

Historie Internetu

Vznik

Historie Internetu je více než 30 let stará a stejně jako mnoho jiných technologií vděčí za svůj původ vojenskému průmyslu. Amerika si uvědomila, že v jaderné válce lze velmi snadno zlikvidovat centrální velení, což by ohrozilo celý systém fungování obrany státu. Společnost RAND na základě objednávky americké vlády vyřešila tento problém velmi originálně. Vytvořila systém sítě, která je schopna fungovat i za okolností, že některé z jejích uzlů budou zničeny. Zprávy probíhající mezi uzly rozdělila na jednotlivé pakety, které budou obsahovat informaci o svém cílovém určení a putovat samostatně. Konkrétní cesta paketu sítě nebude důležitá, jednotlivé uzly si ji budou moci předávat mezi sebou a zpráva se bude kompletovat až v cílovém počítači. V tomto případě je komunikace možná i v okamžiku, kdy bude část sítě vyřazena z provozu.

Postupný rozvoj

Během 60-tých let začal Pentagon prostřednictvím agentury ARPA (Advanced Research Projects Agency) financovat projekt sítě pojmenované ARPANET. Tato v roce 1969 spojovala čtyři univerzity na západě USA a sloužila k předávání vědeckých dat a sdílení výpočetních prostředků tehdejších počítačů. V roce 1971 se síť rozrostla na 15 a o rok později na 37 počítačů. Během doby se na ARPANETu mimo vědeckých dat začala objevovat i data s vědou nesouvisející. Jednalo se o soukromé elektronické zprávy uživatelů, začaly vznikat i první „mailing listy“, neboli automaticky rozesílané identické zprávy pro větší počet přihlášených odběratelů.

Síť ARPANET umožnila rychlý nárůst připojených počítačů nejen díky své decentralizované struktuře, ale i díky použitému komunikačnímu paketové-orientovanému protokolu NCP, který je nezávislý na typu připojeného počítače. Tento byl později nahrazen propracovanějším standardem protokolů TCP/IP. Tento nový protokol se díky volné dostupnosti stal základem mnoha nově vznikajících sítí, které dnes tvoří Internet.

S růstem počtu připojených počítačů si tato forma komunikace získávala stále větší oblibu. Páteř sítě v USA začaly tvořit další vládní subjekty jako NASA, National Institutes of Health, Department of Energy a další. Jednotlivé americké uzly byly rozčleněny dle svého zaměření. Vzniklo šest základních domén: gov (vládní instituce), mil (vojenské instituce) edu (vzdělávací instituce), com (komerční organizace), org (nevýdělečné organizace) a net (pro brány mezi jednotlivými sítěmi). Pro mimo-americké uzly bylo zavedeno rozlišení dle jejich geografického umístění.

Česká republika se připojila na tuto celosvětovou síť v roce 1991. V následujícím roce (1992) vznikla společnost zabývající se historií, vývojem, trendy a etikou Internetu. Došlo k masivnímu nárůstu připojených počítačů - tento nárůst překročil jeden milion. V roce 1993 vzniká instituce, která udržuje adresářové a databázové služby účastníků Internetu, provádí registraci doménových jmen a přidělování adres. Na Internetu můžeme sledovat audio i video vysílání prvních televizních a rozhlasových stanic. Připojují se obchodní společnosti a média, které neměly s počítači nic společného.

V dnešní době v České republice existuje už nespočet poskytovatelů Internetu, jako např. *Telecom, Czechia ...* a *Ces.Net* který je administrátorem domény cz a akademické sítě TEN155.

* Gymnázium Arabská, Arabská 14, Praha 6, makovice@nipax.cz, krikava@vtech.cz

Čím je Internet dnes ?

Tuto otázku si položilo již mnoho lidí a je na ni mnoho různých technicky přesných i filozofických odpovědí, pravděpodobně si na ni však bude muset odpovědět každý podle sebe.

Nejprve bychom rádi vyvrátili mylnou domněnku stále ještě velké skupiny lidí, kteří si spojují Internet se službou WWW (World Wide Web). Je to pouze jedna z několika set složek, které dnes tvoří boom Internetu. Mezi další patří například elektronická pošta, přenos souborů, vzdálené pracování na počítači, rozhovory (chaty)... Internet je veřejná síť, neexistuje žádný ústřední orgán kontrolující její činnost a směr, kterým se odebírá. Internet zaštiťuje mnoho různých oblastí, každý ho používá ke svému účelu, firmy poskytují své prezentace, zpravodajské centrály vystavují nové informace, utlačované skupiny vedou mediální válku, postižení se s jeho pomocí mohou lépe zapojit do běžného života, poskytuje rozsáhlé poznatky snad ze všech oblastí lidského vědění, a to jak legálně tak ilegálně. Stále větší popularitu získávají internetové konference, telefonování přes Internet a v neposlední řadě také elektronické obchodování. Internet je natolik globální, známý a využívaný, že se jeho název zkracuje na „síť“ nebo anglicky „net“. Jelikož i ty nejstručnější informace o Internetu by zabraly jednu celou knihu (a to nikterak tenkou), budou následující údaje pouze jakýmsi odrazovým můstkem po dalším zkoumání.

Základy sítě Internet

Některé obecné pojmy

Internet je postaven na protokolu TCP/IP (Transmission Control Protocol/Internet Protocol), který vychází z modelu ISO/OSI (International Organization for Standardization Open System Interconnection). Tento protokol je čtyřvrstvý a pro přenos využívá „okénkovou metodu“. Nám ale bude stačit vědět, že se skládá ze dvou subprotokolu: první z nich (IP) zajišťuje přenos na nižší úrovni, tj. vyhledání příslušného počítače a zaslání dat, které předá protokol vyšší úrovně (TCP). Ten také kontroluje, zda data byla doručena správně.

Data, která připravuje protokol TCP se nazývají pakety. Ty si lze představit jako obyčejná poštovní psaní, které je dané do obálky. Každá obálka je označena adresou odesílatele a adresou příjemce. Jakmile příjemce zásilku obdrží, pošle adresátovi informaci o jejím doručení. Pokud ne, adresát ji pošle znovu.

IP adresa

Adresa se v sítích TCP/IP značí takzvanou IP adresou, což je číslo tvořené čtyřmi slovy (čísla 0...255) mezi sebou navzájem oddělenými tečkou. Některé z nich jsou vyhrazeny pro speciální účely. Příklad: 193.85.233.78. To znamená, že každý počítač zapojený v Internetu musí mít svou vlastní unikátní IP adresu. Pokud se tedy budete chtít připojit na libovolný počítač v síti, stačí znát jeho adresu.

