

O jednom nekonvenčním vědecko-informačním centru

Petr Klán, Ústav informatiky AV ČR, Praha, pklan@cs.cas.cz

Abstrakt: Na Univerzitě Pardubice bylo založeno vědecko-informační centrum chemické informatiky. Účelem je zejména rychlý přístup k chemickým informačním zdrojům na WWW a také jejich efektivní vytváření. Smyslem příspěvku je diskutovat programové vybavení centra hlavně vzhledem k používání WWW, neboť vybavení a přístupy jsou zde stejné bez ohledu na obor. Podělit se o zkušenosti, co je všechno potřeba k plnohodnotnému využívání WWW na počítačích takového centra instalovat a o čem všem by měl mít informační pracovník minimálně povědomí. Lépe však znalosti podložené vlastními zkušenostmi.

Několik technických údajů o centru

Vědecko-informační centrum chemické informatiky vzniklo na základě finanční podpory projektu MŠMT LI200008. Centrem se rozumí multifunkční učebna s devíti multimediálními klientskými počítači a serverem sloužící zejména ke zpřístupňování chemických informací a chemických informačních zdrojů v národním i mezinárodním měřítku. Slouží univerzitním pracovníkům, studentům a odborné chemické veřejnosti. Běží zde výuka několika předmětů a konají se zde vědecké a odborné semináře v souladu se zaměřením centra. Centrum umožňuje snadný a rychlý přístup pracovníků univerzity, studentů i chemické veřejnosti k chemickým informačním zdrojům Webu (WWW).

Základ technického vybavení centra tvoří počítače firmy Dell typu Dell Optiplex GX 115 MT, jejichž duši tvoří procesor Intel Pentium III s kmitočtem 733 MHz a rychlou 256 kB Cache pamětí. Každý počítač má dále paměť SDRAM o velikosti 128 MB, pevný disk 10 GB, floppy jednotku, kvalitní grafickou kartu ATI All in Wonder 128, 32 MB a velký, přehledný 17'' monitor Dell E770. Rychlý přístup na Internet zajišťuje integrovaná síťová karta s přenosovou rychlostí 100 Mb/s. Součástí centra je také společná laserová tiskárna, kde je možné získané informace bezprostředně tisknout. Server je typu DELL PowerEdge 1400 s procesorem Intel Pentium III, 800 MHz s 512 MB RAM, CD a floppy mechanikou.

Počítače jsou zatím vybaveny operačním systémem Windows 98. Ochranu proti virům zajišťuje antivirový program AVP (firmy PCS software) a systémovou politiku chránící instalované programy před poškozením a přesně vymezující práva uživatelů program OptimAccess (firmy SODAT software).

Páteřním programem pro lokální aplikace je ChemOffice, což je ve světě standardně používaný soubor programů pro práci s chemickými informacemi, který zahrnuje zejména profesionální kreslení chemických struktur ChemDraw (výstupy tohoto programu běžně akceptují chemické časopisy ve světě), databázový program ChemFinder s 250 000 chemickými reakcemi, 180 000 sloučeninami a program Chem3D Ultra pro vizualizace a výpočty.

Podstatný však je rychlý a efektivní přístup k WWW z následujících úhlů pohledu:

- ❖ WWW je moderní a dominantní informační zdroj. Množství dat na Webu se zvyšuje exponenciálně.
- ❖ WWW je účinný komunikační prostředek. Na WWW je možné provozovat elektronickou poštu a přenášet počítačové soubory.
- ❖ WWW je prostorem k publikování informace. Především HTML, XHTML, JavaScript a PHP tvoří pilíře pro publikování na WWW.

O činnosti centra

Centrum je prvotně zaměřeno na poskytování výuky studentům a provozování informačního WWW zdroje „Česká chemie na Internetu“.

Výuka

Doménou centra je výuka Chemické informatiky. Výuku vede autor příspěvku a běží v zimním i letním semestru a je z jedné poloviny zaměřena na práci se stávajícími informačními zdroji a z druhé poloviny na tvorbu vlastních informačních zdrojů. Jinak řečeno, studenti se polovinu semestru seznamují se současnými chemickými WWW zdroji – stávají se jejich uživateli – a druhou polovinu se stávají tvůrci vlastních WWW zdrojů do té míry, že umí pracovat s HTML, XHTML a získají základy jazyka JavaScript zaměřené na tvorbu vyhledávacích databází.

