avaluació

1.1 Identifica les característiques generals dels llenguatges de marques.

1.2 Reconeix els avantatges que proporcionen al tractament de la informació.

1.3 Classifica els llenguatges de marques i identifica els més rellevants.

1.4 Diferencia els àmbits d'aplicació dels llenguatges de marques.

1.5 Reconeix la necessitat i els àmbits específics d'aplicació d'un llenguatge de marques de propòsit general.

1.6 Analitza les característiques pròpies de diferents llenguatges de marques.

1.7 Identifica l'estructura d'un document i les regles sintàctiques.

1.8 Contrasta la necessitat de crear documents ben formats i la influència en el processament.

1.9 Identifica els avantatges que aporten els espais de noms.

2. Introducció al llenguatge de marques

- El llenguatge de marques es una forma de **codificar** un document, mitjançant **text**, **etiquetes**, **marques i/o anotacions**.
- Aquestes **etiquetes**, **marques i/o anotacions** aporten informació **relativa a l'estructura del text** o al **format de presentació**.
- Tot el llenguatge de marques està definit en un document denominat: **DTD** (***Document Type Definition***), en aquests es defineix:
 - Marques: els elements utilitzats en el llenguatge que toqui.
 - Etiquetes i atributs.
 - Sintaxis
 - Normes d'utilització.

3. Exemple codi llenguatge de marques

Informació “normal”	Traducció al llenguatge de marques
17/09/2024 Bon dia, Avui el Llorenç no podrà anar a classe, està malalt. Atentament, la mare del Llorenç.	<carta> <data> 17/09/2024 </data> <salutació> Bon dia, </salutació> <contingut> Avui el Llorenç no podrà anar a classe, està malalt. </contingut> <firma> Atentament, la mare del Llorenç. </firma> </carta>

4. Tipus de llenguatge de marques

Presentació	Híbrid o procedimental	Descriptiu o semàntic
<ul style="list-style-type: none">• És el que indica el format del text. (mida lletra, font, etc)• És important per “maquetar” la presentació d’un document. (llibres, manuals, etc).• Útil per projectes petits. Amb projectes grans és difícil de mantenir.• L’usuari de l’aplicació no veu aquest tipus de llenguatge.• Exemples: LaTeX (deriva de TeX el primer llenguatge de marques, dels anys 70).	<ul style="list-style-type: none">• Enfocat, també, en la presentació del text.• En aquest cas és visible per l’usuari.• El programa interpretarà el codi en ordre, és a dir, si volem posar en negreta una paraula, haurà d’indicar abans d’escriure la paraula, que aquesta estarà en negreta. I una vegada volem deixar la negreta i haurà una etiqueta de tancament.• També pot definir les diferents parts en les que s’estructura el document.• Exemples: HTML (Deriva del SGML), XHTML (deriva del XML i l’HTML), etc.	<ul style="list-style-type: none">• Utilitza etiquetes per descriure els fragments del text.• No s’especifica com ha de ser representat aquest text.• Només indica les parts en les quals s’estructura el document: Per exemple: introducció, cos del text i conclusions. (invent total).• Exemples: XML i derivats d’aquest com: EBML, RDF, XFML, etc.

[Vídeo explicatiu sobre els diferents tipus de llenguatges de marca.](#)

5. Exemples tipus de llenguatge de marca

Codi extret de: [Enllaç](#)

Llenguatge: LaTeX

```
\title{Python 2.*} Títol del document  
\author{Ondiz Zarraga} Autor
```

```
%%%%%%%%%%
```

```
\begin{document} Indica inici del codi  
\maketitle
```

```
\begin{abstract} Indica que s'escriurà un breu resum de la guia  
Este documento es una pequeña guía de Python  
\end{abstract}
```

```
\tableofcontents
```

```
\section{Sobre el lenguaje}
```

```
\begin{itemize} Especifica com será l'índex.  
  \item Interpretado  
  \item Indentación obligatoria  
  \item Distingue mayúsculas - minúsculas  
  \item No hay declaración de variables (\textit{dynamic typing})  
  \item Orientado a objetos  
  \item Garbage colector: quita los objetos a los que no haga referencia nada  
\end{itemize} Indica que acaba l'índex
```

```
\end{document} Indica que acaba el document
```

