ZÁPADOČESKÁ UNIVERZITA V PLZNI

geomatika

Základy XML – struktura dokumentu

(včetně testových otázek)

Otakar Čerba

Oddělení geomatiky

Katedra matematiky

Fakulta aplikovaných věd

Západočeská univerzita v Plzni

Přednáška z předmětu Počítačová kartografie (KMA/POK)

Datum vytvoření dokumentu: 1. 9. 2011

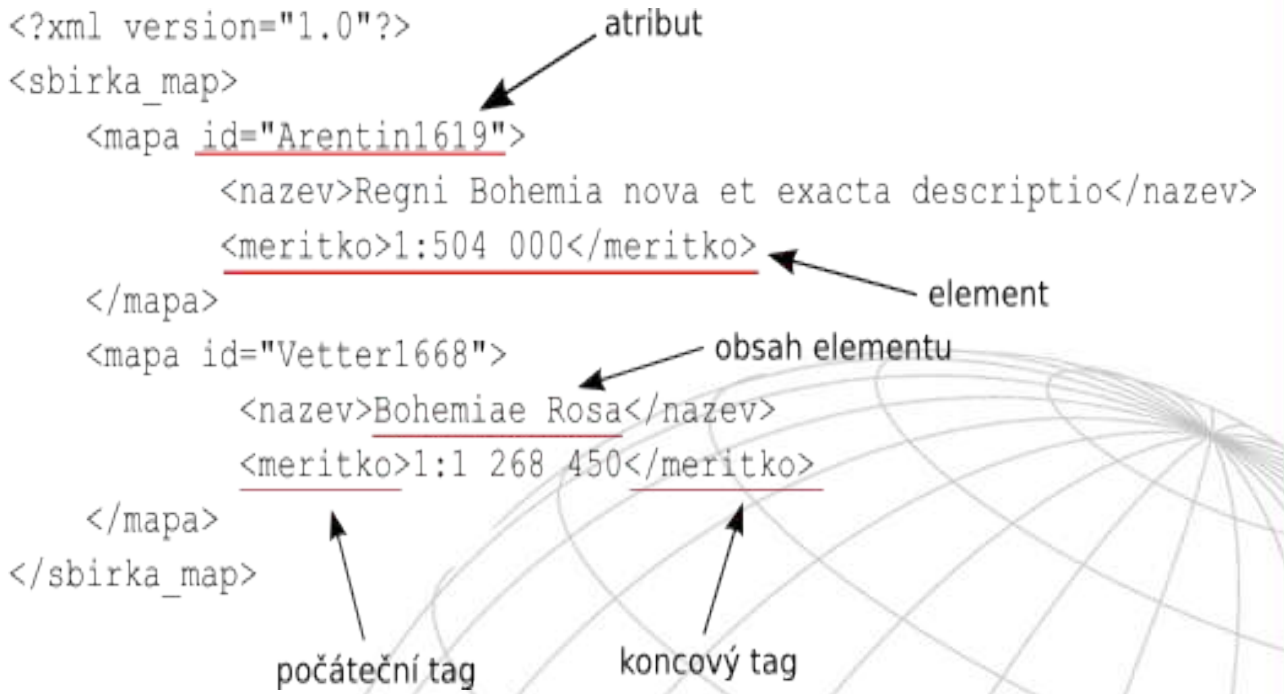
Datum poslední aktualizace: 22. 9. 2011

Extensible Markup Language

Jazyk **Extensible Markup Language (XML)** se řadí do skupiny značkovacích jazyků (markup languages), tedy metajazyků, které označují význam jednotlivých částí dokumentů a nikoli jejich vzhled – hovoří se také o tzv. samopopisných jazycích - o jazycích, které kromě vlastního dokumentu dokáží popsat i jeho strukturu.

