

STANDARDUL SGML PENTRU DIFUZAREA CUNOȘTINTELOR ȘTIINȚIFICE

Mat. Brândușa Trușcă

Institutul de Cercetări în Informatică

Rezumat: Acest articol trece în revistă câteva aspecte ale normei SGML (Limbajul Standard Generalizat de Marcare) și alte proiecte de norme pentru descrierea documentelor pe calculator: DSSSL pentru punerea în pagină și SPDL pentru vizualizare.

Cuvinte cheie: SGML, DSSSL, SPDL, standard ISO.

1. Tendințe existente pe plan mondial

Editarea în lumea științei suferă din cauză că anumite domenii cheie și anumite căi pe care se poate merge, de la foi volante la o revistă serioasă, sunt lungi și sinuoase. Spunând că drumul unui articol către un editor reprezintă un adevărat tunel, nu există dimensiuni și perspectivă în această privință.

Încercările de a îmbunătăți această situație au condus la necesitatea folosirii calculatoarelor și aceasta datorită faptului că anumiți oameni de știință comunică între ei pînă la o limită maximă a capacității lor; puține lucruri s-au realizat pentru a-i sprijini în munca lor. Oricum, calculatoarele sînt greoi utilizate, atît de autori, cît și de editori, neexistînd o legătură clară între ambele capete ale acestui proces.

1.1. Necesarul unui format standard

Dacă vrei să fi pe locul întîii în lume, trebuie să ai un singur tip de calculator, un singur tip de sistem de operare, un singur pachet software, pentru a putea manipula formulele matematice și figurile, ținînd cont că, în același timp, trebuie păstrată legătura între autori și editori.

Un autor trebuie să pregătească un manuscris lansabil local, care să corespundă rețelei consacrate pînă la editor. Editorul face corecția manuscrisului și-l retransmite prin rețea referenților care îl retransmit cu comentariile aferente. Pe urmă, editorul discută cu autorul, se pun de acord, după care editorul dă manuscrisul spre publicare; publicistul stochează articolul în baza de date, creează sumarul, aranjează prezentarea în pagină și publică revista fără a mai recodifica textul.

În lume există o mare varietate de calculatoare, procesoare și rețele care servesc autorii, editorii, referenții și publicității. Soluția optimă o reprezintă renunțarea la toate sistemele existente, cu excepția unuia singur, datorită costurilor foarte mari investite într-un astfel de sistem.

Singura soluție viabilă este de a accepta o mașină independentă standard, care să fie compatibilă cu hardware-ul existent, cu software-ul procesorului de text și cu cerințele rețelei de transmisie.

În sistemul tradițional, editorul este cel care adnotează manuscrisele cu instrucțiuni detaliate,

referitoare la reprezentarea în pagină, fonte, spațiere, numerotare etc. Această marcă reprezintă primul pas între diferitele formate de manuscris și publicația finală. În cadrul procesării electronice, un limbaj de marcă poate fi definit ca un cadru principal de specificare pe caractere, care emite comenzi pentru fiecare acțiune. De exemplu - \bf - implică faptul că următorul text trebuie tipărit cu aldine.

Există o marcă în cadrul fiecărui procesor: secvența - \bf text \rm - plasată în mijlocul unui rînd "spune" sistemului să pună [textul] în stil aldine și, pe urmă, să scrie continuarea în stil roman. Acest tip de marcă se numește **marcă specifică**.

Marcarea de acest tip poate avea mai multe forme: se poate aplica punctual (ca în exemplu), prin intermediul unui caracter aflat la început de rînd (cum ar fi '.'), prin intermediul unei "pagini tip" la un început de paragraf sau cu ajutorul unor macroinstrucțiuni aflate la început de manuscris; pot exista combinații între aceste variante. Toate aceste forme sînt restrictive la un singur limbaj de procesor. Dezvoltarea lor a condus la necesitatea separării nete între conținutul și forma unui manuscris: în timpul scrierii, autorul nu trebuie să aibă grijă de forma de prezentare.

Deci, ce se cere este un sistem de marcă generală, adică un sistem care permite autorului să indice scopul logic al unui element textual, fără a se ține seama de apariția sa fizică. De exemplu, în loc de - \bf text \rm -, autorul ar spune - \high light text - (text subliniat) și ar lăsa sistemul să decidă cum să se scrie acest text.

Identificatori ca - \high light - sînt denumiți fișiere de atribute cu nume unice pe care le putem aplica pe paragrafe de text - tag-uri. Un astfel de limbaj de marcă generică scutește autorul să aibă grijă de prezentare și face ca documentul să devină mai puțin dependent de aparatură, din moment ce tag-urile pot produce aldine pe un sistem și pot scrie un text italic pe un altul. Un astfel de limbaj reprezintă o extensie naturală a procesoarelor de astăzi. Din păcate, codurile de marcă sînt încă prea dependente de pachetul de procesare de texte, existent astăzi.

În scopul "conexiunii" unui astfel de limbaj de marcă cu un limbaj de procesare de text este necesar un **program de aplicație**, care să traducă simbolurile de marcă în simboluri care să fie recunoscute de limbajul specific fiecărui procesor. Figura 1 indică relația dintre un "candidat" la un **meta-limbaj**, SGML, și diferite limbaje de marcă a procesorului de cuvînt precum și rolul de program clar de aplicație. Direcția este într-un singur sens: în urma generării unui manuscris marcat, autorul poate "ajunge" la unul sau mai multe limbaje de procesare prin utilizarea unui program de aplicație.

Un scenariu pentru marcarea unei publicații științifice este următorul: autorul pregătește manuscrisul conform satisfacției personale pe propriul

procesor și odată ce a terminat textul, el îl modifică conform unui limbaj de marcare universal acceptat (Tabelul 1). Manuscrisul este apoi procesat și înaintat, prin poștă pe dischetă, sau printr-o rețea de transmisie, editorului de revistă. Pretipărirea pentru o distribuție locală limitată, în avansul publicației propriu zise, se realizează prin intermediul software-ului de aplicație, aferent procesorului autorului, adică al institutului sau universității de care aparține.

