# Dokumentu motaren definizioa (Document Type Declaration)

### Kasu praktikoa:

Mariak bazkide den enpresaren informazioa gainerako lankideekin partekatzeko beharra ikusten du, enpresak dituen dendek informazioa eguneratuta mantentzeko, eta informazio hori dokumentu mota batean modu ordenatuan gordetzeko modurik ote dagoen pentsatzen du.

Dokumentu horiek nolako egitura bete behar duten aztertzen du, eta ziurtatu beharra dagoela mota bereko dokumentu guztiek egitura bera dutela, bazkide guztiek informazioa era berean interpretatu ditzaten.

Jonek, enpresako informatika arazoak konpontzeaz arduratzen zen teknikariak, Mariari esan dio XML irtenbide ona izan daitekeela bere beharretarako.

Jonek azaldu dio XML dokumentuetan datuak modu errazean idazten direla eta egitura komun bat mantenduz gordetzeko aukera ematen duela, eta era honetan, enpresako lankideen artean datuak partekatzeko aukera izango dutela.

Jonen arabera, metodo errazena dokumentuak normalizatzen saiatzeko, dokumentuek bete behar dituzten hiztegiak definitzean datza.

Hauei **D**okumentu **M**otaren **D**efinizioa esaten zaie (DTD). Gainera, XML lengoaia ez den arren, sintaxi erraza du Mariak eta gainontzeko lankideek ulertzeko.

Dokumentu motaren definizioek (DTD) XML dokumentu baten egitura zehazten dute. DTD bat XML dokumentu baten egitura definitzen duen dokumentu bat da, hau da, ager daitezkeen elementuak, ager daitezkeen aldi kopurua, zein izan daitezkeen elkarren seme-alabak, edukiak, atribututak eta hauen ezaugarriak etab.

XML prozesadoreak DTDan ezarritako murrizketak erabiltzen ditu XML dokumentu bat baliozkoa den egiaztatzeko, hau da, dokumentuak DTDren arauak betetzen dituen egiaztatzeko.

Balidazioa honako hauek egiaztatzeaz arduratzen da:

- Datuen zuzenketa: eskema baten aurka baliozkotzeak % 100ean datuak zuzenak direla bermatzen ez badu ere, formatu baliogabeak edo mailaz kanpoko balioak detektatzeko aukera ematen ditu.
- Datuen osotasuna: baliozkotzean, derrigorrezko informazio guztia dokumentuan dagoela egiaztatzen da.
- Datuen ulermen partekatua: baliozkotzearen bidez egiaztatzen da igorleak eta hartzaileak dokumentua modu berean jasotzen dutela, eta horren ondorioz berdin interpretatzen dutela ere.

## DTD baten erreferentzia XML dokumentu batean

XML dokumentua baliozkotzeko XML prozesadoreak erabili behar duen DTDa DOCTYPE etiketaren bidez adierazten da.

DTD lengoaia ez da XMLa, SGML sintaxitik eratorria da.

DTDa dokumentuan bertan sartuta egon daiteke, kanpoko dokumentua izan daiteke edo biak konbinatu daitezke.

• DTDa dokumentuan bertan sar daiteke, sintaxi honekin:

```
<!DOCTYPE izena [
    lehen adierazpena
    lehen adierazpena
    .
    .
    azken adierazpena
]>
```

• DTDa kanpoko dokumentu batean egon daiteke, eta <u>aplikazio bakar batek bakarrik erabiliko</u> <u>badu</u>, sintaxia honako hau da:

```
<!DOCTYPE izena SYSTEM "uri">
```

• Kanpoko DTD bat barneko DTD batekin konbina daiteke, sintaxi honekin:

• DTDa kanpoko dokumentu batean egon daiteke, eta <u>hainbat aplikaziok erabiliko badute</u>, sintaxia honako hau da:

```
<!DOCTYPE izena PUBLIC "fpi" "uri">
```

• Kanpoko DTD bat barneko DTD batekin konbina daiteke, sintaxi honekin:

### Kasu guztietan:

- **izena**: XML dokumentu motaren izena da, eta bat etorri behar du XML dokumentuaren erroko elementuaren izenarekin.
- uri: lokalean kokatutako DTD dokumenturainoko bidea (absolutua edo erlatiboa).
- **fpi**: publikoki eskuragarri dagoen estadar bat adierazten du (**F**ormal **P**ublic **I**dentifier). Adibidez: <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">

# Adierazpenak

DTDek adierazpenen bidez deskribatzen dute XML dokumentuen egitura.

