

DTD

Document Type Definition

Validácia XML dokumentov

- základné pojmy
- DTD
- základné bloky XML dokumentu z pohľadu DTD
- deklarácia elementov
- deklarácia atribútov
- entity

Validácia

overenie platnosti XML dokumentu
voči (nejako zapísaným) **pravidlám**

pravidlá určujú typ (**triedu**) dokumentov
konkrétny dokument je **inštanciou** triedy

Zápis pravidiel: XML Schema, DTD, ...

Document Type Definition, DTD

pôvod pri SGML - neXML syntax
(nepodporuje namespace)

definícia jazyka pre XML dokument

DTD + sémantika
pri komunikácii, výmene dokumentov
definovaný štandard, spoločná platforma

XML dokument z pohľadu DTD

“stavebné bloky” XML dokumentu:

- elementy
- značky
- atribúty
- entity
- PCDATA
- CDATA

XML dokument z pohľadu DTD (pokr.) - elementy

• elementy

- značky hlavné stavebné bloky XML dok.
- atribúty rôzny obsah - prázdny, text, ...
- entity môžu mať “deti”
- PCDATA definovanie možných určenie poradia
- CDATA

XML dokument z pohľadu DTD (pokr.) - atribúty

- elementy
- značky doplnková informácia o elementoch
- **atribúty** začiatočná značka elementu
- entity dvojica meno - hodnota
- PCDATA vždy odlíšené “”, ‘’
- CDATA

XML dokument z pohľadu DTD (pokr.) - entity

- elementy
- značky definovanie nahradzovaného textu
- atribúty “entity reference”
- **entity** v texte nahradzuje parser
- PCDATA < &
- CDATA

XML dokument z pohľadu DTD (pokr.) - PCDATA

- elementy
- značky “parsed character data”
- atribúty interpretované parserom
(nahradzovanie entít a pod.)
- entity
- **PCDATA**
- CDATA

XML dokument z pohľadu DTD (pokr.) - CDATA

- elementy
- značky “character data”
- atribúty NEinterpretované parserom
- entity
- PCDATA *Poznámka: XML CDATA*
- **CDATA**

Príklad XML + DTD

```
<?xml version="1.0"?>
<!DOCTYPE note [
  <!ELEMENT note (to,from,body)>
  <!ELEMENT to (#PCDATA)>
  <!ELEMENT from (#PCDATA)>
  <!ELEMENT body (#PCDATA)>
]>
<note>
  <to>validator</to>
  <from>author</from>
  <body>Is this note valid?</body>
</note>
```

Deklarácia DTD

Interné DTD

<!DOCTYPE root-element [element-declarations]>

Externé DTD

<!DOCTYPE root-element SYSTEM “filename”>

XML

```
<?xml version="1.0"?>
<!DOCTYPE note SYSTEM "Note.dtd">
<note>
```

DTD

```
<!ELEMENT note (to,from,body)>
<!ELEMENT to (#PCDATA)>
<!ELEMENT from (#PCDATA)>
```

Deklarácie elementov

<!ELEMENT *element-name* *category*>

kategórie: **EMPTY, ANY**

<!ELEMENT *element-name* (*element-content*)>

#PCDATA, sekvencia detí

Deklarácie elementov - príklady

<!ELEMENT *element-name* EMPTY>

DTD **<!ELEMENT** br EMPTY>

XML **
**

<!ELEMENT *element-name* ANY>

DTD **<!ELEMENT** oslovenie ANY>

XML **<oslovenie>**

Vážený/á **<adresát>**správca**</adresát>**,
</oslovenie>

Poznámka: bez kontroly obsahu

Deklarácie elementov - príklady

<!ELEMENT *element-name* (#PCDATA)>

DTD **<!ELEMENT** oslovenie (#PCDATA)>

XML **<oslovenie>**

Vážený správca,
</oslovenie>

reťazec znakov (bez možnosti kontroly dátového typu)

