

XML

СЕМИНАРСКИ РАД



МАТЕМАТИЧКИ ФАКУЛТЕТ
УНИВЕРЗИТЕТ У БЕОГРАДУ

СТУДЕНТ: Миљковић Андрија

БРОЈ ИНДЕКСА: 1025/2012

ПРЕДМЕТ: МЕТОДОЛОГИЈА СТРУЧНОГ И НАУЧНОГ РАДА

ШКОЛСКА ГОДИНА: 2012/2013

ПРОФЕСОР: др Владимир Филиповић

Садржај

Шта је XML?	3
Историјат	4
XML чињенице	4
Предности XML-а	5
Шта XML није?	5
Како ради XML?	6
Еволуција XML-а?	8
Основна намена XML-а	8
Терминологија	9
Литература	10

Шта је XML?



XML је стандардни скуп правила за дефинисање формата података у електронској форми. Прописан је од стране W3C. Следећи правила XML стандарда, корисници дефинишу сопствене (XML) формате података, које могу користити за њихово складиштење, обраду и размену. XML је скраћеница за **E**xtensible **M**arkup **L**anguage, односно прошириви (мета) језик за означавање текстуалних докумената. Идеја је била да се створи језик који ће и људи и рачунарски програми моћи једноставно да читају. XML дефинише општу синтаксу за означавање података помоћу одговарајућих етикета које имају познато или лако разумљиво значење. Формат који обезбеђује XML за рачунарске елементе може се прилагодити најразличитијим областима, као што су електронска размена података, чување података, одвајање података од презентације, векторска графика, системи гласовне поште, израда нових специјализованих језика за означавање. XML је и скуп сродних технологија и сам за себе не представља нешто посебно већ тек са сродним технологијама даје пуне резултате. Сродне технологије су:

- DTD
- CSS
- XLS
- DOM
- ADO
- XLink
- XFragments
- JSON
- XPointy

У неким од поменутих технологија XML се понаша као клијент док је за неке сервер а може бити и једно и друго истовремено. Један од највећ их проблема када је у питању трансфер информација је њихов садржај у логичком смислу те речи. XML служи као контернер за трансфер јер у себи поред информације има и њену позицију у односу на остале информације – XML поред информације описује и структуру.

Историјат

Прво се у IBM-у појавио SGML (Standard Generalized Markup Language) као одговор на проблем пребацивања докумената са једне на другу платформу. Тим Бернерс Ли (Tim Berneres Lee - творац интернета) изабрао је мали скуп етикета из SGML-а и применио их на форматирање докумената. Тако настаје HTML (HyperText Markup Language), који је примењени SGML. HTML је имао коначан скуп етикета које су описивале основне делове докумената. Али, пошто је намењен за израду веб страница, није подесан за употребу изван ове области. Због тога је SGML употребљаван за друге апликације које користи интернет, а нису просте веб странице. Затим се крајем осам десетих у CERN-у (европској лабораторији за физику атомских честица) појавио HTML а када је после успешне промоције на интернету постало очигледно да HTML не може баш све родила се идеја о XML -у негде 1996 године. Међутим XML није еволуирани HTML. Он је комплемент HTML-у и дизајниран је не да замени HTML, већ да понуди оно што HTML не може.

