

DocBook a jeho využití

Tomáš Pitner, Jan Pavlovič, FI MU

XML značkování *DocBook*¹, původně určené k pořizování programové dokumentace, se stává v současnosti velice populárním formátem platformově přenositelných dokumentů "o čemkoli". Nejlepším důkazem jeho kvalit je široké uplatnění v open-source komunitách, akademické sféře i v průmyslu - využívají ho společnosti Novell a SUN, vydavatelství O'Reilly stejně jako komunity kolem projektů PHP, Gnome, KDE, BSD, Linux aj.

DocBook je formátem zavedeným; v SGML podobě, dnes již téměř neznámé, existuje někdy od roku 1991. Od ustavení XML jako standardu v roce 1998 je DocBook rozvíjen již převážně v XML podobě. Hlavní výhodou je kromě totální platformové nezávislosti také "čistota" v něm pořízeného textu. Dokument označkováný podle standardu DocBook nenesé téměř žádné explicitní pokyny k formátování - je tedy stejně dobře následně prezentovatelný na monitoru počítače, na displeji PDA nebo ve vysoce kvalitní tiskové podobě. Nejsme v něm schopni - a tudíž ani nemůžeme - zabývat se tím, jak velké písmo a jaký font se použije na nadpis, kde bude zalomena stránka

Převážná většina značek nese informace o *sémantice* toho, co je značkou ohraničeno. Např. značka `<chapter>` signalizuje, že její vnitřek představuje jednu kapitolu knihy nebo článku, `<title>` je její nadpis; podobně `<para>` ohraničuje odstavec textu a na řádku uvedený element `<emphasis>` vymezuje zdůrazněný úsek textu. O přesné formátování se starají až nadstavby, které například pomocí XSLT transformací a případně dalším následným zpracováním z původně platformově nezávislého dokumentu vytvoří např. PDF soubor s již přesným rozvržením na stránky, vhodnými fonty atd.

I ve srovnání s $\text{T}_{\text{E}}\text{X}$ em znamená DocBook ještě výraznější příklon k *logickému* charakteru značkování. Na druhou stranu je třeba připustit, že pro *stoprocentně* typograficky kvalitní výstup DocBookových dokumentů se manuálním zásahům v jisté fázi zpracování nevyhneme.

¹<http://www.oasis-open.org/docbook>

Ukažme si, jak v DocBooku označujeme jednoduchou knihu se dvěma kapitolami (příčemž si odpouštíme např. informace o autorovi, datum vzniku, copyrightu, atd.):

