

# XML A XWEB JAKO NÁSTROJE PRO TVORBU WEBOVÉHO SÍDLA S VELKÝM MNOŽSTVÍM KŘÍŽOVÝCH ODKAZŮ

Vlastimil Čevela

664 42 Modřice, Benešova 279, tel. 547 216 183,  
<http://www.volweb.cz/cevelavl/>, e-mail: [cevelavl@vol.cz](mailto:cevelavl@vol.cz)

## Abstrakt

Příspěvek se zabývá metodikou projektování a programovým řešením informačního systému (IS), který zajišťuje vytvoření, aktualizaci a udržení konzistence webového sídla s velkým rozsahem vzájemných vnitřních informačních vazeb, při malé četnosti obsahových změn.

## 1. Historie a koncepce řešení

Především je vhodné uvést, že projekt takového IS není nový, ale navazuje na demo-disketu se seznamem referátů Programování 75-91 z roku 1992 a na webovou prezentaci obsahu sborníků Programování a TSW z roku 1999 [1]. Ta obsahovala pouze autorský rejstřík a při pokusech o doplnění tématického rejstříku došlo mj. vinou nestandardního formátu vstupů k nezvládnutému zvýšení složitosti programového řešení i aktualizace, takže v roce 2001 přestala být udržována. S blížícím se 30. ročníkem TSWa PG vznikla v roce 2002 myšlenka, která vzbudila zájem u řady lidí, že by mohla být užitečná databáze obsahů i s texty referátů na CD. Vzhledem k dosavadním nepříliš povzbudivým zkušenostem však pořád nešlo o nic konkrétního.

### 1.1 XML - XWEB - HTML

Přímým inspirativním impulzem k novému řešení byl referát Šárky Ocelkové [2] na TSW 2003. Velice názorně totiž ukázal možnosti XML pro všeobecně srozumitelná pravidla uživatelského formátování vstupu i složitě strukturovaných dat. A tak se dalo uvažovat nejen o autorském, tématickém a firemním rejstříku, ale i kombinovaně přesouvat vstupní data do rejstříkových komentářových textů a tím výrazně zvýšit jejich vypovídací schopnost. Vzhledem k poměrně specializovanému zadání (těžiště v množství vazeb) a dlouholetým zkušenostem s generováním zdrojových programů Cobol i HTML [3] a [4], nebylo zvoleno řešení transformací XSL, ale vlastním programovým vybavením XWEB © Čevela 2003. Současně s jeho tvorbou bylo možno velice rychle vytvořit základní strukturu vstupních XML dat a ihned ji využívat a upřesňovat při prototypovém ladění a ověřování nové verze DB sborníků TSW a PG Ostrava [5].

Základem IS je databáze (DB) vstupních dat v souborech formátu XML s předepsanou strukturou, ze kterých jsou pomocí aplikačních příkazů a programového vybavení XWEB automaticky vygenerovány soubory HTML s potřebnými vazebními odkazy. V dalším textu jsou uvedeny příklady na sestavení modelu informačních vazeb (*model-iv*), definici struktury XML (*ds-xml*) i ukázky aplikačních příkazů XWEB (*ap-příkazů*) pro tvorbu relací a generování HTML výstupů. Rovněž v případě modelování informačních vazeb a celé struktury webového sídla bylo nutno přistoupit k vlastnímu řešení, protože metodika, která by splňovala představy nebyla nalezena.

## 2. Model informačních vazeb

Dlouhou dobu nebylo stanoveno, jak bude forma modelu pro návrh řešení IS vlastně vypadat. Bylo jasné, že potřebné složité vazby se musí nějak zaznamenat, ale v existujících „malých“ podmínkách nebylo možné využít nějakou nákladnou grafickou technologii. Také zde existovala základní představa - *co nejjednodušší formou a na jednom místě* definovat nejen vazby a jejich směry, ale i pravidla pro zobrazení jejich datové reprezentace - tj. co uživatel musí vidět pro úspěšnou navigaci. Nakonec byla vytvořena vlastní metodika modelu-iv, vycházející z funkčního popisu, rozhraní OO modulů a dalších zkušeností autora [3].

Podstatou definování návrhu IS jsou tedy modely-iv jednotlivých typů struktur datových souborů, ve kterých jsou pomocí tzv. *dekompozičních kódů* a jednoznačných textových identifikátorů (*id*) popsány všechny typy návěští a psaných (prvotních) i generovaných (zpětných) odkazů, které definují navigační síť webového sídla.

### 2.1 Dekompoziční kódy a konvence k popisu struktury

*F* = definice id souboru (*file*), *B* = definice návěští bloku (*konkretní id*)/(*a-číslo* = automaticky vypočítané pořadí), *D* = definice id relace, *L* = psaný odkaz na definici (*link*), *G* = generovaný zpětný odkaz, *C* = psaný textový doplněk (*complement*), *T* = generovaný text, *W* = přímý odkaz na web/soubor/mail a pod, *\** = komentářový řádek, */\*detail\*/* = popis detailu je uveden jinde, *data* = povinný výskyt a pořadí, */* = alternativa, *?* = nepovinně 1x, *+* = povinně minimálně 1x, *\** = nepovinně 1 nebo vícekrát, *[data]* = ve výstupním HTML nezobrazovaná, *{data}* = automaticky generovaná do HTML výstupu, *()* = závorkování dle běžných zvyklostí.

