

## PREZENTAREA DOCUMENTELOR ELECTRONICE UTILIZAND LIMBAJUL HTML

*Conf. dr. ing. Doina Banciu,  
Anca Doloc-Mihu*

Noțiunea de document electronic se utilizează pentru a exprima modul în care este prezentată informația ce alcătuiește documentul respectiv: informație digitală. Conform standardelor ISO, un document electronic este un "document existând sub formă electronică într-o manieră care îl face accesibil prin dispozitive de prelucrare a datelor".

Documentele electronice sunt rezultatul prelucrării informației cu ajutorul unor echipamente care permit transformarea acestora în informație digitală. Documentele electronice pot fi produse prin mai multe metode:

- prelucrarea unor date deja stocate pe medii digitale în vederea obținerii unui document cu o anumită formă și structură de prezentare;
- introducerea unui text, tabel etc, de la tastatură prin intermediul unui procesor de texte (WRITE, WORD, WORD PERFECT etc.);
- scanarea unui document sau a unei părți din acesta cu un scanner;
- "codificarea mișcării" pentru sunete și imagini animate.

Documentele electronice create ca imagini, fotografiile ale unui document aflat inițial pe un alt suport decât cel digital, devenite documente în mod imagine, nu pot fi prelucrate cu

calculatorul în sensul intervenției pe conținutul propriu-zis al acestora. Este cazul documentelor realizate prin scanare, al documentelor grafice "în mod raster" (documente care nu sunt reprezentate în mod vector), documentelor sonore și/sau de imagini animate, fișierelor obținute prin transmisiuni fax. Pentru a putea fi folosite în prelucrări ulterioare, aceste documente trebuie supuse unor procese specifice de codificare.

Astfel, textele pot fi tratate cu un soft special pentru recunoașterea optică a caracterelor denumit OCR (Optical Character Recognition) care codifică fiecare caracter conform codului ASCII). Procedul este, în general, costisitor, întrucât softul de tip OCR nu asigură întotdeauna performanțele dorite, mai ales în cazul unor volume mari de date (la codificare se introduc de regulă multe erori care trebuie eliminate prin intervenția omului). În mod similar, documentele grafice pot fi codificate în format vectorizat.

Documentele electronice, în mod imagine, necesită spații de stocare mult superioare celor codificate.

Documentele electronice sunt disponibile fie în forma în care sunt create, fie prelucrate pentru o prezentare grafică, cât mai accesibilă utilizatorilor. Un instrument informatic cu care se poate realiza un astfel de document este limbajul HTML (HyperText Markup Language).

HTML nu este un limbaj de programare; el lucrează doar cu tag-uri (marcatori) și cu regulile de inserare ale acestora, astfel încât acestea să permită scrierea documentelor într-o formă ce poate fi interpretată de către programe specializate pentru accesul în rețeaua Internet. Aceste programe numite și navigatoare (sau browsers în engleză), pot rula pe orice platformă hard sau soft, adică pot fi afișate la fel pe un PC, pe un Mac sau pe un sistem Unix, cu condiția ca acestea să fie conectate la Internet. Dintre ele, cele mai cunoscute sunt : Netscape, Internet Explorer și Mosaic.

Navigatoarele permit accesarea documentelor stocate pe servere Web. World Wide Web, sau WWW, sau prescurtat Web-

ul este un sistem ce permite punerea la dispoziție a informațiilor sub forma unor documente stocate pe servere cu astfel de sisteme.

Pentru a putea accesa un anumit document pe Web, trebuie cunoscută în prealabil localizarea lui fizică, sau adresa lui, exprimată sub o anumită formă :

protocol/nume\_server/cale/nume\_document

Aceasta mai este numită și URL-ul (Uniform Resource Locator) documentului respectiv; protocolul folosit este HTTP (HyperText Transfer Protocol). Mai exact, adresa unui document va avea următoarea formă :

[http://nume\\_server/cale/nume\\_document](http://nume_server/cale/nume_document)

Caracteristicile publicațiilor Web sunt existența legăturilor hypertext și a unor procedee care permit realizarea unui dialog real între client și server. Legăturile hypertext sunt indicate prin marcarea diferită de restul documentului a unor cuvinte, imagini sau simboluri grafice, care, odată selectate, determină ca navigatorul "să aducă" documentul respectiv, indiferent unde anume se află acesta pe Internet. Ansamblul documentelor electronice care fac referință unele la altele a condus la numele Web (țesătură). Procesul de aducere pe sistem a documentelor cu ajutorul navigatoarelor se numește navigare pe Web.

