

XML (Extensible Markup Language)

XML edo Etiketa Lengoaia Hedatua, **World Wide Web Consortium**-ek garatutako etiketa **metahizkuntza** da, eta SGML du inspirazio iturri.

XML ez da lengoaia bat, etiketa multzo bat definitzen ez duelako. XML metahizkuntza bat da, hizkuntzak sortzeko aukera ematen duelako:

- XMLk arau batzuk ematen ditu edonork bere etiketa eta atributu multzoa definitzeko eta etiketa horien artean dauden erlazioak adierazi ahal izateko.

Bestalde, HTML, XHTML, MathML, RSS edo SVG lengoaiak dira, etiketa eta atributu multzo bat definituta dutelako lan egiteko:

- HTML SGML arauetan oinarrituta dago, eta aipatu diren beste lengoaiak, berriz, XMLn oinarrituta daude.

XMLk informazioa gordetzeko eta trukatzeko modu benetan indartsua ematen die webguneei eta aplikazioei. Gainera, formatu unibertsala bihurtu da, eta sistema eragile eta gailu mugikor mota guztiek bereganatu dute.

XML asko erabiltzen da bai web-zerbitzuetan eta APIetan informazioa transmititzeko, baita ere konfigurazio-fitxategietarako.

- Euskarri gehien ematen dion programazio-lengoaietako bat Java da.

HTMLn bezala, XML dokumentu bat testu-dokumentu bat da, kasu honetan ".xml" luzapenarekin, zuhaitzetan egituratutako etiketa bikotez osatua, eta dokumentuaren antolamenduan funtzio bat deskribatzen dutenak.

Jarraian, XML erabilia, egitura hierarkiko eta malgu baten bidez liburu bati buruzko oinarritzko informazioa nola irudika genezakeen azaltzen da:

```
<katalogoa>
  <liburua>
    <egilea>Raúl González Duque</egilea>
    <izenburua>Python guztiontzat</izenburua>
    <generoa>Konputazioa</generoa>
  </liburua>
</katalogoa>
```

XML dokumentu bat sortzeko prozesuak hainbat etapa ditu, eta horietako bakoitzaren arrakasta aurrekoaren kalitatean oinarritzen da.

- Betekizunak zehaztea.
- Etiketak diseinatzea.
- Dokumentuak markatzea.

Markatzea nahi bezain aberatsa izan daiteke, eta interesgarria da etorkizuneko beharrak detektatzea eta erraz eguneratu daitekeen egitura duten dokumentuak sortzea.

XMLko oinarrizko tresnak

XMLn lan egiteko, beharrezkoa da dokumentuak editatzea eta gero prozesatzea; beraz, bi tresna mota ditugu:

XML editoreak:

XML dokumentu konplexuak sortzeko eta datuak gehitzen joateko, XML editore bat erabiltzea komeni da, hauek dokumentuetan erabiltzen diren elementuen egiturak eta etiketak sortzen laguntzen bait digute.

Gainera, editore batzuek beste elementu batzuk sortzeko laguntza ematen dute, hala nola DTD, CSS edo XSL estiloko orriak,...

Erabiltzaileari XML lengoia dokumentuak bistaratzeko, baliozkotzeko eta editatzeko aukera ematen dioten aplikazioak erabil ditzakegu, hala nola Editix eta Oxygen, baina ohikoena da Notepad++ edo VSCode bezalako editore bat erabiltzea eta funtzionalitate horiek lortzeko beharrezkoak diren pluginak gehitzea.

XML prozesadoreak:

XML kodea interpretatzeko edozein nabigatzaile erabil daiteke. XML prozesadoreek XML dokumentuak irakurtzeko eta haien edukia eta egitura ikusteko aukera ematen dute.

Prozesadore bat software-moduluen multzo bat da, eta horien artean XMLren parser edo analizatzaile bat dago, dokumentuak ireki ahal izateko ezarritako arauak betetzen dituela egiaztatzen duena.

Prozesadoreek XML datuak irakurtzeko duten modua W3C-k ezarritako XML gomendioan deskribatuta dago.