Dalším předmětem, který se v centru vyučuje je webové programování. Studenti zde získávají základní webovou gramotnost tvůrce WWW zdrojů v podobě jazyků XHTML, JavaScript a PHP. Tento předmět trvá jeden semestr a vede ho rovněž autor příspěvku. Vzhledem k tomu, že jde o chemické aplikace, je výuka zaměřena hlavně na základy tvorby WWW zdrojů v podobě databází a práce s nimi a seznámení se s různými datovými formáty.

V centru však běží i další předměty zaměřené zejména na matematické modelování chemických procesů jako automatické řízení, identifikace systémů, fuzzy řízení apod. Zde studenti pracují hlavně s programem MATLAB, jehož výuku centrum rovněž zajišťuje. Uvažuje se rovněž o sobotních kurzech webového programování pro odbornou chemickou veřejnost. Hlavní pomůckou by se měla stát publikace [1].

Česká chemie na Internetu

Centrum zajišťuje provoz a průběžné aktualizace tohoto informačního zdroje na WWW adrese chemie.upce.cz. Zdroj se snaží poměrně podrobně mapovat chemické informační zdroje v ČR. Součástí je i rozsáhlá vyhledávací databáze chemikálií vyráběných v ČR, která vzniká ve spolupráci se Svazem chemického průmyslu. Součástí je i zpravodaj „MONITOR“ snažící se mapovat nové významnější chemické WWW zdroje u nás i v zahraničí. Vytváření takového zdroje vyžaduje programovou podporu jak ze strany HTML (XHTML), tak JavaScriptu a PHP. Autor dokonce uvažuje o jisté automatizaci, kdy by data doplňovali samotní uživatelé, neboť aktualizace jsou časově poměrně náročným procesem.

Zároveň je sledována a analyzována návštěvnost centra (viz chemie.upce.cz/usage/), která činí průměrně 100-110 návštěv denně. Analyzuje se i návštěvnost jednotlivých WWW stránek tohoto informačního zdroje. Největší zájem je jednoznačně o vyhledávání v databázi chemikálií. Celkovou měsíční návštěvnost a počet vyhledávání v databázi chemikálií za rok 2002 uvádí následující tabulka:

Měsíc (2002)	Návštěvnost celkem	Počet vyhledávání celkem
Únor	2085	1125
Březen	2411	1202
Duben	2357	1244
Květen	2768	1832
Červen	2073	1188
Červenec	1780	742
Srpen	1997	786
Září	2559	1170
Říjen	3114	2310
Listopad	3282	2857

Prosinec	2802	1383
----------	------	------

Z toho je vidět, že průměrně každý 1,5tý až 2hý návštěvník vyhledává v databázi chemikálií. Z pohledu návštěvnosti jednotlivých WWW stránek je největší zájem o WWW stránky škol a univerzit nabízejících chemické vzdělání, WWW stránky s chemickou informatikou, chemickým průmyslem a MONITOR. Hodně se využívají také WWW stránky mapující význačné světové informační zdroje chemiků.

Uživatelé centra a informační zdroje Webu

Uživatelé informačních zdrojů Webu ke své činnosti potřebují nějaký WWW prohlížeč. Pomocí WWW prohlížeče mohou:

- ❖ Prohlížet informační WWW zdroje (WWW stránky).
- ❖ Zřizovat na poštovním WWW serveru schránky elektronické pošty. Po té posílat a přijímat elektronické dopisy. Účastnit se elektronických diskusních skupin.
- ❖ Pracovat s interaktivními WWW zdroji. Vyplňovat a odesílat elektronické formuláře (cvičný příklad elektronického formuláře je na WWW adrese www.cs.cas.cz/~pklan/formular.htm).
- ❖ Posílat elektronické dopisy jako reakci na obsah WWW zdrojů.
- ❖ Hledat informace v databázích vyhledávačů (podle klíčových slov) a dalších specifických databázích (viz databáze chemikálií na WWW adrese chemie.upce.cz).
- ❖ Provádět matematické výpočty podle vlastních dat (cvičný příklad na odborný výpočet je na WWW adrese www.cs.cas.cz/~pklan/odbvyp.htm).
- ❖ Posílat SMS zprávy na mobilní telefony (viz SMS brána chemie.upce.cz/sms.htm).
- ❖ Nahrávat počítačové programy (downloads).
- ❖ Používat službu FTP (File Transport Protocol) k přenosu počítačových souborů.

Jako nejpoužívanější činnosti uživatelé informačních WWW zdrojů uvádějí vyhledávání informací, používání elektronické pošty a posílání SMS zpráv.