De remarcat că în prezent majoritatea aplicațiilor Web sunt publicații electronice și aceasta datorită posibilităților pe care Web-ul le oferă: o informare rapidă și la un preț redus (de fapt doar costul abonamentului la legătura Internet); informația este structurată, interactivă, rapid actualizată și practic gratuită. Aceste lucruri au condus la larga răspândire a site-urilor Web.

## Descrierea limbajului HTML

Cu ajutorul HTML se pot realiza :

- documente independente de platformă
- legături la alte documente și/sau resurse de pe Internet
- inserții de grafică, sunet și video
- dialog între cititorul documentului și aplicația respectivă

HTML este un limbaj pentru descrierea documentelor structurate. Deoarece majoritatea documentelor au elemente comune (de exemplu: titluri, liste) s-a pornit de la ideea definirii unui set de elemente, urmată apoi de marcarea elementelor documentului cu tag-urile corespunzătoare și de folosire a acestora cu ajutorul legăturilor (fără a mai fi deci nevoie de scrierea de fiecare dată a textului respectiv, textul odată introdus de fiecare dată se vor face doar trimiteri la el; făcându-se astfel și economie de timp și spațiu).

Tag-urile (marcatorii) HTML sunt de forma:

`<nume_tag>Textul afectat de tag</nume_tag>`

și indică navigatorului elemente de structura documentului, formatare, hypertext sau alte elemente (imagini, fișiere sonore). Documentele HTML conțin doar textul propriu-zis și tag-urile HTML, sursele lor putând fi văzute de orice navigator. Când navigatorul încarcă un document HTML, el citește tag-urile, le interpretează, formatează textul și imaginea și le afișează pe ecran. Acesta este motivul pentru care același document HTML apare puțin diferit dacă este privit cu navigatoare diferite.

Fiecare document, uneori referit și ca pagină, este structurat în două părți: început sau antet (head), care conține declarații despre document, și conținut (body), care este (hyper)textul documentului.

Exemplu:

```
<HTML>
```

```
<HEAD><TITLE>Titlul documentului</TITLE></HEAD>
```

```
<BODY>Textul documentului</BODY>
</HTML>
```

Toate tag-urile au: nume scrise între paranteze unghiulare și posibil câteiva atribute care pot lua anumite valori. Tag-urile și atributele pot fi editate atât cu caractere mici cât și cu caractere mari și au același efect. De exemplu, atributul SRC (sau src, care vine de la search = caută) este un atribut al tag-ului IMG (imagine). Atributele au următoarea formă: NUME = "valoare". De exemplu, <IMG SRC = "poza.gif"> semnifică inserarea unei imagini.

Prin HTML se pot realiza legături ale documentului prelucrat cu un alt document aflat pe același calculator sau pe un alt calculator, aflat la distanță (cele două calculatoare însă trebuie să fie conectate în rețea), dar și legături la anumite părți ale documentului respectiv. Acest tip de legături sunt legături hypertext.

Pentru marcarea legăturilor se folosește tag-ul <A> împreună cu numele documentului legat, astfel:

```
<A HREF = "document_legat.html" > Text marcat diferit (cu albastru) </A>
```

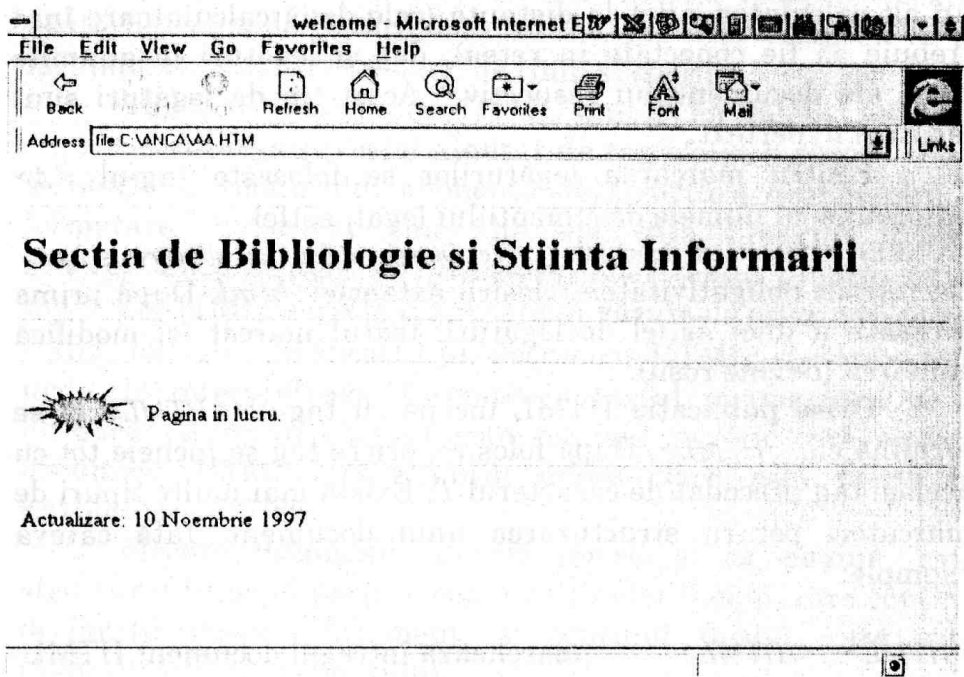
De reținut obligativitatea folosirii extensiei *.html*. După prima accesare a unei astfel de legături, textul marcat își modifică culoarea (devine roșu).