### 2.2 Příklady modelu-iv

Tabulky č. 1, 2, 3

iv	HLAVNÍ STRÁNKA WEBOVÉHO SÍDLA	index.xml,htm
<b>F</b>	id	[ <b>hlavní</b> ]
<b>W</b>	přihlášky a informace	/*kontakty*/
...	...	...
<b>B</b>	členové organizačního výboru	[a-číslo]
<b>D</b>	id	[ <b>člen_org_výboru</b> ]
<b>G</b>	seznam členů	{jméno}* [ <b>člen_org_výboru</b> ]
<b>B</b>	obsahy a texty sborníků	[a-číslo]
<b>D</b>	id	[ <b>hlavní-r</b> ]
<b>G</b>	seznam ročníků	{ročník}* [ <b>hlavní-r</b> ]
<b>B</b>	jmenný rejstřík	[a-číslo]
<b>D</b>	id	[ <b>hlavní-jm</b> ]
<b>G</b>	seznam rejstříkových stránek	{jm-abc}* [ <b>hlavní-jm</b> ]
...	...	...

iv	OBSAH SBORNÍKU	rrrr.xml,htm
----	----------------	--------------

<b>F</b>	id	[ ročník ]
L	odkaz na index	hlavní-r
...	...	...
<b>B</b>	seznam referátů	[ strana ]*
*	odkazy na text referátu	(/)
L	začátek referátu ve sborníku (ročníky 1975 až 1998)	strana.htm
*	podrobnosti /*iv obálka skenované stránky sborníku*/	-
L	kompletní text (ročníky 1999 a další)	strana.doc/pdf
C	referát není dostupný (důvod uveden jako poznámka)	strana
L	seznam autorů referátu	(jméno)+
L	vysílající organizace	firma
C	název referátu	text
L	seznam (max 7) témat, o kterých referát pojednává	(téma)+
C	komentář k tématu	text
L	poznámka do tématického rejstříku	(téma=poznámka)?
C	komentář dle potřeby	text
...	...	...

iv	JMENNÝ REJSTRÍK	jm-cc.xml,htm
<b>F</b>	id	[ jm-abc ]
L	odkaz na index	hlavní-jm
...	...	...
<b>B</b>	seznam osob	[ a-číslo ]*
C	příjmení a plné jméno	text
<b>D</b>	id	[ jméno ]
G	seznam referátů	{ročník[strana]}*
T	seznam témat	{téma}+
C	dřívější organizační činnost	(text)?
W	osobní odkazy	(/*kontakty*/)*
L	aktuální členství ve výboru	(člen_org_výboru)?
...	...	...

### 3. Definice struktury XML

Na zjednodušených příkladech modelu-iv máme definovány typy souborů, jejich informační obsah a třídy identifikátorů (jméno, firma, téma, ročník a další), které specifikují navržené relace. V další etapě projektu - viz následující příklady, je ke každému modelu-iv vytvořena ds-xml, ve které jsou k definicím a odkazům přiřazeny konkrétní tagy a dále popsány i všechny ostatní tagy, určující celkový datový obsah IS (např. záhlaví, nadpisy, a pod.).

Tabulky č. 4, 5, 6

ds	HLAVNÍ STRÁNKA WEBOVÉHO SÍDLA
<pre> &lt;index&gt;hlavní   &lt;z&gt;záhlaví HTML - titulek stránky&lt;/z&gt;   &lt;x1&gt;1. řádek nadpisu&lt;/x1&gt;&lt;radek/&gt;  &lt;x2&gt;2. řádek nadpisu&lt;/x2&gt;&lt;linka/&gt;   &lt;x&gt;Příhlášky a informace:&lt;/x&gt;  &lt;web&gt;software.tanger.cz&lt;/web&gt;  &lt;mail&gt;kupka@tanger.cz&lt;/mail&gt;   ...   &lt;hx&gt;Členové organizačního výboru:&lt;o&gt;člen_org_výboru&lt;/o&gt;&lt;/hx&gt;&lt;radek/&gt;   &lt;hx&gt;Obsahy a texty sborníků:&lt;s&gt;hlavní-r&lt;/s&gt;&lt;/hx&gt;&lt;radek/&gt;   &lt;hx&gt;Jmenný rejstřík:&lt;s&gt;hlavní-jm&lt;/s&gt;&lt;/hx&gt;&lt;radek/&gt;   ... &lt;/index&gt; </pre>	

ds	OBSAH SBORNÍKU
----	----------------

```

<rok max="celkový počet stran ve sborníku">ročník
...
<h>hlavní-r</h> ...
<ref str/doc/pdf/nic="strana">
  <j>jméno<f>firma</f></j>
  * možno opakovat
  <x>Název referátu (i na několik řádků)</x>
  <t>téma<k>komentář k tématu</k></t>
  * možno opakovat max 7x
</ref>
...
</rok>