Jaký WWW prohlížeč?

Každý uživatel preferuje jiný, důvody bývají racionální i iracionální. Proto se osvědčilo současně instalovat více WWW prohlížečů. Nejpoužívanějšími WWW prohlížeči jsou Netscape Navigator a Internet Explorer. Podrobnější informace lze získat na WWW adresách home.netscape.com/ a www.microsoft.com/ie/. Existují však i další volně dostupné WWW prohlížeče, které uživatelům pomáhají:

- ❖ **Lynx** je jednoduchý WWW prohlížeč (tzv. nongraphical browser - nezobrazuje grafiku) popsán na WWW adrese lynx.browser.org. Je volně dostupný na WWW adrese www.crl.com/~subir/lynx/.
- ❖ **Mozilla** je jádrem Netscape Navigatoru na Web adrese www.mozilla.org. Obsahuje celkem příjemný HTML editor.
- ❖ **Opera** je na WWW adrese www.opera.com/download/. Češtinu pro tento oblíbený WWW prohlížeč lze doplnit na www.sw.cz/pspad/opera.htm.
- ❖ **Amaya** je moderní WWW prohlížeč konsorcia W3C. Amaya plní úlohu referenčního WWW prohlížeče, např. při experimentech s XHTML. Co umí Amaya by měl umět i jiný WWW prohlížeč. Amaya také obsahuje HTML editor. WWW prohlížeč/editor Amaya je volně dostupný k instalaci na WWW adrese www.w3.org/Amaya/ a uživatelská příručka je na WWW adrese www.w3.org/Amaya/User/Manual.html. Přístup k prohlížení i editování Web stránek je zde v rámci jednoho procesu. Amaya je kromě podpory XHTML známá ještě jedním nástrojem. Podporou značkovacího jazyka MathML (Mathematical Markup Language), který umožňuje umístění matematických vzorců do WWW zdrojů.

Porovnání WWW prohlížečů a další zdroje lze nalézt na WWW adrese browserwatch.internet.com/browsers.html.

Jak na WWW hledat?

WWW však svým uživatelům slouží především jako informační zdroj. Lépe informační ráj. Problém může být ve způsobu, jak rychle zacílit hledanou informaci a neztratit přitom orientaci v tom závratném množství dat. Uživatelé většinou „brouzdají“ nebo „vyhledávají“.

Pokud uživatelé nemají vyhraněný zájem o konkrétní informaci a jde jim spíše o informace z nějaké oblasti, zpravidla používají způsob prohlížení Webu, kterému se říká „brouzdání“. Spočívá v náhodné navigaci mezi WWW stránkami podle okamžitého zájmu. S brouzdáním se zpravidla začíná na nějaké specifické WWW stránce. Na Webu k tomuto účelu existuje řada tzv. startovních WWW stránek (starting points, starting point pages). Takovými startovními stránkami bývají např. WWW stránky na adrese www.stpt.com. U nás třeba Český informační systém na www.czis.cz.

Asi nejčastější je situace, kdy uživatelé hledají nějaké konkrétní informace a touží je najít co nejrychleji. Cílem je objevit WWW zdroj, kde se hledaná informace nachází. K tomu účelu má Web tzv. *vyhledávače* (search engines). Vyhledávače především umožňují *katalogové* vyhledávání. Sbírají a ukládají informace o WWW zdrojích do svých katalogů. U každého zdroje evidují jeho WWW adresu a stručnou charakteristiku obsahu. Kromě toho vyhledávače ještě nabízejí tzv. *fulltextové* vyhledávání, které umožňuje vyhledávat v celém obsahu WWW zdroje.

Uživatelé centra kromě českých vyhledávačů používají zejména vyhledávač *Google* na WWW adrese www.google.com. Vyhledávač podle IP adresy rozpozná zemi původu uživatele a podle toho volí i komunikační jazyk. V našem případě češtinu. Podobným vyhledávačem této kategorie je i *AlltheWeb* na WWW adrese www.alltheweb.com. Vyhledávač Google umí také pracovat jako obrázkový vyhledávač. Vyhledává pouze obrázky. Tato možnost je k dispozici na WWW adrese images.google.com nebo přímo v Googlu použitím ikony Obrázky (Images).

V souvislosti s vyhledávači je dobré zmínit ještě jeden, velmi efektivní a u uživatelů oblíbený způsob vyhledávání. Poskytují jej tzv. meta vyhledávače. Ty umožňují souběžné vyhledávání v několika vyhledávacích najednou. Asi „nejmocnější“ a uživateli nejpoužívanější je tzv. „třicetsedmička“ na WWW adrese www.37.com. Vyhledává současně v 37 vyhledávacích.