XML dokumentu bat **bi mailatan** zuzena izan daiteke:

Ondo prestatua:

XML dokumentu bati eskatzen zaizkion ezaugarriak edo estilo-gida betetzen dituenean, ondo idatzita dagoela esan daiteke.

Baliozkoa:

XML dokumentu bat ondo prestatuta dagoela egiaztatu ondoren, datuen deskribapena edo erabiltzen ari den txantiloia betetzen duela egiaztatu behar da.

Hau da, eskatzen diren elementuak eta ezaugarriak dituela, proposatutako egituraren arabera antolatzen direla, eta hauetan ezarritako baldintzak betetzen dituztela, besteak beste. Horretarako, DTD edo XML Schema erabiltzen dira.

XML egitura era sintaxia

XML dokumentuak bi zatitan egituratzen dira: aukerakoa den **sarrera** zati batekin eta beharrezkoa den **gorputza** zati batekin.

Sarrerak dokumentu osoari lotutako informazioa jasotzen du, eta gorputzak informazioarekin berarekin dituen elementuak biltzen ditu. Prozesatuko ez diren iruzkinak ere sartzeko aukera ematen du.

Sarrera

Dokumentua prozesatzeko ardura duen interpretari bere lana egiteko behar dituen datu guztien berri emango dio. Bi zati ditu:

XML definizioa

Honetan dokumentua XML-a dela adierazten da, bertsioa eta erabiliko den kodeketa honako atributu hauen bidez:

- *version*: XML bertsioa.
- *encoding*: Dokumentuan erabiltako karaktere kodifikazioa.
- *standalone*: Dokumentua independentea den edo baliozkotzea egiteko kanpoko DTD bat dagoen adierazten du. Bai (independentea da) eta ez (balioztapeneko DTD batekin lotuta dago) balioak onartzen ditu. Atributu hau hautazkoa da.

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

Dokumentu motaren deklarazioa

Behar bezala prozesatzeko sortzen ari garen dokumentu mota definitzen du.

Dokumentu-motari buruzko adierazpen oro kate honetatik hasten da: `<!DOCTYPE gorputza_izena ...>` (*gorputza_izena* lantzen ari den XML dokumentuaren gorputzaren berdina izan behar du.)

Gorputza

XML dokumentu baten zatirik garrantzitsuena da, **elementu** habiatuak osatuta dokumentuaren benetako datuak biltzen bait ditu.

- Elementuak XML dokumentu baten egitura definitzeko aukera ematen duten informazio-blokeak dira.
- HTML elementuen oso antzekoak dira, irekiera-etiketa batek eta itxiera-etiketa batek mugatzen bait dituzte, XMLaren kasuan derrigorrezkoa dena.
- Era berean, elementuak beste elementu eta/edo atributu batzuek osatuta egon daitezke.

```
<!-- erro elementua -->
<katalogoa>
  <!-- edukiontzi elementua -->
  <liburua>
    <!-- elementu sinpleak -->
    <egilea>Raúl González Duque</egilea>
    <izenburua>Python guztiontzat</izenburua>
    <generoa>Konputazioa</generoa>
  </liburua>
</katalogoa>
```

XML dokumentu baten elementuen arteko erlazioak honela izendatzen dira:

- Erro-elementu bakarra dago.
- Elementu baten zuzeneko elementu bakoitzari haur deitzen zaio (*children*).
- Elementu baten aurreko elementu zuzenari bere gurasoa deitzen zaio (*parent*).

- Guraso komuna duten elementuei anai-arrebak deitzen zaie (*sibling*).

```
<katalogoa>
  <liburua>
    <egilea>Raúl González Duque</egilea>
    <izenburua>Python guztiontzat</izenburua>
    <generoa>Konputazioa</generoa>
  </liburua>
  <liburua>
    <egilea>Fabien Potencier</egilea>
    <izenburua>Symfony 6: The Fast Track</izenburua>
    <generoa>Programazioa</generoa>
  </liburua>
</katalogoa>
```

Aurreko XML dokumentua oniarri hartuta:

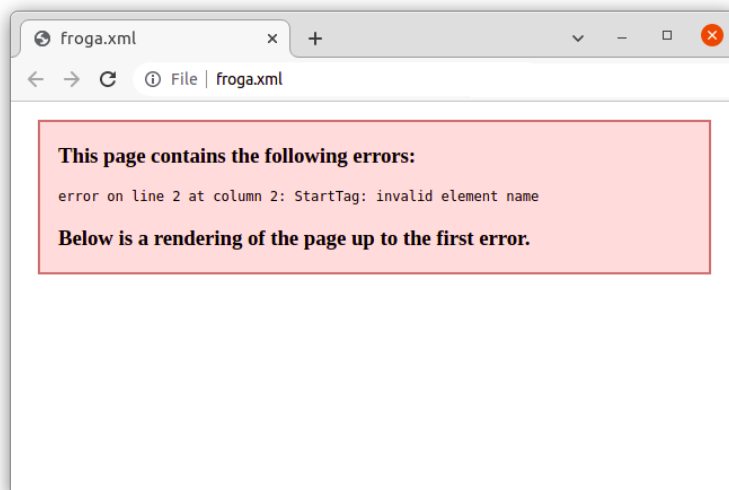
- `<katalogoa>` orri-elementua da.
- `<liburua>` elementuak `<katalogoa>` elementuaren haurrak dira (*children*).
- `<katalogoa>` elementua `<liburua>` elementuen gurasoa da (*parent*).
- `<liburua>` elementuak elkarren anai-arrebak dira (*sibling*).
- `<egilea>`, `<izenburua>` eta `<generoa>` elementuak `<liburua>` elementuen haurrak dira.

Elementu izenak esleitzeko arauak hauek dira:

- Maiuskulak eta minuskulak bereizten dira.
- Letra batekin edo azpi marra batekin (`_`) hasi behar da.
- Elementu-izenak berdinak izan behar dira hasierako eta ixteko etiketetan.
- Izenak karaktere alfanumerikoz, marraz, azpi marraz eta puntuz eratuak izan daitezke.
- Izenek ezin dute espaziorik izan, lehenengo hitza elementuaren izen gisa interpretatzen da eta gainerakoak parametro gisa.

Ondorengo elementuak errore bat sortuko du web-arakatzaille batekin baliozkotzen denean:

```
<?xml version="1.0" encoding="UTF-8"?>
<1color>blue</1color>
```



Atributuak

Atributuek dokumentu baten elementuei propietateak gehitzeko aukera ematen dute.

Atributuak adierazteko arauak hurrengoak dira:

- Balio bat esleitua izan behar dute.
- Balioak komatxoaren artean sartzen dira beti, komatxo sinpleak edo bikoitzak onartuz (itxierako komatxo motak hasierakoarekin bat etorri behar du).
- Maiuskulak eta minuskulak bereizten dira.
- Karaktere edo azpi marra batekin hasi behar dira.
- Karaktere alfanumerikoz, marraz, azpi marraz eta puntuz eratuak izan daitezke.

Adibidez, XML dokumentu baten elementu bateko atributuaren adierazpen ereduak:

```
<liburua isbn="978-84-616-4591-6">
```

Dokumentuaren gorputzan zehar elementuak zuhaitzeko egitura baten arabera banatzen dira, hau da, elementuak habia egin dezakete, baina ezin dira entrelatu.

Atributuak ezin dira hierarkian antolatu, ezin dute beste elementu edo atributurik eduki eta ez dute egitura logikorik izaten.

Ez dago arau zehatzik informazioa elementu edo atributu moduan dokumentuetan erabiltzeko.

Zer irizpide erabil ditzakegu egituratu nahi den agiriaren datu bat elementu edo atributu baten bidez adierazi behar den erabakitzeko?

Datua elementu bat izango da ondorengo baldintzetako bat betetzen badu:

- Azpiegiturak baldin baditu, hau da, ez da atributurik erabili behar zatitzeko moduko informazioa edukitzeko.
- Tamaina handikoak badira.
- Horren balioa maiz aladatu daiteke.
- Horren balioa erabiltzaile edo aplikazio bati erakutsi nahiko zaio.