Orice publicație HTML începe cu tag-ul <HTML> și se termină cu </HTML>. După folosire, oricare tag se încheie tot cu același tag precedat de caracterul '/'. Există mai multe tipuri de marcatori pentru structurarea unui document. Iată câteva exemple:

<HTML>...</HTML>	marchează întregul document HTML
<HEAD>...</HEAD>	marchează partea de început a documentului
<BODY>...</BODY>	marchează conținutul propriu-zis al documentului
<TITLE>...</TITLE>	marchează titlul documentului,

<B>...</B>, <I>...</I>	setează stilul textului cu caractere bold sau italic,
<Hi>...</Hi>	titlu de ordin <i>i</i> ,
<HR>	inserează o linie,
<OL>...</OL>	listă ordonată,
<UL>...</UL>	listă neordonată,
<FONT>	urmat de size, color, etc setează stilul fontului (dimensiune, culoare),
<CENTER>	centrează textul
<P> ...</P>	marchează un paragraf
<CITE>...</CITE>	marchează un citat
etc.	

Exemplu de pagină Web, văzută cu Internet Explorer:



Sursa paginii de mai sus este :

```

<HTML>
<HEAD>
<TITLE>welcome</TITLE>
</HEAD>
<BODY>
<HR>
<H1>Sectia de Bibliologie si Stiinta Informarii</H1>
<HR>
<HR>
<P>
<IMG src="file:///c:/centru/htm/gif/hota.gif" align="Middle" >
Pagina in lucru.<HR>
<P>
<A NAME="UQHTML0"></A>Actualizare: 10 Noembrie 1997
</BODY>
</HTML>

```

### Imagini în documente HTML

În sistemele de informare și documentare, care trebuie să ofere documente în forme și structuri cât mai facile de receptat de către utilizatori, este important ca specialiștii în știința informării să poată cunoaște posibilitățile informatice pe care le au la dispoziție. Prezentăm succint principiile care stau la baza prezentării imaginii în documentele HTML.

Imaginile afișate în documente realizate cu HTML pot fi **inline** (încărcate odată cu documentul) sau imagini **externe** (încărcate la cererea utilizatorului). Formatele utilizate sunt **GIF** și **JPEG**. Deși navigatoarele recunosc doar aceste formate, se pot folosi aplicații auxiliare ce pot manipula și alte tipuri de imagini: PCX, XBM, PICT, etc., convertindu-le spre aceste formate.

Formatul **GIF** se caracterizează prin explorarea întretesută a imaginii; fișierul imagine nu este în format linie cu linie, ci în format întretesut (se salvează fiecare a opta linie,

apoi fiecare a opta linie începând cu a patra, apoi fiecare a patra începând cu a treia, etc.), obținându-se astfel o afișare cu rezoluție gradual crescătoare. Aceasta permite utilizatorului stoparea încărcării imaginii după ce acesta i-a văzut conținutul (încărcarea poate dura destul de mult timp).

Afișarea imaginilor inline se face folosind tag-ul `vid` `<IMG>` urmat de un atribut și de numele fișierului imagine (inclusiv extensia `.gif` sau `.jpeg`) pus între ghilimele. Spre exemplu pentru afișarea imaginii "carte.gif" fișierul sursă HTML conține linia :

```
<IMG SRC ="/..cale../carte.gif">
```

Calea se poate omite în cazul în care imaginea și sursa se află în același director.

Atributele ce pot fi folosite sunt:

--- SRC --- specifică localizarea resursei imagine, și este urmat de numele acesteia.

--- ALT --- permite afișarea textului ce îl urmează, în loc de imagine în cazul navigatoarelor negrafice.

--- ALIGN --- aliniaza imaginea liniei alăturate de text astfel: TOP (aliniere la partea superioară a rândului), MIDDLE (la partea de mijloc) și BOTTOM (la partea de jos) precum și LEFT și RIGHT (aliniere a imaginii în stânga respectiv dreapta paginii).

--- WIDTH și HEIGHT --- specifică dimensiunile la care imaginea va fi afișată păstrând scara (imagine scalată).

---- VSPACE și HSPACE --- specifică distanțele (în pixeli) între imagine și textul vecin pe verticală, respectiv pe orizontală.

