

Standardul SGML si implementarile sale

Asist. Carmen STANCIU

Catedra de Informatica Economica, A.S.E. Bucuresti

Pentru crearea si întretinerea paginilor Web active s-au dezvoltat în ultimii ani o multitudine de standarde, limbaje si instrumente. Unele dintre ele, care au fost mai vechi, s-au pastrat, altele s-au îmbunatatit, iar altele noi au fost create pentru a raspunde cerintelor din ce în ce mai ridicate ale programatorilor de aplicatii distribuite pe Internet. Articolul de fata încearca sa acopere o arie cât mai întinsa implementarile standardului SGML, fara a-si propune sa faca o descriere exhaustiva a acestora.

Cuvinte cheie: SGML, XML, HTML, VRLM.

SGML (Standard Generalized Markup Language), aparut în 1969, este un limbaj de marcare care își are originea în tehnicile editoriale si tipografice traditionale.

În 1986 a fost adoptat ca standard ISO prin versiunea elaborata de IBM, numita **GML** (Generalized Markup Language).

Chiar daca a aparut în urma cu 30 de ani si se desfasoara într-o lume a informaticii foarte dinamica, acest standard este si astazi în plina glorie si chiar dezvoltare, prin implementarile sale moderne HTML, XML.

Principalele caracteristici ale SGML-ului sunt:

- **deschiderea** - adica independenta de orice platforma si produs software de procesare a documentelor;
- **independenta de date** – se bazeaza pe “declaratie SGML” si utilizeaza un set de mapare, translatore a seturilor de caractere;
- **caracterul descriptiv** – realizeaza doar numirea si delimitarea partilor importante ale continutului unui document;
- **tipul de document** – este descris formal si riguros într-o definitie numita *DTD* (*Document Type Definition*).

SGML este de fapt un meta-limbaj, adica un limbaj cu care se descriu alte limbaje, o specificatie a modului de specificare.

Un document SGML este compus din mai multe parti, dintre care cele mai importante sunt:

- **prologul** – cuprinde declaratia SGML si definitia tipului de date (DND);

- **instanta documentului** – reprezinta continutul propriu-zis, marcat conform tipului definit în DND si poate contine, text, marcare si eventual declaratii de entitati.

Pe piata exista o serie de aplicatii care lucreaza cu documente SGML, cum ar fi: parser-e, programe destinate formatarei textului pentru a fi scos la imprimanta pe baza unor specificatii clare ale fiecarui element, sisteme de baze de date orientate text, sisteme hypertext etc.

În continuare vor fi prezentate implementarile actuale ale SGML-ului, sunt sub forma HTML, XML, DHTML, VRLM, precum si compilatoarele HTML.

1. HTML 4.0

HTML 4.0 este noua versiune emisa de W3C (World Wide Web Consortium) ca fiind un înlocuitor al HTML 3.2 si care îmbunatateste HTML-ul cu mai multe optiuni multimedia, limbaje de scripting, facilitati de tiparire, simboluri matematice, internationalizare.

Aceasta versiune are la baza câteva principii fundamentale, cum ar fi:

- **interoperabilitate** – documentele HTML trebuie sa functioneze fara probleme pe orice browser si orice platforma;

- **globalizare** – documentele pot fi scrise în orice limba din lume datorita adoptarii

standardului ISO/IEC 10646 ca set de caractere implicit pentru HTML si încorporarea RFC 2070 care trateaza globalizarea HTML-ului;

- *accesibilitate* – se adreseaza persoanelor cu disfunctionalitati fizice, si includ stule-sheet-uri în locul tabelelor de formatare a paginii, alternative textuale de descriere a imaginilor (pentru nevazatori), etichete active pentru câmpurile din formulare, crearea de grupuri ierarhice pentru câmpurile formularelor;

- *tabele* – noul model de tabel, bazat pe RFC 1942, permite un control mai larg asupra structurii si formatarii tabelelor;

- *documente compuse* – prin elementul OBJECT, alaturi de IMG si APPLET, ofera un mecanism standard pentru înglobarea obiectelor multimedia generice si a aplicatiilor în documentele HTML ;

- *style-sheet-uri* – simplifica scrierea codului HTML, oferind controlul complet asupra documentelor;

- *scripting* – permite crearea paginilor Web dinamice si a formularelor inteligente.