Datua atributu bat izango da ondorengo baldintzetako bat betetzen badu:

- Datua tamaina txikikoa da eta horren balioa ez omen da aldatzen.
- Datuak balio finko batzuk izan ditzake soilik.
- Datuak XML prozesamendua gidatzen du, baina ez da erakutsiko.

CDATA atala

Prozesatu edo aztertu gabeko edukia gehi daiteke atal honetan, iruzkin baten antzekoa.

XML dokumentuaren sustrai edo erro elementuaren barruan agertu behar dira. Adibidez:

```
<?xml version="1.0" encoding="UTF-8"?>
<katalogoa>
  <![CDATA[
    <h1>Liburuen zerrenda</h1>
```

```

        <hr>
        <a href="esteka">zerrenda bistaratu</a>
    ]]>
    <liburua>
        <egilea>Raúl González Duque</egilea>
        <izenburua>Python guztiontzat</izenburua>
        <generoa>Konputazioa</generoa>
    </liburua>
</katalogoa>

```

Aurreko kodeak HTML kodea barne duen XML dokumentu bat erakusten du.

HTML barruan XML arauak betetzen ez dituen `<hr>` elementu bat dago, ez baitu itxiera etiketarik. Hori dela eta, beharrezkoa da HTML kodea CDATA atalean sartzea dokumentua ondo osatuta egon dadin.

[HTML eta XML-aren arteko desberdintasunak](#)

Izen espazioak

XML izen espazioak W3Cren gomendio bat da, XML fitxategi batean izen berdingabea duten elementuak eta atributuak izateko.

XML fitxategi batek hiztegi desberdinetatik datozen elementu edo atributu izenak bildu ditzake, hiztegi bakoitzari izen espazio bat ematen bazaio (biltzen dituen terminoak zerrendatzen diren URI bati erreferentzia eginez).

Izen espazio bat erreserbatutako ***xmlns*** atributua erabiltzen adierazten da, eta horren balioa baliabidearen identifikatzaile berdingabe eta uniforme bat izan behar da, adibidez: `xmlns = "http://www.w3.org/1999/xhtml"`.

- Kontuan izan behar da URIa ez dela benetan helbide gisa irakurtzen; XML analizatzaileak testu-kate gisa tratatzen du.
- `http://www.w3.org/1999/xhtml` berak ez du inolako koderik, XHTML izen-espazioa deskribatzen die giza irakurleei.

URL egitura bat erabiltzeak izen-eremu bat identifikatzeko, kate soil bat erabili ordez, izen-eremu ezberdinek identifikatzaile berdinak erabiltzeko aukera murrizten bait duelako egiten da.

Hau da, XML dokumentu batean zehaztutako URIek ez dute zertan ezer eduki behar, haien funtzioa bakarra izatea da.

Hala ere, informazioa URI batean bistaratu daiteke, egokituz jotzen bada. Adibidez:

- <http://www.w3.org/1999/xhtml/>
- <http://www.w3.org/1999/XSL/Transform>
- <http://www.w3.org/2000/svg>

Hurrengo XML dokumentuak, bezero bati eta hark eskatutako produktu bati buruzko erreferentziak instantziatu ahalko ditu.

Bai bezeroa irudikatzen duen elementuak, bai produktua irudikatzen duenak, "*kode*" izeneko haur elementua izan dezakete.