Deklarácie elementov - sekvencia detí

<!ELEMENT *element-name* (*child-name*,*child-name*,...)>

DTD **<!ELEMENT** note (to,from,body)>

XML **<note>**

<to>...</to>

<from>...</from>

<body>...</body>

</note>

“and” sekvencia - presne **určené poradie**

Sekvencia detí - kardinalita

deklarovaná početnosť

zápis priamo za deklarovaným menom dieťaťa

card	výskyt
	práve jeden
+	aspoň jeden
?	najviac jeden
*	ľubovoľne

Sekvencia detí - kardinalita (príklad)

príklad - jednoduchý odborný článok

```
<!ELEMENT clanok (nazov,autor+,anotacia?,odstavec+)>
<!ELEMENT nazov (#PCDATA)>
<!ELEMENT autor (titul*,meno+,priezvisko,titul*)>
<!ELEMENT anotacia (odstavec+)>
<!ELEMENT odstavec (obrazok?,text)>
<!ELEMENT obrazok EMPTY>
<!ELEMENT text (#PCDATA)>
<!ELEMENT titul (#PCDATA)>
<!ELEMENT meno (#PCDATA)>
<!ELEMENT priezvisko (#PCDATA)>
```

Sekvencia detí (pokr.)

kontext "bud'/alebo" zápisom "|"

```
<!ELEMENT odstavec (obrazok|text)>
<!ELEMENT note (to,from,header,(message|body))>
```

zmiešaný obsah

```
<!ELEMENT note (#PCDATA|to|from|header|message)*>
```

Deklarácia atribútov

```
<!ATTLIST elem-name att-name att-type default-value>
```

elem-name: meno elementu
att-name: meno atribútu
att-type: typ atribútu
default-value: určená hodnota (resp. ďalšie vlastnosti)

Poznámka: bez hierarchie

Deklarácia atribútov - typ atribútu

<!ATTLIST *elem-name att-name att-type default-value* **>**

CDATA	reťazec znakov
(en1 en2 ...)	enumerácia
ID	unikátny reťazec
IDREF	odkaz na ID
musí byť zabezpečená jednoznačnosť ID v rámci dokumentu XML	
NMTOKEN	(name) - pravidlá pre pomenovávanie

Ďalšie: NMTOKENS, NOTATION, ...

Deklarácia atribútov - default value

<!ATTLIST *elem-name att-name att-type default-value* **>**

<i>value</i>	default hodnota
#DEFAULT <i>value</i>	default hodnota
#REQUIRED	vyžadované
#IMPLIED	nevyžadované
#FIXED <i>value</i>	fixná hodnota

Deklarácia atribútov (príklady)

DTD **<!ELEMENT** square EMPTY
<!ATTLIST square width CDATA "0"

XML **<square** width="100"/>

prázdny element **square**, atribút **width**
reťazec znakov, s default hodnotou "0"

- ak neuvedený, nadobúda default hodnotu
- ak uvedený, nastavuje hodnotu

Deklarácia atribútov (príklady)

DTD **<!ATTLIST** contact fax CDATA #IMPLIED

XML **<contact>**
 *<name>*Roman*</name>*
 <contact fax="01/1234"
 *<name>*Ilko*</name>*

atribút **fax** k elementu **contact**
reťazec znakov, nevyžadovaná hodnota
(neexistuje default hodnota)

- ak neuvedený, atribút bez hodnoty
- ak uvedený, nastavuje hodnotu

Deklarácia atribútov (príklady)

DTD `<!ATTLIST contact phone CDATA #REQUIRED>`

XML `<contact phone="01/4567">
 <name>Roman</name>
</contact>`

atribút **phone** k elementu **contact**
reťazec znakov, vyžadovaná hodnota (bez default)

- ak neuvedený, dokument je neplatný
- ak uvedený, nastavuje hodnotu

Deklarácia atribútov (príklady)

DTD `<!ATTLIST square color CDATA #FIXED "red">`

XML `<square width="100" color="red"/>`

atribút **color** k elementu **square**
reťazec znakov, fixná hodnota "red"

- ak neuvedený, default fixná hodnota
- ak uvedený, platný len s fixnou default hodnotou

Deklarácia atribútov (príklady)

DTD `<!ATTLIST payment type (check|cash) "cash">`

XML `<payment type="check"/>
<payment type="cash"/>`

atribút **type** k elementu **payment**
enumerácia, default hodnota

- ak neuvedený, default hodnota "cash"
- ak uvedený, platný len s hodnotou z enumerácie

! hodnota reťazec **bez whitespace** znakov

Deklarácia atribútov - ID, IDREF

atribút typu ID obsahuje unikátnu hodnotu (XML dok.)