În HTML 4.0 au fost introduse o serie de tag-uri noi, dupa cum urmeaza:

- `<Q>...</Q>` - pentru fragmente scurte de text, care nu necesita împartirea în paragrafe;

- `<INS>...</INS>` - marcheaza portiunea de pagina Web care a fost adaugata dupa o anumita data, specificata prin parametrul *datetime*;

- `...` - idem, dar pentru stergere;

- `<COLGROUP>` - specifica un grup de coloane cu caracteristici comune, dintr-un tabel;

- `<LEGEND>` - permit atasarea unei explicatii fiecarui grup definit de FIELDSET;

- `<FIELDSET>` - permite gruparea controalelor unui formular pe criterii tematice;

- `<BUTTON> ...</BUTTON>` - este asemanator unui control INPUT de tip BUTTON.

Evenimentele, oficial acceptate în versiunea 4.0, reprezinta actiuni pe care le poate desfasura utilizatorul asupra unui element al paginii Web, pot declansa executarea unui script si se ataseaza imaginilor, link-urilor, form-urilor. Exemple de evenimente implementate în HTML 4.0 sunt urmatoarele: *onchange, onsubmit, onreset, onkeypress, onkeydown, onkeyup, onload, onunload, onclick, ondbclick, onmouseover, onmouseup, onmousemove, onmouseout, onfocus, onblur*.

2. DHTML

DHTML (Dynamic HTML) reprezinta o combinatie între HTML, stilul paginilor în cascada (CSS –Cascading Style Sheets) si scripturi, toate unite prin DOM, în idea de a obtine un dinamism accentuat al paginilor Web, fara applet-uri Java si controale ActiveX.

Prin DOM (Document Object Model), DHTML furnizeaza un model si o interfata de programare a aplicatiilor care permit manipularea mai usoara a limbajului HTML si CSS.

Pentru vizualizarea paginilor DHTML este necesar Netscape 4.01 cel putin sau Internet Explorer 4.

CSS, tehnica aparuta cu Internet Explorer 3 si aprobata de catre W3C, ofera patru cai de încorporare a componentelor de stil în paginile Web: *descriere de stil externa, importul unei descrieri de stil externe, tag-uri pereche pentru modificare si attribute de stil între tag-urile unui document* .

DOM, modelul documentelor de obiecte, permite fiecarui element dintr-o pagina de Web sa fie inserat, ascuns, deplasat pe ecran, fara a reîncarca întreaga pagina.

În continuare vor fi prezentate pe scurt câteva dintre elementele noi introduse de DHTML:

- un nou obiect *event*, care urmareste succesiunea evenimentelor; proprietatile lui permit identificarea evenimentului, starea tastaturii, pozitia mouse-ului, starea butoanelor acestuia; exemplu: *onclick*, *onmouseover*, *onload* etc.
 - exista mai mult de 60 de **attribute de stil** care permit controlul paginii de Web, iar tag-urile HTML traditionale au devenit obiecte DHTML; se utilizeaza JavaScript sau VBScript pentru manipularea metodelor si proprietatilor metodelor; evenimentele permit selectarea optiunii pentru declansarea codului care raspunde actiunilor utilizatorului sau evenimentelor de pe Web; fiecare tag DHTML are si un corespondent de script.
 - **stiluri dinamice**, care se schimba în functie de evenimentele sistem sau utilizator (proprietatea *className*);
 - **pozitionare dinamica** printr-un sistem de coordonate tridimensional pentru controlul afisarii elementelor HTML; modificarea dimensiunii unui element HTML se realizeaza prin attribute dinamice, cum ar fi: *top*, *left*, *width*, *height*, *zIndex*;
 - **efecte multimedia** atât prin *filtre* (transforma imaginile si textele), cât si prin *tranzitii* (controleaza felul în care doua imagini îsi schimba pozitiile);
 - **conectarea datelor** – accelereaza si îmbunatateste afisarea, actualizarea si manipularea datelor pe Internet; obiectele sursa de date, drivere care fac legatura cu datele din paginile de Web, realizeaza trei functii: *furnizeaza date în mod sincron* (afiseaza pagina de pe server mai repede), *executa pe calculatorul clientului a operatiilor de filtrare si stocare*, *actualizeaza baza de date de pe server cu ajutorul unui formular dintr-un browser*.
- În concluzie, DHTML ofera dezvoltatorilor de pagini Web posibilitatea de a obtine viteze mai mari, interactivitate mai buna, iar utilizatorilor de Web posibilitati mai bogate de navigare.