Calculatorul editorului citește manuscrisul și folosește software-ul de aplicație, aferent procesorului său, pentru a obține o copie lizibilă, adică fără a mai fi nevoie de tag-uri. Apoi, el poate să adauge propriile comentarii versiunii marcate și să transmită manuscrisul revizuit unuia dintre referenți care folosește la rândul său, propriul program de aplicație pentru a obține și el o copie lizibilă. Comentariile sale sînt adăugate manuscrisului original marcat, și acesta este retransmis, via editor, autorului.

O dată procesul de editare completat, manuscrisul marcat în forma finală este transmis publicistului. Acesta utilizează structura din figura 1. Un program de aplicație conduce la "arhivarea" manuscrisului, acesta fiind clasificat și stocat într-o formă permanentă. Un alt program copiază sumarul articolului - dacă el există - într-un fișier de sumare. Un al treilea traduce manuscrisul pentru un procesor local, în limbajul acestuia, pentru a obține o versiune tipărită, iar un al patrulea program generează o copie de prezentare a revistei.

În cazul publicării unei cărți, există și alte programe, pentru a genera tabela de index etc.

1.2 Cerințe aferente unui astfel de format

Pentru obținerea unui format standard sînt necesare următoarele:

- a) să fie compatibil cu toate sistemele de calcul existente;
- b) transferul să se realizeze prin intermediul unei rețele eterogene;
- c) să manipuleze relații matematice și grafice;
- d) să fie independent de formatul unei reviste.

2. Definirea unui model conceptual în concordanță cu standardele ISO-SGML

2.1. Ce este SGML ?

SGML - Limbaj Standard Generalizat de Marcare - este o normă internațională ISO (Organizația Internațională de Standarde) pentru descrierea documentelor (ISO 8879)[1]. El este proiectat special

¹⁾Diferența dintre procesarea cuvîntului și procesarea textului îl reprezintă faptul că zețarul afișează rezultatele pe ecran și permite interacțiunea, în timp ce în al doilea caz nu avem interacțiune.

pentru a permite transferul de texte, în primul rînd în domeniul publicațiilor științifice și al publicațiilor în general, dar are aplicații și în birotică și inginerie. Documentele au o structură sigură, care poate fi analizată de calculator și ușor înțeleasă de factorul uman. Prin text se înțelege o parte a unui document și faptul că el este disponibil sub formă electronică pe calculator sau pe orice periferic al acestuia, cum ar fi banda magnetică, disc, dischetă sau memorie. Un document electronic este stocat într-un fișier pe calculator. Prin permutare de texte și document se înțelege transferul de texte sau documente de la un sistem de calcul la altul (prin intermediul rețelei, dischetei sau benzii) cu aceeași aplicație. Mai poate să însemne deplasarea unui document la o altă aplicație, cum ar fi de la marcator la o bază de date. Marcarea textului și bazei de date trebuie să fie pe același calculator [2].

Prelucrarea textelor reprezintă o serie de procese efectuate asupra unui document, cum ar fi: pregătirea documentului, editarea, marcarea și publicarea. **Marcatorul de text și procesorul de cuvinte**¹⁾ sînt programe de calculator care grupează cuvintele în paragrafe, selectează fontele, calculează poziția cuvintelor în pagină, realizează automat marcarea paginii și culegerea literelor.

Istoric vorbind, marcarea este procesul prin care un redactor introduce mărcile într-un manuscris pentru a indica tehnoredactorului cum trebuie compus manuscrisul (*lumina paginii*). În majoritatea cazurilor este vorba de comenzi scrise de mîna, de exemplu: "Compuneți titlul cu Helvetica-Medium-Italic, corp 12 și între linii 14 puncte, pe o lărgime de 22 picas, bolduite, 1/2 cadran stînga și 0 la dreapta".

O dată cu apariția calculatoarelor, s-au putut codifica aceste comenzi cu ajutorul unui sistem de "encodage" special. Fiecare fotocompoziție are limbajul său propriu, specific și unic, pe care-l numim limbaj de marcare prin analogie cu sistemul manual. În acest timp, cînd autorii și beneficiarii editurilor încep să aibă acces la documentele lor, situația se va complica. Marcarea documentului nu se poate face decît dacă se cunoaște formatul de compoziție utilizat de editor. Dacă se utilizează mărcile unui sistem dat, sînt șanse ca acestea să nu fie compatibile cu cele utilizate de un alt sistem de compoziție.

Situația nu se va îmbunătăți cînd încep să se folosească calculatoarele pentru pregătirea documentelor. În cazul sistemelor de fotocompoziție, documentele erau marcate cu **comenzi specifice** de nivel jos ca: "treci la linia următoare", "centrează textul următor", "treci la pagina următoare".

Pentru a ameliora compatibilitatea între sistemele de marcare existente, este necesară crearea unui limbaj de marcare standard. Acest limbaj de marcare comun este de tip generic. O marcare generică descrie funcția logică a diferitelor elemente ale unui text în contextul