"Kode" elementuari buruzko erreferentziak anbiguoak izan daitezke, non eta elementuak, izen berarekin baina esanahi desberdinarekin agertuko dira, izen desberdineko espazioetara eramaten ez badira elkarren artean bereizteko.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<bezeroa:bezeroa xmlns:bezeroa="http://www-my-uri.eus/bezeroa"
    xmlns:eskaera="http://www-my-uri.eus/eskaera">
  <bezeroa:kodea>123456</bezeroa:kodea>
  <bezeroa:izena>Pernando Amezketarra</bezeroa:izena>
  <bezeroa:telefono-zenbakia>123123123</bezeroa:telefono-zenbakia>
  <eskaera:eskaera>
    <eskaera:kodea>5555555</eskaera:kodea>
    <eskaera:artikulua>Python guztiontzat</eskaera:artikulua>
    <eskaera:prezioa>55,95</eskaera:prezioa>
  </eskaera:eskaera>
</bezeroa:bezeroa>
```

Ondo prestatutako XML dokumentuak

Elementuen eratzeak arau jakin batzuk bete behar ditu, behar bezala definituta gera daitezen eta XML dokumentua XML prozesadoreek interpretatu ahal izan dezaten, akats larriak sortu gabe.

Arauak hurrengoak dira:

- XML dokumentu guztietan erro-elementu bakar bat egon behar da.
- Elementu guztiek hasierako eta amaierako etiketa bana dute. Dokumentuan elementu hutsak badaude, hasierako eta amaierako etiketen orde elementu huts bat jar daiteke.
 - Hau da, <elementu></elementu> etiketak <elementu/> etiketa bakarrarekin ordeztu daitezke.
- Elementuak habiatzean, kontuan izan behar da ezin dela itxi oraindik itxi ez den beste elementuren bat duen elementurik.
 - Adibidez, <elementu><tag></elementu> adierazpena ez da zuzena, <tag> etiketa itxi beharra dago <elementu> etiketa itxi baino lehen.
- Elementu beraren hasierako eta amaierako etiketen izenak berdinak izan behar dira, letra larriak eta xeheak errespetatuz.
- Lehenengo karaktereak hizkia edo gidoi baxua (_) izan behar da.
- Letra xeheak, letra larriak, zenbakiak, puntuak (.), erdiko gidoiak (-) eta gidoi baxuak (_) izan ditzakete.
- Era berean, bi puntu (:) izan ditzakete. Hala ere, izenen espazioak zehazten direnerako gordetzen da hauen erabilera.
- Elementuen edukiak ezin du "]] >" katea eduki, SGML-KO bateragarritasuna bete dadin. Gainera, ezin dira zuzenean erabili baino handiagoa (>), baino txikiagoa (<), anpersand (&), komilla bikoitzak ("), eta apostrofe (') karaktereak.
- Etiketen izenak autodeskribagarriak izan behar dira, horrek lana errazten du.
- Atributuak komillen artean adierazi behar dira.

Bestalde, kontuan izan behar da etiketa baten izenaren atzean espazio zuri bat edo lerro-jauzi bat idazteko aukera dagoela. Adibidez, sintaktikoki zuzena da idaztea:

```
<ciudad >Pamplona</ciudad>
```

Hala ere, ezin da lerro-jauzirik edo zuriunerik egon etiketa baten izena baino lehen:

```
<ciudad>Pamplona</ ciudad>
```

Ondorengo elementuak ez daude behar bezala idatzita sintaxi-arauren bat ez betetzeagatik:

```
<Hiria>Irunea</hiria>
<agurt'erdi/>
<hilabetea>6<hilabetea/>
<hiria>Irunea</hiriamaiera>
<_gorria>
<2kolore>Gorria eta laranja</2kolore>
< Zaletasunak >Zinema, Dantzatzea, Igeri egitea</ Zaletasunak >
<pertsona><izena>Elsa</pertsona></izena>
<kolore gogokoena>urdina</kolore gogokoena>
```

Horren ordeez, zuzena litzateke honako hauek idaztea:

```
<Hiria>Irunea</Hiria>
<agurEtaErdi/>
<hilabetea>6</hilabetea>
<hiria>Irunea</hiria>
<_gorria/>
<kolore2>Gorria eta laranja</kolore2>
<Zaletasunak >Zinema, Dantzatzea, Igeri egitea</Zaletasunak >
<pertsona><izena>Elsa</izena></pertsona>
<kolore-gogokoena>urdina</kolore-gogokoena>
<kolore_gogokoena>urdina</kolore_gogokoena>
<kolore.gogokoena>urdina</kolore.gogokoena>
```

Erdiko gidoia (-) eta puntuari (.) dagokionez, etiketak izendatzeko ere baimenduta dauden arren, horiek ez erabiltzea gomendatzen da.

- Gidoi ertaina ken zeinuarekin nahastu daitekeelako
- Puntua, adibidez, kolore.gogokoena idaztean, gogokoena kolore-objektuaren ezaugarri dela interpretatu daitekeelako.