Součásti XML dokumentu

- Tagy
- Elementy
- Atributy
- Znakové a textové entity
- CDATA
- Komentáře
- Procesní instrukce
- Hlavička dokumentu



Test

Které prvky nejsou součástí XML dokumentu?

A Tagy

B Koncepty

C Atributy

D Komentáře

Tag je značka, která umožňuje strukturování XML dokumentu. Tagy jsou uzavřeny do ostrých závorek. V dokumentech rozlišujeme počáteční a koncové tagy. Koncový tag se od počátečního liší znakem „lomítko” (/), který je umístěný bezprostředně před názvem tagu. Jména tagů a obecně jména v XML, musí začínat písmenem nebo podtržítkem. Kromě těchto znaků smí dále obsahovat čísla, tečky, dvojtečky a pomlčky. Písmena mohou kromě anglické abecedy pocházet i z množiny tzv. ideografických znaků, do kterých patří i znaky české abecedy. Mezery ani jiné znaky nejsou povoleny. Ve jménech jsou rozlišována velká a malá písmena (XML je case-sensitive).

Test

Jak může vypadat tag?

A /tag/

B (tag)

C {tag}

D <tag>

Test

Jak nemůže vypadat tag?

A <1tag>

B <tag1>

C <Tag1>

D <TAG1>

Test

Jak vypadá koncový tag?

A <koncovyTag>

B </koncovyTag>

C <koncovy/tag>

D <KONCOVYTAG>

Test

Jak nemůže vypadat tag?

