

Pagini Web în HTML

Paginile de Web sunt scrise într-un limbaj numit HTML (Hypertext Markup Language limbaj de marcare hipertext). HTML permite utilizatorilor să producă, pagini care includ texte, grafică și indicatori la alte pagini de Web. Vom începe să studiem HTML eu acești indicatori, pentru că ei reprezintă tocmai mecanismul care ține Web-ul conectat

URL- Uniform Resource Locators

O pagină de Web poate să conțină referințe la alte pagini. Să explicăm cum sunt implementate aceste referințe. Încă de la crearea Web-ului a fost clar că pentru a avea o pagină care să indice spre altă pagină este necesar un mecanism care să permită numirea și regăsirea paginilor. În particular sunt trei întrebări la care trebuie să se răspundă înainte de a se putea afișa o pagină:

1. Cum se numește pagina ?
2. Cum este localizată pagina ?
3. Cum se face accesul la pagină ?

Dacă fiecare pagină ar avea un nume unic, atunci nu ar exista nici o ambiguitate în identificarea paginilor. Totuși, problema nu este încă rezolvată. Să considerăm de exemplu o paralelă între oameni și pagini. În SUA aproape fiecare persoană are un număr de asigurare socială; care este un identificator unic, astfel încât nu există două persoane cu același număr. Totuși, cunoscând numai numărul respectiv nu există nici o posibilitate de a găsi adresa persoanei respective, și sigur nu se poate afla dacă persoanei respective trebuie să i se scrie în engleză, spaniolă sau chineză. Web-ul are practic același fel de probleme.

Soluția aleasă identifică paginile într-un mod care rezolvă toate trei problemele în același timp. Fiecare pagină are un URL (Uniform Resource Locater - adresa uniformă pentru localizarea resurselor) care funcționează ca nume al paginii general valabil. Un URL are trei componente: protocolul, numele DNS al mașinii pe care este memorat fișierul și un nume local, care indică în mod unic pagina (de obicei numele fișierului care conține pagina). De exemplu, URL-ul departamentului din care face parte autorul este:
`http://www.cs.vu.nl/welcome.html`

Acest URL este format din trei componente: protocolul (`http`), numele DNS al serverului (`www.cs.vu.nl`) și numele fișierului (`welcome.html`), cu semnele de punctuație corespunzătoare.

Se utilizează notații care reprezintă prescurtări standard. De exemplu `user/` poate să fie pus în corespondență cu directorul WWW al utilizatorului `user`, folosind convenția că o referință la directorul respectiv implică un anumit fișier, de exemplu `index.html`. De exemplu pagina autorului poate să fie referită ca:

`http://www.cs.vu.nl/~ast/`

chiar dacă de fapt numele fișierului este diferit. Pe multe servere un nume de fișier indică implicit pagina organizației căreia îi aparține serverul.

Acum ar trebui să fie clar cum funcționează hipertextul. Pentru a face o porțiune de text selectabilă, cel care scrie pagina trebuie să furnizeze două elemente: textul prin care se face selecția și URL-ul paginii care trebuie adusă dacă textul este selectat. Când se face selecția, programul de navigare caută numele serverului utilizând DNS-ul. Pe baza adresei IP a serverului, programul de navigare stabilește o conexiune TCP spre server. Utilizând această conexiune, se transmite numele fișierului utilizând protocolul specificat. Bingo. Acum sosește pagina.

Protocolul `http` este protocolul nativ pentru Web; el este utilizat de către serverele de HTTP.

Protocolul `ftp` este utilizat pentru accesul la fișiere prin FTP (File Transfer Protocol - protocol pentru transferul de fișiere), protocolul Internet de transfer de fișiere. FTP este utilizat de peste douăzeci de ani și este foarte răspândit. Numeroase servere de FTP din toată lumea permit ca de oriunde din Internet să se facă o conectare și să se aducă orice fișier plasat pe un server FTP. Web-ul nu aduce schimbări aici, face doar ca obținerea fișierelor să se facă mai ușor, pentru că FTP are o interfață mai puțin prietenoasă. În viitor probabil că FTP-ul va dispărea, deoarece nu există nici un avantaj pentru o organizație să aibă un server de FTP în loc de un server de HTTP, care poate să facă tot ce face un server de FTP, și chiar mai mult (desigur mai există niște argumente legate de eficiență).