# Elementu terminalen moten adierazpenak.

Terminal motak zuhaitz egituraren orriekin bat datozen elementuak dira.

Elementuen adierazpenek honako sintaxi hau dute:

```
<!ELEMENT izena (edukia)>
```

- izena: elementuaren izena da.
- **edukia**: elementuaren edukia deskribatzen duen adierazpena.

Elementuaren edukia definitzeko hurrengo terminoak erabil daitezke, edo esamolde konplexuagoak idatzi:

#### EMPTY

Elementua hutsa dela esan nahi du, hau da, ezin duela edukirik izan.

Elementu hutsak irekiera eta itxiera etiketekin idatz daitezke, haien artean ezer egon gabe, ezta espaziorik ere, edo etiketa huts batekin.

EMPTY parentesirik gabe idatzi behar da.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua EMPTY>
]>

<eredua></eredua>

<eredua/>
<eredua>Hau eredu bat da</eredua>
<!-- ERROREA: testu bat dauka -->

<eredua><a></e></eredua>
<!-- ERROREA: <a></eredua>
<!-- ERROREA: <a></e> elementu bat dauka -->
```

#### • (#PCDATA)

Elementuak testua izan dezakeela esan nahi du.

#PCDATA parentesi artean idatzi behar da.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua (#PCDATA)>
]>
<eredua/>
<eredua>Hau eredu bat da</eredua>
<eredua><a></a></eredua>
<!-- ERROREA: <a> elementu bat dauka -->
```

#### ANY

Elementuak edozer eduki dezakeela (testua eta beste elementu batzuk) esan nahi du.

ANY parentesirik gabe idatzi behar da.

Elementu batek beste elementu batzuk izan ditzakeela edo eduki behar dituela adierazteko, hurrengo konektora eta aldagailak erabiltzen dira:

• , (koma): elementuak adierazitako ordenan elementuak dituela esan nahi du.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua (a, b)>
    <!ELEMENT a EMPTY>
    <!ELEMENT b EMPTY>
]>

<eredua><a/>>b/></eredua>
<eredua><a/><b/><!-- ERROREA: <c> elementu bat dauka -->
<eredua><a/><!-- ERROREA: <b/><!-- ERROREA: elementuen ordena ez da zuzena -->
```

(logikoa): elementuak definitutako elementuetako bat duela esan nahi du.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua (a | b)>
    <!ELEMENT a EMPTY>
    <!ELEMENT b EMPTY>
]>

<eredua><a/></eredua>
<eredua><b/></eredua>
<eredua><a/></eredua>
<eredua><a/></eredua>
<!-- ERROREA: bi elementuak ditu -->
<eredua></eredua>
<!-- ERROREA: ez du elementurik -->
```

(galdera ikurra): esan nahi du elementua ager daitekeela edo ez, baina behin bakarrik.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua (a, b?)>
    <!ELEMENT a EMPTY>
    <!ELEMENT b EMPTY>
]>

<eredua><a/></eredua>
<eredua><a/></eredua>
<eredua><b/></eredua>
<!-- ERROREA: <a> elementua falta da -->
<eredua><b/><!-- ERROREA: <b> elementua birritan agertzen da -->
```

\* (asteriskoa): elementua ez agertzea, behin edo gehiagotan agertu daitekela esan nahi du.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua (a*, b)>
    <!ELEMENT a EMPTY>
    <!ELEMENT b EMPTY>
]>

<eredua><b/>
<eredua><b/>
</eredua>
```

```
<eredua><a/></a><b/></eredua>
<eredua><b/><a/></eredua>
<!-- ERROREA: <b> elementua <a> elementua baino lehen agertzen da -->
```

• + (batura ikurra): elementuak behin edo gehiagotan agertu behar duela esan nahi du, ezin da ez agertu.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua (a+, b)>
    <!ELEMENT a EMPTY>
    <!ELEMENT b EMPTY>
]>

<eredua><a/><b/></eredua>
<eredua><a/></eredua>
<eredua><b/></eredua>
<!-- ERROREA: <a> elementua falta da -->
```

• () parentesiak: adierazpenak taldekatzeko aukera ematen du.