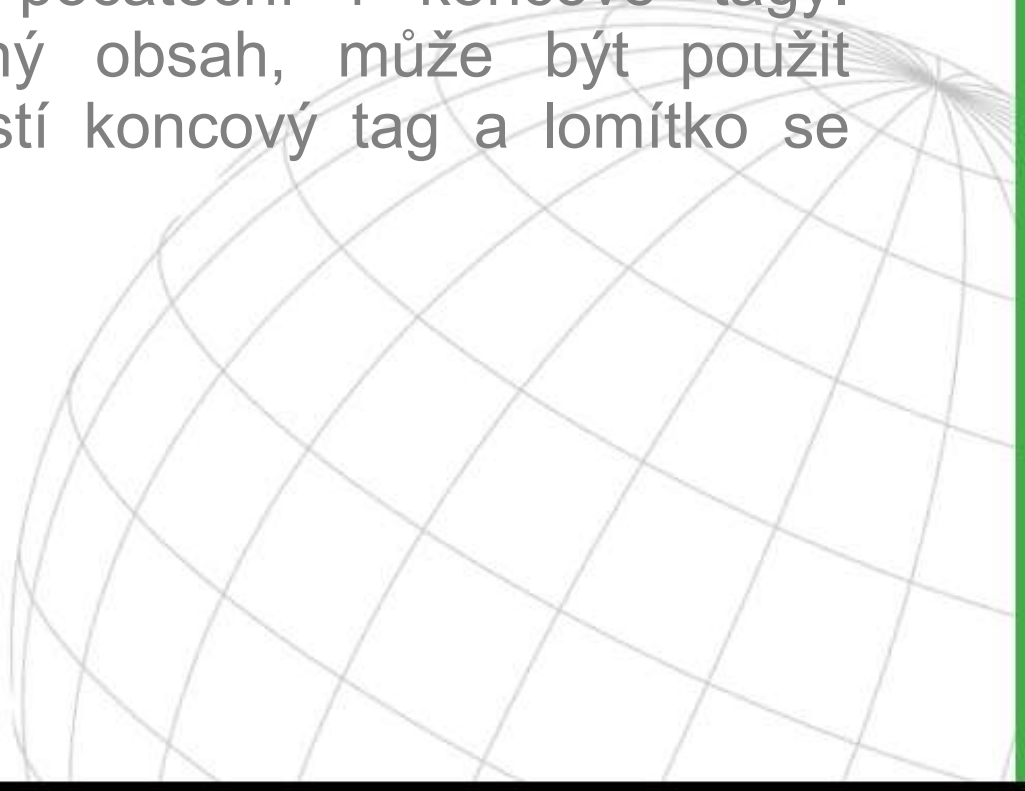
A <můjtag>

B <tagЙ>

C <Tagگ>

D <TAG<>>

Elementy jsou považovány za základní kámen XML. Elementy jsou ohraničeny tagy, na rozdíl od HTML jsou ovšem striktně vyžadovány počáteční i koncové tagy. Jestliže element nemá žádný obsah, může být použit zkrácený zápis, kdy se vypustí koncový tag a lomítko se doplní za jméno elementu.



geomatika

Test

Který element nemá žádný obsah?

A <můjtag/>

B <tag>element</tag>

C <Tag></tag>

D <TAG<>

Test

Musí být počáteční i koncový tag elementu zapsány stejně?

A Musí být zcela shodné.

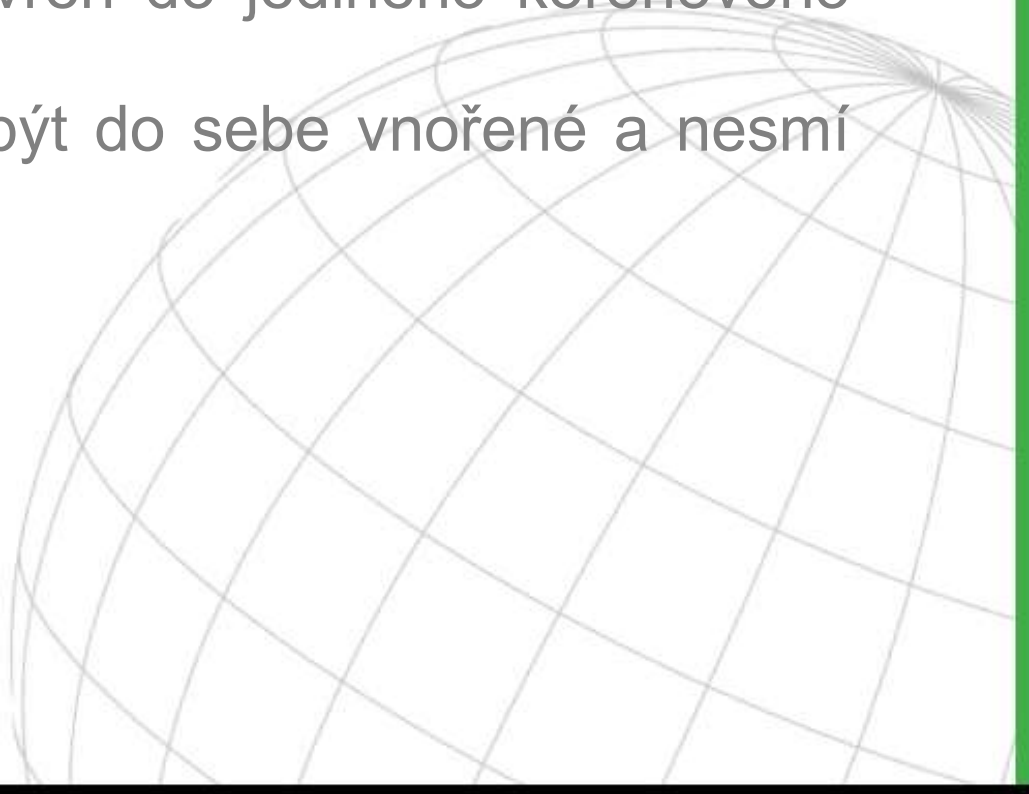
B Liší se pouze lomítkem u koncového tagu.

C Odlišné je lomítko a může být jiná také velikost písmen.

D Záleží pouze na prvním znaku, který nesmí být číslo.