Tipy pro vyhledávání

Na základě zkušeností uživatelů lze formulovat několik tipů, jak informaci na WWW vyhledávat.

- ❖ Specifikovat cíl vyhledávání do jednoho či několika výstižných klíčových slov nebo fráze.
- ❖ Hledání začínat v obecně nebo celosvětově zaměřených vyhledávacích. Vhodnými jsou např. Google, 37, AlltheWeb.
- ❖ Pokračovat v tématicky nebo místně orientovaných vyhledávacích. U nás je to např. Centrum nebo Seznam.
- ❖ Pokud se podaří klíčovými slovy přesně vyjádřit povahu hledané informace, je vhodnější katalogové vyhledávání. Při fulltextovém vyhledávání lze zase vyhledat dokumenty, ve kterých se hledaná informace někde v obsahu vyskytuje, avšak nemusí být v katalogovém záznamu. Vyhledaných fulltextových odkazů je proto zpravidla několikanásobně více.
- ❖ Nalezené informační zdroje či informace zaznamenávat úměrně tomu, kde je informace potřeba. Pokud je určena k veřejnému publikování, je třeba odpovídající zdroj citovat.
- ❖ Záznam provádět v původním jazyce, později nebudou pochybnosti o překladu.
- ❖ Při vyhledávání se osvědčuje pečlivá a systematická práce, přesný záznam WWW adres a zaznamenávání více nezávislých zdrojů vzhledem k ověření kvality informace. WWW servery rychle vznikají, ale také často rychle zanikají.

Aby se vyhledávání stalo co nejvíce pohodlným, je přímo součástí některých WWW prohlížečů. Příkladem je WWW prohlížeč Opera, kde je vyhledávání přímo v pruhu pro adresu WWW stránky vpravo. Tento WWW prohlížeč je velmi oblíbený hlavně mezi studenty.

Nejčastěji používané informační zdroje

Následuje několik klíčových vyhledávacích informačních zdrojů, které uživatelé centra statisticky nejvíce používají:

- ❖ Katalogy našich předních knihoven: Národní knihovny Praha na WWW adrese www.nkp.cz, postupně položky „Katalogy a databáze“ a „Elektronický katalog NK-báze NKC“ a Státní technické knihovny Praha na WWW adrese www.stk.cz, položka „On-line Katalogy“. Další knihovny jsou na WWW adrese chemie.upce.cz/a_liter.htm#knihovny. Podobný systém vyhledávání funguje i ve vysokoškolských knihovnách. Např. pro knihovny ČVUT je katalog na WWW adrese knihovny.cvut.cz, položka „Katalog“.
- ❖ *Web of Science* na WWW adrese wos.cesnet.cz. Jde o možnost prohledávat obrovskou databázi časopiseckých článků a citací amerického institutu ISI (Institute for Scientific Information). Jsou zde abstrakta článků z mnoha světových i našich časopisů. Při řešení odborného, technického nebo vědeckého problému je vždy prvním krokem rešerše v dostupné literatuře. *Web of Science* může být v této fázi skvělým pomocníkem. Databázi je možné prohledávat formou různých dotazů (topic, author, source title, address). Tak lze velmi rychle zjistit stav daného oboru, nejnovější literaturu apod. Jak bylo zmíněno, jsou tam abstrakta článků včetně dalších údajů. Pokud se zjistí, že je určitý článek potřeba úplný, je to žádost na nejbližší knihovnu.
- ❖ *Science@ Direct* na WWW adrese www.sciencedirect.com. Slouží pro bezprostřední přístup k plným verzím časopiseckých článků. Jde sice o placenou službu, ale i těch časopisů, jež jsou přístupné zdarma je zajímavý počet. Rozhodně stojí za to se podívat, zda tam uživatel nenalezne časopis pro sebe. Seznam časopisů je pod položkou „Journals“. Plné texty článků jsou v PDF formátu. Je možné je přímo tisknout.
- ❖ Pokud uživatel hledá nějakou dostupnou literaturu o WWW, jako první zkusí některý vyhledávací stroj (např. AlltheWeb) a frázi „WWW Primer“. V češtině existuje na Webu a o Webu velmi pěkná publikace P. Satrapy: *WWW pro čtenáře, autory a misionáře* na WWW adrese www.kit.vslib.cz/~satrapa/docs/wwwbook/. Na domovské stránce P. Satrapy www.kit.vslib.cz/~satrapa/ je možné najít i další související texty a literaturu. „*Téměř vše o WWW*“ je také název informačních WWW stránek Jirky Koska na www.kosek.cz s mnoha čtivými a populárními články na toto téma. Velmi pěkný tutoriál o Webu v anglickém jazyce je na WWW adrese www.webteacher.org, položka „Web basics“.