```
<!DOCTYPE eredua [
 <!ELEMENT eredua (a, (a|b))>
 <!ELEMENT a EMPTY>
 <!ELEMENT b EMPTY>
]>
<eredua><a/></eredua>
<eredua><a/><b/></eredua>
<eredua><a/></eredua>
<!-- ERROREA: <a> edo <b> elementua falta da \rightarrow
<!DOCTYPE eredua [
 <!ELEMENT eredua ((a,b)|(b,a))>
 <!ELEMENT a EMPTY>
  <!ELEMENT b EMPTY>
]>
<eredua><a/></eredua>
<eredua><b/><a/></eredua>
<eredua><a/></a></eredua>
<!-- ERROREA: <a><b> edo <b><a> soilik izan onartzen dira -->
```

### Atributuen adierazpena

Atributuen deklarazioak honako sintaxi hau du:

<!ATTLIST elementuIzena atributuIzena atributuMota atributuBalioLehentsia>

#### non:

- **elementuIzena**: atributua dagokion elementuaren izena da.
- atributuIzena: atributuaren izena da.
- atributuMota: datu mota bat da.
- **atributuBalioLehentsia**: atributuaren balio lehenetsia da (nahiz eta beste gauza batzuk ere adieraz ditzakeen).

Elementu berean hainbat atributu definitzeko, atributuen adierazpen bat edo gehiago erabil daitezke. Honako adibide hauek baliokideak dira:

```
<!ATTLIST elementuIzena atributuIzena1 atributuMota1 atributuBalioLehentsia1>
<!ATTLIST elementuIzena atributuIzena2 atributuMota2 atributuBalioLehentsia2>
<!ATTLIST elementuIzena
  atributuIzena1 atributuMota1 atributuBalioLehentsia1
  atributuIzena2 atributuMota2 atributuBalioLehentsia2
```

Atributu motak honako hauek dira:

• **CDATA**: atributua karaktereak ditu (mugarik gabe).

• **NMTOKEN**: atributuak letrak, digituak, eta hurrengo karaktereak baino ez ditu: puntua (.), gidoia (-), azpimarra (\_) eta bi puntu (:).

```
<!DOCTYPE eredua [
    <!ELEMENT eredua EMPTY>
    <!ATTLIST eredua kolorea NMTOKEN #REQUIRED>
]>

<eredua kolorea=""></eredua>
<eredua kolorea="1"></eredua>
<eredua kolorea="urdin-iluna"></eredua>
<eredua kolorea="urdin iluna"></eredua>
<!-- ERROREA: espazio bat agertzen da -->
<eredua kolorea="#000033"></eredua>
<!-- ERROREA: # karakterea agertzen da -->
```

• **NMTOKENS**: atributuak NMTOKEN motak dituen balioak izan ahal ditu eta baita espazio zuriak ere.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua EMPTY>
    <!ATTLIST eredua kolorea NMTOKENS #REQUIRED>
]>
<eredua kolorea="urdin-iluna"></eredua>
```

```
<eredua kolorea="urdin iluna"></eredua>
<eredua kolorea="#000033"></eredua>
<!-- ERROREA: # karakterea agertzen da -->
```

• **balioak**: atributuak zerrenda bateko terminoetako bat bakarrik eduki dezake. Zerrenda parentesi artean idazten da, terminoak barra bertikal baten bidez bereizita "|".

```
!DOCTYPE eredua [
    <!ELEMENT eredua EMPTY>
        <!ATTLIST eredua kolorea (urdina|zuria|gorria) #REQUIRED>
]>
eredua kolorea=""></eredua>
<eredua kolorea="urdina"></eredua>
<eredua kolorea="berdea"></eredua>
<!-- ERROREA: berdea ez da agetzen balioen zerrendan -->
```

• **ID**: atributuaren balioa (ez izena) bakarra izan behar da, eta ezin da errepikatu beste elementu edo atributu batzuetan. Balioak ezin du zenbaki batekin hasi.

```
<!DOCTYPE eredua [
 <!ELEMENT eredua (liburua*)>
 <!ELEMENT liburua (PCDATA)>
  <!ATTLIST liburua kodea ID #REQUIRED>
]>
<eredua>
 liburua kodea="L1">Poema de Gilgamesh</liburua>
  liburua kodea="L2">Los preceptos de Ptah-Hotep</liburua>
</eredua>
<eredua>
 liburua kodea="1">Poema de Gilgamesh</liburua>
  <!-- ERROREA: atributuaren balioak ezin du zenbaki batekin hasi -->
</eredua>
<eredua>
  liburua kodea="L1">Poema de Gilgamesh</liburua>
  liburua kodea="L1">Los preceptos de Ptah-Hotep</liburua>
  <!-- ERROREA: ezin da ID motako atributu baliorik errepikatu -->
</eredua>
```

• **IDREF**: atributuaren balioak bat etorri behar du beste elementu baten ID atributuaren balioarekin.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua (liburua|autorea)>
    <!ELEMENT liburua (PCDATA)>
    <!ATTLIST liburua kodea ID #REQUIRED>
    <!ELEMENT autorea (PCDATA)>
    <!ATTLIST autorea liburua IDREF #REQUIRED>
]>

