

XML a nové trendy v publikování na Webu

4IZ228 – tvorba webových stránek a aplikací

Jirka Kosek

Poslední modifikace: \$Date: 2010/05/13 17:56:13 \$

Copyright © 2000-2008 Jiří Kosek

Obsah

Úvod	3
Nové požadavky na web	4
XML a podpora různých koncových zařízení	5
Problém	6
Řešení problému = XML	7
Základní charakteristika XML	8
Formátování a zobrazování dokumentů XML	9
Výhody použití stylů	10
Stylové jazyky	11
Využití XSLT na webu	12
Výměna dat a integrace aplikací v prostředí webu	13
Přeměna webu	14
Webové služby	15
Schéma práce s webovou službou	16
SOAP	17
WSDL	18
UDDI	19
Další vývoj	20
Lepší vyhledávání na webu	21
Jak XML usnadňuje vyhledávání	22
Proč to nefunguje?	23
Sémantický web (Definice)	24
Sémantický web (Idea)	25
Příklad RDF	26
Perspektiva sémantického webu	27
Mikroformáty	28
Další zdroje informací	29
Další zdroje informací	30

Úvod

Nové požadavky na web	4
-----------------------------	---

Nové požadavky na web

- podpora různých zařízení (PC, MT, PDA, set-top box, ...)
- snadná výměna dat libovolného druhu a spolupráce aplikací v celosvětovém měřítku
- snazší a lepší vyhledávání (Internet je dnes zahlcen informacemi)

XML a podpora různých koncových zařízení

Problém	6
Řešení problému = XML	7
Základní charakteristika XML	8
Formátování a zobrazování dokumentů XML	9
Výhody použití stylů	10
Stylové jazyky	11
Využití XSLT na webu	12

Problém

- množina koncových zařízení se stále mění
 - webový prohlížeč (HTML/XHTML)
 - mobilní telefon (WML/cHTML/XHTML Basic)
 - PDA (HTML/XHTML navržené s ohledem na menší obrazovku)
 - tištěný výstup (PDF/RTF)
- ruční převod a údržba informací v několika paralelních variantách je pracná, pomalá a chybová (→ drahá)
- automatický převod není možný nebo je nekvalitní, protože obsah dokumentů je svázán s konkrétním výstupním médiem

Řešení problému = XML

- informace je uložena jen v jednom primárním zdroji
- informace je uložena v podobě, která umožňuje její další snadné zpracování
- informace je uložena v XML
- důležité je zachytit podstatu/význam/sémantiku informace:

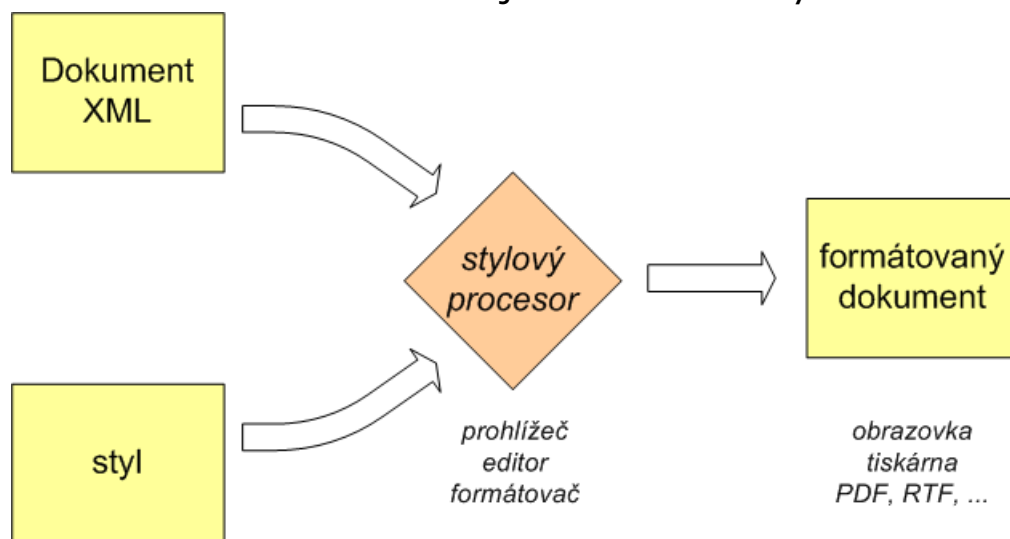
XML	HTML
<pre><ceník> ... <položka kategorie="CD" ▶ kód="04400148712"> <název>Snadné je žít</název> <interpret>MIG 21</interpret> <cena měna="Kč">399<cena> </položka> ... </ceník></pre>	<pre>... <table> ... <tr> <td>MIG 21: Snadné je žít</td> <td align="right">399 Kč</td> <td><a ▶ href="pridej.php?kod=04400148712">Objednat</td> </tr> ... </table> ...</pre>