DNS a domény

Toto je však jen jeden ze způsobů identifikace počítačů zapojených v TCP/IP, na rozdíl od druhé varianty je toto označení povinné. Většina z vás obvykle používá právě druhou variantu (což nejde vždy). Když se díváte na příklad na nějakou WWW stránku, místo nějakého podivného čísla píšete rovnou slovní identifikátor (jméno počítače), například *www.seznam.cz*. Této formě služby se říká DNS (Domain Name System/Server). Princip je jednoduchý, číslu se přiřadí identifikátor (doménové jméno). Nejenom toto však poskytuje DNS, říká například: jakým způsobem se bude předávat pošta dál, informace o serveru (typ hardware i software) a další. Systém DNS je založený na doménovém stromu, v němž jsou zde hierarchicky uspořádány domény první, druhé, třetí a dalších úrov-

ni. Domén první úrovně bylo při vzniku tohoto systému navrženo šest. Nyní při velkém rozmachu stále nových serverů existuje ještě navíc geografické označení. Každá země dostala příponu vytvořenou z jejího názvu (Česká republika má *.cz*, Slovenská republika *.sk*). Původní označení by mělo být užíváno pouze pro subjekty v USA. V následující tabulce je uveden přehled základních šesti domén první úrovně:

.mil	Military	Používá armáda USA
.gov	Government	Je vyhrazena pro úřady a instituce vlády USA
.com	Commercial	Pro komerční subjekty
.net	Network	Pro poskytovatele připojení ISP, síťová zařízení, směrovače
.edu	Educational	Pro vzdělávací instituce, zejména pro vysoké školy
.org	Organizations	Popisuje různé neziskové organizace

V současné době se projednávají další možná rozšíření:

.firm	Pro obchodní společnosti a firmy
.store	Pro společnosti zabývající se prodejem zboží a služeb
.web	Subjekty zabývající se různými aktivitami na síti World Wide Web
.arts	Pro subjekty zapojené do kulturních aktivit
.rec	Pro organizace rekreačního zaměření
.info	Pro subjekty poskytující informační služby
.nom	Pro jednotlivce

Doménové jméno by se tedy dalo napsat obecně asi takto:

[<...>].[<3. úroveň>].<2. úroveň>.<1. úroveň>

Domény u nás spravuje firma EuNet s.r.o., obecně NIC CZ (www.nic.cz), kde si svou doménu můž zaregistrovat kdokoli. Poplatek je přibližně 1700 Kč jednorázově a asi 800 Kč ročně.

Porty a služby

Nyní se podíváme na služby Internetu. Jelikož na jednom serveru jich může běžet několik, jednoduchá IP adresa ani doménové jméno nemůže přesně určit jakou z nich požadujeme, proto byly zavedeny takzvané porty. Jedná se o čísla z intervalu 0...65535, z čehož každý specifikuje určitou službu. Jednotlivá přidělení nejsou nijak zvlášť standardizována, avšak pro některé často používané služby se zavedly „zažité“ porty. Číslo se připojuje dvojtečkou za IP adresu nebo doménové jméno. Příklad: *193.85.233.75:8080*. „Zažité porty“ běžných služeb:

Port	Služba
80	HTTP (WWW)
21	FTP
25	SMTP (odesílání pošty)
110	POP3 (příjem pošty)

URL

Tímto se dostáváme ke kompletnímu označení požadavku v síti. Specifikuje se do takzvané URL (Uniform Resource Location) adresy. Tato adresa má obecně následující tvar:

<schéma>:<jméno podle schématu>

Pro běžné služby v internetu nás však bude více zajímat tento tvar:

<protokol>://<uživatel>:<heslo>@<hostitel>:<port>/<cesta>

Příklady: <http://www.misto.cz/special/index.html>
<ftp://jan.tajne@ftp.misto.cz:2121/soubor.zip>

Model klient-server

V sítích TCP se používá komunikace přes model označený klient-server. Systém je velice jednoduchý. Máme jeden počítač označený jako server, na němž běží některé ze služeb. Doma na svém počítači spustíme program - klienta, který se s tímto počítačem spojí a vyšle žádost. Pokud proběhly v pořádku všechny formality (jako je autorizace apod.), dostane klient odpověď, kterou například zobrazí jako HTML stránku.

Směrování, brány,...

Nyní bychom se krátce zmínili o směrování (routing) a souvisejících problémech, a to jen ve zkratce.

Podobně jako když jedete autem podle mapy to nějaké neznámé vesničky, nejdříve směřujete na nějaké větší město, tam najdete ukazatel na jiné, až se konečně propracujete k cíli. V Internetové síti existují jakési směrovací tabulky - takzvané brány - v nichž požadavek hledá cestu k cílové adrese přes jiné adresy, které jej k cíli postupně dovedou. IP adresy, které začínají určitým číslem půjdou sem, ostatní zase tam. K tomuto také slouží *síťová maska*. Ta říká, která část IP adresy udává směr - takzvanou „nadsíť“ - a která adresu počítače.

Virtuální hosté

Při prozkoumání IP adresy zjistíme jednoduchou matematikou, že adres není neomezeně, že jich je dokonce vzhledem k dnešní spotřebě příliš málo. Proto se zavádějí různé možnosti, jak obejít skutečnost připojení pouze jednoho doménového jména na jednu IP adresu. Používají se takzvaní virtuální hosté, to znamená, že k jedné IP adrese přiřadíme více názvů. Typické je například, že server v Internetu poskytuje služby ftp (File Transport Protocol) a http (www), přitom se jedná o jeden počítač, i když je označen dvěma doménami.

Správní organizace působící na Internetu aneb nositelé standardu

- Internet Society (ISOC)
Jedná se o společnost internetových profesionálů a expertů, kteří koordinují Internet a jeho schopnosti přizpůsobit se novým podmínkám (nové protokoly, standardy,...).
- Internet Architecture Board (IAB)
Technická poradní skupina při společnosti ISOC. Dohlíží nad architekturou a vývojem internetových protokolů.
- Internet Engineering Task Force a Internet Engineering Steering Group (IETF)
Zabývá se návrhem a vývojem protokolů pro Internet.
- Internet Research Task Force a Internet Research Steering Group
Je pracovní skupina při IAB, skládá se z různých skupin, které pracují v oblasti vývoje protokolů, aplikací, architektur a technologií pro Internet.
- Internet Network Information Centre (IncrNIC)
Zabezpečuje registraci Internetových domén.
- Internet Assigned Numbers Authority
Přiřazuje IP adresy a zabezpečuje jedinečnost všech adres domén.

Kde najít informace

Dokumenty RFC a Internet-Drafts

Technické dokumenty o Internetu existují ve dvou hlavních formátech: jednak jsou to dokumenty RFC (Request for Comment) a jednak takzvané Internet-Drafts. Dokumenty

RFC existují již od roku 1969, to jest od dob sítě ARPANET. Obsahují základní sérii technické dokumentace k Internetu. Tyto dokumenty nemají žádné speciální jméno, ale pouze se číslují (například RFC 2000 obsahuje „Internet Official Protocol Standards“). Tento dokument nahrazuje některé dřívější, ty se pak označují jako zastaralé. Proto se před čtením některého RFC dokumentu přesvědčte, zda je nejnovější.

Internet-Drafts představují další řadu technických dokumentů věnovaných Internetu, ale na rozdíl od RFC slouží spíše pro účely pracovních skupin jako je IETF.