<eredua>
    liburua kodea="L1">Python para todos</liburua>
    <autorea liburua="L1">Charles Russell</autorea>
</eredua>
```

```
<eredua>
  liburua kodea="L1">Python para todos</liburua>
  <autorea liburua="L2">Charles Russell</autorea>
  <!-- ERROREA: balioa ez da elementu baten IDa -->
</eredua>
```

• **IDREFS**: atributuaren balioa espazioka bereizitako balio batzuk dira, beste elementu batzuen ID atributuaren balioarekin bat datozenak.

```
<!DOCTYPE eredua [
 <!ELEMENT eredua (liburua|autorea)>
 <!ELEMENT liburua (PCDATA)>
 <!ATTLIST liburua kodea ID #REQUIRED>
 <!ELEMENT autorea (PCDATA)>
  <!ATTLIST autorea liburua IDREFS #REQUIRED>
]>
<eredua>
  liburua kodea="L1">Python para todos I</liburua>
  liburua kodea="L2">Python para todos II</liburua>
  <autorea liburua="L1 L2">Charles Russell</autorea>
</eredua>
<eredua>
 liburua kodea="L1">Python para todos I</liburua>
 liburua kodea="L2">Python para todos II</liburua>
  <autorea liburua="L1">Charles Russell</autorea>
</eredua>
<eredua>
  liburua kodea="L1">Python para todos I</liburua>
 liburua kodea="L2">Python para todos II</liburua>
 <autorea liburua="L1 L2 L3">Charles Russell</autorea>
  <!-- ERROREA: L3 balioa ez da elementu baten IDa -->
</eredua>
```

Atributuen hasierako balioak honako hauek izan daitezke:

• #REQUIRED: atributua nahitaezkoa da, baina ez da balio lehenetsirik zehazten.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua EMPTY>
        <!ATTLIST eredua kolorea CDATA #REQUIRED>
]>

<eredua kolorea=""></eredua>
    <eredua kolorea="urdina"></eredua>
    <eredua kolorea="urdina"></eredua>
    <eredua kolorea="urdin iluna #000033"></eredua>
    <eredua/>
    <!-- ERROREA: kolorea atributua falta da -->
```

• **#IMPLIED**: atributua ez da nahitaezkoa eta ez da balio lehenetsirik zehazten.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua EMPTY>
    <!ATTLIST eredua kolorea CDATA #IMPLIED>
]>
```

```
<eredua kolorea=""></eredua>
<eredua kolorea="urdina"></eredua>
<eredua kolorea="urdin iluna #000033"></eredua>
<eredua/>
```

• #FIXED: atributuak balio finkoa du.

• **Balioa**: atributuak balio lehenetsia du.

```
<!DOCTYPE eredua [
    <!ELEMENT eredua EMPTY>
    <!ATTLIST eredua kolorea CDATA "berdea">
]>

<eredua kolorea="berdea"></eredua>
<eredua kolorea="urdina"></eredua>
```

# Entitateen adierazpenak

Entitate bat izen bat eta bere balioa da (programazio-lengoaietan konstanteen antzekoak).

XML prozesadoreak entitateak aipatzen direnean, bere ordez horren balioak erabiltzen ditu dokumentua prozesatu aurretik.

Entitatea definitu ondoren, horren erreferentzia bat idatziz erabil daiteke dokumentuan.

Erreferentzia hori "&" izaerarekin hasten da, entitatearen izenarekin jarraitzen du eta ";" -rekin amaitzen da. Hau da: &entitateIzena;

Jarraian, barne entitateak deklaratzeko sintaxia:

```
<!ENTITY entitateIzena "entitateBalioa">
```

#### non:

- entitateIzena: entitatearen izena da.
- entitateBalioa: entitate izenerako balioa du.

#### Eredua:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<!DOCTYPE address [
<!ELEMENT address (#PCDATA)>
<!ENTITY name "Tanmay patil">
<!ENTITY company "TutorialsPoint">
```

```
<!ENTITY phone_no "(011) 123-4567">
]>
<address>
    &name;
    &company;
    &phone_no;
</address>
```

Gorde fitxategi hau sample.xml gisa eta zabaldu edozein web arakatzailetan, nabarituko duzu name, company, phone\_erako entitate balioak ordezkatzen diren.