Základní charakteristika XML

- XML = eXtensible Markup Language
- jednoduchý značkovací jazyk
- do dokumentů můžeme přidat sémantickou informaci pomocí značek (elementů) s libovolně zvolenými názvy
- podporuje mnoho jazyků, protože je založen na Unicode
- pro dokument můžeme vytvořit schéma (DTD, XML Schema, RelaxNG), které definuje přípustnou strukturu dokumentu
- aplikací XML/slovníků značek (značkovacích jazyků založených na XML) dnes existují tisíce
 - většinou jsou formálně definovány pomocí schématu
 - příklady: XHTML, WML, SVG, MathML, RSS, ...

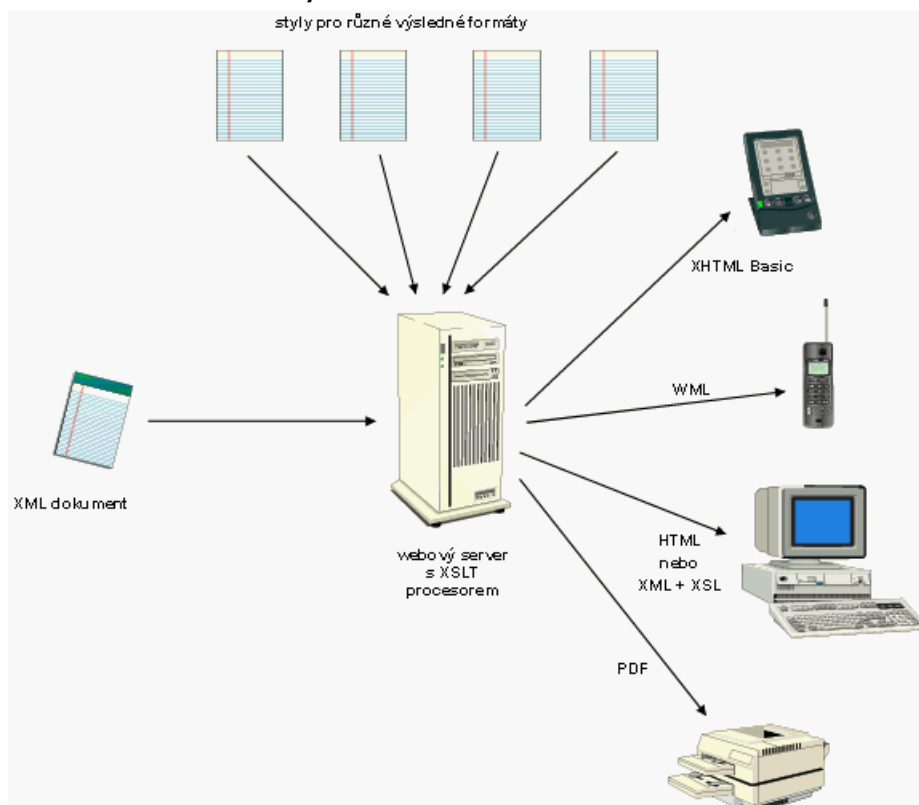
Formátování a zobrazování dokumentů XML

- XML nedefinuje vzhled, ale význam
- zobrazení dokumentu XML je definováno stylem



Výhody použití stylů

- jednotné formátování mnoha dokumentů
- možnost formátovat jeden dokument několika různými styly
 - pro každé výstupní zařízení zvláštní styl
- můžeme tak snadno z jednoho zdroje dat obsloužit několik různých druhů klientských zařízení