Exemplu:

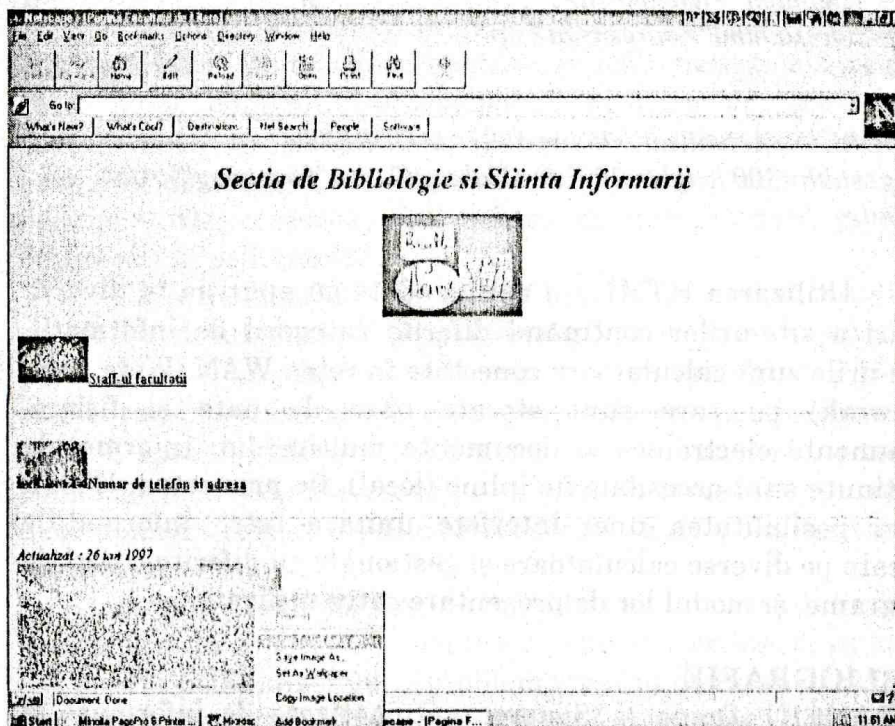
```
<div align=center><a href= "...*.html">  
<img WIDTH =100 HEIGHT=100 src= "lake.gif" alt= "fara imag"  
hspace=3 vspace=2 border=3></a>  
</div>
```

Se va afișa imaginea `.gif` scalată la dimensiunile `width` și `height` în pixeli, centrată, link-editată (legată) la fișierul "...html", la distanța `hspace` și `vspace` pixeli de textul învecinat

pe orizontală și verticală, mărginită în chenar de grosimea *border*; pentru navigatoarele negrafice afișează textul ce urmează atributului ALT.

**Observații:**

- dacă nu se specifică dimensiunile *width* și *height* imaginea este afișată la dimensiunea reală din fișier. În cazul specificării doar a uneia din cele două dimensiuni, se va afișa o imagine scalată la dimensiunea respectivă.
- pentru afișarea imaginii la dimensiune reală se apasă butonul din dreapta mouse-ului (cursorul mouse-ului fiind pe imagine) și se selectează item-ul *View this Image* din Netscape (vezi imaginea de mai jos).



Pentru documentul de mai sus, sursa HTML arată astfel:

```

<html>
<TITLE>Pagina Facultatii de Litere </TITLE>
<center>
<h1><i>Sectia de Bibliologie si Stiinta Informarii</i></h1>
<a href="http://pluto/cgi-bin/script/img.map">
</a><p>
</center>
<h4>
<a
href="staff.html">Staff-ul facultatii</a><p>
<hr><p>
Numar de <a
href="telef.html">telefon</a> si <a
href="adresa.html">adresa</a><p>
<hr><p>
</h4>
<ADDRESS>Actualizat : 26 ian 1997</ADDRESS>
</a>
</html>

```

Utilizarea HTML s-a extins odată cu apariția în diverse locuri a site-urilor conținând diferite categorii de informații. Site-urile sunt calculatoare conectate în rețea WAN (Wide Area Network) pe care sunt stocate baze de date și fișiere, documente electronice și documente multimedia. Informațiile conținute sunt accesibile fie inline (local), fie prin rețea. HTML oferă posibilitatea unei interfețe unitare între informațiile stocate pe diverse calculatoare și gestionate cu diferite tipuri de programe, și modul lor de prezentare către utilizator.

#### **BIBLIOGRAFIE**

1. BANCIU, Doina – “Sisteme automatizate de informare și documentare”, București, Editura Tehnică, 1997.
2. RĂDOIU, Dumitru – “HTML-Publicații Web”, Computer Press Agora s.r.l., 1996.