Reference URL

<http://info.pvtinet.cz>

<http://info.isoc.org>

<http://www.ietf.org>

<http://www.cunet.cz>

<http://info.vtech.cz> – od června 2000

<http://www.isi.edu/iab>

<http://www.internic.net>

Reference knižní

Asi nejúplnější knihou, která se dá u nás na trhu o této tematice sehnat je kniha Dilipa C. Naika „Internet standardy a protokoly“ [Computer Press 1999, ISBN 80-7226-146-0].

Služba WWW (World Wide Web)

Nyní se dostáváme k hlavnímu tématu tohoto článku. Tato světová „pavučina“ je jedna ze služeb, která změnila asi nejvíce podobu Internetu (další taková je například elektronická pošta). Nedá se přesně říci zda k lepšímu, či horšímu. Nyní můžete jednoduše najít téměř vše co potřebujete, problémem je, že dnešní požadavky na multimedia jsou tak velké, že se občas ironicky říká také World Wide Wait.

Proxy server a firewall

Tyto pojmy jsou další novinkou několika posledních let. Jejich využití si ukážeme na následujícím příkladu:

Představme si modelový podnik, který se skládá z několika desítek počítačů. Obsahuje firemní databázi, WWW a FTP server. Všichni zaměstnanci mají přístup do Internetu. Podnik je do něj připojen pevnou linkou o kapacitě 64 Kbps (kilobound per second). Jakmile současně „brouzdá“ na Webu několik zaměstnanců, rychlost se prudce snižuje. Proto se tamější administrátoři rozhodli instalovat tzv. *proxy server*. Dříve šel každý požadavek od každého klienta do Internetu, i když se oba dva dívali na tutéž stránku. Nyní jsou všechny požadavky směřovány nejprve na proxy server. Ten zjistí, jestli požadovaný dokument u sebe nemá. Jestliže ano, pošle jej jako odpověď, když ne, pak se spojí do Internetu, dokument si uloží k sobě a odešle jej klientovi. Tímto se ušetří mnoho zbytečných spojení „ven“ (tj. mimo podnik) a spousta transakcí může probíhat interně.

Tento problém tedy správci vyřešili, ale nyní stojí před druhým: jak zajistit, aby se na pracovní stanice místních zaměstnanců, kteří mají IP adresu, nepřipojovali nezvaní návštěvníci. Na to jim poslouží právě firewall. Tato „protipožární zeď“ se umí dívat do paketu a buďto jejich požadavkům vyhovět, nebo je odmítnout.

Existuje několik typů firewallů, lišící se zejména cenou. V zásadě platí, že čím dražší, tím větší bezpečnost poskytuje. Mezi základní typy se řadí filtrační a aplikační zeď.

- **Filtrační:** dokáže se dívat cílovou a zpáteční adresu, na základě tohoto se může rozhodovat. Cena je od freeware až po 20 000 Kč.
- **Aplikační:** dokáže se dívat i do obsahu paketu a zjišťovat, zda nejde třeba o vir nebo nějaký útok. Cena je 50 000 Kč až po milióny.

Programování na Internetu

Pro upřesnění si ukážeme, jakým způsobem vlastně komunikuje Webový server s klientem. V nejstarší verzi protokolu HTTP/0.9, byla komunikace více než jednoduchá. Klient se pouze připojil na příslušný port serveru a zaslal dotaz ve formě:

```
GET /index.html
```

Jako odpověď obdržel obsah souboru index.html. Můžete si to sami vyzkoušet a zahrát si tak na prohlížeč. Stačí si jen v nějaké terminálové aplikaci připojit na server poskytující WWW a jako port zadat 80. (Malé upozornění: cesta za GET nesmí být prázdná, musí obsahovat alespoň „/“ tj. GET /).

Možná vás napadlo, jakou stránku nám server vrátí, když jako cestu zapíšeme jenom /. Odpověď závisí na individuálním nastavení každého serveru. Většinou to bývá tak, že vrátí soubor index.html (záleží na nastavení), pokud neexistuje, tak vypis adresáře (pokud je prohlížení adresáře povoleno).

Novější verze protokolů mají již syntaxi požadavků o něco komplikovanější. Obecný tvar vypadá takto.

```
<metoda> <URL dokumentu> <verze protokolu HTTP>  
<hlavičky>  
//prázdná řádka>
```

Příklad nejjednoduššího požadavku protokolu HTTP ve verzi 1.0 je:

```
GET /index.html HTTP/1.0  
// prázdná řádka
```

Metod je již více. Nejpoužívanější je stále GET, ale existují i další. Následující tabulka obsahuje metody protokolu HTTP/1.1:

Metoda	Popis
GET	Získání daného dokumentu ze serveru
POST	Odesílání formulářových dat na server, informace jsou v těle dokumentu.
HEAD	Zaslání hlaviček
PUT	Uložení stránky na server
DELETE	Odstranění stránky ze serveru
CONNECT, OPTIONS, TRACE	Slouží k analyzování spojení

Tato verze (HTTP/1.1), která má jednu důležitou vlastnost navíc. Jelikož je možné na jednu IP adresu přiřadit několik doménových jmen, je nutné říci Webovému serveru, ze kterého „hosta“ chceme příslušný dokument. Na to nám ve verzi 1.1 přibyla nová povinná hlavička „host“. Příklad nejjednoduššího požadavku protokolu HTTP ve verzi 1.1 je:

```
GET /index.html HTTP/1.1  
Host: www.misto2.cz  
// prázdná řádka
```

Nyní se ještě krátce zmíníme o formátu odpovědi. Na rozdíl od HTTP/0.9 je v novějších verzích odpověď poněkud „ukecanější“:

```
<protokol> <stavový kód> <stavové hlášení>  
<hlavičky>  
prázdná řádka  
<obsah odpovědi>
```

Protokol je verze protokolu HTTP, stavový kód je trojmístné číslo, které indikuje, jak se nám operace vydařila, přičemž hlášení je textový komentář k vrácenému kódu. Jistě se již každý z vás (mnohokrát) setkal s hlášením typu:

HTTP/1.1 404 Not Found

Neboli: požadovaný dokument nebyl nalezen. Stavové kódy protokolu HTTP lze rozdělit do několika základních kategorií:

Kategorie	Číslo kódu	Popis kategorie
Informační	100-199	Zprávy definované konkrétní aplikací
Úspěch	200-299	Požadavek byl úspěšně zpracován
Přesměrování	300-399	Klient musí pro konečné zpracování požadavku vykonat určitou další činnost
Chyba klienta	400-499	Problémy na straně klienta
Chyba serveru	500-599	Problémy na straně serveru

Nejčastěji se vyskytující stavové kódy:

Číslo	Slovní identifikátor	Popis kódu
200	OK	Operace proběhla bez chyb
201	Created	Požadavek metody typu POST byl kladně vyřízen
202	Accepted	Požadavek byl správně akceptován, ale dopovídající činnost se nemusela nutně provést
204	No Content	Požadavek byl úspěšný, ale klient nemá co zobrazit
300	Multiple Choices	Požadovaný zdroj se dá získat z několika různých míst
301	Moved Permanently	Požadovaná adresa URL je trvale přesunuta na novou
302	Moved Temporarily	Požadovaná adresa URL je dočasně přesunuta na novou
304	Not Modified	Podmíněný požadavek GET se zpracoval správně, dokument však od té doby nebyl modifikován.
400	Bad Request	Server nerozumí požadavku klienta
401	Unauthorized	Pokud byl původní klient anonymní, musí se autentizovat
403	Forbidden	Nesprávná autorizace
404	Not Found	Server nenašel zadanou adresu URL
500	Internal Server Error	Na serveru došlo k neočekávané chybě
501	Not Implemented	Tento požadavek není implementován v serveru
502	Bad Gateway	Proxy server nebo brána obdrželi od serveru neplatnou odpověď
503	Service Unavailable	Server dočasně nemůže nebo nechce zpracovávat požadavek; přetížení serveru, nebo údržba

Ve verzi protokolu HTTP 1.0 je definováno 17 hlaviček, ve verzi 1.1 je jich ještě o něco více. Hlavičky slouží k zaslání další doplňujících informací o dokumentu, jeho typu, datu poslední modifikace, jazyku, ... Pro přehled uvádíme jen ty nejdůležitější:

Hlavička	Popis
Date	Datum a čas vytvoří zprávy
From	Elektronická adresa uživatele, který pracuje s klientem
Host	Doménová adresa počítače, kterému klademe dotaz
Server	Server posílá svou identifikaci
Content-type	Udává typ přenášených dat
Cache-control	Zda se má dokument uložit do vyrovnávací paměti pro pozdější akce
Authorization	Informace potřebné pro autentifikaci uživatele

Jazyk HTML aneb jazyk Webu

Zatím jsme zde hovořili o HTML, ale ještě jsme si vůbec neřekli, co to vlastně je. Jak mnozí z vás jistě tuší, jedná se o základní stavební kámen Webu. Všechny stránky, které dnes můžete najít, jsou psány v HTML (HyperText Markup Language). Jazyk HTML se skládá z několika desítek entit (tagů), pomocí kterých můžeme formátovat text, vytvářet tabulky, odkazy, obrázky, formuláře, seznamy, atd. Veškeré programy, které jako své rozhraní (interface) používají WWW, generují kód obsahující formátovací příkazy ve tvaru HTML tagů. Psaní HTML stránek by se dalo přirovnat k psaní dokumentu například v systému $\text{T}_{\text{E}}\text{X}$. Tam je ovšem třeba výsledný dokument vytvořit kompilací, než jej bude možno zobrazit ve speciálním prohlížeči. Stránky HTML lze zobrazovat webovskými prohlížeči přímo, bez potřeby kompilace. Tyto prohlížeče fungují jako interprety. V této chvíli narazíme na problém, se kterým si láme hlavu spousta vývojářů. Každý prohlížeč si totiž interpretuje HTML po svém. Na jednoduchých stránkách to příliš nepoznáme, ale jakmile použijeme například styly CSS (viz. níže), pak se setkáme s dosti rozdílnými zobrazeními. V současnosti jsou jistým standardem dva konkurenční prohlížeče: Internet Explorer 5.0 od firmy Microsoft a Netscape Navigator firmy Netscape. První je šířen především spolu s operačním systémem Windows na počítačích třídy PC, zatímco druhý lze používat i v prostředí unixových operačních systémů.

Jednotlivé verze jazyka HTML jsou číslovány a aktuální verzi našeho dokumentu bychom měli uvést v hlavičce zdrojového textu stránky. Nejpoužívanější je dnes verze DHTML 4.0. Tato zkratka označuje *dynamické HTML*. Souvisí to s rozvojem javascriptu (viz. níže) a jednoduše řečeno, to znamená možnost změny vzhledu stránky interaktivně během jejího zobrazení. Novější verze jazyka podporují navíc takzvané kaskádové styly neboli CSS (Cascading Style Sheets), o kterých se také zmíníme později.

Další informace lze nalézt na řadě WWW stránek, z nichž si uveďme například <http://www.w3c.org> nebo <http://www.kosek.cz>.

Programování na straně serveru

První technologii, která umožňovala automatické generování stránek na straně serveru, byly CGI-skripty, jejichž název je odvozen z rozhraní CGI (Common Gateway Interface). Toto rozhraní se používá pro předávání dat mezi WWW-serverem a skriptem.

CGI

Když v devadesátých letech byl vytvořen jazyk HTML jakožto základ Webu, bylo nutné vyřešit problém, jakým způsobem předávat data přes Web, a tím vytvářet aplikace. Jinak řečeno, jakým způsobem poslat vyplněný formulář na server, tam spustit program a výsledek opět předat zpět klientovi. Právě pro tento účel bylo vytvořeno CGI, což je jakési základní rozhraní pro tyto operace.

Nenabízí nám mnoho, pouze proměnnou `QUERY_STRING`, v níž je uložena hodnota vstupních polí formuláře. Pokud chceme klientovi něco poslat zpět, pošleme nejprve hlavičku a pak příslušný obsah.

Budeme-li chtít klientovi například vrátit novou html stránku, pošleme jako hlavičku (*content-type*) na standardní výstup (*stdout*): `text/html`. Tyto hlavičky se označují jako takzvané MIME-TYPE (Multipurpose Internet Mail Extensions) a souvisejí s kódováním pošty. Existuje ještě několik dalších hlaviček, které můžete klientovi zaslat, nebo klient serveru (vaší aplikaci).

Content-type	Popis
Text/html	HTML stránka
Text/plain	Obyčejný text
Image/gif	Obrázek typu GIF
Application/msexcel	Tabulka aplikace MS-Excel
Application/x-wav	Hudební soubor WAV
Application/x-compressed	Komprimovaný soubor .ZIP .TGZ

Pro snazší pochopení uvádím následující příklad pro generování obrázku (výpis souboru `getobr.cgi`):

```
#!/bin/bash

SOBRAZEK="/usr/local/share/obrazek.gif"

echo "Content-type: image/gif"
cat $SOBRAZEK
```

Tímto řekneme klientovi, že následující data jsou definice obrázku typu gif.

Do HTML stránky jej pak vložíme následovně (výpis souboru `obr.html`):

```
<HTML>
<HEAD>
  <TITLE>Ukázka 1</TITLE>
</HEAD>
<BODY>
  <IMG SRC="getobr.cgi">
</BODY>
</HTML>
```

Díky řešení, které nám poskytuje CGI, můžete používat jakýkoliv programovací jazyk, který umí posílat data na standardní výstup a číst systémové proměnné operačního systému.

PHP

Historie systému PHP začíná zhruba v roce 1994, kdy Rasmus Lerdorf napsal jednoduchý systém pro evidenci přístupu ke stránkám. Tento systém se postupně vyvíjel až se z něj stal plnohodnotný skriptovací jazyk s velice bohatou knihovnou funkcí. Poslední plnou verzi PHP je 3.0, ostrá verze je k dispozici od června 1998. Přibližně jednou za měsíc je uvolněna nová podverze, která obsahuje opravy chyb předešlých verzí a často přidává podporu pro další nové knihovny. Aktuální nejnovější verze je PHP verze 4 beta 2. V porovnání s ostatními skriptovacími jazyky je PHP jedním z nejrychlejších (o verzi 4 se dokonce mluví jako o nejrychlejším skriptovacím jazyku). Koneckonců během půl roku se budeme moci přesvědčit sami, tou dobou by totiž měla vyjít konečná verze.