- vstupní dokument XML nemusí být statický, může být generován dynamicky na základě požadavku uživatele – přístup lze použít i pro dynamické webové aplikace

Stylové jazyky

- definují vzhled dokumentu
- nejvíce se používají CSS a XSL
- CSS – pouze pro velice jednoduché aplikace
- XSL
 - XSLT – transformační část umožňuje konverzi XML do XML, HTML nebo čistého textu
 - XSL FO – formátovací objekty abstraktním způsobem popisují vzhled objektů, které tvoří zformátovanou podobu dokumentu
- většina jazyků podporovaných koncovými zařízeními je založena na XML (XHTML, WML, FO, SVG) a pro jejich generování se používá XSLT

Využití XSLT na webu

- „on-the-fly“ obsluha požadavků
 - podpora více koncových zařízení
- předgenerování statických stránek
 - ušetření práce
 - automatické generování navigace
- příklady implementací
 - Cocoon¹
 - Forrest²
 - AxKit³

¹ <http://cocoon.apache.org/>

² <http://xml.apache.org/forrest/index.html>

³ <http://axkit.org/>

Výměna dat a integrace aplikací v prostředí webu

Přeměna webu	14
Webové služby	15
Schéma práce s webovou službou	16
SOAP	17
WSDL	18
UDDI	19
Další vývoj	20