3. XML

Explozia de aplicatii bazate pe Web i-a determinat pe dezvoltatori sa recunoasca slaba flexibilitate a structurii de baza, adica a limbajului HTML, XML câstigând teren, ca liant universal al tuturor elementelor dispartate.

XML (eXtensible Markup Language) este unul dintre standardele emise de W3C pentru descrierea structurilor de date, menit sa transforme Internetul într-un loc propice comert electronic.

La fel ca si HTML-ul, XML este o varianta de SGML, standardul pentru descrierea documentelor si poate fi folosit pentru definirea altor limbaje, fiind si un meta-limbaj.

XML are o sintaxa de auto-descriere, este de câteva ori mai complex decât HTML-ul si mai mic decât SGML. HTML se ocupa cu prezentarea informatiei, iar XML-ul descrie datele (nu formatarea). Din pacate XML nu este azi compatibil cu HTML, deoarece XML foloseste o sintaxa usor diferita de HTML si este mai riguros în privinta regulilor de sintaxa. Pentru a deveni compatibile cu XML, documentele HTML vor avea nevoie de adaptari minore. Aceasta convertire a documentelor HTML la XML ar avea ca avantaje:

- o structurare mai puternica a informatiei, ceea ce ar ajuta motoarele de cautare sa returneze rezultate mai consistente,
- introducerea legaturilor bidirectionale si a celor stocate extern, care usureaza întretinerea si ofera stabilitate mai mare aplicatiilor Web.

XML are mai multe parti, dintre care cele mai importante sunt doar trei:

- **DTD** (Document Type Definition) - limbajul de structura, care stabileste gramatica documentului, elementele de pagina, attributele lor, legaturile dintre ele;
- **XML** (eXtensible Style Language)- versiunea XML pentru listele de stiluri;

- **XLL** (eXtensible Link Language) - sistem de manevrare a legaturilor, care depaseste "HREF"-ul din HTML.

Jon Bosac, care conduce grupul XML din cadrul W3C, considera ca aplicatiile care vor determina acceptarea XML-urile sunt de 4 categorii:

- aplicatii care cer clientului Web medierea între doua sau mai multe baze de date;

- aplicatii care încearca sa distribuie clientilor Web o parte semnificativa a procesarilor care ar reveni server-ului Web;

- aplicatii care cer clientului Web sa prezinte la utilizatori diferiti viziuni diferite ale acelorasi date;

- aplicatii în care agentii Web inteligenti încearca sa descopere informatii relevante în functie de interesele specifice ale unor utilizatori individuali.

XML se pare, dupa parerea specialistilor, ca va avea cea mai mare aplicabilitate în domeniul comertului electronic. Astfel, comunitatile de **comert electronic** care l-au acceptat sunt:

- consorțiul format din AT&T, HP, IBM, Oracle si Open Market Inc., au lansat protocolul OTP (Open Trading Protocol), care foloseste XML ca transport de baza pentru conectarea datelor cu diferite aplicatii, cum ar fi autentificarea produselor, ordine de plata si de comanda, completarea comenzii;

- grupul XML/EDI sustine înglobarea XML în aplicatiile comertului electronic;

- consorțiul Commerce-Net sustine folosirea XML în cataloagele bazate pe Internet si aplicatiile de afaceri.

4. VRLM

VRLM (Virtual Reality Modeling Language) este limbajul care descrie mediul 3D pe Web, adica reprezentarea si interactiunea obiectelor virtuale prin intermediul calculatoarelor.