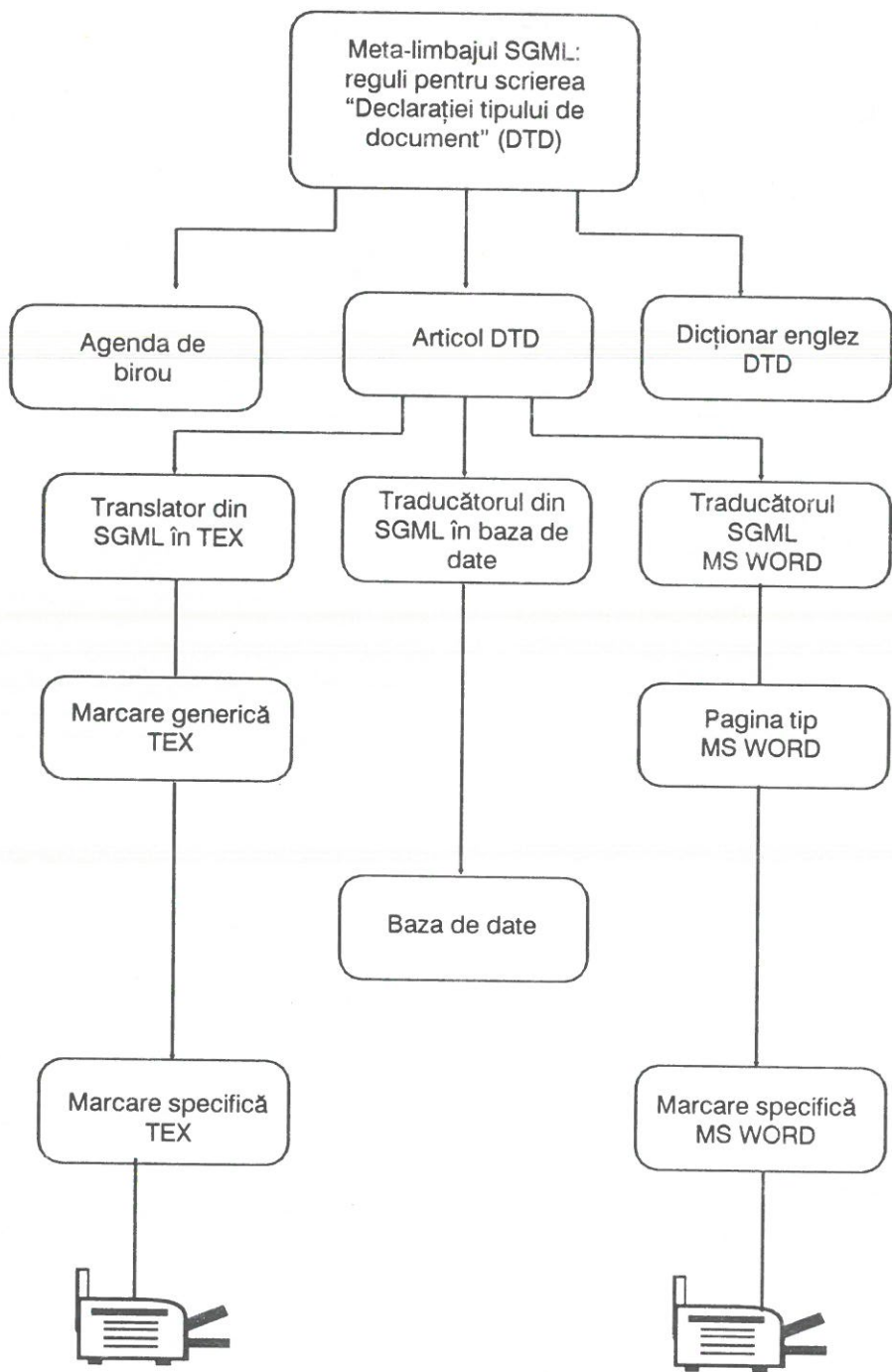


Figura 1 SGML și aplicațiile sale

structurii documentului. De exemplu, se indică prin marcatori paragrafele, titlurile, notele de subsol. Maniera în care un element va fi reprezentat pe pagină nu este specificată în text, ceea ce dă posibilitatea trecerii de la un format la altul.

LATEX este un exemplu de limbaj de marcare generic. Tabelul 1 arată o comparație între mai multe limbaje de marcare generice: SGML cu DTD al Asociației Publiciștilor Americani - American Association of Publishers[4], Book Master[5] al IBM, LATEX [6] și VAX Document [7]. După cum se observă, diferențele între comenzi pentru numeroasele intrări sînt triviale (de exemplu: `<FN>`, `:FN.`, `\footnote` și `<FOOTNOTE>` sînt folosite pentru a marca o notă de subsol). Ceea ce înseamnă că o conversie de la un sistem la altul trebuie să fie posibilă cu condiția ca limbajul să descrie aceeași structură generică.

Printre alte avantaje ale marcării generice, semnalăm faptul că documentul devine independent de sistemul de calcul, de punerea în pagină și că stilul este definit într-un singur loc, în foaia de stil.

Activitățile inițiale în domeniul marcării generice au fost coordonate de *Graphic Communications Association* (GCA), un grup profesional care a creat limbajul de marcare generic `<Gen Code>` pentru sistemele de compoziție. Gen Code, LATEX, Book Master și VAX Document sînt toate exemple de "modele de stare" (*state models*): fiecare schimbare de stare a sistemului e provocată de un cod format dintr-un șir de caractere. De exemplu, comanda **bold** schimbă starea formatului, imprimînd caractere groase, pînă ce întîlnește o altă comandă care repune formatul în starea inițială. Așa cum arată tabelul 1, codurile folosite sînt proprii unui limbaj particular.

Dacă considerăm ansamblul structurilor posibile ca o arborescență, atunci limbajul de marcare descrie un document ca o limbă străină, în care regulile de producție sînt descrise printr-o gramatică, așa zisă de *context-free* în clasificarea lui N. Chomsky (regulile de producție nu depind de contextul utilizat). Definind gramatica, se poate stabili clasa structurilor posibile de documente. Astfel, SGML nu este un limbaj de marcare în același sens ca TEX, el este un limbaj specific sau un meta-limbaj, care permite definirea regulilor pentru crearea unui număr infinit de limbaje de marcare. Un principiu fundamental în concepția SGML este **independența** limbajului de marcare față de sistemul de calcul sau de sistemul de compoziție. Documentele marcate cu SGML pot fi schimbate între diferitele sisteme de calcul. O consecință a folosirii marcării generice, este **inexistența** marcatorelor pentru punerea în pagină, în SGML, pentru că acesta nu conține structura logică de document. Nu există o comandă care să indice că dorim "pagină nouă", "o linie nouă" sau trasarea unei linii de subliniere. Aceste elemente de paginație sînt caracteristice anumitor componente; astfel, un titlu de nivelul I poate să înceapă la pagină nouă, și o linie subțire poate separa titlul de corpul documentului.