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE book PUBLIC
  "-//OASIS//DTD DocBook XML V4.2//EN"
  "http://www.oasis-open.org/docbook/
  xml/4.2/docbookx.dtd">
<book lang="cs">
  <title>První kniha v DocBooku</title>
  <chapter>
    <title>Nadpis první kapitoly</title>
    <para>První odstavec první kapitoly.
    </para>
    <para>Druhý odstavec první kapitoly.
    </para>
  </chapter>
  <chapter>
    <title>Nadpis druhé kapitoly</title>
    <para>Odstavec 1. se
    <emphasis>zdůrazněným textem</emphasis>
    </para>
    <para>Odstavec 2.</para>
  </chapter>
</book>
```

Celkově DocBook nabízí přes 400 druhů elementů (značek), v praxi většinou vystačíme se znalostí několika málo desítek z nich. Široký repertoár je nabízen pro ty, kdo vyžadují jemné sémantické vyznačení takových prvků dokumentu, jako jsou názvy tříd, metod, tlačítek v uživatelském rozhraní programů, příkazů systému, atd. Prvotní účel - prostředek k pořizování dokumentace - se zde jednoznačně odráží.

1 Oblasti využití

Nasazení DocBooku v akademické sféře lze rozdělit do těchto hlavních kategorií:

- tvorba webových prezentací (slidy předmětu)
- tvorba odborných článků a publikací
- tvorba bakalářských a diplomových prací

1.1 Webové prezentace (slidy)

Norman Walsh připravil pod názvem *DocBook Slides*² derivát zjednodušeného DocBooku (*Simpli-*

²<http://docbook.sourceforge.net/projects/slides>

fi ed DocBook³) určený k přípravě prezentačních podkladů (slidů). Současně k nim poskytuje rozsáhlou sadu XSL stylů pro jejich vizualizaci.

Studenti často vyžadují, aby slidy dostali k dispozici i v tisknutelné podobě. Pro tyto případy lze použít buďto existující styly formátující slidy do tzv. *formátovacích objektů*⁴ (jsou k dispozici na http://sourceforge.net/project/showfiles.php?group_id=21935 vedle stylů generujících HTML) a výsledný soubor převést např. procesorem Apache FOP⁵ do PDF souboru a ten vytisknout. Velmi jednoduchým XSLT stylem je také možno převést slidy do normálního DocBooku - např. jako knihu (*book*), případně doplňovat atd. jako běžný DocBook dokument. Tento způsob se osvědčil nejlépe - ne každý student totiž chce celý soubor tisknout, někomu vyhovuje jeho prohlížení - a to i např. na počítačích třídy PDA, pro něž mnohdy neexistuje prohlížeč PDF.

Uveďme pro představu krátkou ukázkou zdrojového tvaru slidů ve značkování DocBook Slides:

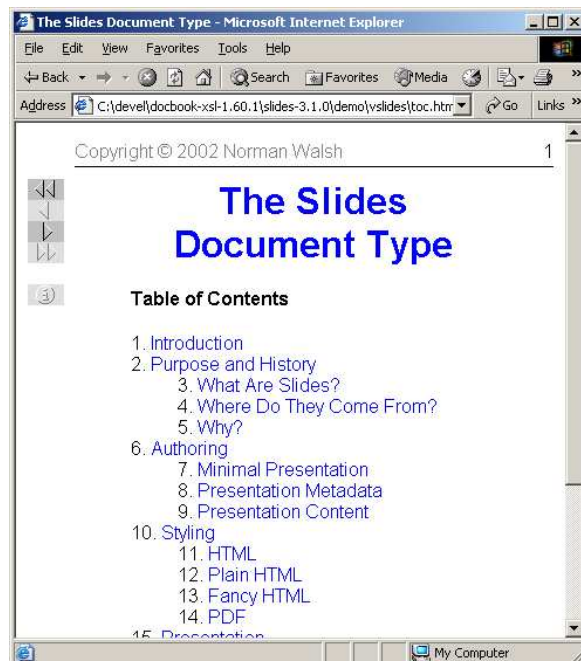
```
<?xml version='1.0'?>
<!DOCTYPE slides SYSTEM
  "../..../schema/dtd/slides.dtd">
<?dbhtml graphics-dir="../graphics"
  css-style-sheet-dir="../browser"?>
<?dbhtml script-dir="../browser"?>
<slides>
<slidesinfo>
  <title>The Slides Document Type</title>
  <titleabbrev>Slides</titleabbrev>
  <author><firstname>Norman</firstname>
    <surname>Walsh</surname></author>
  <pubdate>Tuesday, 25 June 2002</pubdate>
  <releaseinfo role="version">
    3.1.0</releaseinfo>
  <copyright><year>2002</year>
    <holder>Norman Walsh
    </holder></copyright>
</slidesinfo>

<foil><title>Introduction</title>
<para>This is the introductory slide.</para>
<para>If you use foil groups (previously
called sections), you can have introductory
slides before the first group.</para>
```

³<http://www.oasis-open.org/docbook/xml/simple>

⁴XSL:F0, <http://www.w3.org/TR/xsl>

⁵<http://fop.apache.org>



Obrázek 1: Slidy připravené k prezentaci v HTML prohlížeči