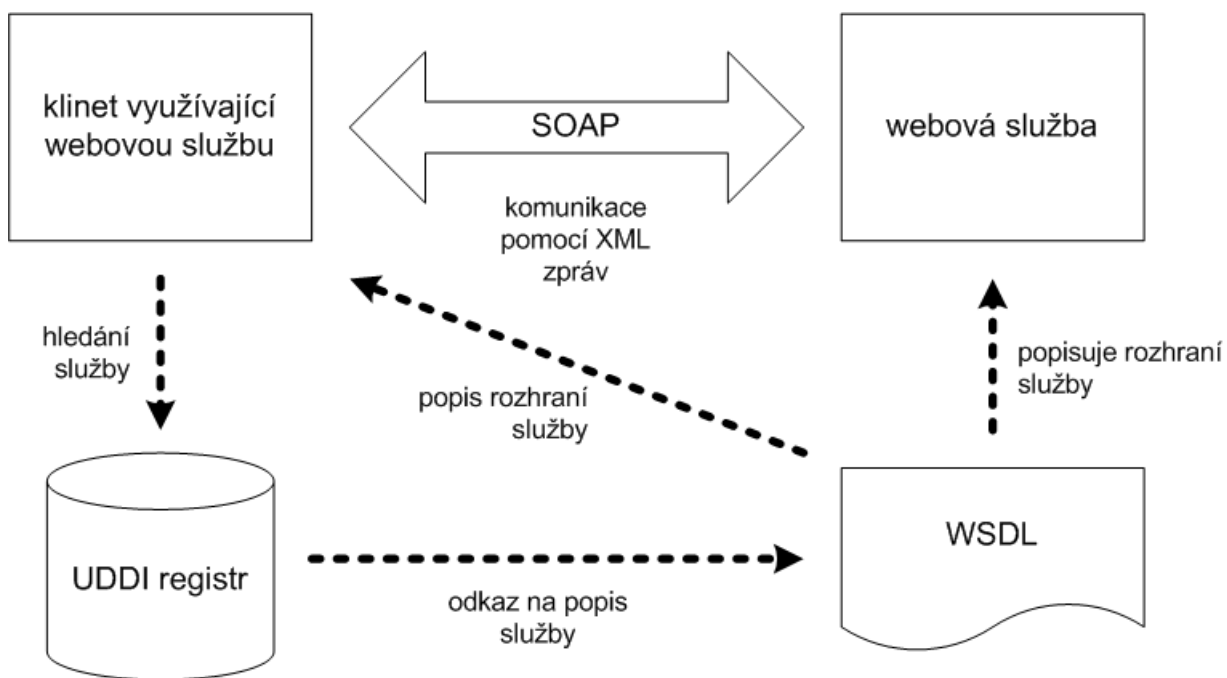
Přeměna webu

- původně byl web navržen jako prostředí pro prezentování statických dokumentů propojených odkazy
- později se (díky CGI skriptům) web začal používat jako rozhraní pro aplikace, se kterými pracuje uživatel
 - internetové obchody, vyhledávače, rozhraní do IS
 - tzv. B2C – business-to-customer aplikace
- nyní je potřeba, aby spolu mohly komunikovat aplikace
 - výměna objednávek, rezervace letenky, syndikace zpráv, ...
 - tzv. B2B – business-to-business aplikace
 - „další progresivní označení“: A2A (application-to-application), EAI (enterprise application integration), ...
 - web nabízí komunikační infrastrukturu postavenou na otevřených standardech (HTTP, XML, kryptografie)

Webové služby

- webová služba = web-service
- webové služby – souhrnné označení pro sadu technologií umožňujících komunikaci mezi aplikacemi
- webová služba je jednoduchá komponenta nabízející určitou službu
 - převod měn, zjištění kurzu akcie, zpracování objednávky, překlad textu
- webové služby vs. Corba, DCOM, RMI
 - zcela nezávislé na použité platformě
 - dostupné globálně, umožňují snadnou integraci různých prostředí
 - postaveny na jednoduchých technologiích jako XML a HTTP
 - méně efektivní než „nativní“ protokoly
- tři základní pilíře
 - SOAP
 - WSDL
 - UDDI

Schéma práce s webovou službou



SOAP

- SOAP = Simple Object Access Protocol
- vzdálené volání služby (funkce) je definováno jako jednoduchá XML zpráva, která se obvykle přenáší pomocí HTTP protokolu
- SOAP definuje základní strukturu XML obálky a způsob mapování datových typů do XML
- existuje několik knihoven pro snadnou implementaci klienta i serveru
- <http://www.soapware.org> – seznam dostupných implementací

WSDL

- WSDL = Web Services Description Language
- přesný popis rozhraní webové služby dostupné přes SOAP
- z WSDL definice lze automaticky generovat klientský kód, který umí službu využívat, i skeleton serverového kódu

UDDI

- UDDI = Universal Description, Discovery and Integration
- univerzální adresář obsahující seznam a popis dostupných webových služeb
- umožňuje automatické nalezení požadované služby

Další vývoj

- „SOAPové“ webové služby
 - doplňování o další a další vrstvy řešící „vše“ (transakce, bezpečnost, ...)
 - ve výsledku velice složité
 - nástroje softwarových firem se snaží o dokonalé odstínění od technologických detailů
- REST
 - „návrat ke kořenům“
 - posílání čistého XML přes HTTP protokol
 - s takovými webovými službami lze snadno pracovat i bez podpory mohutných SW nástrojů

Lepší vyhledávání na webu

Jak XML usnadňuje vyhledávání	22
Proč to nefunguje?	23
Sémantický web (Definice)	24
Sémantický web (Idea)	25
Příklad RDF	26
Perspektiva sémantického webu	27
Mikroformáty	28

Jak XML usnadňuje vyhledávání

- jednotlivým částem dokumentu je přiřazen jejich význam pomocí značek:

```
<ceník>
  <výrobek>
    <název>Příklepová vrtačka</název>
    <typ>PX-KR 897</typ>
    <cena>3750</cena>
  </výrobek>
</ceník>
```

- důraz se klade na význam informací, ne na jejich vzhled (velký rozdíl oproti většině dnešních webových stránek)
- nad bází XML dokumentů (v budoucnu nad celým Webem) můžeme pokládat kontextové dotazy:

Najdi *výrobek*, který má *název rotační vrtačka*
a má *nejnižší cenu*.

- vhodným označováním dokumentu získáme informačně bohatší dokumenty
- potřeba standardizovaných sad značek (DTD, XML Schémata)

Proč to nefunguje?