XML чињенице

- XML је метод за смештање структурираних података у текстуални фајл XML има податке и структуру у текстуалном фајлу. То га квалификује за рад са базама података.
- XML личи на HTML али није HTML Иако користи тагове као и HTML он се суштински разликује од HTML -а јер произвољно проширив таговима које сами измишљате.
- XML је текст али није намењен читању од стране људи већ машина Иако је текст није намењен читању већ парсирању од стране рачунара.
- XML је породица технологија XML чини читава породица технологија и он сам по себи не представља посебно много функционалности али у садејству са осталим технологијама (CSS, HLink, HPointer, HFragments, HSL ...итд) даје одличне резултате.
- XML за последицу има обиман фајл али то није проблем Иако је XML фајл обиман због употребе ознака то није проблем. Ствар се компензује тиме што добијате флексибилност у примени.
- XML је нов али није баш толико нов XML је релативно нов али његови корени сежу у почетке осамдесетих. Он представља еволуцију идеје а не еволуцију језика.
- XML се може користити за развој нових језика XML је претеча WAP-а и WML-а. Wireless Markup Language (WML), који се користи на пример у обележавању Интернет апликација за мобилне телефоне је заправо уско специјализовани XML. Слична ствар је и са WAP-ом.
- XML не припада никоме, не зависи од платформе и добро је подржан XML је само спецификација W3 конзорцијума. Уједно је и препорука признатог ауторитета. Нема профитну позадину и слободан је за употребу. Да бисте га користили нисте ничим обавезани, ни платформом, ни произвођачима, ни лиценцама и уговорима. Схватите његову суштину и користите га онако како вама одговара.

Предности XML-а

Подаци се смештају у XML документе у облику знаковних низова (енгл. Strings), између текстуалних ознака које их описују. Основне јединице података и ознака у XML -у називају се елементи. XML спецификација прецизно дефинише синтаксу које се морате придржавати при писању ознака: како су елементи разграничени ознакама, како ознака изгледа, каква имена елементи могу имати, где се стављају атрибути итд. Површно гледано, ознаке XML документа много личе на ознаке HTML документа, али међу њима постоје битне разлике.

Најважније је да је **XML језик за мета-означавање**. То значи да он нема фиксан скуп ознака и елемената који би требало да задовоље свачије потребе у свим областима и заувек. Сваки покушај да се направи коначан скуп таквих ознака, осуђен је на неуспех. Уместо тога, XML омогућава програмерима и писцима да узимају потребне елементе онда кад им затребају. Хемичари могу употребљавати елементе који описују молекуле, атоме, везе, реакције и остале ентитете који се срећу у хемији.

Шта XML није?

XML је језик за означавање и ништа више од тога. То треба запамтити. Прича о XML –у се толико надувала да има људи који од XML -а очекују да ради све и свашта, ако треба и кола да опере.

Пре свега, XML није програмски језик. Не постоји компајлер XML-а који чита XML датотеке и даје извршни код. Евентуално бисте могли дефинисати скрипт језик који користи формат XML-а као матични, а интерпретира га неки бинарни програм, али би чак и та примена била необична. XML се може употребити као формат наредби у програмима који нешто раде, као што и традиционални програми могу читати текстуалне конфигурацијске датотеке и предузимати разне акције на основу прочитаног. Заиста нема разлога да конфигурацијска датотека не буде у формату XML, уместо у формату неструктурираног текста. Неки новији програми употребљавају XML конфигурацијске датотеке; али је увек програм тај који предузима акцију, а не сам XML документ.

XML документ не ради ништа, он само постоји. Друго, XML није протокол за мрежни пренос. XML не шаље податке преко мреже, као ни HTML. Подаци послати преко мреже HTTP-ом, FTP-ом, NFS-ом или неким другим протоколом, могу бити кодирани у XML-у; али опет мора постојати неки софтвер изван XML документа који ће послати документ. Најзад, да поменемо пример где приче најчешће сакривају истину, XML није база података. XML-ом не можете заменити ORACLE или MySQL сервер. База података може садржати XML податке, било као VARCHAR или BLOB или неки наменски XML тип података, али сама база података није XML документ. XML податке можете сместити у базу података на серверу и преузети их из у формату XML, али вам за то треба софтвер написан на правом

програмском језику као што су С или Java. Да би сместио XML податке у базу података, софтвер на клијентској страни даље их серверу помоћу установљеног мрежног протокола као што је TCP/IP. Софтвер на серверској страни прима XML податке, расписује их и смешта у базу. Да бисте преузели XML документ из базе података, по правилу ћете морати да употребите неки посреднички програм (енгл. *middleware product*) као што је Enhydra, која ће направити и послати SQL упите бази података, а скуп резултата форматирати као XML пре него што их врати клијенту. Чињеница је да неке базе података интегришу тај софтвер у своје сервере или за обављање те функције обезбеђују софтверске додатке, као што је Oracle-ов сервлет XSQL. У тим сценаријима, XML служи веома добро као свеprisутан преносни формат, независан од платформе. Међутим, он није база података, нити га тако треба употребљавати.