#### DTD bat sortzeko adibidea

Adibide meduan, denda bateko bezeroen eskaerak gordetzen dituen XML <u>eskaera.xml</u> erabiliko da:

```
<?xml version="1.0" encoding="UTF-8" ?>
<eskaera zenbakia="26" eguna="2017-12-01">
  <bezeroa kodea="C20" bazkidea="x12y5">
    <izena>Felipe</izena>
    <abizena>Garcia</abizena>
  </bezeroa>
  <artikuluak>
    <artikulua>
      <deskripzioa>Yoda Mimobot USB Flash Drive 8GB</deskripzioa>
      <kopurua>5</kopurua>
      <prezioa>38.99</prezioa>
    </artikulua>
    <artikulua>
      <deskripzioa>Darth Vader Half Helmet Case for iPhone</deskripzioa>
      <kopurua>2</kopurua>
      <prezioa>29.95</prezioa>
    </artikulua>
  </artikuluak>
  <totala balioa="254.85"/>
</eskaera>
```

Ereduaren helburua XMLa balidatzeko dagokion DTDa sortzea da, eskaera.dtd deitutakoa.

XMLaren elementuak aztertu eta DTDa pixkanaka sortuko da:

Dokumentuaren erroa <eskaera> elementua da, eta hiru haur ditu:

```
<!ELEMENT eskaera (bezeroa, artikuluak, totala)>
```

También tiene dos atributos, numero y dia, que sólo pueden tener v

Horrez gain, bi atributu ditu, zenbakia eta eguna, espazio ez duten balioak soilik izan ditzaketenak eta, beraz, NMTOKEN izango dira.

Nahitaezko datuak izango dira zerga arrazoiengatik.

```
<!ATTLIST eskaera zenbakia NMTOKEN #REQUIRED>
<!ATTLIST eskaera eguna NMTOKEN #REQUIRED>
```

Behin lehen maila definitutakoan, ondorengo mailak definitzen jarraitzen da.

Batetik, <bezeroa> elementua.

Bezero guztiek era bakarrean identifikatzen dituen kodea atributua behar dute, baina ez dute zertan bazkidea izan behar eskaerak egiteko.

```
<!ELEMENT bezeroa (izena, abizena)>
<!ATTLIST bezeroa kodea ID #REQUIRED>
<!ATTLIST bezeroa bazkidea NMTOKEN #IMPLIED>
```

Bestalde, <totala> elementua, hutsik dagoena baina atributu bat dauka.

```
<!ELEMENT totala EMPTY>
<!ATTLIST totala balioa CDATA #REQUIRED>
```

Era berean, <artikuluak> elementuak bezeroak erosi dituen artikuluen zerrenda izango du.

Artikulurik gabeko eskaera egitea zentzuzkoa ez denez, erabiliko den aldagailua + izango da.

```
<!ELEMENT artikuluak (artikulua+)>
```

Eduki moduan duen <artikulua> elementuak garatzeak ez du arazo askorik.

```
<!ELEMENT artikulua (deskripzioa, kopurua, prezioa)>
```

Amaitzeko, datuak dituzten elementuak baino ez dira geratzen emaitza gisa *eskaera.dtd* fitxategia lortzeko.

```
<!-- "eskaera"-ren elementuak eta atributuak adierazten dira -->
<!ELEMENT eskaera (bezeroa, artikuluak, totala)>
<!ATTLIST eskaera zenbakia NMTOKEN #REQUIRED>
<!ATTLIST eskaera eguna NMTOKEN #REQUIRED>
<!ELEMENT bezeroa (izena, abizena)>
<!ELEMENT artikuluak (artikulua+)>
<!ELEMENT totala EMPTY>
<!-- "bezeroa"-ren elementuak eta atributuak adierazten dira -->
<!ATTLIST bezeroa kodea ID #REQUIRED>
<!ATTLIST bezeroa bazkidea NMTOKEN #IMPLIED>
<!ELEMENT izena (#PCDATA)>
<!ELEMENT abizena (#PCDATA)>
<!-- "artikulua"-ren elementuak eta atributuak adierazten dira -->
<!ELEMENT artikulua (deskripzioa, kopurua, prezioa)>
<!ELEMENT deskripzioa (#PCDATA)>
<!ELEMENT kopurua (#PCDATA)>
<!ELEMENT prezioa (#PCDATA)>
<!-- "totala"-ren atributuak adierazten dira -->
<!ATTLIST totala balioa CDATA #REQUIRED>
```

Azkenik, XMLa balioztatuko duen DTDarekin lotzeko, *eskaera.xml* dokumentuaren bigarren lerroan, kode hau gehituko da:

```
<!DOCTYPE eskaera SYSTEM "eskaera.dtd">
```