Práce s PHP je velice jednoduchá a intuitivní. Na místo psaní programů, které ve výsledku vygenerují HTML, používá PHP opačný přístup. Do HTML stránky můžeme zapsat příkazy PHP, jejichž výsledek se po interpretaci zkombinuje s okolním HTML kódem.

Syntaxe PHP je převážně odvozena z jazyka C. Některé rysy jsou převzaty z jazyků Perl a z Java -- např. asociativní pole a třídy. Kromě jednoduchého a mocného jazyka je velikou devizou PHP obrovské množství zabudovaných funkcí, které mimo jiné zahrnují:

- matematické funkce všeho druhu, včetně knihovny BC pro práci s dlouhými čísly;
- funkce pro práci s textovými řetězci, z nichž mnohé podporují češtinu (např. převod malá<=>velká písmena);
- podporu regulárních výrazů;
- funkce pro práci s datem a časem, pro převody dat mezi různými druhy kalendářů;
- práci se soubory a adresáři;
- síťové funkce: sockety, čtení záznamu DNS, převod mezi IP a doménovými adresami;
- odesílání elektronické pošty;
- práci s elektronickou poštou pomocí protokolu IMAP;
- přístup do adresářových služeb LDAP;
- generování obrázku GIF pomocí knihovny GD;
- čtení informací ze síťových zařízení pomocí protokolu SNMP;
- podpora souborově orientovaných databází: dBase, dbm a FilePro;
- podpora velkého množství databázových serverů pomocí nativních ovladačů: Oracle, Sybase, Informix, Solid, Adabas D, MS SQL Server, mSQL, MySQL a PostgreSQL;
- přístup k datům pomocí obecného rozhraní ODBC;
- funkce pro zpracování dat z webových formulářů;
- podpora cookies;
- funkce pro spouštění externích programů a zpracování jejich výstupu;
- podpora komprese a dekomprese dat pomocí metody používané programem gzip;
- od verze 3.0.6 bude v PHP zahrnuta podpora XML-parseru expat, takže zpracování XML-dokumentů pomocí PHP nebude žádný problém;
- a mnoho a mnoho dalšího.

Dosud nezmiňovanou a přitom velice podstatnou vlastností PHP je jeho nezávislost na platformě. PHP pracuje pod operačními systémy Unix, Windows 95/98/NT a Macintosh. Existují i projekty, které se snaží o portování PHP na různé starší mainframes. Kromě nezávislosti na operačním systému je PHP nezávislé na použitém WWW-serveru. PHP může spolupracovat s libovolným serverem, který podporuje rozhraní CGI -- to dnes podporují snad všechny servery. Pokud používáme Apache, může být PHP spuštěno jako jeho modul, čímž dosáhneme mnohem vyššího výkonu a větší bezpečnosti. Nynější verze PHP bude zahrnovat i moduly ISAPI (Microsoft), NSAPI (Netscape) a WSAPI (O'Reilly) a umožní běh PHP jako modulu na většině WWW-serverů pod Windows.

Přenositelnost skriptů napsaných v PHP je velkou výhodou. Poměrně často si dnes různé firmy a instituce pořizují server s Windows NT, časem ovšem zjistí, že to z hlediska stability a výkonu nebyla nejlepší volba. Pokud při vytváření dynamických stránek použijí PHP, není žádný problém na počítači nainstalovat nějaký ze systémů Unix podporujících platformu Intel (Linux, FreeBSD, Solaris) a stránky napsané v PHP provozovat vesele dál.

Vytváření jednoduchých skriptů v PHP se příliš neliší od tvorby běžných webových stránek. Pouze stránky ukládáme do souborů s příponou .php nebo .php3, aby se odlišily od stránek, které neobsahují příkazy pro PHP. Příkazy ve skriptu PHP zapisujeme mezi znaky (tagy) '<?' a '?>'.

Formuláře jsou vstupní branou do interaktivního světa Webu. Pomocí formulářů mohou uživatelé na stránkách zadávat objednávky, klást dotazy na různé databáze apod. Aby však nebyl formulář jen černou dírou, musí k němu existovat skript, který uživatelem zadaná data zpracuje a na jejich základě poskytne uživateli adekvátní odezvu.

Opravdovou lahůdkou, kterou nám PHP nabízí, je snadná spolupráce s databázemi. Skripty mohou pomocí jazyka SQL (Structured Query Language) provádět s údaji v databázích libovolné operace.

PHP kontra CGI

Výhodou CGI-skriptu je jejich nezávislost na použitém programovacím jazyce. Stačí, když daný programovací jazyk umí pracovat se standardním vstupem a výstupem. CGI-skripty se tak nejčastěji píšou v Perlu nebo příkazovém shellu případně v C/C++.

Použití CGI-skriptu je však v porovnání s PHP poměrně těžkopádné a přináší mnohem více potencionálních bezpečnostních rizik. Prvním problémem je už samotné dekodování dat získaných z HTML-formulářů. Data odesílaná prohlížečem jsou upravena tak, aby bezpečně „přežila přenos“ pomocí protokolu HTTP. PHP data automaticky rozkóduje do jejich původní podoby. V CGI-skriptu se o převod musíme postarat sami, nebo použít nějakou knihovnu, která tuto práci odvede za nás.

Další nepříjemností je poněkud těžkopádné kombinování HTML-kódu s kódem CGI-skriptu. Veškeré HTML zde musí být vygenerováno použitím příkazu `print` nebo jeho obdoby. To vývojáři přináší spoustu zbytečného psaní.

Asi největším problémem CGI-skriptů, zvláště pro začínající vývojáře, je umění napsat bezpečný skript, který neumožní jeho zneužití hackerem. CGI-skripty psané v příkazovém shellu nebo v Perlu lze ošálit odesláním vhodných dat. Pokud tato data obsahují speciální znaky, které například v shellu ukončují příkazy, může hacker spustit na serveru prakticky libovolný program. Správné ošetření všech pro shell nebezpečných znaků potřebuje notnou dávku zkušenosti.

Poslední argumentem, který hovoří pro PHP, je výkon. Pokud PHP běží jako modul serveru, je neustále zavedeno v paměti a může velice rychle obsluhovat všechny požadavky. Pro každý požadavek, který má být obslužen CGI-skriptem, musí být spuštěn nový proces, což není úkon zrovna nejrychlejší. CGI-skript může být rychlejší v případech, kdy provádí nějaké časově náročné operace -- zkompileované čecho je mnohem rychlejší než interpretované PHP. Nic však nebrání tomu napsat kritické rutiny v C a pomocí PHP API je zpřístupnit jako funkce pro skripty v PHP.

