

# XML

---

## Uvod

**XML** (Extensible Markup Language) je jezik za označavanje (*engl.* Markup language), sličan HTML-u. Služi za definisanje formata podataka (za razliku od HTML-a koji služi za definisanje prikaza podataka). XML standard propisan je od strane W3C. Ovaj (meta) jezik definiše sintaksu za grupisanje podataka koristeći se takozvanim **etiketama** (*engl.* Tags). Mnogi jezici za označavanje se baziraju upravo na XML-u i brojni su softveri i biblioteke koji pružaju podršku za čitanje i pisanje XML-a. Validnost XML dokumenta je lako proveriti, jer XML ima stroga sintakсна pravila.

XML format podataka nezavisan je od platforme i XML dokument može čitati svaka alatka koja rukuje tekstualnim datotekama, s obzirom da su tekstualni ne samo podaci već i etikete. Dakle, XML generiše podatke koji su prenosivi i nezavisni od hardverske ili softverske platforme. Njegova osnovna zamisao je upravo da olakša razmenu podataka kroz informacione sisteme, koji su potencijalno nekompatibilni.

## Struktura i sintaksa

Etikete XML-a nisu predefinisane, tj. moraju se definisati. Etiketom se daje značenje podatku. Njegova proširivost se ogleda u tome da je dozvoljeno definisanje novih etiketa. XML je samodeskriptivan, što znači da svojim etiketama definiše sadržaj obuhvaćen njima. Jedan XML dokument sastoji se iz teksta i etiketa.

U XML dokumentima podatke smeštamo u obliku stringova između etiketa koje ih opisuju i definišu. Ove jedinice se nazivaju **elementima**. Dokument ima **stablo** strukturu. Nepisano pravilo je i da etikete opisuju i semantiku podataka i shodno tome se imenuju. Bitno je zapamtiti da etikete ne sadrže informaciju o tome kako će element biti prikazan.

Hijerarhijska, odnosno stablo, struktura dokumenta, podrazumeva da jedan dokument ima jedan i samo jedan **koreni** element (*engl.* Root). Svi ostali elementi nalaze se unutar korenog elementa. Dozvoljeno je višestruko ugnježdavanje elemenata. Pri pisanju XML dokumenta bitno je voditi računa o indentaciji dokumenta, kako bi se razaznalo kako su elementi međusobno ugnježdjeni.

Primer XML dokumenta:

```
<student>

  <ime>Pera</ime>

  <prezime>Perić</prezime>

  <index>RA46</index>

</student>
```

Prva *student* etiketa predstavlja **otvarajuću etiketu**, dok krajnja *student* etiketa predstavlja **zatvarajuću etiketu**. Svaka etiketa mora biti otvorena i zatvorena. Zatvarajuća etiketa izgleda isto kao otvarajuća sa razlikom u predznaku /. Sve etikete se nalaze unutar uglastih zagrada <>. Isto važi i za etikete *ime*, *prezime* i *index*. Sadržaj unutar etiketa predstavlja **tekst**. Imena etiketa razlikuju velika i mala slova (*engl.* Case sensitive).

## XML atributi

Elementima je moguće pridružiti attribute. XML atributi obezbeđuju dodatne informacije elementima. Vrednost atributa se uvek nalazi između znaka navoda " " ili ' '. Moguće je i kombinovanje u slučajevima kada je potrebno prikazati i znake navoda. Primer korišćenja atributa:

```
element:

<ime>Pera</ime>

atribut:

<nesto ime="Pera">
```

Ne postoji jasno definisano pravilo kada se služiti elementom, a kada atributom, ali preporučljivo je da se elementi koriste kada se opisuje podatak koji je već sam po sebi celovita informacija, a ne jedan njen deo. Oba primera koja slede daju istu informaciju, ali su drugačije predstavljeni. Prvi je u formi atributa, drugi elementa:

```
<student tip="novi">

<ime>Pera</ime>

<prezime>Perić</prezime>

</student>

<student>novi</student>

<ime>Pera</ime>

<prezime>Perić</prezime>
```

Mane korišćenja atributa:

- Teška manipulacija iz programskog koda
- Teška proširivost
- Ne mogu da sadrže višestruke vrednosti
- Ne opisuju strukturu

## XML deklaracija

Svaki XML dokument mora sadržati deklaraciju na samom početku dokumenta. Ova deklaracija (instrukcija obrade) služi da bi se dokument identifikovao kao XML dokument. Atributi koje deklaracija može da sadrži su dati u tabeli ispod.

Atribut	Vrednost atributa	Opis
version	1.0	Korišćena verzija XML standarda.
encoding	UTF-8, UTF-16...	Korišćeno kodiranje karaktera. Podrazumevana vrednost je <b>UTF-8</b> .
standalone	yes/no	Govori da li se dokument oslanja na informacije iz nekog drugog dokumenta. Podrazumevana vrednost je <b>no</b> .

