



Característiques de l'XML

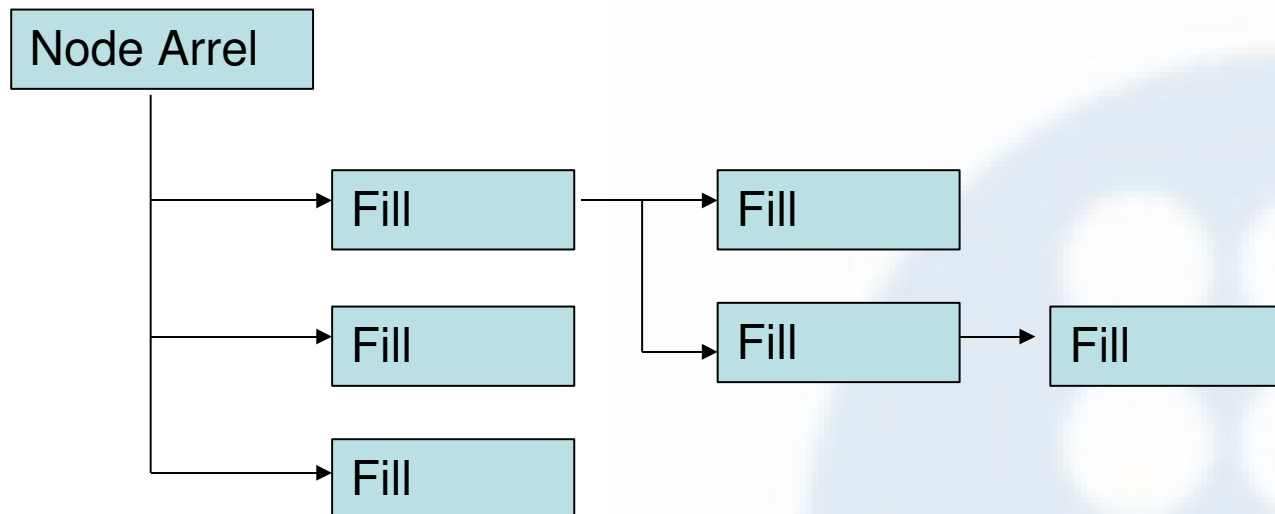
Estructura i característiques de XML.





Estructura d'un XML

- La informació d'un XML s'estructura en forma d'arbre. És a dir:



- On cada node és d'un tipus concret (com ara atribut o element) i pot tenir un conjunt de propietats.



Estructura d'un XML

- Per tant, un document XML es podria veure com:

```
<?xml version="1.0" encoding="UTF-16"?>
```

```
<a>  
  <b x="a">  
    <!-- hola -->  
    hola  
  </b>  
</a>
```