Chytré vyhledávání

Rok 2000 lze (snad ne s příliš velkou nadsázkou) označit za počátek věku Webu a vůbec počátek věku internetového či „sít'ového“. Hmatatelným důsledkem působení Sítě začíná být informační exploze. A těžko uchopitelným problémem, jak využívat data na WWW a jak v nich efektivně vyhledávat.

Vyhledávače AlltheWeb a Google hlásí více než 2 miliardy záznamů v databázi, která se prohledává. Proto nebývá neobvyklé na dotaz obdržet několik tisíc záznamů a více. Zejména při fulltextovém vyhledávání. Vyhledávač se obvykle snaží vyhledané záznamy nějak seřadit. Z pohledu jejich potenciální užitečnosti vzhledem k položenému dotazu. Pokud je dotaz příliš obecný, např. „chemie“, vrátí Google téměř 6 milionů záznamů. Při 30 vteřinách na každou nabídnutou WWW stránku to dělá cca 6 let nepřetržitého prohlížení. Je dobré proto zejména začínající uživatele upozorňovat na možnosti tzv. chytrého vyhledávání, které omezí počet vyhledaných stránek.

Především se ukazuje, že užitečným nástrojem pro „obohacování“ informacemi může být nejen vyhledávání, ale i prosté brouzdání. Dotazem totiž uživatel obdrží pouze tu informaci, na kterou se

zeptá. Ovšem obohatit dokáží i informace, získané náhodně při brouzdání. Získávání těchto náhodných informací je však zatím na Webu mnohem méně propracované než získávání cílené informace. A přitom jde o oblast neméně důležitou.

Pozornost je třeba upřít i na zlepšení vyhledávání informace. Zatím vyhledávání probíhá zpravidla tak, že uživatel vyhledávači formuluje dotaz typu:

Najdi odpovídající WWW stránky, kde se vyskytují tato klíčová slova (fráze)

K tomu se postupně přidává další způsob. Někde existuje soubor WWW stránek určitých autorit (dobře popisujících např. určitý obor). Na tyto WWW stránky se potom odkazuje nějaká skupina zájmových WWW stránek. Lze to zjednodušeně přirovnat ke „klubu fandů“ či k nějaké komunitě, kterou pojí společný zájem. Příkladem může být např. skupina autorizovaných WWW stránek hudebních vydavatelství třeba s relaxační hudbou. A existuje společenství přívrženců této muziky, kteří mají své WWW stránky odkazující se právě např. na jednotlivé skladby na autorizovaných WWW stránkách. A třeba i mezi sebou. A velmi užitečným se při vyhledávání může stát nalezení takového společenství. S podobným zaměřením, které uživatel hledá. V takové komunitě lze totiž objevit spolehlivé a aktualizované informace.

Jiným zdrojem informací se stávají tzv. *weblogy*. Jde o zpracování dotazů uživatelů vyhledávačů. Tady je možné získávat statisticky významné trendy (jakousi sociologii) např. o oblíbenosti jednotlivých osob (populární osoba bude zřejmě častěji vyhledávána), náladách společnosti, poruchovosti výrobků atd.

To vše formuje novou úlohu centra. Centrum budoucnosti jednak může pružně reagovat na „sociologii“ svých uživatelů a jednak hrát podstatně větší úlohu při třídění nalezené informace. A právě zde se jeví jako nezbytné maximálně podporovat tvůrčí přístup k informačním zdrojům a nástroje, které s tím souvisejí.

Uživatelé centra jako tvůrci informačních zdrojů Webu

Tvůrce připravuje informační WWW zdroje za použití značkovacích jazyků HTML nebo XHTML. Pokud nezná HTML nebo XHTML jazyk, je vhodné použít nějaký HTML editor, který může proces tvorby WWW stránek výrazněji urychlit. Profesionálové zpravidla kombinují ruční výrobu s podporou HTML editorů v zádech.