Ajuns la versiunea 2.0, VRLM este standardul industrial de facto pentru repre-

zentarea distribuita si interactiva a lumilor compuse din elemente 3D, multimedia si logica.

Pentru crearea unei lumi virtuale, la dispozitia proiectantului *stau elemente geometrice obisnuite* (cum ar fi sfera, con, cilindru, cub) si *structuri complexe* (ca "ceata", "teren accidentat", "fundal" etc). Implementarea interactiunii dintre obiecte se face prin *structuri predefinite ale limbajului* (ca detectori de coliziune, senzori, senzori temporali) si *structuri definite de proiectantul lumii virtuale*, prin sectiuni de cod speciale si actionate prin intermediul unor evenimente primite din partea mediului, de la senzorii si detectorii implantati.

Informatiile în VRLM sunt o descriere geometrica a tuturor obiectelor dintr-o *scena*, scena care poate fi modificata dupa instructiunile incluse în limbaj. Scena VR este sub forma unui *graf* ce contine *noduri terminale* sau *neterminale*, adica *noduri de grupare si noduri fii*. Se creeaza astfel o ierarhie de noduri:

- fiecare nod de grupare – definind un nou sistem de coordonate pentru toti fii sai;

- fiecare nod fiu – referindu-se la sistemul de coordonare al tatalui.

Nodul este caracterizat prin: *tip, nume, variabile si evenimente asociate*. Un nod poate receptiona un numar de *evenimente* care îi vor modifica anumite valori ale câmpurilor asociate evenimentelor, sau poate genera evenimente care sa modifice starea nodului.

Limbajul permite construirea si controlarea *fluxului de evenimente*, oferind posibilitatea de a construi modele foarte apropiate de realitate, guvernate de legi cauza- efect.

O alta facilitate importanta este *generarea dinamica a rutelor*, prin scripturi din interiorul browser-ului, creând astfel asociieri ce se pot automodela dinamic.

Senzorii sunt noduri care genereaza evenimente bazate pe anumiti stimuli

externi din punct de vedere al limbajului, cum ar fi miscarea mouse-ului sau trecerea timpului. Exista senzori de vizibilitate, de proximitate, de coliziune, de timp etc.

O calitate importanta a limbajului VRLM 2.0 este design-ul deschis si extensibil, implementat prin utilizarea *prototipurilor*. Prototipizarea este mecanismul prin care un set de noduri existente poate fi extins conceptual prin crearea unei specificatii pentru un nou nod. Prototipul este alcatuit din numele noului tip de nod, declaratia de prototip, definirea propriu-zisa a prototipului, care contine o lista de noduri, rute si alte prototipuri.

Pentru *vizualizarea* documentelor VRLM este necesar un browser care suporta în întregime standardul VRLM (SGI, Live 3D, World View) sau unul obisnuit îmbunatatit cu plug-in-uri corespunzatoare. În anexa 4 este prezentat un exemplu de VRLM.

În prezent, VWWW (Virtual WWW) este considerat util datorita:

- efectelor speciale imposibil de atins într-o pagina Web clasica, care fac astfel pagina VRLM mai atractiva;
- functionalitatii paginilor 3D, baza viitoarelor magazine virtuale, care vor constitui lumea comertului electronic.

Folosind VRLM se pot crea reprezentari tridimensionale ale unor scene complexe, pornind de la simple prezentari comerciale interactive si pâna la virtualizari substantiale conceptuale.

5. Editoare HTML

Realizarea paginilor HTML poate deveni extrem de usoara daca se utilizeaza un editor specializat. În acest sens, în continuarea se va face o scurta descriere a celor mai raspândite editoare HTML utilizate astazi.