Node tipus element



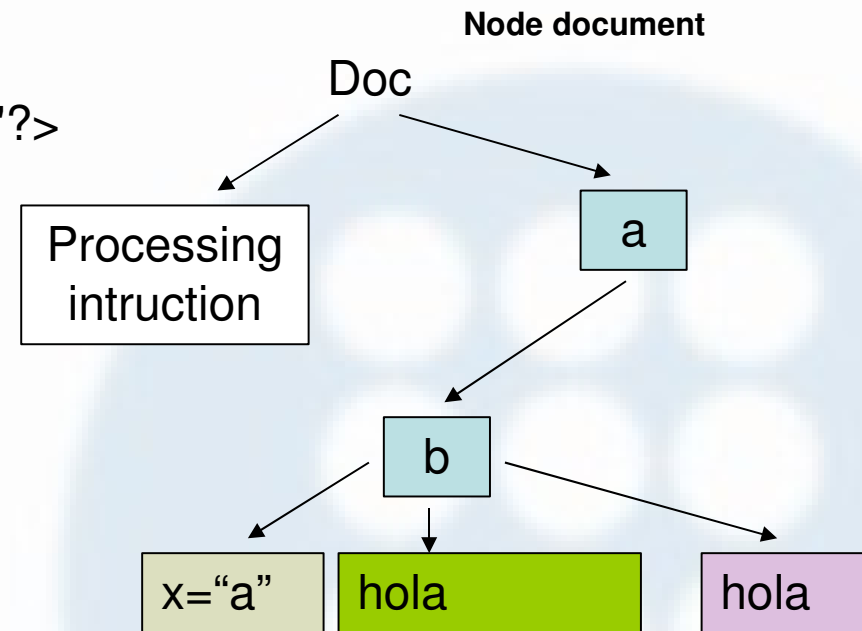
Node tipus text



Node tipus comentari



Node tipus atribut





Estructura d'un XML

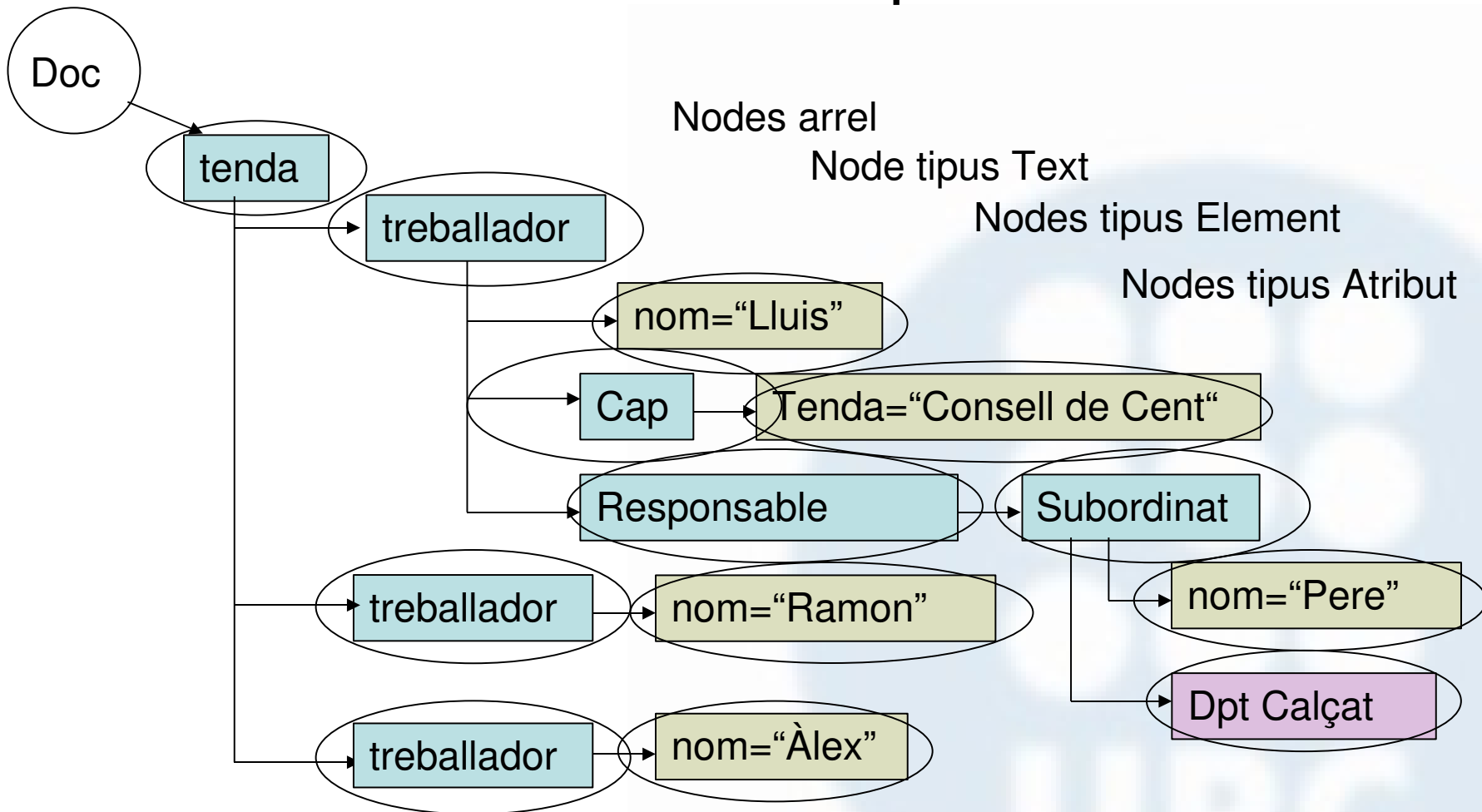
- Un exemple més complet, suposem que tenim el següent XML:

```
<?xml version="1.0" encoding="utf-8"?>
<tenda>
  <treballador nom="Lluís">
    <cap tenda="Consell De Cent"/>
    <responsable>
      <subordinat nom="Pere">
        Dpt Calçat
      </subordinat>
    </responsable>
  </treballador>
  <treballador nom="Ramon"/>
  <treballador nom="Àlex"/>
</tenda>
```