5. Exemples tipus de llenguatge de marca

```
<html>

<!--Definir que es poden utilitzar accents-->
<meta charset="UTF-8"/>
<!--Titol de la pagina-->
<title>Formulari inscripcio</title>
<article>
  <h1> Formulari d'inscripcio</h1>
  <fieldset style="height:100px; width:300px;" style="color:■rgb(136, 0, 171);"> <!-- recuadre per deixar tancada
la següent informacio:-->
    <p><b> Especifico que el text que hi anirà dintre de l'àrea definida, és un paràgraf i que el text està en negreta
    Aquest formulari es un exemple per posar en practica tota la informacio i els coneixements
    explicats durant els exemples realitzats anteriorment.
    </p></b>
  </fieldset>
</article>
```

Codi propi

Llenguatge: HTML

Es crea un àrea (format) on hi anirà un text

Formulari d'inscripcio

Aquest formulari es un exemple per posar en practica tota la informacio i els coneixements explicats durant els exemples realitzats anteriorment.

5. Exemples tipus de llenguatge de marca

Codi extret pròpi

Llenguatge: XML

```
<?xml version="1.0" encoding="UTF-8"?>
<pelicles>
  <pelicula>
    <titol>Alien: Romulus</titol>
    <director edat='&quot;quaranta-sis&quot;'>Federico Àlvarez</director>
  </pelicula>
  <pelicula>
    <titulo>300</titulo>
    <director edad="&apos;cinquanta-vuit&apos;">Zach Snyder</director>
  </pelicula>
</pelicles>
```