Pro zápis elementů platí následující **pravidla**:

I. Celý XML dokument je uzavřen do jediného kořenového elementu.
II. Jednotlivé elementy musí být do sebe vnořené a nesmí se křížit.



Test

Který zápis XML dokumentu je správný?

A <A>aaa</A><B>bbb</B><C>ccc</C>

B <A><B>bbb<C></B>ccc</C></A>

C <A><B>bbb<C>ccc</C></A>

D <A><B>bbb</B><C>ccc</C></A>

Test

Který zápis XML dokumentu je správný?

A <A><c>CcC</C></A>

B <A>bbb1<C>CcC<C></A>

C <A>bbb<C>CcC</C></A>

D <A><B>bbb</B><c>ccc</C></A>

Test

Který zápis XML dokumentu je správný?

A <A>aaa<A>

B </C>

C <B>

D <A/>

Atributy představují doplňkovou informaci k elementům – element může obsahovat i více různých atributů. Atributy se zapisují do počátečního tagu elementu ve tvaru jméno atributu, rovnítko a hodnota atributu zapsaná do uvozovek nebo apostrofů.

```
<odstavec autor="J.Novák" software="XML Mind">Text odstavce...</odstavec>
```

Test

Proč je následující zápis správný?

```
<E a="11" A="11"/>
```

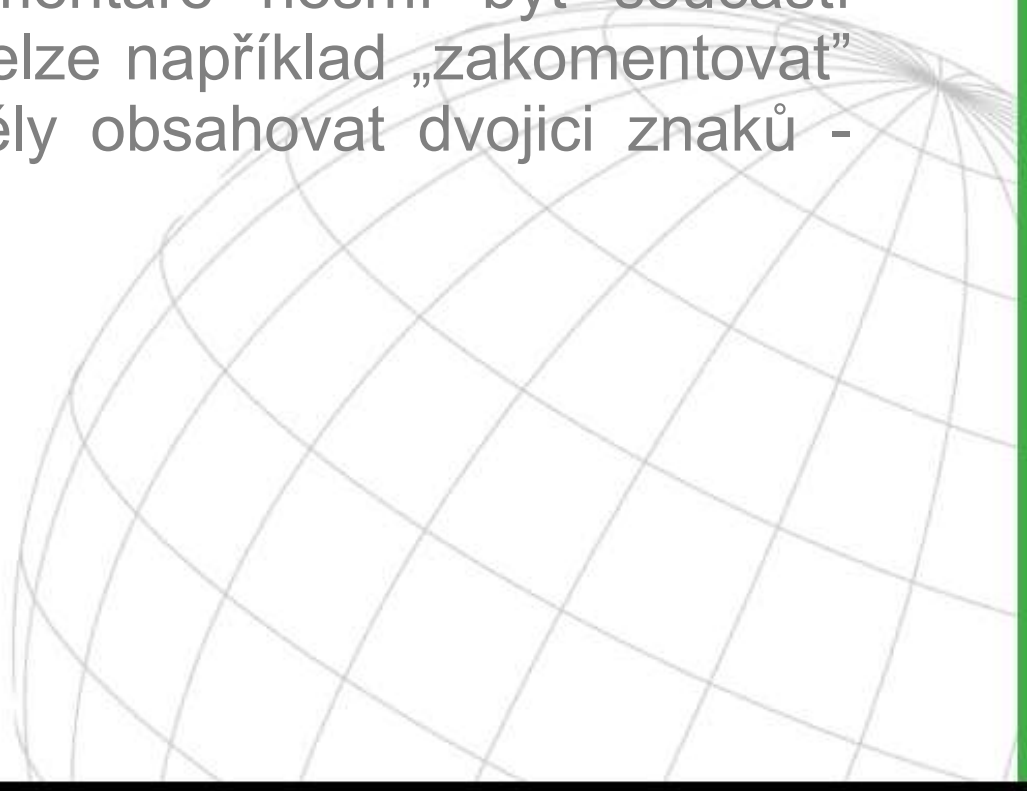
A Protože se jedná o dva elementy vložené do elementu E.