Estructura d'un XML

- L'arbre associat de l'exemple tenda és:





Característiques de sintaxi

- Es obligatori incloure la declaració XML a la primera línia.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

- Nom dels elements escollits pel dissenyador.
 - Els caràcters que es poden usar en els noms dels elements són:
<http://www.w3.org/TR/2004/REC-xml11-20040204/#charsets>
- Els elements s'han de tancar sempre:

```
<llibre>
```

```
  <personatge> Frodo </personatge>
```

```
</llibre>
```

És invalid :

```
<personatge> Frodo
```

- Els elements poden no tenir contingut

```
<llibre></llibre>
```



Declaració de document XML

- Permet saber a l'aplicació encarregada de processar el document que és tracta d'un document XML.
- La sintaxi de la declaració és:

```
<?xml Calificadors ?>
```

- Calificadors:
 - “*version*”: Versió de document. *version*="1.0"
 - “*encoding*”: Tipus de codificació amb la qual s'ha fet el document. *encoding*="utf-8" (o bé ISO-8859-1)
 - *standalone*: indica si s'ha de processar o no la DTD associada al document. *standalone*="yes/no"



Característiques de sintaxi

- Els elements poden tenir elements aniuats:

```
<llibre autor="J.R.R. Tolkien" titol="El Hobbit">  
  <capitol numero="1">Visita inesperada</capitol>  
</llibre>
```

- Els elements poden tenir contingut text

```
<llibre autor="J.R.R. Tolkien" titol="El senyor  
  dels anells">  
  És un gran llibre !  
</llibre>
```




Característiques de sintaxi

- Els elements poden tenir contingut mixt:

```
<llibre autor="J.R.R. Tolkien" titol="El Hobbit">  
  <capitol numero="1">Visita inesperada</capitol>  
  Aquest llibre és increïble !!!  
</llibre>
```

- Els elements poden tenir atributs:

```
<llibre autor="J.R.R. Tolkien" titol="El senyor dels  
  anells"></llibre>
```

- Els elements sense dades es poden tancar al final:

```
<llibre autor="J.R.R. Tolkien" titol="El senyor dels  
  anells"/>
```

En comptes de l'anterior `<llibre></llibre>`



Característiques de sintaxi

- Les marques distingeixen majúscula / minúscula, és a dir:
 - El següent tag és incorrecte:

```
<llibre></Llibre>
```

- Cal expressar-lo:

```
<llibre></llibre> o bé <Llibre></Llibre>
```





Característiques de sintaxi

- Els espais en blanc es tenen en compte en el contingut dels elements, però són normalitzats en els atributs. És a dir:

```
<llibre autor="J.R.R. Tolkien">
El      Hobbit
</llibre>
```

Equival a :

```
<llibre autor="J.R.R. Tolkien">
El      Hobbit
</llibre>
```



Característiques de sintaxi

- Quan declarem atributs, hem de tenir present:
 - Hi ha distinció entre majúscules i minúscules.

- És a dir, no és el mateix:

```
<autor nom="Tolkien"> que <autor nom="tolkien">  
ni <autor nom="Tolkien"> que <autor Nom="Tolkien">
```

- Un element pot tenir un nombre il·limitat d'atributs.
- Un atribut no pot repetir-se dins un mateix element.
 - Creem un XML on un dels elements tingui un atribut repetit.
 - Què diu el processador de l'XML?
- L'ordre dels atributs no influeix, és a dir:

```
<autor nom="Tolkien" ab="J.R.R.T"> equival a  
<autor ab="J.R.R.T" nom="Tolkien">
```



Restriccions d'un document

- Els documents XML estan lligats a certes restriccions:
 - Té un format textual, per tant les dades binàries s'han d'enviar en Base 64 o enviar a part del document XML (via enllaç, com fa l'HTML amb les imatges).
 - Els documents han d'estar ben formats. És a dir, han de seguir les restriccions sintàctiques enumerades en l'especificació. (Comentades anteriorment)
 - Els documents poden ser vàlids. És a dir es pot comprovar si segueixen certes estructures.