2.1.1. Marcatori SGML

Marcatorii SGML sînt introduși într-un document specificînd pentru fiecare element poziția în arborescența descrisă formal prin "Definirea de tip document" (DTD). Un document este o **realizare** a unei clase de documente care au toate aceeași structură. O dată ce un element a fost reperat, marcarea se face cu un **marcator de început** (*start-tag*) și cu un **marcator de sfîrșit** (*end-tag*). Sub forma sa cea mai simplă, o marcare SGML constă din *identificator generic*; exemplu:

`<titlu>`

În acest exemplu, "titlu" este identificatorul generic, el poate fi scris cu majuscule `<TITLU>`, minuscule `<titlu>` sau amestecat `<Titlu>`, în toate cazurile fiind vorba de același identificator.

Un marcator se poate găsi oriunde pe linie, nu neapărat la început. Un element poate avea un conținut vid (exemplu: `<data>` folosită singură înseamnă că data va fi furnizată de sistem) sau poate delimita o parte a textului (exemplu: `<P>` pentru paragraf), care are date de tip caracter în conținut.

Un marcator de sfîrșit are același identificator generic cu cel de început, numai că este precedat de un delimitator al mărcii de închidere `</>`; exemplu: `</TITLU>`.

Figura 2 ne arată componentele folosite la marcare.

În forma sa generală, marcarea SGML constă într-un delimitator de deschidere, un identificator generic, o **listă de atribute**, un delimitator de închidere și un text aparținînd marcatorului. Exemplu:

`<Titlu format=standard statut=public> Acest document...`

Există două sintaxe posibile pentru a specifica atributele unei mărci. Forma curentă este constituită din identificatorul de atribut, un `<=>` și o valoare fără blăncuri. Dar se pot folosi și valorile atributelor singure. Exemplu:

`<Titlu standard public> Acest document...`

Dacă valoarea unui atribut conține și alte caractere decît cele alfanumerice, ea trebuie delimitată prin `" "`, astfel:

`<Titlu format="nestandard" statut="secret" scara="1.5">`

O altă noțiune foarte importantă în SGML este **entitatea**. Un document poate conține elemente care nu pot fi tastate direct la calculator, cum ar fi caracterele grecești sau simbolurile matematice. De asemenea, se pot include în text desene, fotografii sau părți de text din alte documente. Atunci SGML a apelat la o entitate reprezentată printr-un șir de caractere, asociată la elementul respectiv. Exemplu:

`<! ENTITY SRAIT "Asociația SRAIT">`

Pentru ilustrații și documente externe, se procedează analog. În acest caz, cîmpul de definire al entității va conține identificatorul de fișier, în sistemul de calcul în care se găsește, și va indica tipul datelor care se găsesc acolo. În interiorul unui document, o referință la o entitate indică locul unde cantitatea trebuie să fie

Descriere	SGML (AAP-Article DTD)	BookMaster	LaTeX	troff/troff macros ms/mm	VAX Document
Nivel 0	nu e disponibil	:H0	\part	nu e disponibil	<PART>
Nivel 1	nu e disponibil	:H1	\chapter	.NH1/H1	<CHAPTER>
Nivel 2	<sec>	:H2	\section	.NH2/H2	<HEAD1>
Nivel 3	<ss1>	:H3	\subsection	.NH3/H3	<HEAD2>
Nivel 4	<ss2>	:H4	\subsubsection	.NH4/H4	<HEAD3>
Nivel 5	<ss3>	:H5	\paragraph	.NH5/H5	<HEAD4>
Nivel 6	<ss4>	:H6	\subparagraph	.NH6/H6	<HEAD5>
Nivel paragraf	<P>	:P	\par	.PP/Ph	<P>
Text normal	text normal	:HP0.	text normal	text normal	text normal
Text italic	<i1> texte </i1>	:HP1.texte:EH1.	{\it texte}	.I texte	<EMPHASIS> (text)
Text bold	<b2> texte </b2>	:E2.texte:E2.	{\bf texte}	.B texte	<EMPHASIS> (text) \ gras
Text bold-italic	<i3> texte </i3>	:HP3.texte:EH3.	nu se utilizează oricum	nu se utilizează oricum	nu se utilizează oricum
Citat	<O> texte </O>	:O.texte:O.		.OS texte:OE	<QUOTE> (text)
Note de subsol	<FN> texte </FN>	:FN.texte:FN.	\footnote {texte}	.FS texte:FE	<FOOTNOTE> (text)
Listă numerotată	<L1>	:OL.	\begin {enumerate}	[.LS N]/AL	<LIST> (numbered)
Listă nenumerotată	<L2>	:UL.	\begin {itemize}	[.LS B]/BL	<LIST> (unnumbered)
Intrare în listă		:LI.	\item	.LI	<LE>
Listă de descriere	<DL>	:DL.	\begin {description}	-/VL	<DEFINITION_LIST>
Termenul listei	<DT>	:DT.	\item [...]	-/LI	<DEFLIST_DEF>
Definirea listei	<DD>	:DD.	texte	-/texte	<DEFLIST_ITEM>
Caracter grece (de exemplu α)	α	&alpha.	S\alpha	\(*a	<MATH_CHAR> (ALPHA)
Simbol matematic (de exemplu \approx)	≈	&app.	S\approx	\{ap	<MATH_CHAR> (approx)
Formule în text	<f> texte </f>	:F.texte:EF.	S{texte}	# texte #	<MATH> (text)
Punerea în valoare	<fd> texte </fd>	:DF.texte:EDF.	S{texte} sau \[texte]	.EQ texte .EN	<MATH> (display) text
Exponenți	<f> ^{texte} </f>	:SUP.texte:ESUP.	S^{\rm texte} S	# sup roman text #	<SUPERSCRIP> (text)
Indici	<f> <inf> texte </inf> </f>	:SUB.texte:ESUB.	S_{\rm texte} S	# sub roman text #	<SUBSCRIPT> (text)
Accente (de exemplu é)	´	&ac.	\e	e*	<MCS> (small e acute)
Umflaut (de exemplu ü)	ü	&ue.	\u	u*	<MCS> (small u umlaut)
Spațiu indivizibil	 	&rb.	-	\ sau \ (space)	necunoscut
Linie întreruptă	depinde de mărimea paginii	.br	\	.br	<LINE>
Salt la pagină	depinde de mărimea paginii	.pa	\ newpage	.bp	<PAGE>

TABEL 1. Comparație între câteva limbaje de marcare.