B Protože XML je case-sensitive.

C Protože jeden element může obsahovat více stejných atributů.

D Protože oba elementy mají stejnou hodnotu.

Komentáře se zapisují mezi znaky <!-- a -->. Komentáře nejsou součástí programového kódu, tzn. že nejsou zpracovávány programy. Komentáře nesmí být součástí ostatního značkování (např. nelze například „zakomentovat" atributy). Komentáře by neměly obsahovat dvojici znaků - (spojovník).



Test

Jaký komentář je správný?

A <!-- komentář --!>

B <-- komentář -->

C <!-- komentář -->

D <!-- --> komentář </!-- -->

Místo některých znaků (například <, >, &), které z nějakého důvodu nemůžeme do textu zapsat (například tvoří součást zápisu značkování), použijeme tzv. **znakové entity**. Konkrétně znaky pro začátek tagu (<) a začátek entity (&) se do obsahu elementu nebo atributu musí vždy zapisovat pomocí entity. Pro výše uvedené příklady platí znakové entity <, >, &.

Kromě těchto tří znakových entit jsou v XML předdefinované pouze dvě další entity pro apostrofy (') a uvozovky ("). Díky tomu může text obsahovat uvozovky i apostrofy zároveň. Jako znakovou entitu lze vložit libovolný znak pomocí Unicode kódu – před vlastní kód je nutné umístit prefix, jestliže se jedná o decimální kód znak používá se prefix #, v případě hexadecimálního kódu se prefix zapisuje ve tvaru #x.

Test

Které znaky se musí zapisovat pomocí znakových entit?

A >, <, ', &, “

B >, <

C >, <, &

D <, &

Test

Jak zapíšete následující výraz - 1<2 - jako obsah elementu A

A <A1<2/>

B <A>1<<2</A>

C <A><1<2</A>

D <A>1<2</A>

Hlavička dokumentu (XML deklarace) tvoří první řádku XML dokumentu. Je ohraničená ostrými závorkami a otazníky. Skládá se z klíčového slova xml a povinné deklarace verze XML (version).
Nepovinné jsou deklarace použitého kódování (encoding) a parametru standalone, který specifikuje možnost používání externích souborů - hodnoty tohoto parametru jsou yes (defaultní hodnota - soubor existuje sám o sobě a nepoužívá externí soubory) a no. Deklarace kódování se nemusí používat, pokud je XML soubor zapsán v UTF-8. Atributy hlavičky XML dokumentu musí být zapisovány přesně v pořadí uvedeném v následujícím příkladu.

Test

Která hlavička XML dokumentu je zapsaná správně?

A <? xml version=”1.0” coding=“unicode“?>

B <? xml version=”1.0” ?>

C <?-- xml version=”1.0” encoding=“windows-1250“ --?>

D <? xml standalone=“yes“ version=”1.0” ?>

Procesní instrukce (instrukce pro zpracování, prováděcí instrukce, processing instructions, PI) představují speciální mechanismus pro přidávání nestandardních dat ke XML dokumentu. Pomocí procesních instrukcí je možné do XML kódu vložit nejrůznější stylové soubory nebo příkazy skriptovacího jazyka. Prováděcí instrukce se zapisují do špičatých závorek a otazníků - <? procesní instrukce ?>. První slovo v zápisu procesní instrukce definuje cíl příkazu.

```
<?xml-stylesheet      href="mystylesheet.css"
type="text/css" ?>
...
<datum><?php echo Date("d.m.Y")?></datum>
```