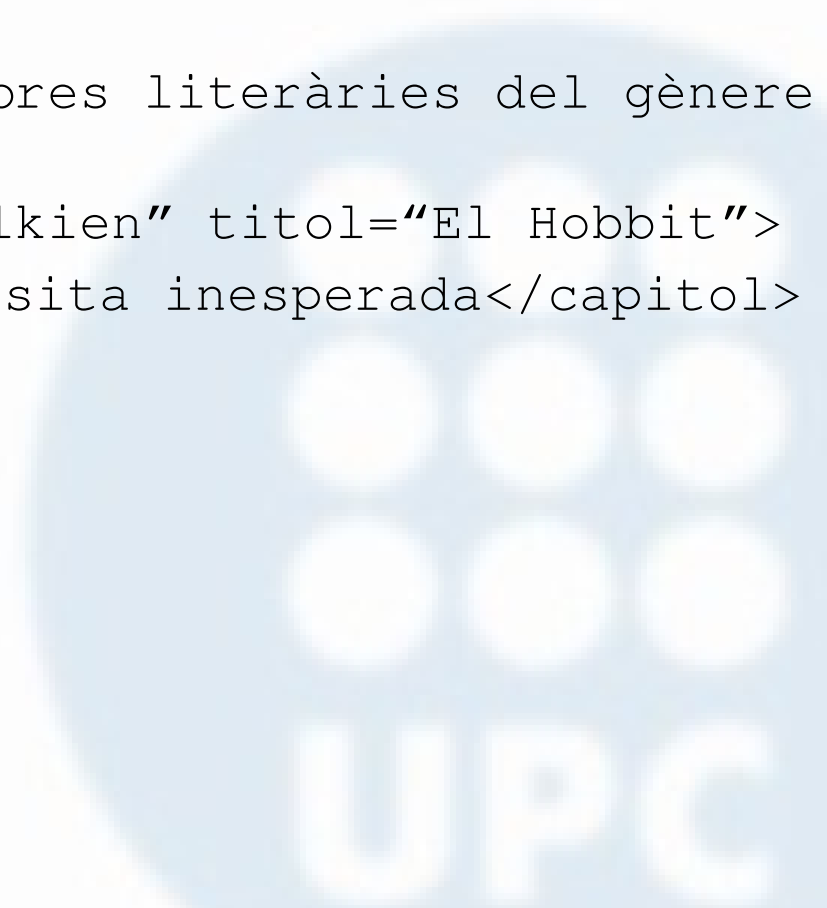


Comentaris en XML

- Podem afegir comentaris als nostres XMLs igual que ho fem amb HTML:

```
<!-- Una de les millors obres literàries del gènere de ciència ficció-->
```

```
<llibre autor="J.R.R. Tolkien" titol="El Hobbit">  
  <capitol numero="1">Visita inesperada</capitol>  
</llibre>
```