```

```

ds JMENNÝ REJSTRÍK
<jmena>jm-abc
...
<h>hlavní-jm</h> ...
<jm>příjmení a plné jméno<j>jméno</j></jm>
  * nepovinné informace
  <k>dřívější organizační činnost</k>
  <tx>osobní kontakty</tx> <web>adresa stránek </web> <mail>e-mail adresa</mail>
  <o>člen_org_výboru</o>
</jm>
...
</jmena>

```

#### 4. Aplikační příkazy a programové vybavení XWEB

Programové vybavení XWEB bylo vytvořeno v jazyku Cobol pomocí zdrojově OO vývojového prostředí [4] s kompilátorem *MX Cobol* © Čevela 1996-2000 a skládá se ze tří logických částí: *analýza vstupních dat*, *generování relací pro indexsekvencní soubor a generování HTML výstupů*. Sborníky ročníků 1975 až 1998 nebyly k dispozici v elektronické podobě, takže je bylo nutno naskenovat. Generování „obálkových“ HTML souborů k listování ve skenovaných GIF stránkách je realizováno samostatně.

##### 4.1 Analýza vstupních dat

Dle seznamu se přečtou vstupní soubory a pro každou datovou položku se vytvoří tzv. *struktura XML tagů* = *posloupnost hierarchické struktury od kořene k aktuálnímu tagu*.

V uvedených tabulkách ds-xml je to např.:

- „*index,web*“ pro webovou adresu kontaktu
- „*index,hx,s*“ pro definici jednoho z typů seznamu na hlavní stránce
- „*rok,ref,j,f*“ pro odkaz na firmu autora referátu

##### 4.2 Aplikační příkazy

Struktura tagů pak slouží pro přiřazení ap-příkazů a maker ke struktuře vstupních dat. Všechny příkazy i makra jsou uloženy v samostatných souborech formátu XML, takže systém je „uživatelsky programovatelný“. Ap-příkazy představují jistou množinu pseudopříkazů, kterou programy XWEB umí interpretovat k vykonání posloupnosti potřebných kroků a je možné je sdružovat do maker a parametrizovat k vykonání shodných činností pro různé struktury tagů.

Syntaxe elementárních ap-příkazů, kterých je asi 30, vychází ze syntaxe Cobolu - např.:

- „initialize POC, add 1 to POC, move POC to ULO“ = ovládání počítačidla
- „move DAT to ULI“ = uložení aktuálních dat do paměti
- „get htm 27, put htm using DAT“ = výběr šablony HTML č. 27 a modifikace daty

### 4.3 Generování

Na základě analyzované struktury tagů se pomocí ap-příkazů nejprve vytvoří relace podle shodných identifikátorů a uloží všechny přesouvané komentáře (C), a potom se s pomocí parametrizovatelných HTML šablon ke každému vstupnímu XML souboru generuje odpovídající výstup s příslušnými hypertextovými odkazy (L a G z modelu-iv) dle potřeby. Kromě toho se generují přímé odkazy (W) a texty (T) i všechny další potřebné HTML tagy, včetně vazeb na kaskádové styly, které jsou pro celé webové sídlo definovány v samostatném CSS souboru.

### 5. Závěr

Použitelnost popisovaného obecného řešení IS dokumentuje konkrétní aplikace "*DB obsahů a textů sborníků Tvorba softwaru a Programování Ostrava*" na webovém serveru Fakulty informatiky Masarykovy univerzity v Brně [5] a *CD-TSW/2004 Tanger Ostrava* k 30. ročníku stejnojmenné akce. Příslušná *DB-obsahů* v asi 50 vstupních souborech XML popisuje 30 sborníků (1975-2004) od přibližně 600 autorů z 200 organizací s 800 referáty, klasifikovanými pomocí více než 100 témat. Při generování výstupních souborů HTML je automaticky vytvářeno asi 7 tisíc křížových hypertextových odkazů. Pro *DB-textů* je pak ze stejných vstupů generováno dalších 6 tisíc "obálkových" souborů HTML k jednotlivým naskenovaným stránkám GIF ročníků 1975 až 1998, s celkovým počtem cca 60 tisíc odkazů, umožňujících různé způsoby navigace.

### Podklady a literatura:

1. <http://www.volweb.cz/cevelav/>
2. Ocelková Š., Návrh a realizace www prezentace ČKR, TSW Tanger Ostrava 2003
3. Čevela V., referáty na seminářích TSW a PG, DT a Tanger Ostrava, 1976 až 2003
4. -"- , Práce s OO generátorem programů Cobol, Programování 1993, DT Ostrava
5. <http://honor.fi.muni.cz/tsw/>