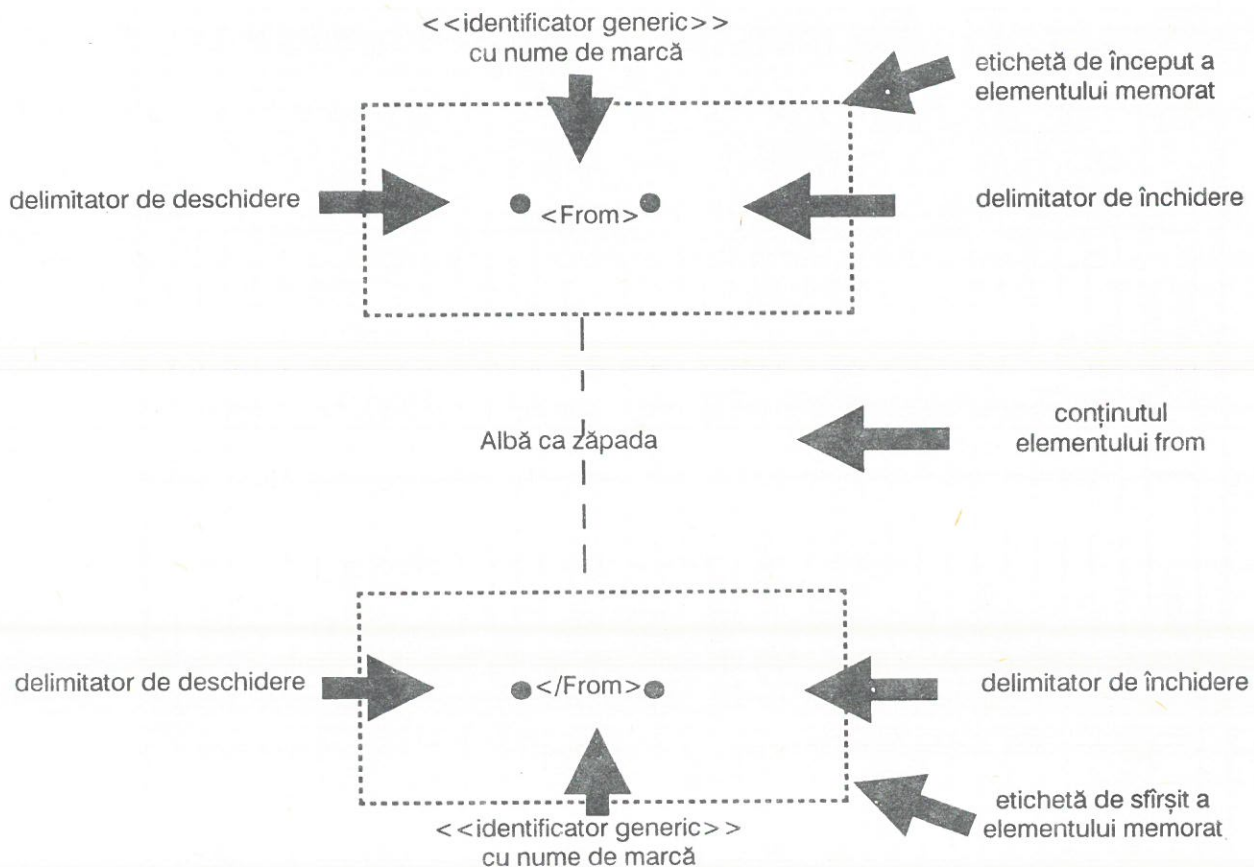


Figura 2. Componentele marcării

inserată, de exemplu: & SRAIT; Notăm (&) și (;) delimitatorii de început și de sfârșit la o referință a unei entități. De reținut ar fi faptul că, numele entităților nu poate fi scris cu majuscule sau minuscule, ca identificatorul generic.

2.1.2. Exemplu de document SGML

Tabelul 2 arată un exemplu de articol marcat cu SGML. Textele puse sub formă SGML sînt independente de sistemul de calcul sau configurația calculatorului, pentru că nu conțin secvențe de caractere de control specifice, și fișierul este în ASCII. Această independență față de un sistem de calcul reprezintă forța și slăbiciunea SGML. Este foarte frumos să ai un document în format SGML, dar pentru a-l putea utiliza (de exemplu lista) este nevoie de un program de traducere a mărcilor SGML în comenzi ale unui sistem de prelucrare de text.

2.1.3. SGML și Hypertextele

Unul din atuurile SGML este că permite aplicațiilor multiple să plece de la același text. Este ușor de adăugat o dimensiune suplimentară unei informații prezentate într-un document SGML, singurul scop nu trebuie să

se limiteze la obținerea unei versiuni imprimare. Considerăm noțiunea **hypertextului** care este foarte la modă în zilele noastre. Faptul că un document este scris în SGML permite extragerea cu ușurință a informației pentru a putea fi pusă sub formă de hypertext.