Takto vytvořené informační zdroje se ukládají na WWW serverech. Tam jsou přístupné celé komunitě Webu. Tím proces jejich publikování končí. Pokud totiž tvůrce umístí informační zdroj na WWW server, stává se veřejným dokumentem. To je důležitá skutečnost. Je otevřený mezinárodní audienci, tj. mohou jej prohlížet WWW prohlížeče odkudkoliv z Internetu. Stává se oknem, které tvůrce reprezentuje vůči světu.

Právě pro tvůrce informačních zdrojů je na serveru centra instalován WWW server. Instalaci WWW serveru je lépe svěřit administrátorovi serveru. Zabývá se tím i řada specializovaných firem. WWW servery běží nepřetržitě 24 hodin denně. Tomu musí odpovídat i příslušné hardware. Klíčovou je rychlost a dostatek paměti. Návrh WWW serveru je dobře popsán např. na WWW adrese www.w3.org/hypertext/WWW/DesignIssues/Overview.html. Rozšířeným a v centru instalovaným WWW serverem je *Apáč* (Apache) z WWW adresy www.apache.org. Je poskytován zdarma. Odhaduje se, že téměř 80% všech WWW serverů je právě Apáč. Nicméně, další používané WWW servery jsou na WWW adrese serverwatch.internet.com/webservers.html. Jejich porovnání zase na WWW adrese webservercompare.internet.com.

Na cestě k vlastním informačním WWW zdrojům se tvůrcům osvědčuje dodržet následující posloupnost kroků:

- ❖ V návrhu vyjít od cíle. Má to být informační zdroj informativní, zábavný, vážný? Nápad je dobré kreslit. Shromáždit potřebný materiál (obrázky, dokumenty, odkazy). Nejzajímavější dát na počátek zdroje. Důležitou položku v návrhu tvoří inspirace jinými zdroji. V této souvislosti lze čerpat např. z následujících WWW stránek: info.med.yale.edu/caim/manual/, (styl, provedení) nebo www.100hot.com (informační obsah, nejnavštěvovanější WWW zdroje).
- ❖ Postupně vytvořit prototyp v jazyce XHTML (HTML). Důležité jsou zejména přitažlivé a přesné názvy, živost, uvolněnost, vyvážený poměr zaplněného a prázdného místa, přesné a výstižné odkazy, zajímavý obsah, střídmost při používání grafiky. Vyváženým zdrojem je z tohoto pohledu např. CNN na www.cnn.com. Dobrým zvykem je uvádět datum poslední aktualizace a spojení na autora.
- ❖ Dbát zejména na optimální používání jazyka HTML nebo XHTML, žádný nadbytečný kód, snadnou navigaci při prohlížení. Příkladem snadné navigace mohou být např. WWW stránky Mezinárodní organizace autorů WWW (HTML Writers Guild) na www.hwg.org.
- ❖ Testovat informační WWW zdroj před jeho zveřejněním. Požádat o názor okolí. Dělat úpravy.
- ❖ Publikovat informační zdroj na WWW. Odpovídající počítačový soubor přehrát na WWW server. Oznámit zdroj vyhledávačům.
- ❖ Registrovat počet uživatelů zdroje. Právě počet uživatelů se považuje za kritérium jeho úspěšnosti.

HTML editory

Ačkoliv je možné informační zdroje připravit v běžném textovém editoru typu „poznámkový blok“, existují specializované editory, které přípravu zdroje v XHTML (HTML) urychlí. To je velmi příjemné. Takovým editorům se říká *HTML editory*. Opět se osvědčilo jich pro tvůrce instalovat několik, řada z nich je volně dostupných:

- ❖ *PSPad* na WWW adrese www.sw.cz/pspad/pspad.htm. Jde o český plnohodnotný textový editor se šablonami pro tvorbu informačních WWW zdrojů.
- ❖ *TextPad* na WWW adrese www.textpad.com. Všechny ilustrační WWW stránky v XHTML, JavaScriptu i PHP v publikaci [1] byly vytvořeny právě v tomto editoru.
- ❖ *HTML-Kit* na WWW adrese www.chami.com/html-kit/.
- ❖ *Arachnophilia* na WWW adrese www.arachnoid.com/arachnophilia/.
- ❖ *Mozilla*, která tvoří jádro Netscape Navigatoru. Je současně WWW prohlížečem i HTML editorem na WWW adrese www.mozilla.org. Podobně jako Amaya.
- ❖ *CoffeeCup HTML Editor* na WWW adrese www.coffeecup.com/software/.
- ❖ *FrontPage*, který je součástí balíku Office v provedení „Premium“.