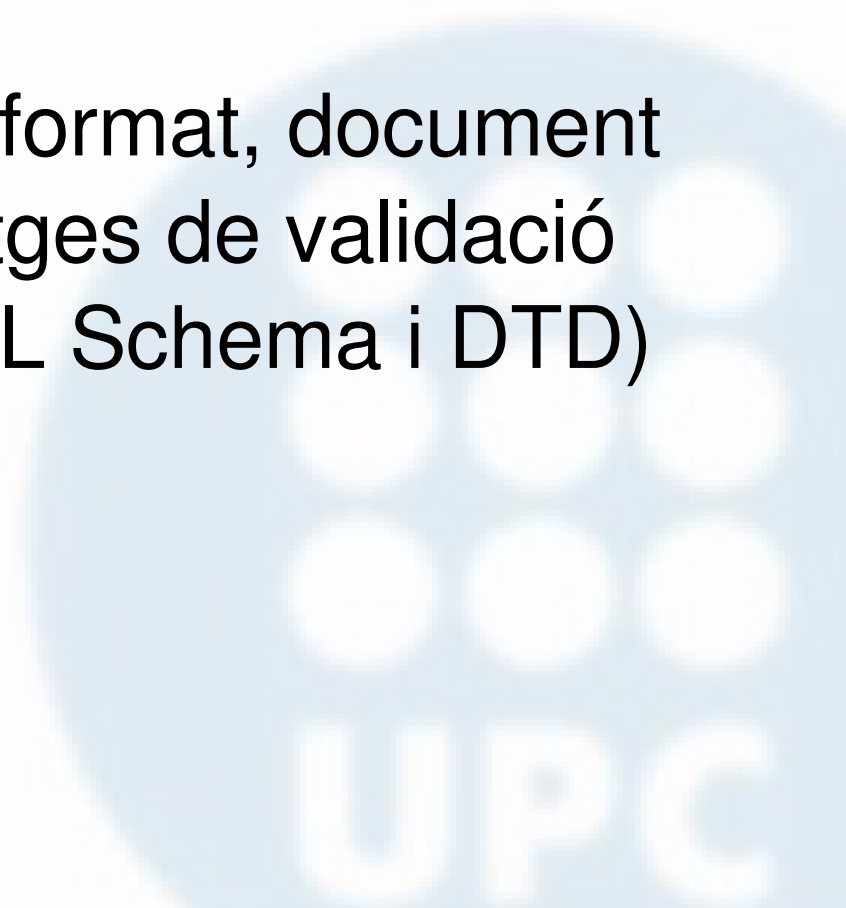
Tipus de dades dels atributs

- En els fitxers XML no s'especifica de quin tipus (enters, caràcters,...) són els valors del atributs, però sí en la seva DTD o XSD corresponent.
- Per tant, a l'hora de crear un XML cal tenir present els tipus definits en la seva DTD o XSD.
- Ara parlarem dels DTD i dels XML Schemas com a eines de validació semàntica de documents XML.



Validació de documents XML

Document ben format, document
vàlid i llenguatges de validació
semàntica (XML Schema i DTD)





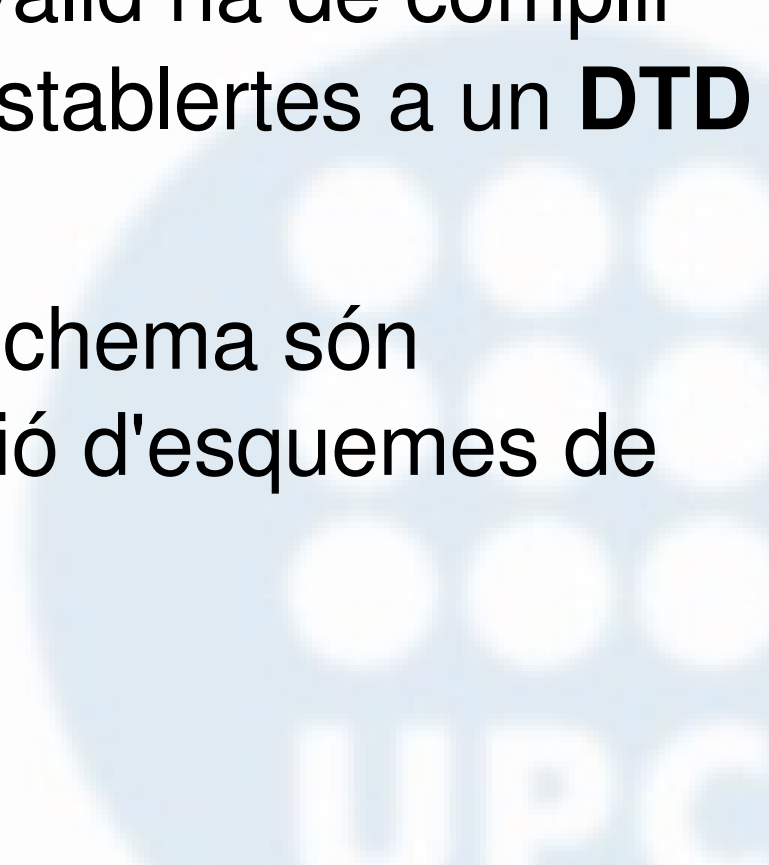
Documents ben formats

- Perquè un document sigui un document XML sintàcticament correcte ha de ser un **document ben format**. Per a que sigui un document ben format ha de tenir estructura d'arbre i ha de complir totes les regles sintàctiques anteriors.