Na velice podobném principu jako PHP pracují i další skriptovací jazyky jako ASP (Active Server Pages) a ColdFusion. Jedná se však o komerční a nikterak levné systémy. Jejich další nevýhodou je malé spektrum podporovaných platforem. Dnes pracují pouze ve Windows a na přenosu do Unixu se teprve pracuje.

Více než Web

Díky velké flexibilitě Linuxu lze s PHP provádět opravdu netušené věci. Naznačíme alespoň některé možnosti. Napsat informační systém, který bude přístupný pomocí dynamických stránek generovaných pomocí PHP, není problém. Ale co když potřebujeme něco vytisknout? Tiskový výstup současných prohlížečů jistě nelze považovat za dostačující. Nic nám ale nebrání v umístění tlačítka *tisk* přímo na stránku. Stiskem tohoto tlačítka můžeme vyvolat skript v PHP, který spustí TeX a předá mu požadovaný dokument k vysázení. Perfektně zformátovaný dokument pak můžeme vytisknout na síťové tiskárně, která je nejbližší uživateli, který stiskem tlačítka vyvolal tisk.

Programování na straně klienta

JavaScript

Tato technologie roku 1996 od firmy Netscape znamenala veliký krok dopředu ve vývoji Webových aplikací. Na rozdíl od zatím standardního řešení pomocí pomalé a zdlouhavé komunikace přes CGI se vytváří odezva nikoliv na serveru, ale u klienta v prohlížeči. Kód javascriptu se dá zabudovat přímo do stránky nebo do speciálního souboru. Jeho revolučnost spočívá v možnosti komunikovat s jednotlivými elementy naší html stránky. Může kontrolovat hodnoty vstupních polí formulářů, měnit pozadí stránky, posouvat text nahoru a dolů, měnit obrázky po určité události (například pohyb myši), zkrátka existuje

nepřeberné množství funkcí, které se přes tento systém dají vyřešit elegantně, navíc spousta z nich ani jinak řešit nejde. Při tomto přístupu dostává uživatel mnohem rychlejší odezvu, díky menšímu počtu potřebného navazování spojení se serverem, než tomu je při použití CGI skriptů. Platí ale také, že spousta věcí se pomocí javascriptu vyřešit nedá, proto je nejlepší najít určitou harmonii a spojit tyto dvě technologie v jednu.

```
<html>
<head>

<script type="text/javascript" language="JavaScript">
<!--
function Potvrd_js() {
vek = document.formular.vek.value;

if ((vek < 1) || (vek > 105)) alert ("To se mi nějak nezdá");
else if (vek < 18) alert ("Mladík");
else if (vek == 18) alert ("Viš o tom že jsi plnoletý?");
else alert ("Už je ti " + vek + " let");
}
-->
</script>
<title>Příklad skriptu v jazyce Javascript</title>
</head>

<body>
<h1 align="center">Žkouska</h1>
<form name="formular" action="/php-bin/ukazka.php" method="get">
<div align="center">
<input type="Text" name="vek" value="0" size="3" maxlength="3">
<br><br><br>
<input type="Button" name="potvrd_js" value="Potvrd přes Javascript"
onClick="Potvrd_js();">
</div>
</form>
</body>
</html>
```

Příklad skriptu v jazyce javascript

Další informace lze nalézt na adresách

<http://developer.netscape.com>, <http://msdn.microsoft.com>, <http://www.kosek.cz>

Flash

Flash (technologie SHOCKWAVE) je grafický systém pracující na principu interaktivních vektorových animací. Pomocí tohoto produktu můžete vytvářet webové stránky, interaktivní animace, výukové pořady, filmy, či dokonce celé hry na webovských stránkách. Toto vše lze vhodným způsobem ozvučit.

Výsledný soubor, který vytvoříme pomocí jednoho z výrobcem dodávaných programů, je do prohlížeče vložen jako objekt a nabízí celou řadu možností dynamického zpracování a propojení s ostatními komponentami stránky. Tento systém není bohužel staršími verzemi prohlížečů podporován přímo, ale až po stáhnutí náležitého „plug-inu“. Příklad poslední verze Netscape Navigatoru, který tento „plug-in“ však již má ve své základní výbavě naznačuje budoucí možnosti.

Obecně platí, že FLASH je primárně určen pro animace, zatím co jiný produkt (firmy MACROMEDIA) DIRECTOR pro tvorbu náročnějších aplikací, her atd. V druhém případě je totiž k dispozici skriptovací jazyk Lingo, který se velmi podobá JavaScriptu či VBScriptu.

V obou případech ale platí, že výstup nemusí být primárně určen pro web, ale také pro multimediální CD-ROM či "jen" pro prezentaci v přehrávači Shockwave Flash (ten je samozřejmě součástí instalace Flash).

Základem pro práci s tímto programem je vektorová grafika, přestože import, animování a částečná úprava rastrové grafiky je podporována. Díky orientaci na vektory je výsledný soubor až neuvěřitelně malý při zachování velmi vysoké kvality (zvětšené vektory vypadají i na 21" monitoru dokonale).

Na závěr lze jen smutně podotknout: „Škoda, že to není freeware.“ Více o tomto produktu lze nalézt na domovské stránce firmy Macromedia na adrese <http://www.macromedia.com>.

Audio-Video

Multimediální prvky se na dnešním Webu objevují stále častěji, neboť jsou velmi dobrým způsobem jak přitáhnout pozornost náhodných návštěvníků vašich stránek.

Nejpoužívanějším formátem pro umístění multimédií na Webové stránky je již zmíněný Flash, který umí vše od obyčejných obrázků a zvuku přes videa až po celé aplikace. Jak je vidět, Flash je jednou z nejlepších variant, ne ovšem jedinou. Existují i další možnosti, jak vložit vaše video, nebo zvuk na Web. Můžete na to použít jednu z funkcí Javascriptu, nebo můžete použít některý nástroj od Microsoftu. Všeobecně ovšem platí, že vše co na stránku vkládáte se vkládá jako objekt, který musí daný prohlížeč umět zobrazit či přehrát.

Co se týká audia, je to téměř stejné. Nejjednodušší je umístění zvuku na pozadí stránky přímo příkazem HTML. Jakmile se někdo na takto upravenou stránku podívá, přehrává se mu tento zvuk automaticky. Podobně jako při výběru videa i v případě audio objektů je třeba věnovat pozornost výběru vhodného formátu. Nynější „čtyřkové“ verze prohlížečů již podporují tyto formáty: WAV, MIDI, MP3, MP2 aj. Nejperspektivnějším se v současné době jeví formát MP3. Tento formát je založený na zkomprimovaném „wavu“. Je to v této době snad nejoblíbenější formát. Díky jeho schopnosti zkomprimovat zvuk až na 1/12 téměř bez ztráty kvality, vám dovoluje umístit na Web mnohem delší skladby než ve formátu WAV a mnohem v lepší kvalitě než ve formátu MIDI. Pro ostatní formáty je třeba, aby měl váš prohlížeč nainstalovaný příslušný „plug-in“ (např. RealAudio).

XML

Co to XML vůbec je?