Jaký způsob tvorby informačních WWW zdrojů zvolit? Předně zkušenost říká, že na vytváření pěkných a svěžích (cool) zdrojů není nutné mít vzdělání na úrovni PhD v matematické informatice. Ani umět programovat. Na druhé straně, jistá znalost jazyka HTML případně XHTML věci velmi prospěje.

Převody do HTML

K informačním zdrojům v XHTML nebo HTML je však možné se dostat také cestou převodu z jiného formátu. Je potřeba proto pamatovat na instalaci odpovídajících programů. Jde o soubor programů, které se souhrnně nazývají „HTML Converters“. Známy jsou např. převodní programy z formátu RTF do HTML (RTFtoHTML, viz www.logictran.com/latest.html) či LaTeX do HTML (LaTeX2HTML). Často jde o vítanou možnost, pokud je formát zdroje již předem daný. Pokud jsou obsahem složité tabulky nebo grafy, převod do HTML nemusí dopadnout úplně perfektně. Hlavní práce s převodem se však vykoná. Užitečný seznam dostupných převodních programů (včetně odkazů) je na WWW adrese www.hypernews.org/HyperNews/get/www/html/converters.html.

Možnost výstupu přímo v HTML mají rovněž některé známé textové editory (např. Microsoft Word2000). Jistě k nim brzy přibude i výstup v XHTML. Výstup spočívá v tom, že se vytvoří úplně klasický textový dokument a při ukládání se vybere možnost uložit jako HTML. Tím se do dokumentu automaticky doplní HTML značky. Kódování však bývá až příliš neúspěšné a do HTML dokumentů se dostane tolik nadbytečných věcí, že neúměrně (i několikrát) zvětší velikost souboru. Ten se stává zcela nepřehledným. Proto se tato možnost využívá spíše pro jednorázové záležitosti.

Dále se nabízí přirozená otázka, co se zdroji připravenými v HTML, když přichází čas jazyka XHTML, který je svým způsobem odlišný. Je zde převodní program nazývaný *HTML Tidy* (jednoduše „Tidy“). Převodní program Tidy je možné získat, včetně postačujícího popisu jeho použití, na WWW adrese www.w3.org/People/Raggett/tidy/.

Uživatelé centra vylepšují informační zdroje Webu

Samotné HTML při tvorbě zdrojů WWW neumožňuje provádět mnoho potřebných věcí. Proto současně s vývojem HTML vznikla i poměrně bohatá jazyků, které HTML vylepšují. Protože WWW je obecně otázkou dvou programů – WWW prohlížeče a serveru, vylepšení se rovněž týkají jak WWW serverů, tak WWW prohlížečů.

Klientské WWW programování

První možností, jak vylepšit informační WWW zdroj je použití nějaký skriptový jazyk, který WWW prohlížeč umí interpretovat. Pokud se informační zdroj umístí na WWW server, stává se potenciálním kandidátem na prohlížení. Proces prohlížení probíhá tak, že se zdroj připravený v HTML přenesou do počítače uživatele (klienta). Aktivita se chopí WWW prohlížeč. Zejména jeho „tlumočnick“ jazyka HTML. Zdroj pro uživatele převede do zobrazitelné podoby.

Informační zdroje připravené pouze v jazyce HTML mají svá omezení. Neumožňují řadu věcí, které, jednoduše řečeno, dávají informačním zdrojům život. Při použití samotného jazyka HTML v informačních WWW zdrojích vážně aktivita, vzájemné působení s uživatelem, automatizovaná výměna dat.

Jednoduchý příklad. Firma obchoduje se zahraničím. Často potřebuje provádět přepočty různých měn. Proč na to nevytvořit vlastní WWW stránku? Je totiž předpoklad, že přepočty bude používat více lidí. V HTML jazyce nejde problém zvládnout. Firma chce mít na WWW vyhledávací databázi svých výrobků a umožnit tak potenciálním zákazníkům seznámení s produkcí. Opět se to nedá zvládnout v HTML. Dalším častým požadavkem je dynamická změna obsahu informačního zdroje v závislosti na okamžitém zájmu uživatele. Např. rozbalování menu, výměna obrázků apod. V HTML je to prakticky nemožné.

Existují však jazyky, které možnosti HTML rozšiřují. Dokáží odstranit výše uvedená omezení. V této souvislosti se hovoří o tzv. skriptování WWW stránek (Web page scripting) a programovacím jazykům, které skriptování umožňují, se říká skriptové jazyky. Předním skriptovým jazykem je JavaScript. Dalším VBScript. Programům ve skriptových jazycích se říká skriptové programy. Krátce *skripty* (scénáře).