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<pelicles>
  ▼<pelicula>
    <titol>Alien: Romulus</titol>
    <director edat=""quaranta-sis"">Federico Àlvarez</director>
  </pelicula>
  ▼<pelicula>
    <titulo>300</titulo>
    <director edad=""cinquanta-vuit"">Zach Snyder</director>
  </pelicula>
</pelicles>
```

6. Àmbits d'aplicació

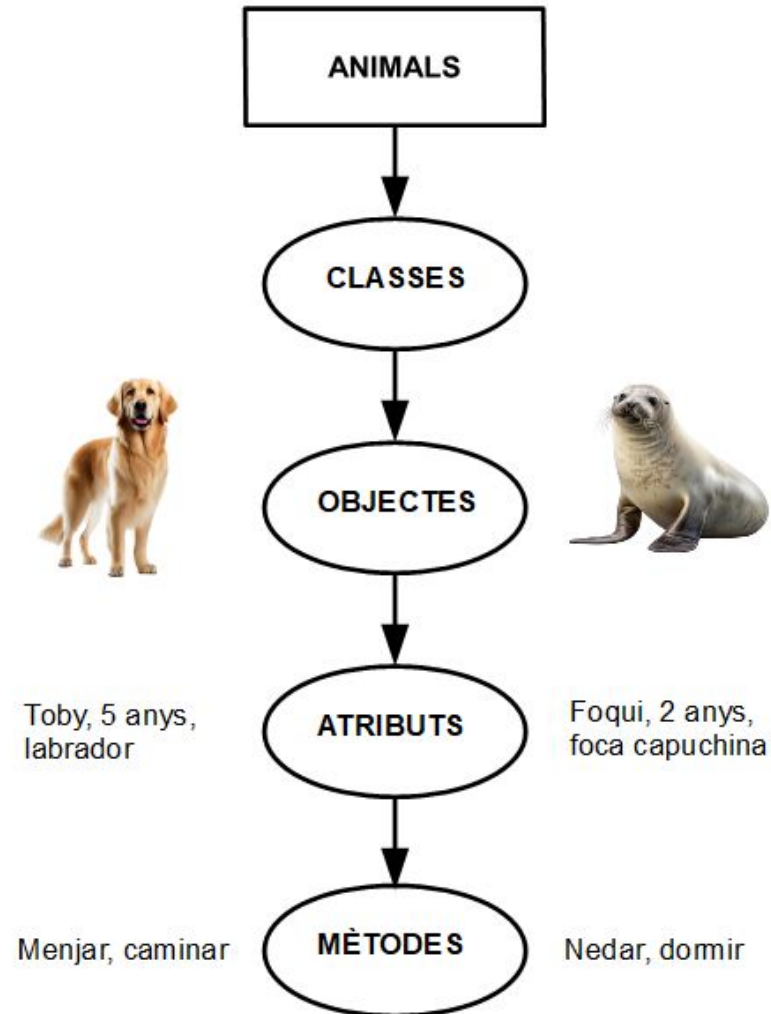
El llenguatge de marques, s'acostuma a utilitzar en els següents àmbits:

Documentació electrònica	Tecnologia Internet	Multimèdia	Altres
<ul style="list-style-type: none">• Markdown: creat a 2004 i utilitzat pel desenvolupament software.• RTF: Permet intercanviar arxius de text en diferents processadors de text.• TeX: Creació d'equacions matemàtiques complexes.• WikiText: permet crear pàgines Wiki (com wikipedia) en servidors específics.	<ul style="list-style-type: none">• HTML: creat per WWW (World Wide Web) per crear pàgines web.• RSS: permet la difusió de contingut web. Utilitzat per exemple per extreure informació de diferents medis de comunicació, blogs, etc.	<ul style="list-style-type: none">• SVG: Permet definir gràfics vectorials en 2D.• MusicXML: permet intercanviar partitures entre els diferents editors/processadors de partitures.	<ul style="list-style-type: none">• MathML: expressa formalismes matemàtics, per a que qualsevol sistema, aplicació o editor el pugui llegir.

7. Introducció al XML

- XML (*eXtensible Markup Language*).
- Conegut per ser un **llenguatge d'etiquetes** o/i de marcat d'estructures.
- Encarregat de guardar la informació de forma ordenada i estructurada.
- Compatible amb qualsevol tipus de sistema operatiu.
- Qualsevol document que verifiqui les regles XML està conforme amb el SGML.
- XML és orientat a objectes, és a dir, la programació està organitzada al voltant de dades i no de funcions i operacions matemàtiques.

7. Introducció al XML



7.1 Estructura d'un document XML

Els documents estan formats per dues parts:

Pròleg	Exemplar
<ul style="list-style-type: none">• És opcional afegir-la.• Si s'afegeix és una part que va sempre abans de l'exemplar.• Aquesta part està formada per dues declaracions:<ul style="list-style-type: none">- Declaració XML: es declara la versió d'XML utilitzada, la codificació de caràcters i la autonomia del document.- Declaració del tipus de document: defineix el tipus de document, per ser processat correctament.	<ul style="list-style-type: none">• És la part del codi, per tant és obligatòria.• Conté les dades reals del document XML.• Està format pels elements anidats.

7.1 Estructura d'un document XML