Како ради XML?

За функционалност који пружа XML потребан је парсер. У Internet Explorer-у 5.0 и надаље постоји уграђен парсер, а за парсирање из нпр. Visual Basic-а потребно је поставити референцу на XML парсер. Парсирање, када је XML у питању значи следеће – рашчлањивање текстуалног фајла и прављење структуре која се рекурзивно пуни елементима XML стабла. То значи да парсер изводи следеће операције:

1. Ишчитава препроцесорски део документа (део на почетку документа између ? знакова) да би дошао од информација које се односе на документ, а нису део самог XML стабла.

На пример:

```
<?xml version="1.0" encoding="windows-1252"?>
```

2. Затим се ишчитава први таг у XML структури и записује његово име – ово је елемент највишег нивоа или стартни таг.
3. Затим се записује име елемента.
4. Затим се ишчитавају остали елементи редом да би се одредило која својства има дати елемент структуре и затим се та својства уписују - ако је у питању елемент уписује се његова вредност или у форми уређених парова атрибут = вредност ако је у питању атрибут.
5. Ако следећи таг након првог нађеног није ознака за затварање ишчитава се следећи таг и он се дефинише као дете тренутног елемента. Онда се парсер враћа на корак 3. Ако је нађени таг ознака за затварање онда је елемент дефинисан.
6. Овај процес се понавља док се не обради читав документ.

У примеру 1 приказан је једноставан XML документ, какав се среће у систему за управљање инвентаром или у бази података складишта. Његови подаци су обележени ознакама (енгл. Tags) и атрибутима који описују боју, величину, број бар-кода, име произвођача, име производа итд.

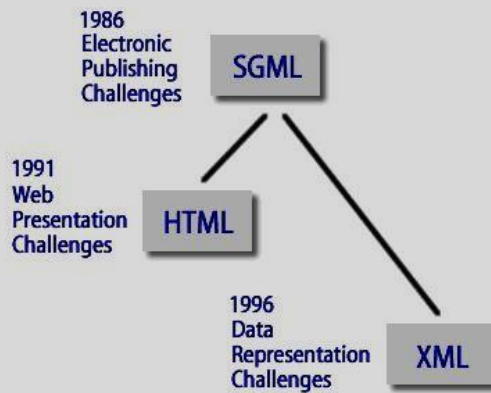
Пример 1

```
XML dokument
<?xml version="1.0"?>
<product barcode="2394287410">
  <manufacturer>Verbatim</manufacturer>
  <name>Datalife MF 2HD</name>
  <quantity>10</quantity>
  <size>3.5" </size>
  <color>crna</color>
<description>diskete</description>
</product>
```

Овај документ је текст, па се може сместити у текстуалну датотеку коју можете уређивати у сваком стандардном програму за обраду текста, као што су BBEdit, jEdit, UltraEdit, Emacs или Vim. За XML вам није потребан специјалан програм за обраду текста. Штавише, сматрам да већина XML едитора опште намене више кошта него што вреди и да их је много теже користити од класичних програма за обраду текста.

Програми који заиста хоће да схвате садржај XML документа – дакле, који с њим имају веће амбиције него да га третирају као било коју другу датотеку – за читање документа употребљавају XML анализатор. Анализатор рашчлањује документ на елементе, атрибуте и остале компоненте. Он апликацији прослеђује садржај XML документа парче по парче. Ако анализатор у било ком тренутку открије да документ крши XML-ова правила о добром обликовању, апликацији ће пријавити грешку и престаће да анализира документ. У неким случајевима, анализатор и после прве пронађене грешке наставља да чита документ, да би могао да открије и пријави и остале грешке. Међутим, након откривања прве грешке у добром обликовању, анализатор увек престаје да прослеђује садржај елемената и атрибута које препознаје.