Primeri XML deklaracije:

```
<?xml version="1.0" ?>

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

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

Karakter **<** u deklaraciji označava instrukciju obrade – poruka za program koji procesira XML dokument.

Za verifikaciju XML dokumenta zadužen je **XML parser**. XML parser čita dokument i vrši konverziju u strukturu stabla.

## Entiteti

Neki karakteri ne mogu da se koriste kao deo teksta u XML. Na primer, karakter **<** unutar XML dokumenta će parser prepoznati kao početak nove etikete.

```
<message>salary is < 1000</message>
```

U primeru iznad, parser će prikazati grešku prilikom parsiranja jer znak **<** prepoznaje kao otvarajuću etiketu, a u XML dokumentu ne postoji zatvarajuća etiketa za njega.

U ovakvim slučajevima u XML dokumentu se koriste **entiteti**. Oni se obrađuju prilikom parsiranja od strane XML parsera. Imena entiteta uvek počinju karakterom **&** za kojim sledi tekst koji opisuje značenje entiteta. Neki od predefinisanih XML entiteta su:

- **&lt;** – manje od (karakter **<**)
- **&gt;** – veće od (karakter **>**)
- **&amp;** – ampersand (karakter **&**)
- **&quot;** – dupli znaci navoda (karakter **"**)
- **&apos;** – apostrof (karakter **'**).

Prethodni primer bi se korišćenjem entiteta napisao na sledeći način:

```
<message>salary is &lt; 1000</message>
```

Samo karakteri **<** i **&** su zabranjeni za korišćenje unutar teksta XML dokumenta, ali je dobra praksa koristiti i ostale entitete.

## Imenski prostori (engl. *Namespaces*)

Imenski prostori se koriste kako bi se izbegli konflikti pri zadavanju imena elementima. Primer:

```
<table>

<tr>

  <td>Apples</td>

  <td>Bananas</td>

</tr>

</table>
```

U XML dokumentu iznad definisana je tabela slično definiciji HTML tabele.

```
<table>

<name>African Coffee Table</name>

<width>80</width>

<length>120</length>

</table>
```

U XML dokumentu iznad definisan je opis stola (engl. Table).

Ukoliko bi se ova dva dela XML dokumenta našla u istom dokumentu pojavio bi se konflikt u nazivu etiketa (oba sadrže etiketu `<table>` međutim unutrašnje etikete se razlikuju). Ovakvi konflikti se rešavaju korišćenjem prefiksa u nazivu etikete:

```
<h:table>

  <h:tr>

    <h:td>Apples</h:td>

    <h:td>Bananas</h:td>

  </h:tr>

</h:table>

<f:table>

  <f:name>African Coffee Table</f:name>

  <f:width>80</f:width>

  <f:length>120</f:length>

</f:table>
```

Kada se u XML dokumentu koriste prefiksi, potrebno je definisati imenski prostor za prefiks. Imenski prostor se definiše atributom **xmlns** u početnoj etiketi elementa. Atribut se definiše na sledeći način: `xmlns:prefix="URI"` (*engl.* Uniform Resource Identifier). Kada se definiše imenski prostor elementa, svi deca – elementi sa istim prefiksom su povezani sa definisanim imenskim prostorom. Imenski prostori se takođe mogu definisati i u korenskom elementu XML dokumenta, koristeći navedenu sintaksu i razdvajajući razmakom.

Primer XML dokumenta sa definisanim imenskim prostorom i prefiksima:

```
<root>

<h:table xmlns:h="html/table">
  <h:tr>
    <h:td>Apples</h:td>
    <h:td>Bananas</h:td>
  </h:tr>
</h:table>

<f:table xmlns:f="furniture/table">
  <f:name>African Coffee Table</f:name>
  <f:width>80</f:width>
  <f:length>120</f:length>
</f:table>

</root>
```

Primer definisanja imenskog prostora unutar korenog elementa:

```
<root xmlns:h="html/table" xmlns:f="furniture/table">
...
</root>
```

Primer jednog složenijeg XML dokumenta, sa upotrebom deklaracije, elemenata sa atributima i imenskim prostorom:

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

<bookstore>

  <author xmlns:daunt="http://www.dauntbooks.co.uk" name="Eric Arthur Blair" pseudonym="George Orwell">

    <daunt:book daunt:title="Animal Farm" daunt:year="1945">

      <price>500.00</price>

    </daunt:book>

    <book title="1984" year="1949"/>

  </author>

  <author name="Samuel Langhorne Clemens" pseudonym="Mark Twain">

    <book title="The Adventures of Tom Sawyer" year="1876"/>

    <book title="A Dog's Tale" year="1904"/>

  </author>

  <author name="Stephen Edwin King">

    <book title="Carrie" year="1974"/>

    <book title="The Shining" year="1977"/>

    <book title="Misery" year="1987"/>

  </author>

</bookstore>
```