

# XML – Podaci o podacima

**Boris Bjelošević**

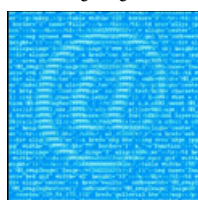
Extensible Markup Language, skraćeno XML, je svojevrstan jezik namenjen za internetsku komunikaciju. Pored toga, on predstavlja i tehnologiju za upravljanje, prikazivanje i organizovanje podataka, a koja se svakim danom sve više razvija.



Od samog početka korišćenja računara, programeri i korisnici su pokušavali da iznađu način za što lakšu razmenu podataka između različitih računarskih programa. Trebalo je pronaći format koji će imati prednosti binarnih fajlova (računar ih lako razume, a sami fajlovi su kompaktni, često malih gabarita), ali biti univerzalan po pitanju razmene, što je prednost klasičnih tekstualnih fajlova.

Jedan od prvih pokušaja takvog objedinjavanja bio je Standard Generalized Markup Language (SGML), jezik baziran na tekstu. Napravljen je tako da predstavlja način obeležavanja podataka za bilo koju namenu sa namenom da se koristi u sistemima za upravljanje i rad sa velikim količinama podataka. Međutim, baš kod takvog rada se pokazala njegova slabost, jer je zahtevao previše stvari o kojima je trebalo voditi računa prilikom programiranja. Ali u složenosti ovog jezika i leže njegove velike mogućnosti.

HTML-jezik je došao kao neka vrsta podskupa jezika SGML. S obzirom da su pravila za izradu SGML-dokumenata bila jasno definisana i dobro postavljena, sam jezik je predstavljao dobru osnovu za izradu specifičnog jezika – HTML-a. Pošto je i HTML, kao podskup SGML-a baziran na tekstu, fajl može napraviti svako ko ume da koristi bilo koji program za unos teksta. Čak i mnogi procesori za obradu teksta nude mogućnost da se sadržaj fajla snimi u formatu HTML-dokumenta.



XML je takođe podskup jezika SGML, nastao iz potrebe da se podaci razmenjuju u računarskoj mreži. Sam SGML-jezik je jako komplikovan i nije prikladan za to namenu, a HTML se pokazao kao ograničenih mogućnosti (ako sama HTML-stranica sadrži informacije o nekom filmu: naziv, ime režisera, glumaca,..., teško da će neki procesor podataka iz tog fajla moći da izvuče samo ime glavnog glumca). XML-jezik je omogućio da se prenesu potrebni podaci, ali je i naznačavao šta koji podatak znači. Praktično XML u sebi "nosi" podatke, ali i podatke o podacima.

Sam XML nije namenjen samo(?) da bi se podaci koje sadrži u sebi prikazali na web-stranicama, nego svima kojima ti podaci nešto znače. Tu je još jedna velika razlika između ovog jezika i HTML-a, koji je namenjen isključivo za kreiranje web-sadržaja. Kad već pominjemo razlike, postoji još jedna bitna – XML je "osetljiv" na veličinu slova: tag koji u HTML-u počinje sa <p> može se zatvoriti tagom </P>, tog kod XML-a to nije slučaj. XML daje slobodu korisniku/programeru da tagovi imaju proizvoljan naziv (uz neka ograničenja

**Omega magazin**

po pitanju karaktera koji se mogu koristiti), ali se strogo mora voditi računa o tome naziv otvorenog i zatvorenog taga bude identičan.

S obzirom da se XML-fajl može koristiti za mnoge stvari, a ne samo za web, postavljena su pravila kako jedan XML-fajl mora da izgleda, a sve to da bi mogao ispravno da se koristi. Program koji učitava podatke iz XML-fajla mora da razume šta znači koji podatak i da ga ispravno tretira. Prema specifikaciji jezika XML (pogledati <http://www.w3.org/TR/1998/REC-xml-19980210#sec-intro>) softverski modul, nazvan XML-ovim procesorom, se koristi za čitanje XML-fajlova i omogućava pristup sadržaju i strukturi samog fajla. Sam XML-ov procesor se naziva parser, a podrazumeva se da ovaj XML-ov procesor svoj posao obavlja za neku drugu konkretnu aplikaciju.

Pored toga što mora ispravno da tretira podatke u XML-fajlu, parser mora i da prepozna greške. Postoje dva tipa grešaka:

1. obične greške (eng. error) – predstavlja nepoštovanje pravila iz same specifikacije jezika XML, pa je krajnji rezultat nedefinisan; prilikom nailaska na ovakvu grešku, parser može ponovo da uspostavi rad i nastavi obrađivanje izvornog fajla;
2. neotklonjive greške (eng. fatal error) – prilikom nastanka ove greške parseru nije dozvoljeno da nastavi rad; neretko proizvođači parsera dozvoljavaju da se u tom slučaju nastavi rad, ali samo da bi se pronašla još neka greška (ako je ima) u izvornom fajlu.



Uvođenje ovakve "strogoće" je posledica nepostojanja iste kod specifikacije za HTML-jezik, pa brauzeri (odnosno njihovi kreatori) imaju veliku međusobnu nekompatibilnost često baš iz razloga što ne znaju šta je kreator HTML-fajla hteo da "kaže" pojedinim delovima svoga koda. Drastičan odnos prema greškama u XML-fajl pomaže i kreatorima parsera, ali i onima koji prave XML-fajlove, jer parser na vreme i na pravi način opiše grešku.

Priča o XML-u ovde tek počinje, kako vezana za njegovo korišćenje na web-u, tako i u ostalim mrežnim aplikacijama. Uz veliku podršku CSS-a i XSL(T)-a, XML pruža neverovatne mogućnosti kreiranja dinamičkih web-stranica. XML se može povezati sa bazom podataka (kako za dobijanje podataka iz jedne ili više tabela, tako i za smeštanje XML-a u bazu), koristiti za kreiranje višeslojnih aplikacija, u elektronskoj trgovini,...