- **FrontPage** – *Microsoft* – este unul dintre cele mai complete produse pentru dezvoltarea paginilor HTML, care prezinta suport si pentru tehnologiile viitoare, ca DHTML, CDF (Channel Definition Files –

permite transmiterea automata a continuturilor actualizate din site-ul Web celor interesati), CSS (Cascading Style Sheets) si themes (tehnica templet-urilor de design – adica seturi de elemente optice de reprezentare). Este un pachet modular care contine Explorer, Editor, Personal Web Server, Server Administration si Image Composer. Pentru crearea formularelor transmise prin e-mail, utilizeaza o rutina proprie, usor de configurat, menita sa înlocuiasca scriptul CGI. Versiunea 98 aduce elemente noi si în domeniul elementelor active, cum ar fi un timer pentru schimbarea imaginilor si “hover-buttons” (butoane care-si modifica aspectul la trecerea mouse-ului deasupra lor). Ca dezavantaj ar fi de semnalat existenta unor elemente care nu functioneaza sub Netscape, cum ar fi tabelele create cu FrontPage 98.

- **PageMill 3.0** – *Adobe* – desi nu ofera facilitati avansate precum CCS (Cascade Style Sheets), DHTML sau scripting, permite crearea paginilor de Web si a unor site-uri atragatoare. Oferă usurinta în utilizare si nu necesita cunostinte avansate de tehnici HTML. Intefata utilizator este aerisita, alcatuita dintr-o fereastră în care se face editarea si previzualizarea, iar în partea stânga a ecranului se afla doua ferestre de lucru, una cu proprietati si una pe post de masa de lucru. Pentru vizualizarea obiectelor direct în pagina de lucru accepta aplicatii plug-ins, compatibile cu Netscape 2.0. Pentru realizarea paginilor de *comert electronic* exista posibilitatea de a introduce butoane de la **ICentral ShopSite Express**, pentru siguranta mai mare a informatiilor. Permite vizualizarea applet-urilor Java si a controalelor ActiveX în timpul editarii. Previzualizarea paginilor create se face cu navigatorul inclus, cu Internet Explorer-ul aflat pe CD-ul de instalare sau cu orice alt browser. Include componenta **SiteMill**, destinata managementului fisierelor unui site, care tine doua copii ale fisierelor (una pe serverul de

Web, iar alta pe calculatorul dezvoltatorului aplicatiei) si care permite schimbarea locatiei fisierelor sau redenumirea lor, determinând schimbarea automata a tuturor legaturilor din întregul site. După construirea întregului site se poate aplica o operatie de upload, care va pune printr-o legatura ftp toate fisierele pe server.

- **Visual Page** –*Symantec* - se aseamana cu Adobe PageMill ca nivel de lucru, având ca avantaj viteza mare de rulare, necesar mai mic de resurse, aranjarea codului HTML în fisierele generate. Contine pachetul de fisiere **Designer's Toolkit** cu sabloane de pagini HTML, pornind de la prezentari si pâna la applet-uri Java si pagini cu formulare pentru aproape orice domeniu. Include un navigator care permite previzualizarea paginilor create. Interfata utilizator este foarte degajata, de pe ecran lipsind obisnuitele ferestre de lucru. Tabelele pot fi create direct cu sau importate din Excel. Permite crearea unui întreg site printr-un sistem de management al fisierelor în directoare locale si prin asigurarea transferului între server si calculatorul local. O alta facilitate consta în aducerea unei întregi pagini HTML de pe Internet, împreuna cu toate componentele sale. Pagina adusa este deschisa direct în editor, iar fisierele depuse automat într-un director specificat în prealabil de catre utilizator.

- **Fusion 3.0** – *NetObjects* – este un editor pentru site care se doresc a fi stilizate, oferind un mod usor de realizare si întretinere a unui site. Permite upgrade-uri spre mai multe pagini în acelasi timp. Ca dezavantaj trebuie amintit faptul ca nu permite modificarea codului sursa, decât în finalul realizarii site-ului. Oferă posibilitatea exportarii site-ului în format cu grafica, pentru legaturile rapide, sau în format text pentru legaturile mai lente.

- **Homesite 3.0** –*Allaire* – este unul dintre cele apreciate editoare de catre profesionalistii paginilor Web, datorita controlului pe care îl ofera utilizatorului

asupra codului HTML. Dispune de toate tehnologiile, începând de la JavaScript si pâna la DHTML. Contine si un editor bazat pe text.