Sekce CDATA (character data) slouží k zápisu velkých částí textu, který obsahuje speciální znaky - například výpisy programového kódu. Používáním CDATA se omezí používání znakových entit, které je dosti komplikované. CDATA se zapisují pomocí této struktury - <[CDATA[...]]>

```
<program typ="SVG">
   <![CDATA[
         <line  x1="140"  y1="100"  x2="150"  y2="170"
style="stroke:black; stroke-width:1"/>
        <polyline  points="50  50,  80  60,  140  100"
style="stroke:black; stroke-width:1; fill:none"/>
          <path  d="M  50  50  L  90  130  150  170"
style="stroke:black; stroke-width:1; fill:none"/>
           <path  d="M  80  60  l  10  70  50  -30"
style="stroke:black; stroke-width:1; fill:none"/>
   ]]>
</program>
```

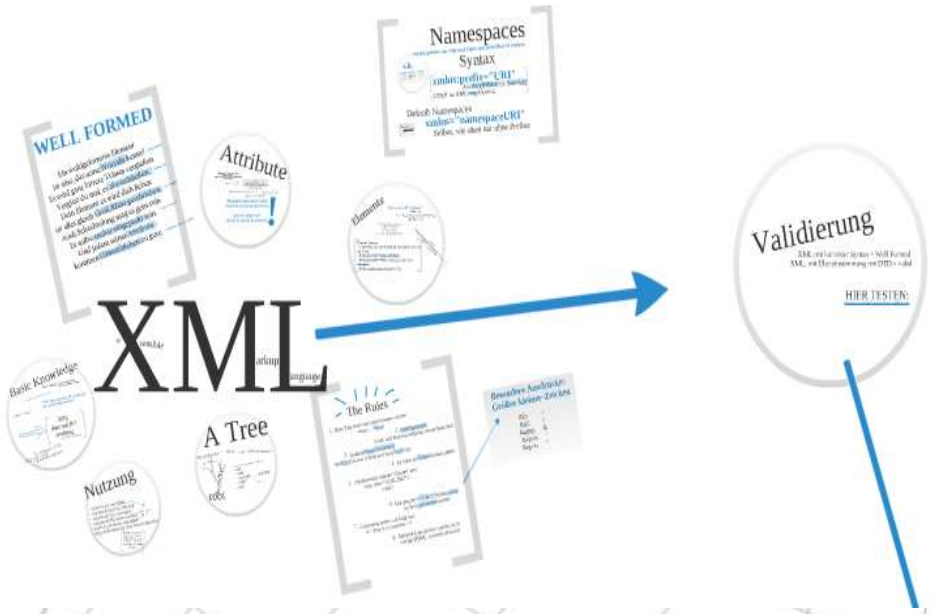
Kromě znakových entit existují také textové entity, které rozdělujeme na interní a externí. Interní textové entity slouží k výraznému zkrácení textu, kdy do entity můžeme uložit text, který se bude v textu často opakovat. Místo tohoto textu se pak uvede pouze název entity. Externí textové entity slouží k vložení částí kódů do XML souborů - tento mechanismus se využívá při modularizaci rozsáhlých XML dokumentů, které pak může editovat více uživatelů současně.

Entity se do dokumentu vkládají pomocí sekvence `&jméno_entity;`.

V souvislosti s entitami, také s jinými mechanismy podporujícími odkazování mezi dokumenty, je vhodné popsat rozdíl mezi XML souborem a XML dokumentem. Dokument zapsaný v XML může být uložen v jediném XML souboru, ale také se může skládat z několika vzájemně provázaných XML souborů. Stejně tak jeden XML soubor může být součástí několika různých XML dokumentů.

Závěr aneb kam dál...

- Stránky na W3C – specifikace a jiné informace
- W3C Schools – tutoriály
- XML a značkovací jazyky – prezentace na dipity.com
- Prezentace na Slideshare, Scribd a Prezi
- Zvon & Kosek – tradiční, léty prověřené české zdroje



geomatika

Děkuji za pozornost a případné dotazy

cerba@kma.zcu.cz

http://cz.linkedin.com/in/otakarcerba

Podpořeno projektem FRVŠ 584/2011