```
</foil>
```

```
<foilgroup>
<title>Purpose and History</title>
```

```
<para>Groups can now have introductory
text as well.</para>
```

```
...
```

Na obr.1 je vidět výsledná HTML podoba zobrazená v prohlížeči.

V praxi se ukázalo, že hlavní předností pořizování slidů k přednáškám i jednorázovým prezentacím v DocBooku je snadná produkce výstupů k prohlížení na obrazovce, promítání z projektoru a pro tisk - vše z jednoho zdrojového souboru.

1.2 Tvorba odborných článků a publikací

Velmi užitečný nástroj je DocBook při tvorbě článků do konferencí a jiných odborných publikací. Většinou se po autorech žádá, aby svůj příspěvek formátovали do předem dané T_EXové šablony. Jestliže své příspěvky hodláme pořizovat v DocBooku, lze pro příslušnou konferenci napsat krátký XSLT styl, který nám zdrojový text v DocBooku převede do požadovaného formátu.

Výchozí zdroje zůstávají v jednotné podobě nezávislé na požadavcích editora daného sborníku, časopisu apod. Díky tomu, že zdrojový soubor ve značkování DocBook nenese (téměř) žádné pokyny pro formátovací a sémantika značek je dána poměrně pevně, není rovněž problém kombinovat zdroje různých autorů - např. v situaci, kdy pracují na společném článku, knize...

1.3 Tvorba závěrečných prací

Jedním z požadavků na závěrečné práce je přiměřená typografická úroveň. Na Fakultě informatiky tradičně existuje L^AT_EXový styl, které závěrečným pracím dává jednotnou podobu. V dnešní době však T_EXová gramotnost mezi studenty velmi klesá a napsat typograficky kvalitní práci ve WYSIWYG editoru není snadné. Práce v DocBooku nám velmi zjednoduší jak samotné pořízení textu práce, tak především přípravu její výstupní podoby - obvykle ve formátech PDF a XHTML.

Tvorba práce v DocBooku je několikafázový automatizovaný proces, viz. obr 2

V průběhu celého procesu k transformaci zdrojového dokumentu v DocBooku do L^AT_EXového kódu. L^AT_EXový procesor nám ve výstupu dává dokument v požadované kvalitě. Jelikož je celý proces automatizovaný, nemusí uživatel znát prakticky nic o L^AT_EXu.

2 Nástroje

2.1 Editory

Bez kvalitních autorských nástrojů by samotné značkování nebylo příliš užitečné. Naštěstí je DocBook dnes koncipován a rozvíjet především jako XML značkování a tudíž lze k editování dokumentů použít jakýkoli XML editor, případně i editor čistého textu. Nejlepší přímou podporu DocBook nabízejí komerční prostředí jako je Epic, XMLSpy nebo Corel XMetaL. Bez velkých omezení je ale použitelný i XMLMind XML Editor⁶, jehož základní verze je dostupná zdarma a je k dispozici i jako modul (xmlmind) na strojích

⁶<http://www.xmlmind.com/xmleditor>

Fakulty informatiky. Rovněž tak lze zdarma použít i javový editor jEdit⁷, "klasiku" Emacs, vim a řadu dalších.

2.2 Modul xslt2

Aby mohl uživatel plně využít předností automatizovaného zpracování dokumentů v DocBooku, byl na Fakultě informatiky vytvořen systém xslt2⁸. Celý systém je napsán v jazyce Perl a je tudíž přenositelný na všechny operační systémy s podporou Perlu.

Na UNIXových strojích FI je systém dostupný pomocí modulu xslt2, který zavedeme běžným způsobem:

```
module add xslt2
```

Po přidání modulu můžeme transformovat XML dokument pomocí XSL šablony. Lze taktéž i specifikovat XSLT procesor. Systém plně podporuje i XML Catalogs⁹, tudíž se nemusíme starat ani o lokální umístění DTD definic odkazovaných ve zdrojových souborech. Například:

```
xslt -c -l -s fi -t pdf dp.xml  
xslt -c -s xhtml dp.xml
```

nám transformuje diplomovou práci (dp.xml) do PDF a XHTML.

2.3 Struktura a možné modifikace xslt2

Samotný DocBook poskytuje dostatečně široký repertoár elementů pro napsání věcného obsahu závěrečné práce, bylo však nutné jej rozšířit o elementy umožňující specifikovat informace náležící k závěrečné práci - zejména k jejímu autorství, příslušnosti k fakultě atd. Architektura gramatiky DocBooku je naštěstí konstruována jako modulární a tudíž bylo rozšíření pro účely studentských závěrečných prací relativně přímočaré, psané jako DTD modul fithesis.mod¹⁰. Analogicky jako gramatika byly rozšířeny i XSLT styly, byť samozřejmě jinou technikou.

Chování DocBook XSLT stylů docbook-xsl¹¹ a db2latex¹², které systém xslt2 používá, lze

⁷<http://jedit.org>

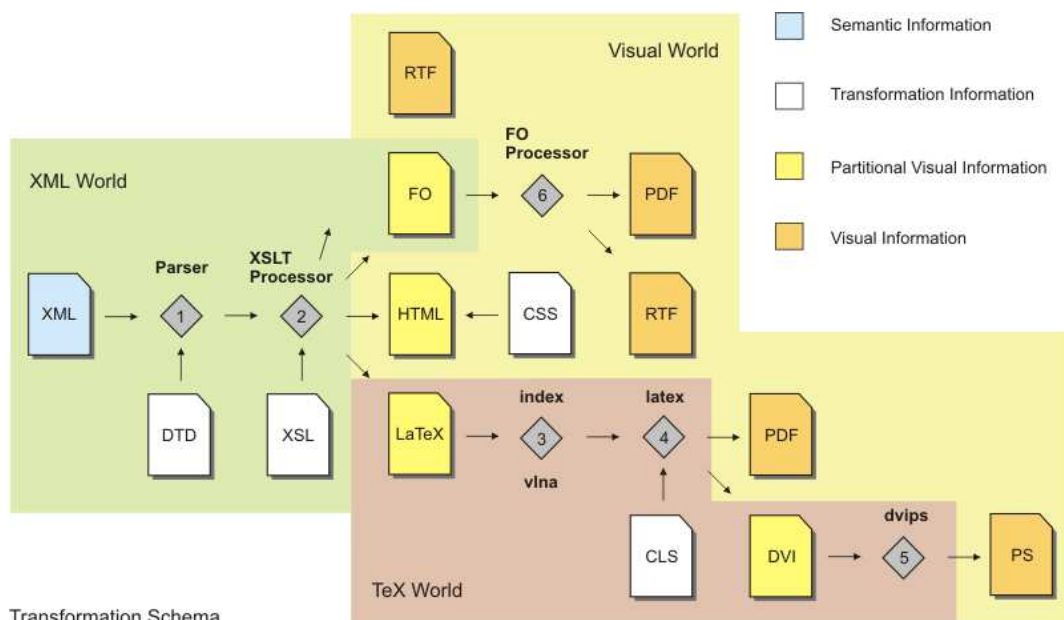
⁸<http://www.fi.muni.cz/~xpavlov/xml>

⁹<http://www.oasis-open.org/committees/entity/spec-2001-08-06.html>

¹⁰<http://www.fi.muni.cz/~xpavlov/fithesis.mod>

¹¹<http://sf.net/projects/docbook>

¹²<http://sf.net/projects/db2latex>



Obrázek 2: Transformace DocBooku

ovlivňovat nastavením parametrů. Pro rozsáhlejší modifikace, nepokryté parametry, nabízejí XSLT styly mnohem silnější techniku. Tou je překrývání (nahrazování) již existujících šablon novými. Velkou výhodou je, že nemusíme zasahovat do originálních stylů. Veškeré úpravy uvedeme ve vlastním stylovém souboru, kde v jeho začátku importujeme pomocí `xsl:import` originální hlavní XSLT soubor.

Jako příklad vezmeme *db2latex* styly, které originálně sestavují uvozovky kolem slova pomocí jednotlivých znakových entit daného jazyka. V češtině se spíše používá pro uvozovky makro `\uv{}`. Proto vytvoříme vlastní definici šablony pro element `quote`, která v procesu transformace nahradí originální definici. Příklad:

```
<?xml version="1.0" encoding="utf-8"?>
<xsl:stylesheet xmlns:xsl=
  "http://www.w3.org/1999/XSL/Transform"
  version='1.0'>

<xsl:import href="..docbook.xsl"/>

<xsl:template match="quote">
  <xsl:text>\uv{</xsl:text>
  <xsl:call-template name="inline.charseq"/>
  <xsl:text>}</xsl:text>
</xsl:template>

</xsl:stylesheet>
```

3 Matematika a DocBook

Jelikož podpora MathML není zatím na přijatelné úrovni, je lepší v současné době matematické vzorce v DocBooku zapisovat pomocí \LaTeX ové syntaxe. Do výsledného \LaTeX ového kódu se zápis matematiky vloží přímo a do HTML se vloží odkaz (uvedený v elementu `imagedata`) na obrázek, který vznikl konverzí z EPS souboru vyprodukovaného \LaTeX em. K zápisu matematických kódů slouží v DocBooku tři elementy: `inlineequation`, který se používá k vložení matematiky přímo do textu, `informalequation`, který matematiku zobrazí v samostatném odstavci a `equation` zobrazující matematiku v odstavci s uvozovacím titulkem. Příklad:

```
<equation>
  <title>Rovnice hmoty a energie</title>
  <mediaobject>
    <imageobject>
      <imagedata fileref="math1.png"/>
    </imageobject>
    <textobject role="tex">
      <phrase>e=mc^2</phrase>
    </textobject>
  </mediaobject>
</equation>
```

4 Kam dál?

DocBook je široce použitelné značkování, k němuž existuje celá řada kvalitních nástrojů a informačních zdrojů. Zájemce můžeme odkázat na v úvodu zmíněnou základní stránku standardu DocBook při konsorciu OASIS, stránku ke knize N. Walshe *DocBook: The Definitive Guide*¹³, stránky wiki k DocBooku dostupné tamtéž nebo příslušnou mailovou skupinu při OASIS¹⁴. Kromě toho se v polovině dubna 2004 na Fakultě informatiky konala úvodní přednáška k DocBooku vedená Jiřím Koskem¹⁵, členem OASIS TC DocBook. Videozáznam přednášky je dostupný na <http://video.fi.muni.cz/ostatni.shtml>. □

¹³<http://docbook.org>

¹⁴<http://lists.oasis-open.org/archives/docbook>

¹⁵<http://www.kosek.cz>