Este posibil să se facă acces la un fișier local ca la o pagină de Web, fie utilizând protocolul file (fișier), sau pur și simplu utilizând numele fișierului. Această abordare este similară utilizării protocolului FTP, dar nu implică existența unui server. Desigur funcționează numai pentru fișiere locale.

{Protocolul news permite unui utilizator de Web să citească un articol din știri ca și cum ar fi o pagină de Web. Aceasta înseamnă de fapt că un program de navigare este în același timp și un cititor de știri. De fapt multe programe de navigare au butoane sau elemente de meniu care permit citirea știrilor USENET mai ușor decât dacă se utilizează cititoare de știri obișnuite.

Protocolul news admite două formate. Primul format specifică un grup de știri și poate să fie utilizat pentru a obține o listă de articole de la un server de știri preconfigurat. Al doilea format cere identificatorul unui articol, de exemplu AA0134223112@csutah.edu. Programul de navigare aduce articolul de la serverul corespunzător utilizând protocolul NNTP.

Protocolul gopher este utilizat de sistemul Gopher, care a fost proiectat la universitatea Minnesota. Numele este cel al echipei atletice a universității, the Golden Gopher, de asemenea acest nume este utilizat în argou pentru “go for”; adică o comandă de aducere Gopher-ul a precedat Web-ul cu câțiva ani. Este o metodă de regăsire a informației, similar conceptual cu cea utilizată de Web, dar acceptând numai text și imagini. Când un utilizator se conectează la un server de Gopher, va avea la dispoziție un meniu de fișiere și directoare, fiecare intrare putând să fie conectată la orice meniu de Gopher oriunde în lume.

Marele avantaj al Gopher-ului în raport cu Web-ul este că funcționează foarte bine cu terminale ASCII care afișează 25 x 80 caractere și din care există încă foarte multe în lume. Pentru că este bazat pe text, Gopher este foarte rapid. De aceea în lume există multe servere de Gopher. Utilizatorii de Web pot să acceseze un server de Gopher și să vadă fiecare meniu Gopher ca o pagină Web cu intrări selectabile. Dacă nu ați lucrat cu Gopher-ul utilizați mașina favorită de căutare în Web pentru a căuta cuvântul “gopher”.

Este posibil să se trimită cerere de căutare completă unui server de Gopher utilizând un protocol gopher+. Ce se va afișa este rezultatul cererii transmise serverului de Gopher.

Ultimele două protocole nu sunt de fapt protocole pentru aducerea unor pagini de Web și nu sunt suportate de orice program de navigare, dar sunt utile. Protocolul mailto permite transmiterea de poștă dintr-un program de navigare. Pentru a face această operație se selectează butonul OPEN și se specifică un URL constând din mailto: urmat de adresa destinatarului. Majoritatea programelor de navigare vor răspunde cu un formular care are intrări corespunzătoare subiectului și altor informații din antet și spațiu pentru mesaj.

Protocolul telnet este utilizat pentru stabilirea unei conexiuni pe o mașină aflată la distanță. Se utilizează în același fel ca și programul Telnet, ceea ce nu constituie o surpriză, deoarece majoritatea programelor de navigare utilizează programul Telnet. }

Pe scurt URL-urile au fost proiectate nu numai pentru a permite utilizatorilor să navigheze prin Web, dar și pentru a utiliza FTP, news, Gopher, email și telnet, ceea ce face inutile interfețele specializate pentru aceste protocole integrând astfel într-un singur program, navigatorul în Web, aproape toate tipurile de acces în Internet. Dacă metoda nu ar fi fost proiectată de un fizician ar fi putut să pară produsul departamentului de publicitate al unei companii de software.

În ciuda tuturor acestor proprietăți, creșterea Web-ului scoate în evidență și o slăbiciune a metodei utilizării URL-urilor. Pentru o pagină care este foarte des referită, ar fi de preferat să existe mai multe copii pe servere diferite, pentru a reduce traficul în rețea. Problema este că URL-urile nu oferă nici o posibilitate de indicare a unei pagini fără să se specifice unde este localizată pagina respectivă. Nu există nici o metodă pentru a spune ceva de genul: “Vreau pagina xyz, dar nu mă interesează de unde o aduci”. Pentru a rezolva această problemă și a permite multiplicarea paginilor IETF lucrează la un sistem de URL (Universal Resource Identifiers - identificatori universali de resurse). Un URL poate să fie privit ca un URL generalizat. Acest subiect este în curs de cercetare.

Documentație: Andrew Tannenbaum – “Rețele de calculatoare”