V angličtině tato zkratka znamená eXtensible Markup Language. Prakticky je to nový standard pro tvorbu webových stránek (nástupce HTML). O XML se mluví jako o velmi slibném formátu s velkou budoucností. Ne každý ovšem ví, co všechno formát XML zahrnuje.

Výhoda XML spočívá v tom, že autor stránky může používat vlastní tagy, čímž přesněji určí význam prezentované informace. To má význam především při vyhledávání informací a dnešní Web informacemi přímo přetéká, takže pokusy nalézt konkrétní údaj jsou stále obtížnější, mnohdy až nemožné. Tento problém nevyřeší sebelepší vyhledávací servery, pokud jim autoři stránek pomoci XML nepomohou.

V jazyce HTML musíme pomocí tagů především definovat, jak se má daný text zobrazit, a ne co znamená. V XML naopak určujeme především, co daná část textu znamená. Uvedme si krátkou ukázkou:

```
<knih>
  <název>Česko-anglický slovník</název>
  <autor>Ivan Poldauf</autor>
  <cena>560</cena>
  <měna>Kč</měna>
</knih>
```

Nyní víme, jaký význam mají jednotlivé části stránky, proto není problém zadat vyhledávacímu serveru příkaz k nalezení všech stránek, které obsahují informace o dané knize nebo autorovi.

Oproti jazyku HTML, kde jednotlivé tagy určují jak se má daná část stránky zobrazit, v XML žádná podobná přímá vazba neexistuje, a proto musíme zobrazení jednotlivých částí dokumentu definovat jiným způsobem – tj. pomocí stylů. Ve stylu pak můžeme pro každý element určit jak se má zobrazit (tj. stanovit typ a velikost písma, barvu, zarovnání, umístění apod.).

Velkou výhodou stylů je, že jsou to samostatné soubory které mohou být použity více než jednou stránkou, čímž docílíme jednotného vzhledu mnoha stránek. Pokud potřebujeme pozměnit design stránky, stačí drobně upravit styl a změna se promítne na všech stránkách, které styl používají.

Kaskádové styly (CSS) bylo možné využívat i v HTML. Jejich použití však bylo pouze doplňkem, kterým jsme mohli modifikovat vzhled některých elementů. Autoři stránek mohli zcela beztrápně "zneužívat" jednotlivé HTML tagy pro dosažení požadovaného vzhledu stránky. XML-dokument bez stylu nelze zobrazit, a tak autoři nemohou vytvářet špatně strukturované stránky (jinak dosahující požadovaného vzhledu).

Využití XML

A nyní trochu o tom, kde je vhodné XML nasadit a kde může použití XML aplikovat něco nového. O XML se nejčastěji hovoří jako o novém jazyku pro tvorbu webových stránek. Situace se dnes vyvíjí tak, že XML umožní rozšiřovat množinu elementů, které nám při tvorbě stránek nabízí HTML. Stránky v XML jsou zároveň snazší pro čtení než ty současné v HTML, které obsahují mnoho chyb. To umožní vývoj nových jednoduchých prohlížečů určených zejména pro různá kapesní mobilní zařízení. XML však nezůstává pouze technologií určenou pro Web. Využití nalezne všude, kde je potřeba jedny informace prezentovat v několika formátech: v tištěné podobě na papíře, jako publikaci na CD-ROM, nebo například na Webu. Výhoda XML spočívá v tom, že kromě samotného textu nese i informaci o jeho významu. Konverze do libovolného formátu je pak snadná a

může probíhat zcela automaticky. XML proto nalézá uplatnění při tvorbě technické dokumentace, což v některých oblastech průmyslu znamená práci s tisícistránkovými dokumenty. Příspěvky do různých vědeckých a odborných časopisů je dnes rovněž potřeba publikovat kromě tradiční papírové podoby i elektronicky. XML se nabízí jako otevřený standard pro uchovávání a výměnu tohoto typu dat. Využití XML může být výhodné i při klasickém publikování na papíru. Pokud budou rukopisy odevzdávány v XML, bude jejich import a zalomení v nějakém sázecím systému mnohem jednodušší než když import provedeme z textového editoru, kde může autor zcela neodborně měnit vzhled dokumentu.

Hovoříme-li o XML-dokumentech, může to v nás vyvolat dojem, že se XML hodí pouze pro textové dokumenty -- webové stránky, dopisy, knihy apod. Opak je však pravdou. Největší objem dat, který bude v podobě XML přenášen, budou strukturovaná data, která se dnes ukládají do relačních databází. XML poslouží jako vhodný přenosový formát při komunikaci mezi aplikacemi různých výrobců, mezi webovým serverem a prohlížečem apod. Dalo by se říci, že se jedná o universální databázi, která bude čitelná v každém programu.

V USA se připravuje pilotní projekt, který umožňuje odeslat daňové přiznání ve formátu XML bernímu úřadu. Díky tomu, že je definován požadovaný datový formát, může být k odeslání daňového přiznání použita libovolná aplikace - stačí když zvládne vyexportovat data do XML. Někdo si daňové přiznání vyplní ručně v obyčejném textovém editoru, někdo si může vygenerovat daňové přiznání přímo z účetního programu, který používá.

XML může obrovským způsobem pomoci při rozvoji elektronického obchodování. Při virtuálních nákupech je potřeba přenášet data mezi uživatelem, internetovým obchodním domem, bankou a certifikačními autoritami. Pro výměnu těchto dat může být použit formát založený na XML. Tím odpadne použití nekompatibilních a drahých řešení.

Již několik let se úspěšně používá formát EDI (Electronic Data Interchange). Problémem EDI je jeho složitost a nekompatibilita. Každá velká firma má svoje vlastní řešení EDI, pokud tedy nějaký dodavatel spolupracuje s několika odběrateli, musí si za velké peníze pořídit tolik EDI aplikací, kolik má odběratelů.

Definice stylu

V úvodu článku jsme naznačili, že je potřeba definovat styl, aby dokument dostal grafickou podobu. Pro definici vzhledu XML-dokumentů se dnes používají tři stylové jazyky: CSS, XSL a DSSSL.

CSS (Cascading Style Sheets) není nic jiného než kaskádové styly používané v HTML. Tento stylový jazyk již podporují dnešní „čtyřkové“ verze prohlížečů. Kaskádové styly byly původně určeny pouze pro prezentování informací na obrazovce. Další parametry formátování, které je potřeba zadat pro tištěný výstup, byly přidány do nové verze kaskádových stylů CSS2. Pro potřeby XML byly do CSS2 přidána i další rozšíření, která umožňují konkrétní styl aplikovat i na základě hodnoty atributu. CSS1 umožňovalo styl aplikovat pouze na určitý element bez návaznosti na obsah jeho atributů.

Opakem k jazyku CSS je DSSSL (Document Style Semantics and Specification Language). DSSSL je tak rozsáhlý jazyk, že jej zatím nepodporuje příliš mnoho aplikací. Pro potřebu elektronického publikování byla definována jeho ochuzená verze DSSSL-O, která se hodí pro méně náročné aplikace.