- je nemožné dohodnout se na celosvětově sdílené sadě značek
- autoři stránek jsou příliš líní a nevzdělaní
- autoři nemají důvod sémanticky značkovat stránky, protože to nepodporují vyhledávače
- vyhledávače nepodporují sémantické vyhledávání, protože stránky neobsahují sémantické značkování
- východisko: sémantický web

Sémantický web

Definice

The Semantic Web is the representation of data on the World Wide Web. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming.

The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

—Tim Berners-Lee, James Hendler, Ora Lassila

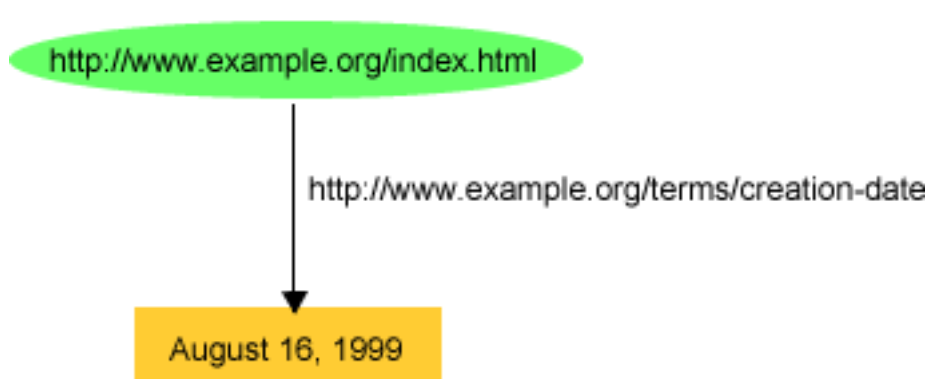
Sémantický web

Idea

- doplnit síť webových stránek sítí výroků
- výroky lze na rozdíl od webových stránek automatizovaně zpracovávat
- výroky se zapisují ve standardizovaném formátu RDF (Resource Description Format)
 - založen na XML
 - výrok = (subjekt, predikát, objekt)
 - jednotlivé části výroku jsou identifikovány URI adresou (případně hodnotou)
- z „webu dokumentů“ se stane „web znalostí“

Příklad RDF

- Stránka `http://www.example.org/index.html` má *datum vytvoření*, jehož hodnota je *16. srpna 1999*.



- RDF reprezentace:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:externs="http://www.example.org/terms/">

  <rdf:Description rdf:about="http://www.example.org/index.html">
    <externs:creation-date>August 16, 1999</externs:creation-date>
  </rdf:Description>

</rdf:RDF>
```

Perspektiva sémantického webu

- na široké nasazení se zatím čeká
- RDF anotace musí být vytvářeny ručně
- pro agregaci RDF metadat a jejich další zpracování je potřeba sjednotit „slovníky pojmů“ (tzv. ontologie)

Mikroformáty

- konvence pro vkládání strukturovaných metainformací do HTML, která využívá pouze stávající vlastnosti jazyka
- ```

 Přednáška o využití XML na webu:
 <abbr class="dtstart" title="2006-04-28">28. dubna</abbr>
 <abbr class="dtend" title="2006-04-28"></abbr>,
 v posluchárně B Vysoké školy ekonomické v ►
Praze


```
- mikroformáty nevdí prohlížečům (mnoho prohlížečů nezvládne zpracovat XHTML dokument s vloženými fragmenty RDF nebo XML)
- mikroformáty lze strojově zpracovávat
- vyvinuly se a navazují na důsledné použití tříd v CSS

# Další zdroje informací

Další zdroje informací .....	30
------------------------------	----

# Další zdroje informací

<http://www.w3.org/>  
specifikace XML, XSLT a dalších základních XML technologií

<http://www.oasis-open.org/cover/>  
odkazy na aplikace, standardy, novinky, ...

<http://www.zvon.org/>  
tutoriály a referenční příručky

<http://www.kosek.cz/clanky/swn-xml/>  
rozsáhlý seriál o XML v češtině

<http://www.w3.org/2001/sw/>  
stránka W3C o sémantickém webu

<http://microformats.org/>  
stránka o mikroformátech

<http://www.kosek.cz/xml/2010znanosti/>  
slidy z přednášky o sémantice ve webových stránkách