Еволуција XML-а?



XML је потомак језика SGML (Standardized Generalized Markup Language). Језик од којег је настао SGML седамдесетих година изумили су Charles F. Goldfarb, Ed Mosher и Ray Lorie из IBM-а, развијало га је више стотина људи широм света, све док га 1986. ISO није прихватио као стандард 8879. SGML је решавао многе проблеме које решава XML, и то на исти начин на који их решава XML. То је семантички и структуриран језик за означавање текстуалних докумената. SGML је изузетно моћан, па је био прихваћен у америчкој војсци, државним службама, авионској, космичкој индустрији и другим областима којима је требао ефикасан начин руковања техничким документима дугачким десетине хиљада страница.

Основна намена XML-а

Циљ XML -а је да олакша аутоматску обраду докумената и података. Идеја је да се оствари такво структурирање информација које омогућава и људима да их читају на web-у и разноврсним апликацијама да их аутоматски обрађују.

Пример 2

```

<?xml version="1.0" encoding="ISO-8859-2"?>
<članak>
<naslov>Širak i Žospen najveći favoriti</naslov>
<datum>5. II 2002.</datum>
<tekst>
<mesto>Pariz</mesto>
Manje od tri meseca uoči predsedničkih izbora u Francuskoj, glavni konkurenti, sadašnji
predsednik, degolista Žak Širak i premijer, socijalista, Lionel Žospen, još nisu istakli svoje
kandidature. Međutim, celokupnoj Francuskoj javnosti je jasno da će ta dvojica rivala iz
prethodnih izbora, 1995, ponovo izaći pred glasače 21. aprila ove godine.
</tekst>
</članak>
  
```

Терминологија

Атрибут

је пар име-вредност који се придружује почетној етикети елемента. Имена се одвајају од вредности знаком једнакости и опционим белинама. Вредности су окружене двоструким или једноструким знацима навода. Вредности су ниске карактера и не могу бити структуриране.

Елемент

је секвенција `<PočetnaEtiketa NiskaAtributa> ...sadržaj... </ZavršnaEtiketa>`. Садржај, то јест све оно што се налази између почетне и завршне етикете, може да буде структуриран и може садржати друге елементе. У HTML -у су све етикете унапред дате (јер се он заснива на SGML -у). XML дозвољава да корисник по вољи додаје нове етикете.

Празан елемент

је елемент код кога је садржај између почетне и завршне етикете празан. За такав елемент се може користити посебна почетна етикета `<PočetnaEtiketa NiskaAtributa/>`. На пример, у XHTML -у елементи за нови ред и хоризонталну линију су `<br/>` и `<hr/>` уместо `<br>` и `<hr>`. Не треба мислити да празни елементи не носе информацију, та информација је садржана у атрибутима.

Корени елемент

је први елемент у XML документу који садржи све остале елементе. У Примеру 1, корени елемент је елемент `<članak>`.

XML имена

се користе за именовање елемената, атрибута и других конструката. Она представљају ниску алфанумеричких карактера, под чиме се подразумевају слова алфавета, цифре од 0 до 9, као и слова и цифре из других алфавета. Дозвољено је и коришћење три интерпункцијска карактера: подвлака, цртица и тачка. Име може почети само словом и подвлаком.

Литература

1. Слободна енциклопедија, Wikipedia <http://www.wikipedia.org>
2. Др Милица Вучковић – Увод у XML и XML технологије
3. XML Tutorial <http://www.w3schools.com/xml/>
4. <http://poincare.matf.bg.ac.rs/~cvetana/kurs-xml/xml-sadr.html>