Zlatou střední cestou mezi oběma jazyky je XSL (eXtensible Stylesheet Language). Tento jazyk má možnosti srovnatelné s DSSSL, jeho syntaxe je však mnohem jednodušší. Zajímavostí je, že XSL se zapisuje v jazyce XML.

Jazyk XSL má řešení i pro náročné, totiž ECMAScript. ECMAScript je standardizovaná verze JavaScriptu a v XSL jej můžeme použít pro manipulaci s dokumentem - například můžeme automaticky vygenerovat obsah dokumentu apod.

Styly lze zapsat pomocí DTD (Document Type Definition). V DTD určíme jaké elementy může dokument obsahovat, v jakém mohou být vzájemném vztahu a jaké atributy může každý element mít.

Dnes existují desítky a možná i stovky DTD, které se běžně používají. Pokud pro naše dokumenty použijeme některé standardní DTD, otevírá se nám možnost využívání dalšího softwaru, který umí s dokumenty daného typu pracovat. Následující přehled obsahuje některé nejběžnější typy DTD a oblasti jejich použití:

Hypertext Markup Language (HTML)

Jazyk HTML není třeba nikomu představovat. Existují DTD pro jeho jednotlivé verze 2.0, 3.2 a 4.0. V současné době se pracuje na rozložení jednoho DTD HTML 4.0 na několik nezávislých částí. Každé DTD pak bude obsahovat určitou část funkčnosti HTML -- např. formátování, formuláře či tabulky.

Channel Definition Format (CDF)

Formát CDF byl poprvé použit pro definici kanálů v prohlížeči Microsoft Internet Explorer 4.0. Kanálová mánie a všechno bláznění kolem push-technologií dnes již poněkud pomínulo, ale na miliónech počítačů po celém světě nalezneme soubory CDF.

Mathematical Markup Language (MathML)

MathML je jazyk určený pro zápis matematických výrazů. Kromě přesného vzhledu výrazu dokáže zachytit i jeho přesný význam. Použití nalezne zejména při zařazování matematických vzorců do webových stránek a při výměně dat mezi různými matematickými programy jako Maple a Mathematica a webovými stránkami.

Resource Description Framework (RDF)

RDF je standardem z dílny konsorcia W3C. Definuje standardní rámec, který je možné využít pro přidávání metainformací k datovým zdrojům. Praktické využití je zejména pro zlepšení práce vyhledávacích strojů a agentů, navigačních aplikací a aplikací pro hodnocení obsahu stránek.

Open Software Description Format (OSD)

Formát OSD slouží pro popis softwarových aplikací a jejich jednotlivých částí. Cílem je usnadnit automatickou distribuci a obnovu (update) softwaru.

Chemical Markup Language (CML)

CML je jazyk určený pro snadný zápis vzorců chemických sloučenin. Kromě standardu pro výměnu dat mezi chemickými aplikacemi umožní zařazování chemických vzorců na webové stránky mnohem úspornějším a efektivnějším způsobem než pomocí bitmapových obrázků.

Synchronized Multimedia Integration Language (SMIL)

SMIL slouží pro tvorbu multimediálních prezentací. Pomocí jazyka můžeme definovat jednotlivé zdroje audio a video dat, jejich umístění na obrazovce a čas, ve kterém budou jednotlivé části prezentace přehrány.

Precision Graphics Markup Language (PGML)

PGML je návrh jazyka určeného pro zařazování dvourozměrné vektorové grafiky na webové stránky. Vychází z PostScriptu a PDF, syntaxe je však upravena tak, aby vyhovovala XML.

UML eXchange Format (UXF)

UXF slouží k výměně dat v jazyce UML. UML (Unified Modeling Language) je standardní jazyk používaný při analýze a návrhu informačních systémů.

DocBook

DocBook je DTD vhodné pro psaní technické dokumentace. Jedná se o velice rozšířené DTD, které je přímo podporováno mnoha aplikacemi.

Scalable Vector Graphics (SVG)

SVG je vektorový formát pro dvourozměrnou grafiku určený zejména pro webové stránky. Tento formát má dnes největší předpoklady stát se standardem pro vektorovou grafiku na Webu. Na jeho vývoji pracuje konsorcium W3C.

Text Encoding Initiative (TEI)

TEI je rozsáhlý projekt, jehož cílem je vytvoření skupiny standardizovaných DTD vhodných pro uchovávání a výměnu knih. Na několika univerzitách ve světě probíhají projekty, které převádějí klasická díla (např. Shakespearova) do elektronické podoby právě pomocí TEI.

Continuos Acquisition and Lifecycle Support (CALS)

Skupina několika DTD a dalších formátů pro výměnu dokumentace mezi Ministerstvem obrany USA a jeho dodavateli. Jedno z DTD je vhodné pro zápis tabulek a v praxi se používá velice často.

Hypertext v XML

Hypertextové odkazy tak, jak je známe z HTML, byly v době svého vzniku převratnou technologií, která se dnes dostala na stoly miliónů uživatelů. Po pár letech jejich používání se však objevily nové požadavky. Problematiku odkazů v XML-dokumentech řeší samostatné jazyky XLink a XPointer, které někdy bývají souhrnně označovány jako XLL.

XLL samozřejmě umožňuje vytváření jednoduchých odkazů, které známe z HTML. Přichází však i s velkou novinkou. Odkaz nyní může vést i na tu část stránky, která neobsahuje žádné návěští. Odkazy v XLL můžeme například odkázat na pátý odstavec, jež se jmenuje „Ze života hmyzu“. Tímto způsobem můžeme vytvářet odkazy i na ta místa dokumentu, kam autor neumístil žádné návěští.

Další novinkou je možnost vytvářet obousměrné odkazy, které jasně vyjadřují souvislost dvou stránek. Problémem nejsou ani odkazy, kdy jeden odkaz ukazuje najednou na několik dalších zdrojů. Uživatel pak má možnost vybrat si zdroj, který jej nejvíce zajímá.

Odkazy lze pomocí XLL ukládat i zcela mimo dokumenty, kterých se týkají. Tímto způsobem můžeme vytvářet například anotace ke stránkám, ke kterým nemáme jinak přístup pro zápis. Na svém počítači nebo na firemním serveru můžeme mít databázi odkazů, se kterou bude spolupracovat náš prohlížeč, a budeme pak mít na cizích stránkách přístupné anotace vlastní i našich kolegů. Není vyloučeno, že se najdou provozovatelé serverů, které budou spravovat veřejné databáze odkazů. Stránky tak půjde snadno doplňovat o další informace, které mohou být užitečné pro všechny uživatele Webu.

Pár slov nakonec o XML

Formát XML má opravdu takové možnosti, že by zabrali minimálně desetinásobek tohoto příspěvku. Toto má být, ale pouze shrnutím základních informací o tomto formátu. Myslím, že s XML se již pravidelně setkáme na Webu již během roku. Je formát s obrovskou možností využití a s hlavní myšlenkou kompatibility systémů a dat, se kterou se v nynější době potýká snad každý. Představa, že si kdokoli z nás sedne k jakémukoli počítači, pustí si jakýkoli program a bude pracovat s daty, která si napsal doma na svém počítači je jisté pro každého příjemná.