```
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE libro>
```

PRÒLEG

```
<llibres>  
  <titol>XML pràctic</titol>  
  <autor>Sebastien Lecomte</autor>  
  <autor>Thierry Boulanger</autor>  
  <editorial>Ediciones Eni</editorial>  
  <isbn>978-2-7460-4958-1</isbn>  
  <edicio>1</edicio>  
  <pagines>347</pagines>  
</llibres>
```

EXEMPLAR

7.1 Estructura d'un document XML

Parts del pròleg:

Versió	<p>Fa referència a la versió d'XML que s'està utilitzant o amb la que s'està escrivint el codi. Nosaltres per regla general sempre definim aquesta versió amb el valor 1.0.</p> <pre><?xml version="1.0" ?></pre>
Codificació de caràcters (Altres codificacions)	<p>La part de l'<i>encoding</i> és la codificació dels caràcters, és a dir, normalment fa referència a l'alfabet o la llengua que s'utilitzarà. Nosaltres agafarem l'unicode que és la que permet l'ús d'accents i caràcters "especials" com ara bé la Ñ.</p> <pre><?xml version="1.0" encoding="UTF-8" ?></pre>
Autonomia del document	<p>Si el document XML depèn d'un altre document XML o d'altre tipus, això s'indicarà a la part del <i>standalone</i>, si indica NO, significa que no depèn, si indica YES, depèn d'altres arxius o documents.</p> <pre><?xml version="1.0" encoding="iso-8859-1" standalone="yes" ?></pre>

7.2 Elements

Les parts més importants, dintre de l'apartat d'**exemplars** són els **elements** i **atributs**.

Els elements: són els blocs d'informació que permeten definir l'estructura del document. Aquests elements poden ser “inventats” pel desenvolupador i han de ser autodescriptius, és a dir que reflexin la informació que conté.

Aquests elements estan definits per una obertura i un tancament. Per exemple:
<llibre>...</llibre>

Dintre d'aquest element (entre l'obertura i el tancament) poden haver-hi més elements. A aquest element principal se l'acostuma a anomenar **element arrel** o **root**.
i només pot haver-hi 1 element arrel.

7.2 Elements

Les parts més importants, dintre de l'apartat d'**exemplars** són els **elements** i **atributs**.

Els elements: són els blocs d'informació que permeten definir l'estructura del document. Aquests elements poden ser “inventats” pel desenvolupador i han de ser autodescriptius, és a dir que reflexin la informació que conté.

Aquests elements estan definits per una obertura i un tancament, el nom dels elements han de ser els mateixos a l'obertura i el tancament, respectant minúscules i majúscules. Els noms dels elements poden contenir (.), (-), (_), o caràcters amb accents, però s'aconsella no utilitzar-les. Per exemple: **<element>...</element>**

Dintre d'aquest element (entre l'obertura i el tancament) poden haver-hi més elements. A aquest element principal se l'acostuma a anomenar **element arrel** o **root**. **i només pot haver-hi 1 element arrel.**

7.2 Elements

Les parts més importants, dintre de l'apartat d'**exemplars** són els **elements** i **atributs**.

Els elements:

Sempre que hi hagi un element buit, per exemple, aquest element ha de passar a anomenar-se **<element/>**, per indicar que està buit.

Existeixen els **elements** **niat** és un element que depèn d'un altre l'element:

```
<element1>  
  <element2> coses </element2>  
</element1>
```

7.2 Elements

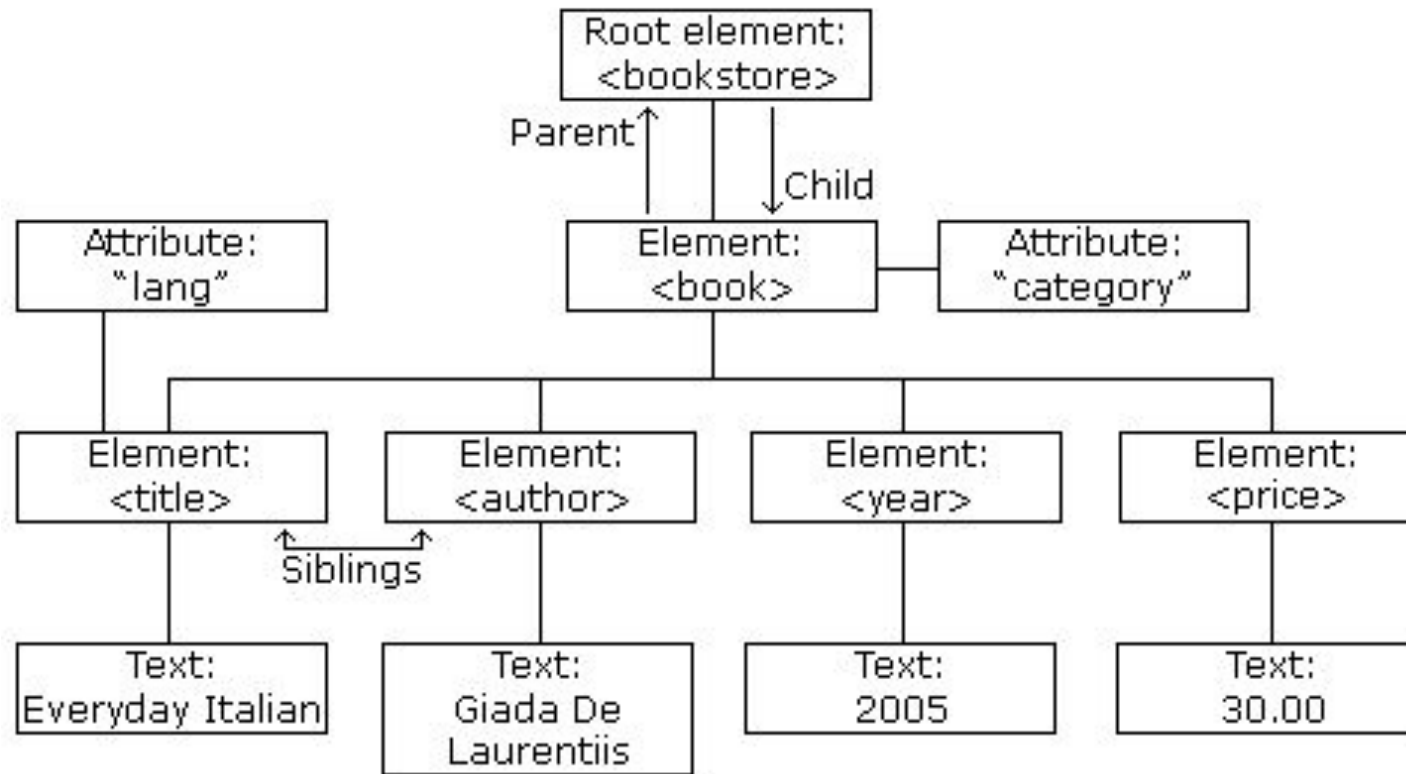
Estructura

d'arbre

dels

elements:

[Enllaç](#)



7.3 Atributs

Les parts més importants, dintre de l'apartat d'**exemplars** són els **elements** i **atributs**.

Els atributs: s'utilitzen per afegir propietats i característiques als elements d'un document.

Els atributs van sempre definits en l'obertura d'un element, i no pot tenir el mateix nom d'un altre element o etiqueta, han de ser **sempre** noms **independents**.

El nom dels atributs han de **complir les mateixen normes** que el nom del elements, és a dir: evitem barrejar majúscules i minúscules, evitem caràcters especials (_), (-), (.)....

<elelemt atribut = "coses"> </element>

7.3 Atributs

```
<?xml version="1.0" encoding="UTF-8"?>
<animals>
  <domestic>
    <tipus>gos</tipus>
    <raça>Bretó</raça>
    <nom>Jaume</nom>
    <data>10/03/2019</data>
    <tipus>gat</tipus>
    <raça>Persa</raça>
    <nom>Bigotitos</nom>
    <data>14/02/2014</data>
  </domestic>
  <salvatges/>
</animals>
```



```
<?xml version="1.0" encoding="UTF-8"?>
<animals>
  <domestic tipus="Gos" raça="Bretó">
    <nom> Jaume </nom>
    <data> 10/06/2019 </nom>
  </domestic>
  <domestic tipus="gat" raça="persa">
    <nom> Bigotitos </nom>
    <data>14/02/2014</data>
  <salvatges/>
</animals>
```

7.4 Espais en blanc

En un document XML, els **espais en blanc**, **tabulacions** i **salts de línia** poden ser tractats d'un mode especial.

Dintre d'un **atribut** o **element** diversos espais en blanc, s'interpreten com un de sol, és a dir, per exemple:

```
<?xml version="1.0" encoding="UTF-8"?>
<animals>
  <domestic tipus="Gos" raça="Bretó">
    <nom> Jaume </nom>
    <data> 10/06/2019

  </nom>
</domestic>
<domestic tipus      ="gat" raça="persa">
  <nom> Bigotitos </nom>
  <data>14/02/2014</data>
<salvatges/>
</animals>
```

=

```
<?xml version="1.0" encoding="UTF-8"?>
<animals>
  <domestic tipus="Gos" raça="Bretó">
    <nom> Jaume </nom>
    <data> 10/06/2019 </nom>
  </domestic>
  <domestic tipus="gat" raça="persa">
    <nom> Bigotitos </nom>
    <data>14/02/2014</data>
  <salvatges/>
</animals>
```

7.4 Espais en blanc

Què passa si els espais els necessito per representar les dades?

Per exemple: vull representar una llista d'estudiants on hi han diverses columnes i files:

```
<?xml version="1.0" encoding="UTF-8"?>

<alumnat>

1.      Garcia, Josep      10
2.      Salguero, Marcos   5
3.      Gomez, Pablo      7
4.      Almansa, Ivan     0

</alumnat>
```

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<alumnat> 1. Garcia, Josep 10 2. Salguero, Marcos 5 3. Gomez, Pablo 7 4. Almansa, Ivan 0 </alumnat>
```

7.4 Espais en blanc

Què passa si els espais els necessito per representar les dades?

En aquest cas, per **conservar** tots els espais, s'ha d'afegir, com atribut: **xml:space = "preserve"**, si volem mantenir l'atribut **xml:space**, però que no mantingui els canvis, haurem d'escriure **xml:space = "default"**.

```
<?xml version="1.0" encoding="UTF-8"?>

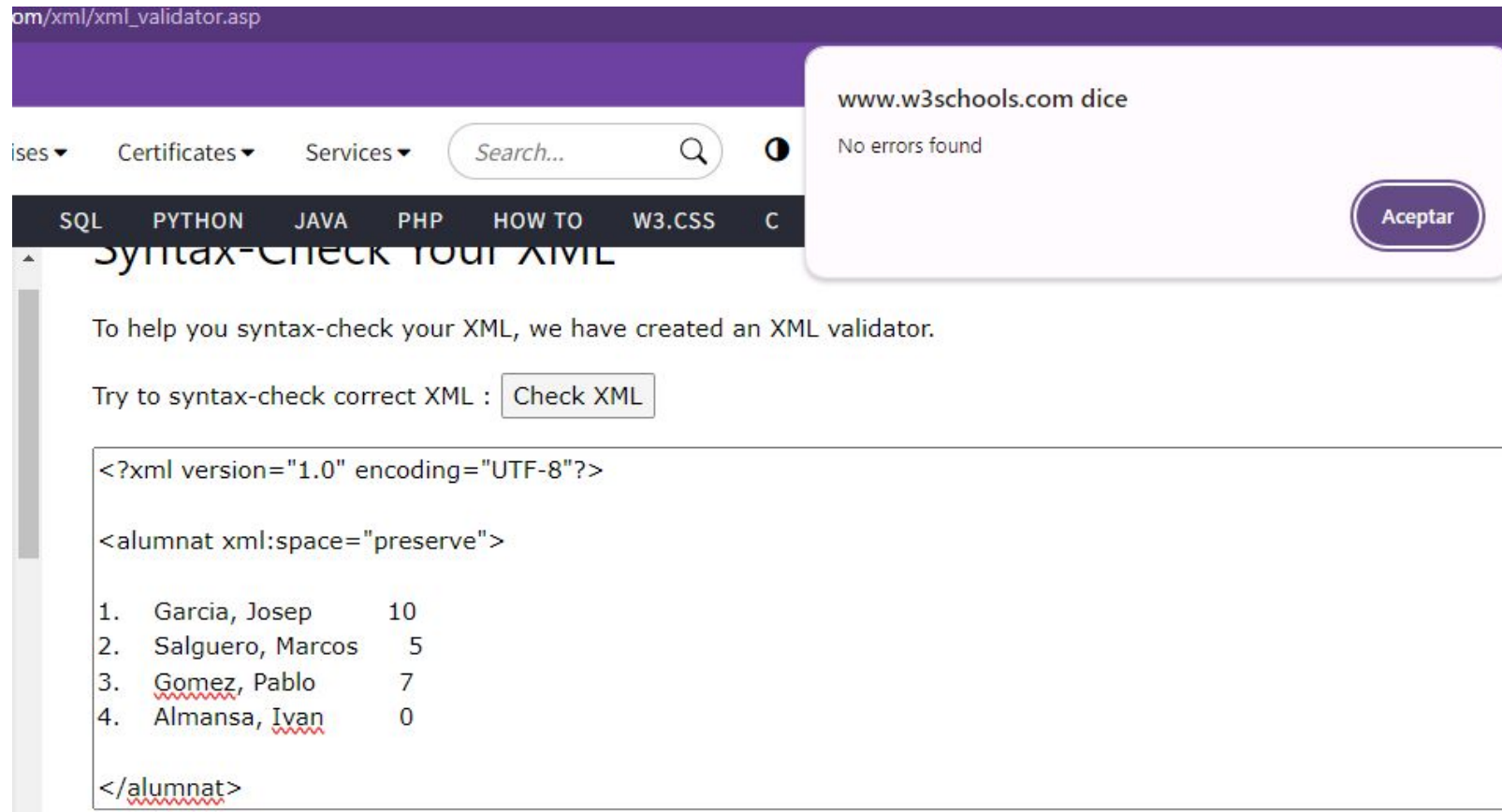
<alumnat xml:space="preserve">

1.      Garcia, Josep          10
2.      Salguero, Marcos       5
3.      Gomez, Pablo           7
4.      Almansa, Ivan          0

</alumnat>
```