Documents vàlids

- Perquè un document sigui un document XML semànticament vàlid ha de complir amb les restriccions establertes a un **DTD** o un **XML Schema**.
- Tant DTD com XML Schema són llenguatges de definició d'esquemes de documents.





Sintaxi vs. Semàntica

- SINTAXI
 - Correcte sintàcticament.
 - Document XML ben format o Well-formed XML Document: que compleix amb les regles sintàctiques de l'especificació.
- SEMÀNTICA
 - Vàlid semànticament.
 - Els documents XML es validen mitjançant llenguatges de definició d'esquemes de documents.



Llenguatges de definició d'esquemes

- Tot document XML hauria de seguir una estructura concreta. Aquesta estructura es pot representar a través d'una DTD o d'una XML Schema.
- Permeten validar si un document XML segueix una estructura determinada.
- No és necessari que un document XML contingui la declaració de tipus de document.
- Existeixen altres llenguatges: RELAX NG, Schematron...



Document Type Definition (DTD)

- La declaració de tipus de document es fa de la següent manera:

```
<!DOCTYPE nom ... >
```
- El **node arrel** al qual volem aplicar la DTD **haurà de tenir el mateix nom** que hem indicat.
- A continuació del nom del document hi podem tenir:
 - La definició de l'estructura del document (**DTD interna**).
 - Una referència a un recurs extern que conté la DTD d'aquest document (**DTD privada**).
 - Una referència a una definició pública (**DTD Pública**).



Declaració de DTD interna

- Declarem la DTD dins el propi document XML.

```
<!DOCTYPE nom [  
    declaració1,  
    declaració2,  
    .. ]>
```

Podeu trobar l'exemple a la carpeta **exemple 4** de la sessió 1.





Declaració de DTD privada

- El terme de declaració privada es fa servir quan la DTD es troba dins el propi directori físic o accessible a través d'HTTP.

<!DOCTYPE nom SYSTEM "camí">

On el camí pot ser:

- Una URL.
- Un camí local a la màquina.

Podeu trobar un exemple amb la seva corresponent DTD a l'**exemple 5** de la sessió 1.



Declaració de DTD pública

- Fem referència a DTDs que són de gran difusió, i poden ser considerades públiques.
- La seva declaració és del tipus:

`<!DOCTYPE nom PUBLIC id url>`

On l'identificador és únic de la forma:

Tipus identificador//organització//DTD descripció//Idioma

- El tipus d'identificador indica si l'identificador està registrat o no:
 - “-” Si no està registrat.
 - “+” Si està registrat. En cas de ser registrat per la ISO com a norma general s'escriurà la cadena “ISO 8879:1986”



Declaració del tipus de document

- **Organització:** El nom de l'organització o persona propietària de la DTD, en el cas que la propietària sigui la pròpia ISO, aquest cap restarà buit.
- **DTD:** paraula reservada per indicar que el recurs públic és una DTD.
- **Descripció:** text descriptiu de la DTD.
- **Llenguatge:** idioma en que està escrit la DTD. L'idioma s'indicarà utilitzant les convencions indicades en la norma ISO 639.

Podeu trobar un exemple amb la seva corresponent DTD a l'**exemple 6** de la sessió 1.



Declaració del tipus de document

- Un exemple més *real* es pot trobar en el document HTML de la següent URL (que conté una definició de document que fa referència a la DTD la qual defineix l'estructura dels documents HTML)
- Podeu trobar un exemple amb la seva corresponent DTD a l'**exemple 7** de la sessió 1.



Exercici

- Penseu un XML que contingui un llistat d'alumnes d'un curs i les seves dades personals.
- Implementeu una XML amb l'estructura que heu pensat.
- Una possible solució la trobeu a la carpeta **exercici 1** de la sessió 1.



XML Schema

- Els XML Schemas ens permetran fer la definició de tipus d'un document XML.
- Avantatges respecte DTD:
 - Són documents XML, per tant usen espais de noms, atributs, elements...
 - Ens permeten modelar de forma molt més acurada.
- Declaració del XML Schema:

```
<node_arrel xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="Path_xml_schema.xsd">
```