

---

# Přehled obsahu předmětu

## Obsah

XML, základní standardy, zpracování XML, navigace v XML, modelování XML dat .....	1
XML .....	1
Struktura XML dokumentů .....	2
Základní standardy .....	2
Odkazování a navigace v XML datech .....	3
Modelování XML dat .....	3
Transformace XML dat, ukládání a dotazování nad XML daty .....	4
Jazyk XSLT - základní principy .....	4
Významné aplikace XML, mikroformáty, ETL nástroje, metadata, RDF .....	4
Významné aplikace XML - UI .....	4
Významné aplikace - formátování .....	5
Významné aplikace - formátování (pokračování) .....	5
Významné aplikace - DocBook .....	6
Mikroformáty .....	6
ETL nástroje .....	7
RDF (Model) .....	7
RDF (Schema, Ontologie) .....	8

## XML, základní standardy, zpracování XML, navigace v XML, modelování XML dat

### XML

- Co je to XML?
  - *XML je metajazyk sloužící k definici vlastních značkovacích jazyků.*
- Základní požadavky na strukturu XML dokumentu
  - XML prolog
  - kořenový element
  - vnořené elementy - nesmí se křížit
  - hodnoty atributů uzavřeny do uvozovek resp. apostrofů.
- Co to znamená, že je dokument well-formed?

## Struktura XML dokumentů

- Logická
  - elementy
  - atributy
  - textové uzly
  - instrukce pro zpracování
  - komentáře
- fyzická
  - Dokument může být ve více fyzických jednotkách - entitách
  - Vždy alespoň jedna entita dokumentu
  - typy entit
    - parsované - soubory se správně utvořeným značkováním
    - neparsované - např. binární data
    - textové
- znaky
  - kódování (Unicode)
  - znakové sady (UTF-8, UTF-16, UTF-32)

## Základní standardy

- XML Base - slouží k vyhodnocování relativních odkazů
- Jmenné prostory
  - deklarace
  - implicitní jmenný prostor
  - prefix jmenného prostoru
- XML Infoset - popisuje, jaké informace obsahuje daný uzel XML dokumentu.
- XML Catalog - sjednocení odkazování parsovatelných entit na různých systémech, na různých místech

## Odkazování a navigace v XML datech

- XPath
  - navigace v XML datech
  - podobná navigaci v souborovém systému
  - Struktura:
    - cesta
    - predikát upřesňující výběr
- XLink
  - obecný standard pro odkazování se mezi dokumenty
  - umožňuje vazby 1:1, 1:n, m:n
  - řešení pomocí atributů ve jmenném prostoru pro XLink
- XPointer
  - adresace v rámci dokumentu
  - využívá se např. ve spojení se standardem XLink

## Modelování XML dat

- Slouží k definici struktury XML dat.
- Validita dokumentu
- Příklady standardů používaných pro modelování struktury:
  - DTD
  - XML Schema
  - Relax NG
    - XML formát
    - kompaktní
  - Schematron
  - ...

- Nástroje pro validaci:
  - DTD, XML Schema - xerces
  - Relax NG - Jing, libxml2
  - ...

## **Transformace XML dat, ukládání a dotazování nad XML daty**

### **Jazyk XSLT - základní principy**

- standard W3C
- slouží ke specifikaci transformace XML dat na jiná XML data, případně HTML resp. textová data;
- funkcionální jazyk
- struktura šablony:
  - jednotlivé vzory (templates)
  - template má
    - výběrovou část - xpath výraz, který udává na které uzly se má daný vzor aplikovat
    - konstrukční část - data, která se mají přenést na výstup.

## **Významné aplikace XML, mikroformáty, ETL nástroje, metadata, RDF**

### **Významné aplikace XML - UI**

- GUI
  - SVG
  - XHTML
  - XForms
- Dialogové rozhraní - W3C VoiceBrowser Activity

- VoiceXML
- SRGS
- SISR

## **Významné aplikace - formátování**

- Úrovně formátování
  - konceptuální
  - logické
  - fyzické
- Nástroje
  - filtrování proudu událostí
  - manipulace s DOM
  - XSLT
  - dotazovací jazyk

## **Významné aplikace - formátování (pokračování)**

- Postupy
  - Web, běžná obrazovka, RTF, ...
    - XSLT
    - CSS
  - Tiskový výstup
    1. XSLT: XML -> XSL:FO
    2. rendering XSL:FO do výsledného formátu
      - Apache FOP, ...
      - PassiveTex/XML Tex
  - pro malé displaye:
    - XSLT do WML

- ...
- hlasový výstup:
  - XSLT na VoiceXML

## Významné aplikace - DocBook

- Formát pro značkování dokumentů
- Původně SGML, nyní převažuje XML
- Základní princip - sémantické značkování textu
- Základní typy prvků:
  - set - kolekce knih resp. dalších kolekcí
  - book - kniha
  - part
  - article
  - chapter
  - appendix
  - ...

## Mikroformáty

- Slouží k přidání sémantického značkování do lidsky čitelných údajů tak, aby je bylo možno programově zpracovávat
- Cíle:
  - získání
  - indexace
  - prohledávání
  - vzájemné odkazování
- Příklady mikroformátů

- hCalendar
- hCard
- XOXO
- ...
- Existují návrhové vzory pro jejich využití

## ETL nástroje

- Extract-Transform-Load
- nástroje pro datovou integraci
- Implementace
  - Clover ETL
  - Microsoft SQL Server Integration Services
  - Octopus Java-XML ETL Tool
  - java-etl
- Její nasazení s sebou nese určité problémy

## RDF (Model)

- RDF Model
  - základem jsou trojice:
    - zdroj
    - vlastnost
    - hodnota
  - Slouží k přiřazení sémantiky danému zdroji.
  - Lze znázornit:
    - graficky
    - jako trojice
    - XML syntaxí

## RDF (Schema, Ontologie)

- RDF Schema:
  - Specifikuje omezení množiny vlastností, jejich definičních oborů a oborů hodnot
  - Modeluje se pomocí RDF
  - Užívaná schémata:
    - Dublin Core
    - Rámec pro metadata ISVS ČR
    - Aplikační profil NMS
- Ontologie
  - prostředek pro popis znalostí
  - používají metadatové rámce, ale mají bohatší a přesnější sémantiku
  - budovány pro specifické domény