7.5 Documents XML ben formats

Com podria saber si el meu document XML està bé? Necessitem un [validador](#).



7.6 Conflictes amb els noms

Què passa si s'ajunten (en un futur ho veurem) dos o més documents XML i algun d'ells **comparteix noms d'elements** amb un altre.

En aquest cas, s'ha d'utilitzar el que s'anomena **ESPAI**. Per exemple: Ajuntem dos documents que tenen l'element: **<marca>...</marca>**.

En aquest cas s'han de definir dos espais, un per cada element repetit, per tant, a partir d'ara, quan adjuntem els dos fitxer: **<e1:marca>...</e1:marca>** i **<e2:marca>...</e2:marca>**

Nota: els elements fills o nets que depenen de l'element arrel ***marca***, també han de tenir aquesta referència.

7.6 Conflictes amb els noms

<pre><marca> <prenda> Sabates </prenda> <talla> 43 </talla> </marca></pre>	<pre><marca> <preu> 34 </preu> <estat> Disponible</estat> </marca></pre>
<pre><e1:marca> <e1:prenda> Sabates </e1:prenda> <e1:talla> 43 </e1:talla> </e1:marca></pre>	<pre><e2:marca> <e2:preu> 34 </e2:preu> <e2:estat> Disponible</e2:estat> </e2:marca></pre>

8. Referències

1. W3Schools.com. (s. f.). https://www.w3schools.com/xml/xml_tree.asp
2. W3Schools.com. (s. f.-b). https://www.w3schools.com/xml/xml_validator.asp
3. X.Sala i C.Martí, *Llenguatge de marques i sistemes de gestió d'informació*, 10/09/2024, IOC
(Institut Obert de Catalunya).