- **WebEdit Pro 3.1** - *Luckman* – este un editor bazat pe text, singurul care previzualizeaza frame-urile si pagina creata în browser-ul principal si care se adreseaza utilizatorilor cu experienta în domeniu. Contine wizard-uri care se ocupa cu realizarea de tabele, forme, link-uri.

- **HotDog Profesional 5** - *Sausage* – este un editor care se adreseaza celor care doresc sa scrie cod HTML complicat, prin extensii CSS, Java, Real Video etc. Templet-urile sunt destul de slabe, interfata destul de confuza. Contine wizard-uri pentru crearea de tabele, forme, frame-uri. Pentru managementul site-ului contine Website Wizard, care permite crearea unei oglinzi a site-ului Web pe hard disk.

- **Dreamweaver 1.2a** -*Macromedia* – unul dintre cele mai puternice pachete care permite crearea paginilor Web cu ajutorul stylesheet-urilor si a DHTML-ului. Contine si un editor bazat pe text de la Homesite. Include toate tehnologiile prezente si viitoare, cum ar fi DHTML, CSS, applet-uri Java, controale ActiveX, plugin-uri Netscape etc.

- **HTML Editor++ 98 5.1** - *Coffeecup* – este un editor bazat pe text care este apreciat în special de amatori datorita numarului mare de scripturi existente în biblioteca încorporata, scripturi ca JavaScript, VBScript, CGI Script. Prezinta de asemenea un suport foarte bun pentru stiluri (stylesheet). Permite inserarea în paginile sale de elemente multimedia. Este la un pret de 40\$, situându-se astfel la un pret performata – pret destul de bun. Din testele efectuate de CHIP s-a ajuns la concluzia ca cel mai bun editor ar fi Allaire Homesite 3.0 si Suasage HotDog Profesional 5 (tabelul 1).

Tabelul 1

	HotDog Professional 5	Home site 3.0	Visual Page 2.0	Front Page 98	Dream weaver 1.2a	HTML Editor++ 98 5.1	Web Edit Pro 3.1	Fusion 3.0	Page Mill 3.0
Producator	Sausage	Allaire	Symantec	Microsoft	Macromedia	Cofeecup	Luckman	NetObjects	Iris
Pret	80\$	100\$	90\$	170\$	300\$	40\$	139\$	285\$	120\$
Ergonomie	5	5	5	4	5	4	4	4	4
Funcționalitate	5	5	4	5	5	5	4	4	3
Mediu de lucru	5	5	5	5	4	5	4	3	3

Bibliografie

- /BLAG98/ BLAGA, A., Dynamic HTML, PC Report, mai, 1998;
- /DOBS97/ DOBSON, R., HTML Dinamic pe înțelesul tuturor, BYTE România, nov, 1997;
- /FOTA97/ FOTACHE, C., Standardul HTML 4.0, PC Report, oct., 1997;
- /GILL97/ GILLMOR, S., Instrumente pentru aplicatii Web, BYTE România, dec., 1997;
- /LANG98/ LANG, M., s.a., Fabrica de pagini web, CHIP, iunie, 1998
- /MATE98/ MATEI, C., Adobe PageMill 3.0, PC Report, apr., 1998
- /NAST98/ NASTASE, F., Internet, World Wide Web, JavaScript, HTML, Java, Editura Economica, Bucuresti, 1998;
- /PANO98/ PANONE, J., Noi lianti pentru aplicatii Web, BYTE România, iul., 1998
- /ȘÂRB97/ ȘÂRBU, M., s.a., SGML – Din nou în actualitate, PC Report, iul., 1998;
- /SION98/ SION, R., Introducere în VRML 2.0, PC Report, mai, 1998;
- /SWOD97/ SWOBODA, N., HTML – based Interfaces, cs.indiana.edu, feb., 1997;
- /VAID98/ VAIDA, C., Publicatii Web, CHIP, nov., 1998;
- /VULC97/ VULCU, C., VRML, cele trei dimensiuni si Internetul, Planeta Internet, iul-aug, 1997.