Un proiect de cercetare interesant este "HyTime", o descriere muzicală utilizînd hypermedia și dimensiunea timpului. În 1986, ANSI a definit Limbajul de Descriere Standard pentru Muzică (SDML), care este o aplicație (DTD) a SGML. Ideea este că un document conține un ansamblu de evenimente care depind de timp, de exemplu audio, video, imagini, dans, emisiuni. Durata fiecărui eveniment este măsurată într-o unitate de timp virtual (vtu). Acestea sînt legate de timpul real printr-o "baghetă" care exprimă timpul virtual ca o funcție generală de timp real. De asemenea, sînt posibile indicații de timp, cum ar fi lent sau rapid. Acest model poate fi descris utilizînd funcționalitatea cea mai simplă a SGML, elementele, entitățile și atributele. Legăturile hypertext, conectînd două puncte într-un document sau în documente diferite, sînt realizate cu ajutorul elementelor speciale numite "link". HyTime poate descrie orice situație ce depinde de timp, cum ar fi simulatoare de conducere sau de zbor, formatul CD-ROM și CD-I. Lucrul pe HyTime este coordonat

```

<!DOCTYPE article SYSTEM "AAPARTCL DTD *" [
<! ENTITY CNLFG! SYSTEM "CNLFIG1 EPS *" CDATA EPS >
.....
<ENTITY CNLFIG1 SYSTEM "CHAPTER1 SGML *"> ]>
<article>
<at1>Une introduction &agrave; SGML et sa r&eacute; lation avec DSSSL et SPDL
<DATE><yr>> 22 juillet 1991
<AU>Michel Goosens et Eric van Herwijnen
<Aff>CERN,
<odv>Division AS,
<cty>Gen&egrave;ve 23
<pc>1211-CH
<cny>Suisse
<pubfm>
<cdn>AS/91-1
</pubfm>
<ABS>
<p>
Cet article passe en revue quelques aspects de la norme ISO 8879 SGML, le langage standard g&eacute; n&eacute; ralis&eacute; de balisage, et la situe par rapport à d'autres projets de normes pour la description des documents sur ordinateur: DSSSL pour la mise en page et SPDL pour la visualisation.
</ABS>
<BDY>
<sec id='h1sgm1'><st>L'Inter&circ;t de la norme SGML </st>
<P>
SGML &mdash; langage standard g&eacute; n&eacute; ralis&eacute; de balisage ou <e3> Standard Markup Language</e3> en anglais &mdash; est un

```

Tabel 2. Exemplu de articol marcat cu SGML

de Charles Goldfarb de la IBM și Steve Newcomb de la Universitatea din Florida [8].

2.2. Definirea Tipului de Document (DTD)

Definirea Tipului de Document (DTD) conține ansamblul complet al tuturor marcatorelor pentru o clasă dată de documente. Ideea este că utilizatorii unei aplicații (cum ar fi schimbarea de texte între autori și una sau mai multe edituri) analizează necesitățile lor comune și din această discuție rezultă un ansamblu de tipuri și de structuri de documente pe care le iau în considerare. Părțile prezente se pun de acord asupra elementelor logice indispensabile, cum ar fi capitolele, paragrafele, notele de subsol etc. De subliniat ar fi că alegerea marcatorelor poate rămâne diferită pentru fiecare aplicație.

Utilizarea DTD implică faptul că problemele asociate numelor specifice mărcilor și structurilor standard pot fi evitate, pentru că toată lumea poate inventa propriul său limbaj de marcare și totuși să profite de instrumentele și produsele SGML, la care numărul crește în fiecare zi. Totodată, o analiză aprofundată se impune înainte ca aplicațiile să poată schimba liber documentele.

Un DTD definește următoarele obiecte:

- *numele* și conținutul tuturor elementelor

dintr-un document al unei clase date;

- *de câte ori* poate apărea un element;
- *ordinea* în care elementele trebuie să apară;
- dacă marcatorele de deschidere și închidere pot fi omiși sau nu;
- *convenții de utilizare* pentru a ușura introducerea marcatorelor. De exemplu, o linie albă urmată de o linie înfundată de 3 spații albe poate fi interpretată ca un marcator de deschidere a unui paragraf;
- *conținutul* tuturor elementelor, numele tuturor identificatorilor generici care pot fi conținute într-un element până la nivelul de date de tip caracter;
- *atributele* posibile pentru fiecare marcă și valoarea lor în lipsă;
- numele tuturor *entităților* predefinite, la care utilizatorul are acces.

Un DTD nu conține informații pentru prelucrarea unui document, nu definește delimitatorii care trebuie utilizați pentru marcare. Un program special numit "parser" verifică structura unui document SGML și controlează dacă marcarea este conformă cu numele și regulile definite în DTD.

Când vorbim de SGML, este foarte important să specificăm ce DTD este utilizat. Câteva produse SGML nu oferă posibilitatea folosirii unui alt DTD

decît aceea de a codifica într-o manieră “dură” în produs. Un exemplu de DTD care a fost dezvoltat pentru articolele științifice este acela al Asociației Edîtorilor Americani (AAP)[4].

2.3. Semantica stilului de document și limbajul de specificare (DSSSL)

SGML descrie structura logică a unui document, dar lasă utilizatorul să-și dezvolte propriile instrumente pentru a reprezenta această structură pe o pagină imprimată. Separarea conținutului și formei permite utilizarea multiplă a acelorași date. Există multe programe de traducere (ad hoc) pentru convertirea marcatorilor SGML în coduri pentru sisteme de prelucrare a textelor. În același timp, este uneori necesar să obținem o corespondență punct cu punct la nivelul ieșirii a două documente transmisibile. Pînă astăzi, nu există metode standard pentru descrierea prezentării externe a elementelor pe un suport de ieșire. DSSSL (*Document Style Semantics and Specification Language*) încearcă să aducă un răspuns la această problemă.

Această asociere se împarte în două etape:

1. o transformare de tip general pentru schimbarea structurii unui document (de exemplu o tablă de materii, o bibliografie, referințe interne și externe): rezultatul acestei transformări este un document intermediar care poate exista sau nu sub formă reală;
2. o transformare specifică bazată pe semantică. Zona de ieșire este definită cu ajutorul limbajului de specificație, ceea ce permite o ieșire ulterioară pe hîrtie sau pe un alt suport.

Un document DSSSL conține:

1. sursa textului în SGML, marcată după regulile DTD ale documentului original;
2. DTD care corespunde cu structura documentului virtual intermediar;
3. specificarea DSSSL, care asociază la fiecare element logic o valoare semantică după arhitectura documentului DSSSL.