- pre typ **ID** musí default byť **#IMPLIED** / **#REQUIRED**
- element môže mať **najviac jeden** atribút typu ID
- hodnota atribútu typu ID musí byť **meno**

atribút typu **IDREF** obsahuje odkaz na ID

- hodnota atribútu typu IDREF musí byť ID niektorého elementu XML dokumentu

Deklarácia atribútov - ID, IDREF (príklad)

DTD `<!ELEMENT document (#PCDATA|link)*>`
`<!ATTLIST document id ID #REQUIRED>`

`<!ELEMENT link (#PCDATA)>`
`<!ATTLIST link doc IDREF #REQUIRED>`

XML

```
<documents>
  <document id="doc001">
    This document is referenced.
  </document>

  <document id="doc002">
    This one is not, but it references
    another <link doc="doc001">one</link>.
  </document>
</documents>
```

Deklarácia entít

nahradzované textové reťazce

interná entita: `<!ENTITY entity-name "entity-value">`

`<!ENTITY meno "Roman Filkorn">`

```
<naapis>AML</naapis>
<autor>&meno;</autor>
<notacia>
```

externá entita: `<!ENTITY entity-name SYSTEM "URL">`

`<!ENTITY meno SYSTEM "http://a.sk/a.xml">`

Deklarácia entít (pokr.)

nahradzovanie reprezentácií znakov

ISO/IEC 10646

desiatkový zápis : `&#kód;`

hexazápis : `&#xkód;`

typický príklad (` `): ` `

Kompletný DTD

```
<!-- Dateiname: Auftrag.dtd, Projekt PROWEB -->
<!ELEMENT Auftrag (Teilauftrag* | Kostenplan )>
<!ATTLIST Auftrag Auftrags_Nummer CDATA #REQUIRED>
```

```
<!ELEMENT Teilauftrag (Kostenplan )>
<!ATTLIST Teilauftrag Teilauftrags_Nummer CDATA #REQUIRED>
```

```
<!ELEMENT Kostenplan (Kostenplan_pro_Monat+ )>
```

```
<!ELEMENT Kostenplan_pro_Monat (Kostenplan_pro_Herkunft_und_Monat* )>
<!ATTLIST Kostenplan_pro_Monat Monat CDATA #REQUIRED>
<!ATTLIST Kostenplan_pro_Monat Rechner_Kosten CDATA #IMPLIED>
<!ATTLIST Kostenplan_pro_Monat Reise_Kosten CDATA #IMPLIED>
<!ATTLIST Kostenplan_pro_Monat Sonstige_Kosten CDATA #IMPLIED>
```

```
<!ELEMENT Kostenplan_pro_Herkunft_und_Monat EMPTY>
<!ATTLIST Kostenplan_pro_Herkunft_und_Monat Herkunft CDATA #REQUIRED>
<!ATTLIST Kostenplan_pro_Herkunft_und_Monat Stunden CDATA #REQUIRED>
```


DTD - zhrnutie

deklarácia elementov

- rôzny obsah (EMPTY, ANY, #PCDATA, ...)
- kardinalita "detí" (+, *, ?)

deklarácia atribútov

- rôzne typy (CDATA, ID/IDREF, ...)
- default hodnota, význam (#REQUIRED, #IMPLIED, ...)

deklarácia entít

nepodporuje namespace

Poznámka: vynechané podmienené štruktúry,
% entity, ...

DTD - Úloha

```
<!-- By David Moisan, www.shore.net/~dmoisan/ -->
<!DOCTYPE TVSCHEDULE [

<!ELEMENT TVSCHEDULE (CHANEL+)>
<!ELEMENT CHANNEL (BANNER, DAY+)>
<!ELEMENT BANNER (#PCDATA)>
<!ELEMENT DAY (DATE, (HOLIDAY|PROGRAMSLOT+))+>
<!ELEMENT HOLIDAY (#PCDATA)>
<!ELEMENT DATE (#PCDATA)>
<!ELEMENT PROGRAMSLOT (TIME, TITLE, DESCRIPTION?)>
<!ELEMENT TIME (#PCDATA)>
<!ELEMENT TITLE (#PCDATA)>
<!ELEMENT DESCRIPTION (#PCDATA)>

<ATTLIST TVSCHEDULE NAME CDATA #REQUIRED>
<ATTLIST CHANNEL CHAN CDATA #REQUIRED>
<ATTLIST PROGRAMSLOT VTR CDATA #IMPLIED>
<ATTLIST TITLE RATING CDATA #IMPLIED>
<ATTLIST TITLE LANGUAGE CDATA #IMPLIED>

]>
```

Ďakujem za pozornosť



WWW Consortium (špecifikácie)
www.w3.org

OpenSource implementácia parsera (Xerces)
xml.apache.org

Tutoriály, materiály
www.zvon.org

iné
www.xml.org, ...