Skriptové programy vykoná WWW prohlížeč. Tedy počítač na straně uživatele. WWW zdrojům to přináší nové pěkné možnosti, než kdyby byly připraveny pouze v HTML. Pokud např. na odpovídající WWW stránce budou nějaké matematické výpočty, provede je WWW prohlížeč na straně uživatele. Vůbec veškerá činnost spojená s interpretací skriptových programů se provádí na tomto počítači (tzv. client-side computing).

Úvod do jazyka JavaScript lze najít např. v [1]. Přínosem je, že k používání takového skriptového jazyka není třeba nic dalšího kromě znalosti jazyka. Tedy žádná instalace jiných programů, úprava WWW serveru nebo WWW prohlížeče apod. Soubor různých aplikací JavaScriptu lze nalézt na WWW adrese www.cs.cas.cz/~pklan/seminjs.htm.

Serverové WWW programování

Předchozí postup, kdy vše vykonává počítač na straně klienta, může být velmi užitečný při menším objemu zpracovávaných dat nebo menším objemu výpočtů. Pro velký objem dat nebo výpočtů má jistá úskalí. Počítače klientů nebývají tak rychlé a paměťově vybavené jako servery. Pokud se např. bude zpracovávat větší množství dat při prohledávání databáze, bude nutné databázi při každém prohlížení přenášet od WWW serveru k WWW prohlížeči, což může být časově a nakonec i paměťově náročné. Nedoporučuje se totiž Webem přenášet data větší než 300 KB. Co v takovém případě?

Existuje řešení. Databázi nebo výpočty neprovádět na straně klientů, nýbrž přímo na WWW serverech. K tomu byly vyvinuty zvláštní skriptové jazyky WWW serverů jako jsou oblíbené PHP či ASP nebo JSP. WWW server je však nutné na provádění takových skriptových programů předem připravit tj. např. PHP, který se v centru používá, na WWW server instalovat. Je možné tak učinit z adresy www.php.net/. Soubor různých aplikací PHP lze nalézt na WWW adrese www.cs.cas.cz/~pklan/seminphp.htm nebo www.cs.cas.cz/~pklan/semindata.html pro databázové aplikace PHP s MySQL.

Organizace např. vyhledávání v databázi je v tomto případě následující. Databáze se umístí na WWW server. WWW stránka vypadá na první pohled úplně stejně jako při zpracovávání na straně klientů. Jinak však proběhne organizace vyhledávání. Databáze zůstane umístěná na WWW serveru a nepřenáší se do počítače uživatele. Od uživatele se přenesou pouze dotazy na databázi. Na WWW serveru se automaticky spustí příslušný skriptový program, např. v PHP [1]. Ten zajistí prohledání databáze na přijatý dotaz. Nalezené záznamy převede do HTML a jako běžný HTML dokument odešle WWW prohlížeči. WWW prohlížeč je zobrazí. Uvedený proces má tu výhodu, že může proběhnout rychle vzhledem k vyšší kvalitě WWW serveru (tzv. server-side computing). Také se nepřenáší velké objemy dat.

Závěr

Příspěvek se snaží mapovat činnost jednoho vědecko-informačního centra. Nekonvenčního v tom, jak je z výše uvedeného patrné, že se aktivně snaží být ne pouze „internetovou kavárnou WWW brouzdačů a smskačů“ i když je brouzdání nepochybnou součástí procesu poznávání, ale také místem skutečných tvůrců chemických informačních zdrojů.

Proces vzdělávání je obojím. Jednak mít znalosti jako uživatel informačních zdrojů a jednak umět řešit problémy jako tvůrce informačních zdrojů. Obě činnosti by měly být vyvážené. Vychovávat pouze „uživatele“ je jistě mnohem jednodušší. I z pohledu internetového informačního pracovníka. Stačí ovládnout WWW prohlížeč.

V případě tvůrce WWW zdrojů je to o poznání složitější. Je třeba znát a mít možnost používat HTML, XHTML, JavaScript, PHP atd. a brát je jako prostředky k vytváření WWW zdrojů. To však vyžaduje mnohem větší úsilí. A také kontrolu, pokud student tápe a neví si s něčím rady. A roli informačních pracovníků v procesu vzdělávání vidím právě tady. Tak hodně štěstí.

Použitá literatura:

1. Klán P., Jindřich J. *WWW pro zelenáče*. Neocortex Praha, 2003.