Cum este cazul cu SGML, avem nevoie de un produs-program pentru a prelucra informația conținută în documentele DSSSL. Un “formator” DSSSL va fi capabil să genereze automat specificația DSSSL și va

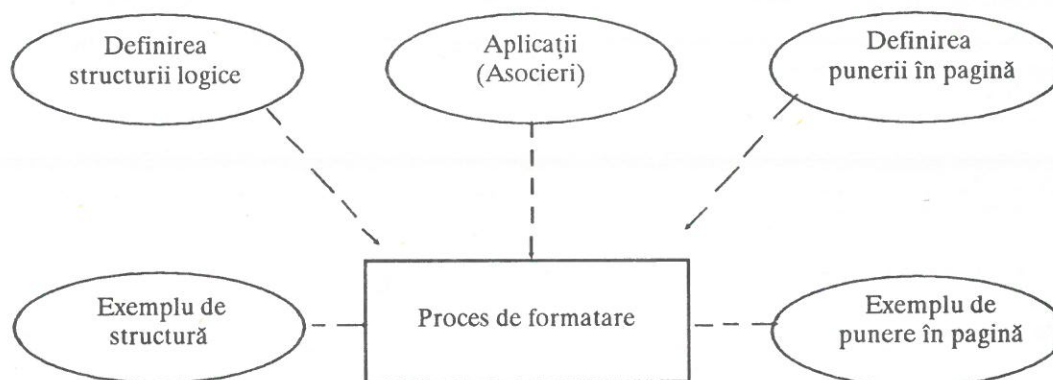


Figura 3. Asocierea informației de prezentare la elementele logice

Unul din obiectivele DSSSL [9] este de a furniza un mijloc formal și riguros de a exprima gama specificațiilor pentru producția de documente, incluzînd și tipografie de înaltă calitate. “DSSSL pleacă de la o aplicație SGML propunînd o arhitectură normalizată pentru specificațiile de prezentare. Mulți utilizatori au nevoie de un acces unificat pentru a schimba informațiile de prezentare”. Se poate imagina extinderea DSSSL la extragerea informației conținute în documentele SGML pentru a o încărca într-o bază de date. Altfel spus, DSSSL propune o modalitate de normalizare a descrierii prezentării externe de documente.

DSSSL asociază specificațiile de prezentare la elementele logice ale unui document SGML cum arată și figura 3.

putea fi aplicat documentelor care sînt conforme cu SGML. La ieșire se generează în etapa de “compoziție” un fișier de intrare pentru un program de prelucrare de texte sau o ieșire pe Post-Script sau SPDL (*limbaj standard de descriere a paginii*). După 1991, DSSSL este în stadiu de proiect standard internațional.

2.4. Limbajul standard de descriere a paginii (SPDL)

SPDL [10] (*Standard Page Description Language*) definește “un limbaj pentru specificarea unui format de vizualizare (de exemplu imprimat sau prezentat pe ecran) de documente electronice, conținînd texte în alb și negru, cu mai multe nivele de gri sau în întregime în

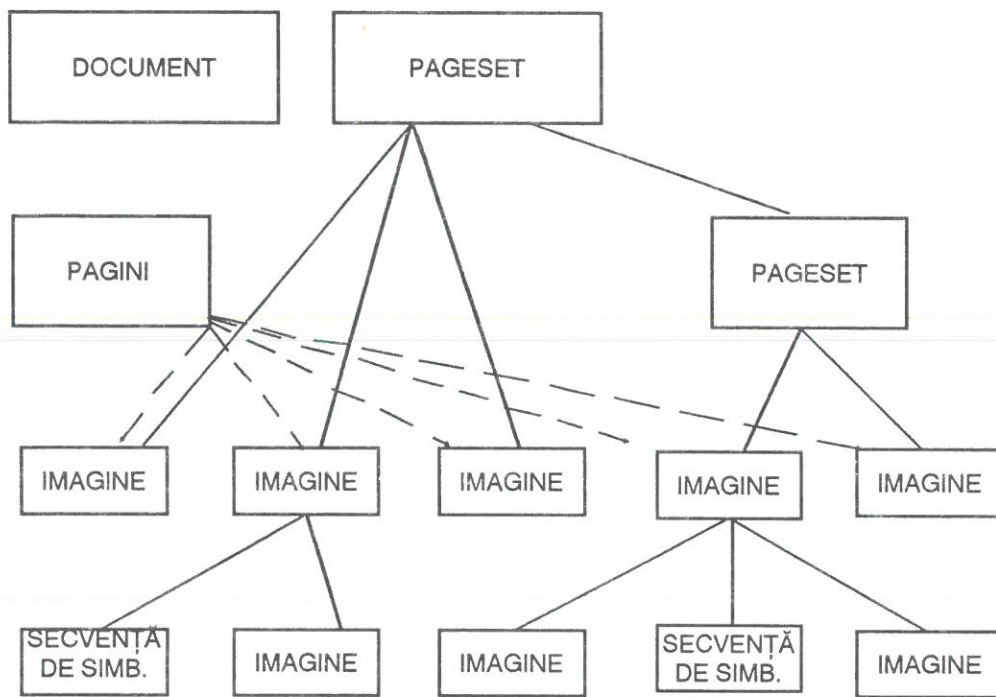


Figura 4. Structura unui document SPDL

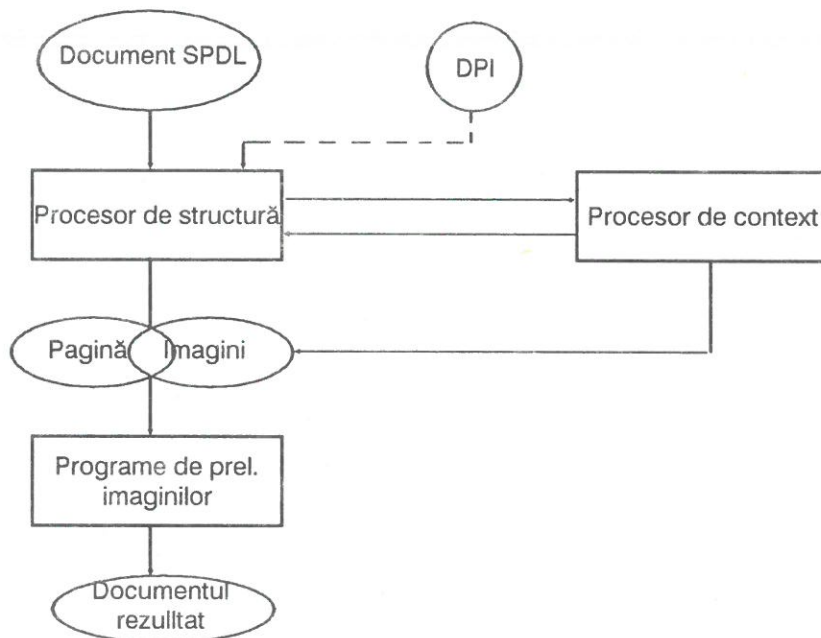


Figura 5. Prelucrarea unui document SPDL.

culori, imagini, forme geometrice". Acesta este un format tipic pentru un document în faza sa finală. SPDL oferă posibilitatea de reproducere într-un mod eficace "reprezentarea" calculată în faza de compoziție și de punere în pagină.

SPDL are multe în comun cu Post Script și cu predecesorul său Interpress. În fine, Adobe și Xerox au dominat efortul de normalizare care a generat SPDL. Acest proiect de norme ISO definește într-un mod independent de suport ieșirea folosind un formalism pentru a descrie prezentarea de documente, conținând material textual și grafic.

Un document SPDL are o structură și un conținut. Structura unui document SPDL este ierarhică, nivelul cel mai înalt fiind documentul. Un document poate conține grupuri de pagini (*pagesets*) și pagini. O pagină poate conține imagini (*pictures*) și secvențe de tokens (*jetoane sau tokensequences*). O descriere a acestei structuri sub formă de DTD SGML este propusă în textul proiectului de norme SPDL și sub formă schematică în figura 4.

În ceea ce privește conținutul unui document SPDL, constă în date sub formă de structuri de tip succesiuni de simboluri ("tokens"). Procesul de interpretare a simbolurilor este realizat cu ajutorul unei mașini în stare virtuală. Obținerea imaginilor este parametrizată în raport cu un sistem de coordonate de întărire și de reguli de "detoumage" (*clipping*) și folosește trei categorii de operații grafice, corespunzând șirurilor de caractere de text, imaginilor în mod bit și formelor geometrice.

Prelucrarea unui document SPDL este prezentată în figura 5.

Prezentarea ieșirii a fost calculată în faza de compoziție și punerea în pagină (*Layout and Composition Process*), care poate corespunde cu un "formator" DSSSL. Aceste decizii cuprind selecția de caractere, plasarea lor în plan și prezentarea grafică. Programul de vizualizare ia documentul în format SPDL și încearcă să garanteze cea mai bună aproximare posibilă ținând cont de restricțiile suportului de prezentare, realizarea calculată în timpul fazei de compoziție și punere în pagină.

Ca și DSSSL, SPDL este încă în stadiul de proiect de normă internațională. Așa că, datorită asemănării cu limbajele de descriere a paginii existente, se așteaptă să se vadă implementarea practică a SPDL, înaintea DSSSL [11].

3. Concluzii

Ansamblul SGML, DSSSL și SPDL ne propune următoarele instrumente:

- o normă (SGML) pentru specificarea structurii logice

a unei clase de documente (DTD);

- o normă (DSSSL) pentru a specifica prezentarea vizuală, asociată la un document aparținând unei clase date de documente SGML;
- o normă (SPDL) pentru a realiza o vizualizare concretă, a unui document specific.

DSSSL nu este încă o normă internațională. În plus, datorită complexității limbajului de specificare, este dificil să codificăm cu mîna. DSSSL nu va fi disponibil decît peste cîțiva ani. Asta nu înseamnă că nu putem începe utilizarea SGML-ului. Toate produsele SGML disponibile vin cu propriul lor sistem de traducere, care transformă marcatorii SGML în comenzile unui sistem de prelucrare text. Utilizatorii SGML-ului pot beneficia de oferta SGML în domeniul portabilității, generalității, independenței de formă și posibilităților de utilizare.

Bibliografie

1. Organisation Internationale de Normalisation - Langage normalisé de balise généralisé (SGML). ISO 8879-1986, Genève, 1986.
2. van HERWIJNEN, E.- Practical SGML. Wolters-Kluwer Academic Publishers, Boston, 1990
3. VIGNAUD, D.- L' édition structurée des documents. Éditions du Cercle de la Librairie, Paris, 1990
4. AMERICAN NATIONAL STANDARDS INSTITUTE. American National Standard for Electronic Manuscript Preparation and Markup. ANSI/NISO Z39.59-1988, 1988
5. IBM Corporation, Boulder - Document Composition Facility BookMaster - Version 3. SC34-5009-03, 1990
6. LAMPORT, L. - LATEX, a document preparation system. Addison Wesley, Reading, 1986.
7. Digital Equipment Co. VAX Document, Using Global Tags. AA-JT84C-TE, 1991
8. GOLDFARB, C.F. - Hy Time. A standard for structured hypermedia interchange. IEEE Computer, August 1991, pp 81-84.
9. ADLLER, S. (Ed.) - International Organization for Standardization. Document Style Semantics and Specification Language. ISO DP 10179, ISO Geneva, 1991.
10. STRASEN, S. and FOWLEY, M. (Ed.) - International Organization for Standardization. Standard Page Description Language. ISO DP 10180, ISO Geneva, 1991.
11. DARDAILLER, D. - Normes et fontes, Cahiers GUTenberg, no 4, Décembre 1991, pp 2-9.